

Independence

1 Independent Events

Suppose that we flip two fair coins simultaneously on opposite sides of a room. Intuitively, the way one coin lands does not affect the way the other coin lands. The mathematical concept that captures this intuition is called *independence*. In particular, events A and B are independent if and only if:

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

Generally, independence is something you *assume* in modeling a phenomenon—or wish you could realistically assume. Many useful probability formulas only hold if certain events are independent, so a dash of independence can greatly simplify the analysis of a system.

1.1 Examples

Let's return to the experiment of flipping two fair coins. Let A be the event that the first coin comes up heads, and let B be the event that the second coin is heads. If we assume that A and B are independent, then the probability that both coins come up heads is:

$$\begin{aligned}\Pr(A \cap B) &= \Pr(A) \cdot \Pr(B) \\ &= \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{4}\end{aligned}$$

On the other hand, let C be the event that tomorrow is cloudy and R be the event that tomorrow is rainy. Perhaps $\Pr(C) = 1/5$ and $\Pr(R) = 1/10$ around here. If these events were independent, then we could conclude that the probability of a rainy, cloudy day was quite small:

$$\begin{aligned}\Pr(R \cap C) &= \Pr(R) \cdot \Pr(C) \\ &= \frac{1}{5} \cdot \frac{1}{10} \\ &= \frac{1}{50}\end{aligned}$$

Unfortunately, these events are definitely not independent; in particular, every rainy day is cloudy. Thus, the probability of a rainy, cloudy day is actually $1/10$.

1.2 Working with Independence

There is another way to think about independence that you may find more intuitive. According to the definition, events A and B are independent if and only if:

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

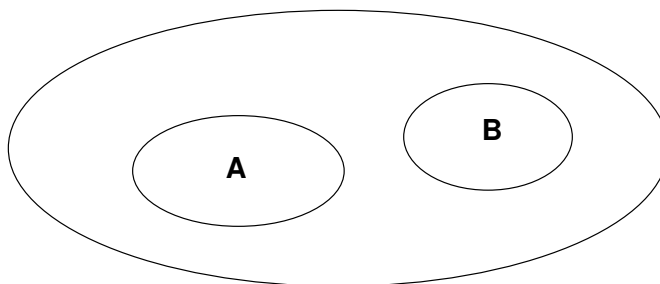
The equation on the left always holds if $\Pr(B) = 0$. Otherwise, we can divide both sides by $\Pr(B)$ and use the definition of conditional probability to obtain an alternative definition of independence:

$$\Pr(A | B) = \Pr(A) \quad \text{or} \quad \Pr(B) = 0$$

This equation says that events A and B are independent if the probability of A is unaffected by the fact that B happens. In these terms, the two coin tosses of the previous section were independent, because the probability that one coin comes up heads is unaffected by the fact that the other came up heads. Turning to our other example, the probability of clouds in the sky is strongly affected by the fact that it is raining. So, as we noted before, these events are not independent.

1.3 Some Intuition

Suppose that A and B are disjoint events, as shown in the figure below.



Are these events independent? Let's check. On one hand, we know

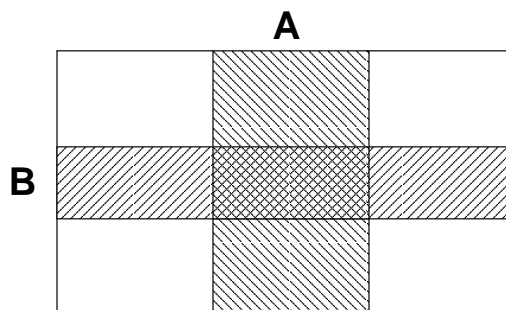
$$\Pr(A \cap B) = 0$$

because $A \cap B$ contains no outcomes. On the other hand, we have

$$\Pr(A) \cdot \Pr(B) > 0$$

except in degenerate cases where A or B has zero probability. Thus, *disjointness and independence are very different ideas*.

Here's a better mental picture of what independent events look like.



