Lecture Notes

1 Independence

The topic of the lecture today is *independence.*

Definition An event A is *independent* of an event B if:

$$\Pr(A \mid B) = \Pr(A)$$

In other words, that fact that event B occurs does not affect the probability that event A occurs.

1.1 Examples with Coins

For example, suppose we flip two fair coins. Let A be the event that the first coin is heads, and let B be the event that the second coin is heads. Since the coins are fair, we have $\Pr(A) = \Pr(B) = \frac{1}{2}$. In fact, the probability that the first coin is heads is still $\frac{1}{2}$, even if we are given that the second coin is heads; the outcome of one toss does not affect the outcome of the other. In symbols, $\Pr(A \mid B) = \frac{1}{2}$. Since $\Pr(A \mid B) = \Pr(A)$, we say that event A is independent of event B.

Suppose that we glue the coins together, heads to heads. Now each coin still has probability $\frac{1}{2}$ of coming up heads; that is $\Pr(A) = \Pr(B) = \frac{1}{2}$. But if the second coin comes up heads, then the first coin must be tails! That is, $\Pr(A \mid B) = 0$. Now, since $\Pr(A \mid B) \neq \Pr(A)$, the event A is not independent of event B.

1.2 Abstract Examples

Suppose that events A and B are disjoint, as shown in Figure 1; that is, no outcome is in both events. In the diagram, we see that $\Pr(A)$ is non-zero. On the other hand:

$$\begin{aligned}\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} \\ &= 0\end{aligned}$$

Therefore, $\Pr(A \mid B) \neq \Pr(A)$, and so event A is not independent of event B. The conclusion is that *disjoint* events are not *independent.*

Figure 2 shows an arrangement of events such that A is independent of B. Assume that the probability of an event is proportional to its area in the diagram. In this example, event A occupies the same fraction of event B as of event S, namely $\frac{1}{2}$. Therefore, the probability of event A is $\frac{1}{2}$ and the probability of event A, given event B, is also $\frac{1}{2}$. This implies that A is independent of B.

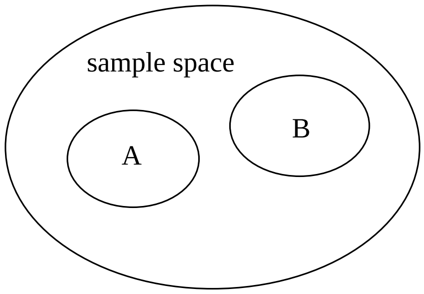

Figure 1: *This diagram shows two disjoint events, A and B. Disjoint events are not independent!*

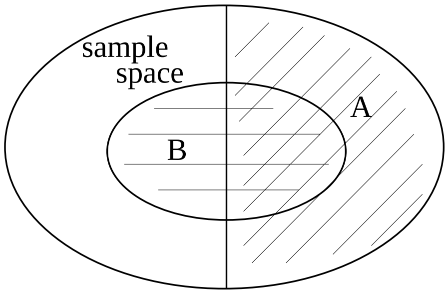

Figure 2: *In this diagram, event A is independent of event B.*

1.3 Symmetry of Independence

As one might expect, independence is a symmetric relation.

Rule 1.1 If event A is independent of B, then B is independent of A.

For this reason, we do not have to say, "A is independent of B" or vice versa; we can just say "A and B are independent".

Proof. By assumption, $\Pr(A \mid B) = \Pr(A)$. We must show that $\Pr(B \mid A) = \Pr(B)$.

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

The first equation uses the Product Rule. In the second equation, we use the Product Rule again to rewrite $\Pr(A \cap B)$. The third equation uses the assumption, $\Pr(A \mid B) = \Pr(A)$. The last step is simplification. ■

1.4 Product Rule for Independent Events

The Product Rule says:

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

If A and B are independent events, then $\Pr(A \mid B) = \Pr(A)$. In this case, the Product Rule simplifies to:

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

This rule is very useful and worth remembering. But also remember that it only holds if A and B are independent events! In fact, many texts *define* two events to be independent if $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$. This is equivalent to our definition, though we will not prove that here.

2 Examples Involving Independence

The first two examples below show that it is not always obvious whether two events are independent or not. The third example shows that carelessness with independence can produce erroneous results.

2.1 An Experiment with Two Coins

Suppose that we flip two independent, fair coins. Let A be the event that the coins match; that is, both are head or both are tails. Let B the event that the first coin is heads. Are these independent events?