The sample space is the whole rectangle. Event A is a vertical stripe, and event B is a horizontal stripe. Assume that the probability of each event is proportional to its area in the diagram. Now if A covers an α -fraction of the sample space, and B covers a β -fraction, then the area of the intersection region is $\alpha \cdot \beta$. In terms of probability:

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

1.4 An Experiment with Two Coins

Suppose that we flip two independent, fair coins. Consider the following two events:

A = the coins match (both are heads or both are tails)

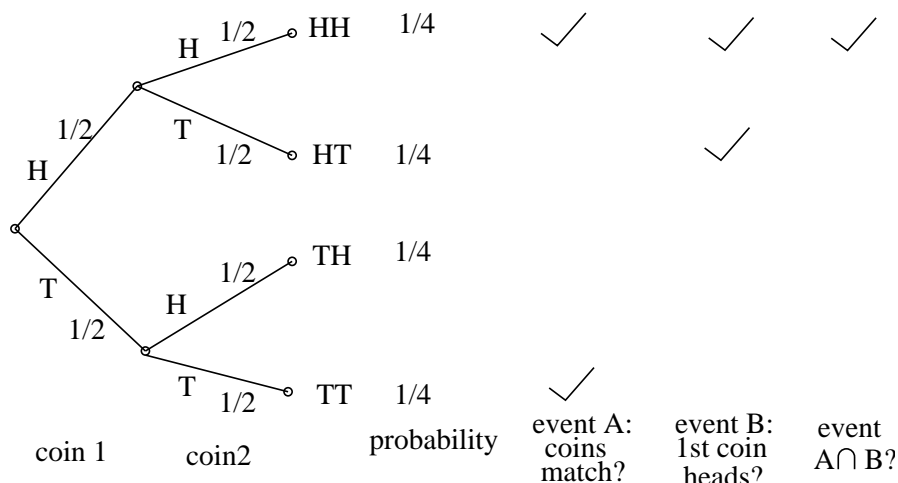
B = the first coin is heads

Are these independent events? Intuitively, the answer is “no”. After all, whether or not the coins match *depends* on how the first coin comes up; if we toss HH , they match, but if we toss TH , then they do not. However, the mathematical definition of independence does not correspond perfectly to the intuitive notion of “unrelated” or “unconnected”. These events actually are independent!

Claim 1. *Events A and B are independent.*

Proof. We must show that $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$.

Step 1: Find the Sample Space. As shown in the tree diagram below, there are four possible outcomes: HH , HT , TH , and TT .



Step 2: Define Events of Interest. The outcomes in event A (coins match) and event B (first coin is heads) are checked in the tree diagram above

Step 3: Compute Outcome Probabilities. Since the coins are independent and fair, all edge probabilities are $1/2$. We find outcome probabilities by multiplying edge probabilities along each root-to-leaf path. All outcomes have probability $1/4$.

Step 4: Compute Event Probabilities. Now we can verify that $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$:

$$\Pr(A) = \Pr(HH) + \Pr(TT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$\Pr(B) = \Pr(HH) + \Pr(HT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

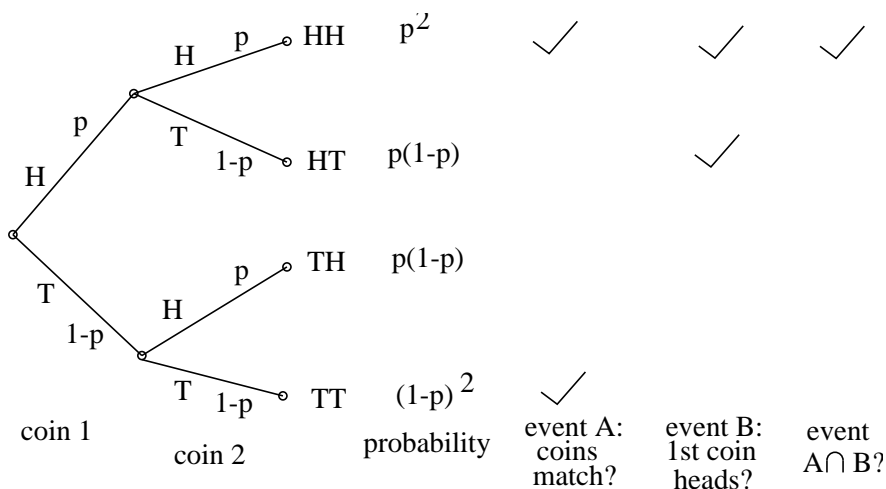
$$\Pr(A \cap B) = \Pr(HH) = \frac{1}{4}$$

Therefore, A and B are independent events as claimed. \square

1.5 A Variation of the Two-Coin Experiment

Suppose that we alter the preceding experiment so that the coins are independent, but not fair. In particular, suppose each coin is heads with probability p and tails with probability $1 - p$ where p might not be $1/2$. As before, let A be the event that the coins match, and let B be the event that the first coin is heads. Are events A and B independent for all values of p ?

The problem is worked out in the tree diagram below.



We can read event probabilities off the tree diagram:

$$\begin{aligned} \Pr(A) &= \Pr(HH) + \Pr(TT) = p^2 + (1-p)^2 = 2p^2 - 2p + 1 \\ \Pr(B) &= \Pr(HH) + \Pr(HT) = p^2 + p(1-p) = p \\ \Pr(A \cap B) &= \Pr(HH) = p^2 \end{aligned}$$

Now events A and B are independent if and only if $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$ or, equivalently:

$$(2p^2 - 2p + 1) \cdot p = p^2$$

Since both sides are multiples of p , one solution is $p = 0$. Dividing both sides by p and simplifying leaves a quadratic equation:

$$2p^2 - 3p + 1 = 0$$

According to the quadratic formula, the remaining solutions are $p = 1$ and $p = 1/2$.

This analysis shows that events A and B are independent only if the coins are either *fair* or *completely biased* toward either heads or tails. Evidently, there was some dependence lurking at the fringes of the previous problem, but it was kept at bay because the coins were fair!

The Ultimate Application

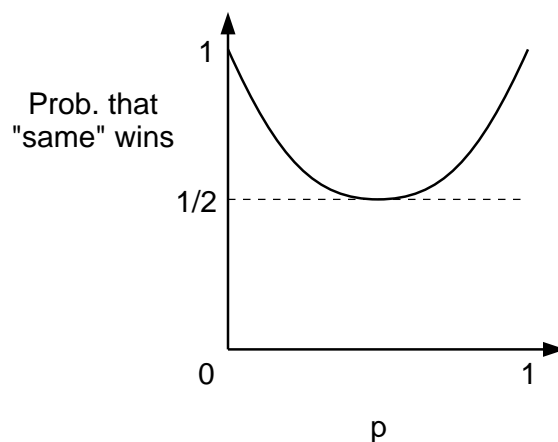
Surprisingly, this has an application to Ultimate Frisbee. Here is an excerpt from the Tenth Edition rules:

- A. Representatives of the two teams each flip a disc. The representative of one team calls "same" or "different" while the discs are in the air. The team winning the flip has the choice of:
1. Receiving or throwing the initial throw-off; or
 2. Selecting which goal they wish to defend initially.
- B. The team losing the flip is given the remaining choice.

As we computed above, the probability that two flips match is:

$$\Pr(A) = p^2 + (1 - p)^2$$

Below we've plotted this match probability as a function of p , the probability that one disc lands face-up.



Frisbees are asymmetric objects with strong aerodynamic properties, so p is not likely to be $1/2$. That plot shows that if there is any bias one way or the other, then saying "same" wins *more* than half the time. In fact, even if frisbees land face up exactly half the time ($p = 1/2$), then saying "same" still wins half the time. Therefore, might as well *always* say "same" during the opening flip!

2 Mutual Independence

We have defined what it means for two events to be independent. But how can we talk about independence when there are more than two events? For example, how can we say that the orientations of n coins are all independent of one another?