At first, the answer may appear to be "no". After all, whether or not the coins match depends on how the first coin comes up; if we toss HH, then they match, but if we toss TH, then they do not.

The preceding observation is true, but irrelevant. The two events are independent if $\Pr(A \mid B) = \Pr(A)$, and we can prove this by the usual procedure.

Claim 2.1 Events A and B are independent.

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

Step 1: Find the Sample Space. The tree diagram in Figure 3 shows that there are four outcomes in this experiment, HH, TH, HT, and TT.

Step 2: Define Events of Interest. As previously defined, A is the event that the coins match, and B is the event that the first coin is heads. Outcomes in each event are marked in the tree diagram.

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

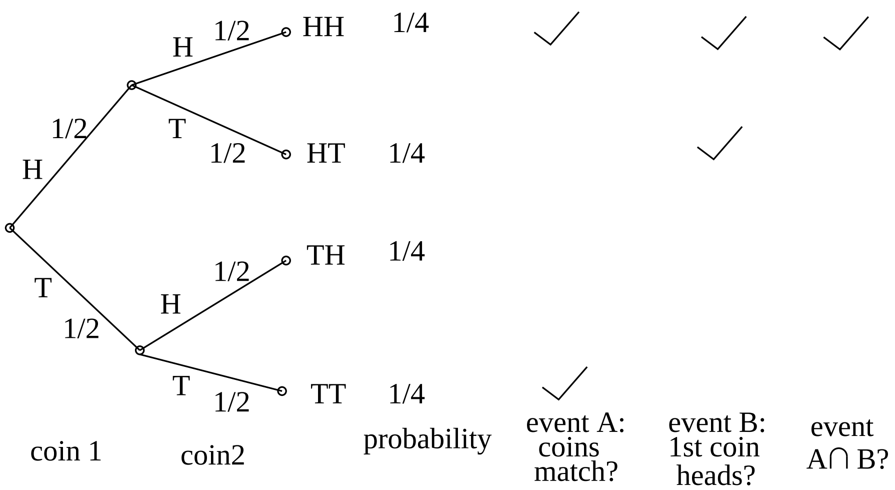

Figure 3: *This is a tree diagram for the two coins experiment.*

Step 4: Compute Event Probabilities.

$$\begin{aligned}
\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} \\
&= \frac{\Pr(HH)}{\Pr(HH) + \Pr(HT)} \\
&= \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{4}} \\
&= \frac{1}{2}
\end{aligned}$$

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

Therefore, $\Pr(A \mid B) = \Pr(A)$, and so A and B are independent events as claimed. ■

2.2 A Variation of the Two-Coin Experiment

Suppose that we alter the preceding experiment so that the coins are independent, but not fair. That is each coin is heads with probability p and tails with probability $1 - p$. Again, let A be the event that the coins match, and let B the event that the first coin is heads. Are events A and B independent for all values of p?

The problem is worked out with a tree diagram in Figure 4. The sample space and events are the same as before, so we will not repeat steps 1 and 2 of the probability calculation.

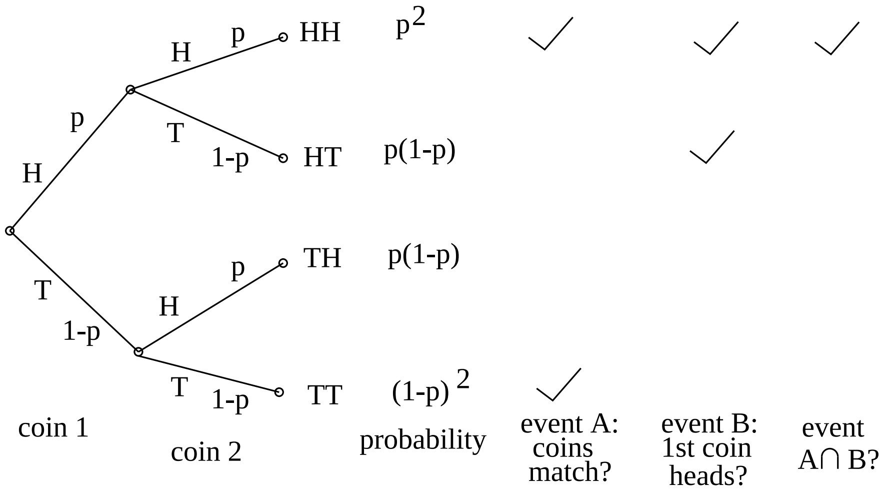

Figure 4: *This is a tree diagram for a variant of the two coins experiment. The coins are still independent, but no longer necessarily fair.*

Step 3: Compute Outcome Probabilities. Since the coins are independent, all edge probabilities are p or $1-p$. Outcomes probabilities are products of edge probabilities on root-to-leaf paths, as shown in the figure.