Events E_1, \dots, E_n are **mutually independent** if and only if for every subset of the events, the probability of the intersection is the product of the probabilities. In other words, all of the following equations must hold:

$$\begin{aligned} \Pr(E_i \cap E_j) &= \Pr(E_i) \cdot \Pr(E_j) && \text{for all distinct } i, j \\ \Pr(E_i \cap E_j \cap E_k) &= \Pr(E_i) \cdot \Pr(E_j) \cdot \Pr(E_k) && \text{for all distinct } i, j, k \\ \Pr(E_i \cap E_j \cap E_k \cap E_l) &= \Pr(E_i) \cdot \Pr(E_j) \cdot \Pr(E_k) \cdot \Pr(E_l) && \text{for all distinct } i, j, k, l \\ &\dots \\ \Pr(E_1 \cap \dots \cap E_n) &= \Pr(E_1) \cdot \dots \cdot \Pr(E_n) \end{aligned}$$

As an example, if we toss 100 fair coins and let E_i be the event that the i th coin lands heads, then we might reasonably assume that E_1, \dots, E_{100} are mutually independent.

2.1 DNA Testing

This is testimony from the O. J. Simpson murder trial on May 15, 1995:

MR. CLARKE: When you make these estimations of frequency— and I believe you touched a little bit on a concept called independence?

DR. COTTON: Yes, I did.

MR. CLARKE: And what is that again?

DR. COTTON: It means whether or not you inherit one allele that you have is not— does not affect the second allele that you might get. That is, if you inherit a band at 5,000 base pairs, that doesn't mean you'll automatically or with some probability inherit one at 6,000. What you inherit from one parent is what you inherit from the other. (*Got that? – EAL*)

MR. CLARKE: Why is that important?

DR. COTTON: Mathematically that's important because if that were not the case, it would be improper to multiply the frequencies between the different genetic locations.

MR. CLARKE: How do you— well, first of all, are these markers independent that you've described in your testing in this case?

The jury was told that genetic markers in blood found at the crime scene matched Simpson's. Furthermore, the probability that the markers would be found in a randomly-selected person was at most 1 in 170 million. This astronomical figure was derived from statistics such as:

- 1 person in 100 has marker A .
- 1 person in 50 marker B .
- 1 person in 40 has marker C .
- 1 person in 5 has marker D .
- 1 person in 170 has marker E .

Then these numbers were multiplied to give the probability that a randomly-selected person would have all five markers:

$$\begin{aligned} \Pr(A \cap B \cap C \cap D \cap E) &= \Pr(A) \cdot \Pr(B) \cdot \Pr(C) \cdot \Pr(D) \cdot \Pr(E) \\ &= \frac{1}{100} \cdot \frac{1}{50} \cdot \frac{1}{40} \cdot \frac{1}{5} \cdot \frac{1}{170} \\ &= \frac{1}{170,000,000} \end{aligned}$$

The defense pointed out that this assumes that the markers appear mutually independently. Furthermore, all the statistics were based on just a few hundred blood samples. The jury was widely mocked for failing to “understand” the DNA evidence. If you were a juror, would *you* accept the 1 in 170 million calculation?

2.2 Pairwise Independence

The definition of mutual independence seems awfully complicated— there are so many conditions! Here’s an example that illustrates the subtlety of independence when more than two events are involved and the need for all those conditions. Suppose that we flip three fair, mutually-independent coins. Define the following events:

- A_1 is the event that coin 1 matches coin 2.
- A_2 is the event that coin 2 matches coin 3.
- A_3 is the event that coin 3 matches coin 1.

Are A_1, A_2, A_3 mutually independent?

The sample space for this experiment is:

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Every outcome has probability $(1/2)^3 = 1/8$ by our assumption that the coins are mutually independent.

To see if events A_1 , A_2 , and A_3 are mutually independent, we must check a sequence of equalities. It will be helpful first to compute the probability of each event A_i :

$$\begin{aligned}\Pr(A_1) &= \Pr(HHH) + \Pr(HHT) + \Pr(TTH) + \Pr(TTT) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{1}{2}\end{aligned}$$

By symmetry, $\Pr(A_2) = \Pr(A_3) = 1/2$ as well. Now we can begin checking all the equalities required for mutual independence.

$$\begin{aligned}\Pr(A_1 \cap A_2) &= \Pr(HHH) + \Pr(TTT) \\ &= \frac{1}{8} + \frac{1}{8} \\ &= \frac{1}{4} \\ &= \frac{1}{2} \cdot \frac{1}{2} \\ &= \Pr(A_1) \Pr(A_2)\end{aligned}$$