Step 4: Compute Event Probabilities.

We want to determine whether $\Pr(A \mid B) = \Pr(A)$.

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

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

Events A and B are independent only if these two probabilities are equal:

$$\begin{aligned}
\Pr(A \mid B) &= \Pr(A) \\
\Leftrightarrow \quad p &= 1 - 2p + 2p^2 \\
\Leftrightarrow \quad 0 &= 1 - 3p + 2p^2 \\
\Leftrightarrow \quad 0 &= (1-2p)(1-p) \\
\Leftrightarrow \quad p &= \frac{1}{2}, 1
\end{aligned}$$

The two events are independent only if the coins are fair or if both always come up heads. Evidently, there was some dependence lurking in the previous problem, but it was cleverly hidden by the unbiased coins!

3 An Application to the OJ Trial

Recently, there was a guy named "O. J." on trial for murder[1]. Strangely enough, a probability problem involving independence came up during the trial.

The issue arose when a prosecution witness claimed that only one in 200 Americans has the blood type found at the crime scene. The witness then presented facts something like the following:

- $\frac{1}{10}$ of people have type O blood.
- $\frac{1}{5}$ of people have a positive Rh factor.
- $\frac{1}{4}$ of people have some other special marker.

The one in 200 figure came from multiplying these three fractions! Is the witness reasoning correctly?

The answer depends on whether or not the three blood characteristics are independent. This might not be true; maybe most people with O^+ blood have the special marker. When the special math-competent defense lawyer asked the witness whether these characteristics were independent, he could not say! He could not justify his claim!

4 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? In the O.J. blood example, the witness wanted to multiply probabilites of *three* events; what sort of independence is necessary to justify this maneuver? The notion we need is called *mutual independence*.

[1] "Yeah, right!", you say, "No real person has a name like *O. J.*!" But, honest! True story!

Definition Events $A_1, A_2, \ldots, A_n$ are *mutually independent* if for all i such that $1 \leq i \leq n$ and for all $J \subseteq [1, n] - \{i\}$, we have:

$$\Pr(A_i \mid \bigcap_{j \in J} A_j) = \Pr(A_i)$$

In other words, a collection of events is mutually independent if each event is independent of the intersection of every subset of the others. The following definition is equivalent, though we will not prove this.

Definition Events $A_1, A_2, \ldots, A_n$ are *mutually independent* if for all $J \subseteq [1, n]$, we have:

$$\Pr(\bigcap_{j \in J} A_j) = \prod^{j \in J} \Pr(A_j)$$

Like the general form of the Inclusion-Exclusion principle, these general definitions of mutual independence are a bit tricky. The following special case of the second definition may be easier to follow.

Definition Three events A_1, A_2, A_3 are mutually independent if all of the following hold:

$$\begin{aligned}
\Pr(A_1 \cap A_2) &= \Pr(A_1) \cdot \Pr(A_2) \\
\Pr(A_1 \cap A_3) &= \Pr(A_1) \cdot \Pr(A_3) \\
\Pr(A_2 \cap A_3) &= \Pr(A_2) \cdot \Pr(A_3) \\
\Pr(A_1 \cap A_2 \cap A_3) &= \Pr(A_1) \cdot \Pr(A_2) \cdot \Pr(A_3)
\end{aligned}$$

To prove a set of three or more events mutually independent, it is *not* sufficient to prove every pair of events independent! In particular, for three events we must also prove that the fourth equality listed above holds.

In the O.J. example, if the three blood characteristics were mutually independent, then the witness was justified in multiplying probabilities by the fourth equality above.