By symmetry, $\Pr(A_1 \cap A_3) = \Pr(A_1) \cdot \Pr(A_3)$ and $\Pr(A_2 \cap A_3) = \Pr(A_2) \cdot \Pr(A_3)$ must hold also. Finally, we must check one last condition:

$$\begin{aligned}\Pr(A_1 \cap A_2 \cap A_3) &= \Pr(HHH) + \Pr(TTT) \\ &= \frac{1}{8} + \frac{1}{8} \\ &= \frac{1}{4} \\ &\neq \Pr(A_1) \Pr(A_2) \Pr(A_3) = \frac{1}{8}\end{aligned}$$

The three events A_1 , A_2 , and A_3 are not mutually independent, even though all *pairs* of events are independent!

A set of events is *pairwise independent* if every pair is independent. Pairwise independence is a much weaker property than mutual independence. For example, suppose that the prosecutors in the O. J. Simpson trial were wrong and markers A , B , C , D , and E appear only *pairwise* independently. Then the probability that a randomly-selected person has all five markers is no more than:

$$\begin{aligned}\Pr(A \cap B \cap C \cap D \cap E) &\leq \Pr(A \cap E) \\ &= \Pr(A) \cdot \Pr(E) \\ &= \frac{1}{100} \cdot \frac{1}{170} \\ &= \frac{1}{17,000}\end{aligned}$$

The first line uses the fact that $A \cap B \cap C \cap D \cap E$ is a subset of $A \cap E$. (We picked out the A and E markers because they're the rarest.) We use pairwise independence on the second line. Now the probability of a random match is 1 in 17,000—a far cry from 1 in 170 million! And this is the strongest conclusion we can reach assuming only pairwise independence.

3 The Birthday Paradox

Suppose that there are 100 students in a lecture hall. There are 365 possible birthdays, ignoring February 29. What is the probability that two students have the same birthday? 50%? 90%? 99%? Let's make some modeling assumptions:

- For each student, all possible birthdays are equally likely. The idea underlying this assumption is that each student's birthday is determined by a random process involving parents, fate, and, um, some issues that we discussed earlier in the context of graph theory. Our assumption is not completely accurate, however; a disproportionate number of babies are born in August and September, for example. (Counting back nine months explains the reason why!)
- Birthdays are mutually independent. This isn't perfectly accurate either. For example, if there are twins in the lecture hall, then their birthdays are surely not independent.

We'll stick with these assumptions, despite their limitations. Part of the reason is to simplify the analysis. But the bigger reason is that our conclusions will apply to many situations in computer science where twins, leap days, and romantic holidays are not considerations. Also in pursuit of generality, let's switch from specific numbers to variables. Let m be the number of people in the room, and let N be the number of days in a year.

3.1 The Four-Step Method

We can solve this problem using the standard four-step method. However, a tree diagram will be of little value because the sample space is so enormous. This time we'll have to proceed without the visual aid!

Step 1: Find the Sample Space

Let's number the people in the room from 1 to m . An outcome of the experiment is a sequence (b_1, \dots, b_m) where b_i is the birthday of the i th person. The sample space is the set of all such sequences:

$$S = \{(b_1, \dots, b_m) \mid b_i \in \{1, \dots, N\}\}$$

Step 2: Define Events of Interest

Our goal is to determine the probability of the event A , in which some two people have the same birthday. This event is a little awkward to study directly, however. So we'll use a common trick, which is to analyze the *complementary* event \bar{A} , in which all m people have different birthdays:

$$\bar{A} = \{(b_1, \dots, b_m) \in S \mid \text{all } b_i \text{ are distinct}\}$$

If we can compute $\Pr(\bar{A})$, then we can compute what we really want, $\Pr(A)$, using the relation:

$$\Pr(A) + \Pr(\bar{A}) = 1$$

Step 3: Assign Outcome Probabilities

We need to compute the probability that m people have a particular combination of birthdays (b_1, \dots, b_m) . There are N possible birthdays and all of them are equally likely for each student. Therefore, the probability that the i th person was born on day b_i is $1/N$. Since we're assuming that birthdays are mutually independent, we can multiply probabilities. Therefore, the probability that the first person was born on day b_1 , the second on day b_2 , and so forth is $(1/N)^m$. This is the probability of every outcome in the sample space.

Step 4: Compute Event Probabilities

Now we're interested in the probability of event \bar{A} in which everyone has a different birthday:

$$\bar{A} = \{(b_1, \dots, b_m) \in S \mid \text{all } b_i \text{ are distinct}\}$$

This is a gigantic set. In fact, there are N choices for b_1 , $N - 1$ choices for b_2 , and so forth. Therefore, by the Generalized Product Rule:

$$|\bar{A}| = N(N - 1)(N - 2) \dots (N - m + 1)$$

The probability of the event \bar{A} is the sum of the probabilities of all these outcomes. Happily, this sum is easy to compute, owing to the fact that every outcome has the same probability:

$$\begin{aligned} \Pr(\bar{A}) &= \sum_{w \in \bar{A}} \Pr(w) \\ &= \frac{|\bar{A}|}{N^m} \\ &= \frac{N(N - 1)(N - 2) \dots (N - m + 1)}{N^m} \end{aligned}$$

We're done!

3.2 An Alternative Approach

The probability theorems and formulas we've developed provide some other ways to solve probability problems. Let's demonstrate this by solving the birthday problem using a different approach—which had better give the same answer! As before, there are m people and N days in a year. Number the people from 1 to m , and let E_i be the event that the i th person has a birthday different from the preceding $i - 1$ people. In these terms, we have:

$$\begin{aligned} \Pr(\text{all } m \text{ birthdays different}) &= \Pr(E_1 \cap E_2 \cap \dots \cap E_m) \\ &= \Pr(E_1) \cdot \Pr(E_2 \mid E_1) \cdot \Pr(E_3 \mid E_1 \cap E_2) \cdots \Pr(E_m \mid E_1 \cap \dots \cap E_{m-1}) \end{aligned}$$

On the second line, we're using the Product Rule for probabilities. The nasty-looking conditional probabilities aren't really so bad. The first person has a birthday different from all predecessors, because there are no predecessors:

$$\Pr(E_1) = 1$$

We're assuming that birthdates are equally probable and birthdays are independent, so the probability that the second person has the same birthday as the first is only $1/N$. Thus:

$$\Pr(E_2 \mid E_1) = 1 - \frac{1}{N}$$

Given that the first two people have different birthdays, the third person shares a birthday with one or the other with probability $2/N$, so:

$$\Pr(E_3 \mid E_1 \cap E_2) = 1 - \frac{2}{N}$$

Extending this reasoning gives:

$$\Pr(\text{all } m \text{ birthdays different}) = \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{m-1}{N}\right)$$

We're done—again! This is our previous answer written in a different way.

3.3 An Approximate Solution

One justification we offered for teaching approximation techniques was that approximate answers are often easier to work with and interpret than exact answers. Let's use the birthday problem as an illustration. We proved that m people all have different birthdays with probability

$$\Pr(\text{all } m \text{ birthdays different}) = \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{m-1}{N}\right)$$

where N is the number of days in a year. This expression is exact, but inconvenient; evaluating it would require $\Omega(m)$ arithmetic operations. Furthermore, this expression is difficult to interpret; for example, how many people must be in a room to make the probability of a birthday match about $1/2$? Hard to say!

Let's look for a simpler, more meaningful approximate solution to the birthday problem. Our expression is a product, but we have more tools for coping with sums. So we first translate the product into a sum by moving it "upstairs" with the rule $u = e^{\ln u}$:

$$\begin{aligned} \Pr(\text{all } m \text{ birthdays different}) &= \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{m-1}{N}\right) \\ &= e^{\ln\left(1 - \frac{1}{N}\right) + \ln\left(1 - \frac{2}{N}\right) + \cdots + \ln\left(1 - \frac{m-1}{N}\right)} \end{aligned}$$

Now we can apply bounds on $\ln(1 - x)$, which we obtained from Taylor's Theorem:

$$\frac{-x}{1-x} \leq \ln(1-x) \leq -x$$

From the upper bound, we get:

$$\begin{aligned} e^{\ln\left(1 - \frac{1}{N}\right) + \ln\left(1 - \frac{2}{N}\right) + \cdots + \ln\left(1 - \frac{m-1}{N}\right)} &\leq e^{-\frac{1}{N} - \frac{2}{N} - \cdots - \frac{m-1}{N}} \\ &= e^{-\frac{m(m-1)}{2N}} \end{aligned}$$

On the last line, we're using the summation formula $1 + 2 + \cdots + (m-1) = m(m-1)/2$.