4.1 Flipping a Set of Coins

Suppose we flip n fair coins. Let A_i be the event that the i-th coin is heads. The outcome of one coin in unaffected by the outcomes of the others. Therefore, for all i, we have:

$$\Pr(A_i \mid \text{ any set of other events}) = \frac{1}{2} = \Pr(A_i)$$

This implies that the events are mutually independent. The probability of flipping all heads can be found by using the second definition of mutual independence:

$$\Pr(\bigcap_{i=1}^{n} A_i) = \left(\frac{1}{2}\right)^n$$

4.2 A Red Sox Streak

The Boston Red Sox baseball team has lost 14 consecutive playoff games. What are the odds of such a miserable streak?

Suppose that we assume that the Sox have a $\frac{1}{2}$ chance of winning each game and that the game results are mutually independent. Then we can compute the probability of losing 14 straight games as follows. Let L_i be the event that the Sox lose the i-th game. This gives:

$$\begin{aligned}
\Pr(L_1 \cap L_2 \cap \ldots \cap L_{14}) &= \Pr(L_1)\Pr(L_2)\cdot\ldots\Pr(L_{14}) \\
&= \left(\frac{1}{2}\right)^{14} \\
&= \frac{1}{16,384}
\end{aligned}$$

The first equation follows from the second definition of mutual independence. The remaining steps use only substitution and simplification.

These are pretty long odds; of course, the probability that the Red Sox lose a playoff game may be greater than $\frac{1}{2}$. Maybe they're cursed.

4.3 An Experiment with Three Coins

This is a tricky problem that always confuses people! Suppose that we flip three fair coins and that the results are mutually independent. 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 these three events mutually independent?

For once, we will dispense with the tree diagram. The sample space is easy enough to find anyway; there are eight outcomes, corresponding to every possible sequence of three flips: HHH, HHT, HTH, ... We are interested in events A_1, A_2, and A_3, defined as above. Each outcome has probability $\frac{1}{8}$.

To prove that the three events are mutually independent, we must prove a sequence of equalities. It will be helpful first to compute the probability of each event A_i:

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

By symmetry, $\Pr(A_2) = \Pr(A_3) = \frac{1}{2}$. 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, $\Pr(A_1 \cap A_3) = \Pr(A_1)\Pr(A_3)$ and $\Pr(A_2 \cap A_3) = \Pr(A_2)\Pr(A_3)$ must hold as well. We have now proven that every pair of events is independent. But this is not enough to prove that A_1, A_2, and A_3 are mutually independent! We must check the fourth condition:

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

The three events A_1, A_2, and A_3 are not mutually independent, even though all pairs of events are independent! When proving a set of events independent, remember to check all pairs of events, *and* all sets of three events, four events, etc.

4.4 Pairwise Independence

Suppose we have a set of events. We know that all pairs of events are independent, but we know nothing about the independence of subsets of three or more events. This situation arises often enough that there is a special term:

Definition Events $A_1, A_2, \dots A_n$ are *pairwise independent* if A_i and A_j are independent events for all $i \neq j$.

Note that mutual independence is much stronger than pairwise independence. That is, if a set of events is mutually independent, then it must be pairwise independent, but the reverse is not true. For example, the events in the three coin experiment of the preceding section were pairwise independent, but not mutually independent.

In the O.J. example, suppose initially that we know nothing about independence. Then we can only say that the probability that a person has all three blood factors is no greater than the probability that a person has blood type O, which is $\frac{1}{10}$.

If we know that the three blood factors in the O. J. case appear pairwise independently, then we can conclude:

$$\begin{aligned}
\Pr(\text{person has all 3 factors}) &\leq \Pr(\text{person is type } O \text{ and Rh positive}) \\
&= \Pr(\text{person is type } O)\Pr(\text{person is Rh positive}) \\
&= \frac{1}{10} \cdot \frac{1}{5} \\
&= \frac{1}{50}
\end{aligned}$$

Knowing that a set of events is pairwise independent is useful!

However, if all three factors are mutually independent, then the witness is right; the probability a person has all three factors is $\frac{1}{200}$.

The point is that we get progressively tighter upper bounds as we strengthen our assumption about independence.

5 The Birthday Paradox

There are 365 possible birthdays, ignoring leap-days, and there are 120 students in the lecture hall. What is the probability that two students have the same birthday? 50%? 90%? 99%? In fact, the probability is greater than 99.99999997%! There is less than one chance in four billion that everyone has a different birthday!

There are two big assumptions underlying this assertion. First, we assume that all birthdates are equally likely. Second, we assume that birthdays are mutually independent. Neither assumption is really true. Birthdays follow seasonal patterns and are often related to major events. For example, nine months after a blackout in the 70's there was a sudden increase in the number of births in New England. (Try counting back nine months from your birthday!) Nevertheless, we'll stick with these assumptions!

Suppose we perform the following experiment. We ask one student at a time for his or her birthday. How many students must we ask before we find two with the same birthday? Surprisingly, the answer is usually in the mid-twenties. This seems odd! There are 12 months in the year. At a point when we've only collected about two birthdays per month, we have usually already found two students with exactly the same birthday!

Here is the intuition. The probability that a pair of students have the same birthday is only $\frac{1}{365}$. This is very small. But by the time we have a couple dozen birthdays, we have lots of pairs of students! Therefore, the probability that some pair have the same birthday is really pretty good.

In general, suppose there are m students and N days in the year. We want to determine the probability that at least two students have the same birthday. There are at least two good ways to solve this problem. We will show only one here.

Step 1. Find the Sample Space

Finding the sample space with a tree diagram is difficult. Each internal node will have N children, and the tree will have depth m. For the values of N and m of interest, this tree will be huge!

We have to find the sample space a different way. There are N possible birthdays for the first student, N birthdays for the second student, and so on for all m students. In fact, we can regard an outcome as a vector of m birthdays. The sample space is the set of all such vectors:

$$S = \{< b_1, b_2, \ldots, b_m >| \; b_i \in [1, N] \text{ for all } i\}$$

By the Product Rule for the cardinality of a product of sets, the size of the sample space is N^m.

Step 2: Define Events of Interest

Let A be the event that two or more students have the same birthday. That is, event A is the following set of outcomes:

$$\{< b_1, b_2, \ldots, b_m >| \; b_i = b_j \text{ for some distinct } i \text{ and } j\}$$

Step 3: Compute Outcome Probabilities

The probability of outcome $< b_1, b_2, \ldots, b_m >$ is the probability that the first student has birthday b_1 and the second student has birthday b_2 and the third student has birthday b_3, etc. The i-th person has birthday b_i with probability $\frac{1}{N}$. Since birthdates are independent, we can multiply probabilities to get the probability of a particular outcome:

$$\Pr(< b_1, b_2, \ldots, b_m >) = \frac{1}{N^m}$$

In principle, we could obtain the same result from a tree diagram. Since each students is equally likely to have any one of N birthdays, independent of all other students, we would label every edge in the tree with probability $\frac{1}{N}$. Then we would compute the probability of an outcome by multiplying probabilities on the corresponding root-to-leaf path. Since the tree has depth m, we would find that every outcome has probability $\frac{1}{N^m}$, exactly as above.

5.0.1 Digression: Uniform Sample Spaces

Let's digress from the analysis of the birthday problem for a bit to discuss a nice feature that will arise in other problems. The remaining task in the birthday problem is to compute the probability of event A. This is particularly easy because the sample space is *uniform.*

Definition A sample space S is *uniform* or *equiprobable* if $\Pr(s)$ is the same for all $s \in S$.

This definition has the following consequence. Since the sum of the probabilities over all outcomes is 1, we must have $\Pr(s) = \frac{1}{|S|}$ for all outcomes s in a uniform sample space S.

The probability of an event in a uniform sample space is very easy to compute. We only need to count the number of outcomes in the event. This amounts to counting the number of items in a set— we know all about that!

Theorem 5.1 If A is an event in a uniform sample space, then

$$\Pr(A) = \frac{|A|}{|S|}$$

Proof.

$$\begin{aligned} \Pr(A) &= \sum^{a \in A} \Pr(a) \\ &= \sum^{a \in A} \frac{1}{|S|} \\ &= \frac{|A|}{|S|} \end{aligned}$$

The first equation follows because the probability of an event is defined to be the sum of the probabilities of the outcomes it contains. The second equation comes from the observation above that every outcome in a uniform sample space has probability $\frac{1}{|S|}$. The final equation holds because the sum has $|A|$ terms. ■

Now let's return to the birthday problem.

Step 4: Compute Event Probabilities

We want to compute the probability of A, the event that two or more students have the same birthday. Since the sample space is uniform, we need only count the number of outcomes in the event. This can be done with inclusion-exclusion, but the calculation is involved.