The lower bound on $\ln(1 - x)$ implies:

$$\ln\left(1 - \frac{k}{N}\right) \geq \frac{-\frac{k}{N}}{1 - \frac{k}{N}} = \frac{-k}{N-k}$$

Substituting this fact into the birthday probability gives:

$$\begin{aligned} e^{\ln\left(1 - \frac{1}{N}\right) + \ln\left(1 - \frac{2}{N}\right) + \cdots + \ln\left(1 - \frac{m-1}{N}\right)} &\geq e^{-\frac{1}{N-1} - \frac{2}{N-2} - \cdots - \frac{m-1}{N-(m-1)}} \\ &\geq e^{-\frac{1}{N-m} - \frac{2}{N-m} - \cdots - \frac{m-1}{N-m}} \\ &= e^{-\frac{m(m-1)}{2(N-m)}} \end{aligned}$$

On the second line, we increase all the denominators to make them equal so that we can sum them.

So now we have closely-matching bounds on the probability that m people have different birthdays:

$$e^{-\frac{m(m-1)}{2(N-m)}} \leq \Pr(\text{all } m \text{ birthdays different}) \leq e^{-\frac{m(m-1)}{2N}}$$

The difference between the exponents is $m^2(m-1)/2N(N-m)$, which goes to zero provided $m = o(N^{2/3})$. So, in the limit, the ratio of the upper bound to the lower bound is 1. Since the exact probability is sandwiched in between these two, we have an asymptotically tight solution to the birthday problem:

$$\Pr(\text{all } m \text{ birthdays different}) \sim e^{-\frac{m(m-1)}{2N}}$$

3.4 Interpreting the Answer

Let's use all this analysis get some concrete answers. If there are $m = 100$ people in a room and $N = 365$ days in a year, then what is the probability that no two have the same birthday? Our upper bound says this is at most:

$$\Pr(100 \text{ are different}) \leq e^{-100 \cdot 99 / (2 \cdot 365)} < 0.0000013$$

So the odds 100 people have different birthdays are around 1 in a million!

In principle, there could be $m = 365$ people in a room, all with different birthdays. However, the probability of that happening by chance is at most:

$$\Pr(365 \text{ are different}) \leq e^{-365 \cdot 364 / (2 \cdot 365)} = e^{-182} < 10^{-79}$$

Not gonna happen!

In fact, our upper bound implies that if there are only $m = 23$ people in a room, then the probability that all have different birthdays is *still less than half*:

$$\Pr(23 \text{ are different}) \leq e^{-23 \cdot 22 / (2 \cdot 365)} < 1/2$$

In other words, a room with only $m = 23$ people contains two people with the same birthday, more likely than not!

3.5 The Birthday Principle

How many people must be in a room so that there's a half chance that two have the same birthday?

Setting our approximate solution to the birthday problem equal to 1/2 and solving for m gives:

$$m \approx \sqrt{(2 \ln 2)N} \approx 1.18\sqrt{N}.$$

This answer (which is asymptotically tight) is called the *birthday principle*:

If there are N days in a year and there are $\sqrt{(2 \ln 2)N}$ people in a room, then there is about an even chance that two have the same birthday.

An informal argument partly explains this phenomenon. Two people share a birthday with probability $1/N$. Therefore, we should expect to find matching birthdays when the number of *pairs* of people in the room is around N , which happens when $\binom{m}{2} = N$ or $m \approx \sqrt{2N}$, which roughly agrees with the Birthday Principle.

The Birthday Principle is a great rule of thumb with surprisingly many applications. For example, cryptographic systems and digital signature schemes must be hardened against “birthday attacks”. The principle also says a hash table with N buckets starts to experience collisions when around $\sqrt{(2 \ln 2)N}$ items are inserted.