A simpler method is to use the trick of "counting the complement". Let $\bar{A}$ be the complementary event; that is, let $\bar{A} = S - A$. Then, since $\Pr(A) = 1 - \Pr(\bar{A})$, we are done if we can determine the probability of event $\bar{A}$.

In the event $\bar{A}$, all students have different birthdays. The event consists of the following outcomes:

$$\{< b_1, b_2, \ldots, b_m >|\ b_i \neq b_j \text{ for all distinct } i \text{ and } j\}$$

The notation above is a little confusing, but we have seen sets like this before. The set $\bar{A}$ consists of all m-permutations of the set of N possible birthdays! We can now compute the probability of event $\bar{A}$:

$$\begin{aligned}\Pr(\bar{A}) &= \frac{|\bar{A}|}{|S|} \\ &= \frac{|\bar{A}|}{N^m} \\ &= \frac{P(N,m)}{N^m} \\ &= \frac{N!}{(N-m)!\ N^m}\end{aligned}$$

The first equation uses Theorem 5.1 to reduce the probability problem to a counting problem. In the second step, we substitute in the size of the sample space S. In the third step, we use the observation that $\bar{A}$ is the set of all m-permutations of an N-element set. In the last step, we substitute the value of $P(N, m)$ that we worked out a few weeks ago.

We now know $\Pr(A)$:

$$\Pr(A) = 1 - \frac{N!}{(N-m)!\ N^m}$$

This is the probability that at least two students in a room of m have the same birthday in a year with N days.

Using this formula, if we have $m = 22$ students and a year with $N = 365$ days, then at least two students have the same birthday with probability $P(A) \approx 0.476$. If we have $m = 23$ students, then the probability rises to $P(A) \approx 0.507$. Therefore, in a room with 23 students, the odds are better than even that at least two have the same birthday!

6 Approximating the Answer to the Birthday Problem

We now know that $\Pr(A) = 1 - \frac{N!}{(N-m)!\ N^m}$. This formula is hard to work with because it is not a closed form. Evaluating the expression for, say, $N = 365$ and $m = 120$ is a lot of work!

There is another reason to want a closed form. We might want to know how many students are needed so that at least two have the same birthday with probability $\frac{1}{2}$. We computed this for a year with $N = 365$ days, but what is the answer as a function of N?

6.1 Approximating $\Pr(\bar{A})$

Unfortunately, there is no closed form for $\Pr(A)$. There is a nice approximation for $\Pr(\bar{A})$ using Stirling's formula, but some ugly math is required to find it! Hold on tight!

$$\begin{aligned}
\frac{N!}{(N-m)!\,N^m} &= \frac{\sqrt{2\pi N}\left(\frac{N}{e}\right)^N e^{a_N}}{\sqrt{2\pi(N-m)}\left(\frac{N-m}{e}\right)^{N-m} e^{a_{N-m}}\,N^m} \\
&= \sqrt{\frac{N}{N-m}}\;\frac{N^N e^{N-m} e^{a_N-a_{N-m}}}{(N-m)^{N-m} e^N N^m} \\
&= \sqrt{\frac{N}{N-m}}\;\frac{e^{a_N-a_{N-m}}}{e^m\left(1-\frac{m}{N}\right)^{N-m}}
\end{aligned}$$

In the expressions above, the symbol a_i denotes some value between $\frac{1}{12i+1}$ and $\frac{1}{12i}$. In effect, a_i captures the error in Stirling's approximation.

The last expression is useful for approximating $\Pr(\bar{A})$. For example, substituting $N = 365$ and $m = 23$ gives approximately:

$$\begin{aligned}
\Pr(\bar{A}) &= 1.033\,\frac{0.99998}{2.097} \\
&= 0.492\ldots
\end{aligned}$$

Substituting $N = 365$ and $m = 120$ gives:

$$\begin{aligned}
\Pr(\bar{A}) &= 1.221\,\frac{0.99999}{5.00 \cdot 10^9} \\
&= 2.442 \cdot 10^{-10}
\end{aligned}$$

The main difference between these two cases is the denominator of the fraction. For $m = 23$, the denominator is about 2. For $m = 120$, it is about 5 billion!

Since the denominator is the main term, let's try to simplify it some more. The first step looks arbitrary; we apply the identity $x = e^{\ln x}$, seemingly at random:

$$e^m\left(1-\frac{m}{N}\right)^{N-m} = e^m\left(e^{\ln\left(1-\frac{m}{N}\right)}\right)^{N-m}$$

The motivation for the preceding step is that there is a nice Taylor series expansion for $\ln(1-x)$:

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots$$

We can now continue by replacing $\ln(1-\frac{m}{N})$ with the Taylor series:

$$
\begin{aligned}
e^m \left(e^{\ln(1-\frac{m}{N})}\right)^{N-m} &= e^m e^{(-\frac{m}{N}-\frac{m^2}{2N^2}-\frac{m^3}{3N^3}-\dots)(N-m)} \\
&= e^m e^{(-m-\frac{m^2}{2N}-\frac{m^3}{3N^2}-\dots)+(\frac{m^2}{N}+\frac{m^3}{2N^2}+\frac{m^4}{3N^3}+\dots)} \\
&= e^{\frac{m^2}{2N}+\frac{m^3}{6N^2}+\frac{m^4}{12N^3}+\dots} \\
&\geq e^{\frac{m^2}{2N}}
\end{aligned}
$$

Now we have simplified the denominator in our expression for $\Pr(\bar{A})$. Substituting the simplified denominator back into this expression gives:

$$\Pr(\bar{A}) \leq \sqrt{\frac{N}{N-m}}\, e^{-\frac{m^2}{2N}}$$

This is the pretty approximation we were aiming for!

6.2 Nice Features of the Formula for $\Pr(\bar{A})$

The formula derived above for $\Pr(\bar{A})$ has several nice features. First, it shows why the probability that all students have distinct birthdays drops off so incredibly fast as the number of students grows. The reason is that the number of students, m, is squared and then exponentiated! Therefore m has a huge effect on the probability.

Second, this formula actually has some intuitive justification! The number of ways to pair m students is $\binom{m}{2} \approx \frac{m^2}{2}$. The event that a pair of students has the same birthday has probability $\frac{1}{N}$. *If these events were independent*, then the probability that no pair of students had the same birthday would be:

$$
\begin{aligned}
\Pr(\text{no pair have same birthday}) &\approx \left(1-\frac{1}{N}\right)^{\frac{m^2}{2}} \\
&\approx e^{-\frac{1}{N}\cdot\frac{m^2}{2}} \\
&= e^{-\frac{m^2}{2N}}
\end{aligned}
$$

(We use the approximation $1-x \approx e^{-x}$.)

This is almost the same answer that we derived above! Of course, this calculation is for intuition's sake only; it is blatantly not correct, since we assumed that events were independent when they are not!

6.3 Making the Probability of a Match $\frac{1}{2}$

A final nice feature of our approximation for $\Pr(\bar{A})$ is that we can determine the number of students for which the probability that two have the same birthday is $\frac{1}{2}$.

First, notice that if $m = o(N^{\frac{2}{3}})$, then our approximation is really very good. The reason is that the major source of error was truncating the Taylor series for a logarithm. The terms we threw out were:

$$\frac{m^3}{6N^2} + \frac{m^4}{12N^3} + \dots$$

Under the assumption that $m = o(N^{\frac{2}{3}})$, these terms all go to zero as N becomes large. Furthermore, under this same assumption the term $\sqrt{\frac{N}{N-m}}$ in our formula tends to 1 as N grows large. Therefore, with this assumption we can conclude:

$$\Pr(\bar{A}) \sim e^{-\frac{m^2}{2N}}$$

All that remains is to set the probability that all birthdays are distinct to $\frac{1}{2}$ and solve for the number of students.

$$\begin{aligned} e^{-\frac{m^2}{2N}} &\sim \frac{1}{2} \\ e^{\frac{m^2}{2N}} &\sim 2 \\ \frac{m^2}{2N} &\sim \ln 2 \\ m &\sim \sqrt{2N \ln 2} \\ &\approx 1.177\sqrt{N} \\ &= \Theta(\sqrt{N}) \end{aligned}$$

We are interested in values of m that are $\Theta(\sqrt{N}) = o(N^{\frac{2}{3}})$; therefore, our earlier assumption is justified and we can expect our approximation to be good!

For example, if $N = 365$, then $1.177\sqrt{N} = 22.49$. This is consistent with out earlier calculation; we found that the probability that at least two students have the same birthday is $\frac{1}{2}$ in a room with around 22 or 23 students!

Of course, one has to be careful with the $\sim$ notation; we may end up with an approximation that is only good for very large values. In this case, though, our approximation works well for reasonable values!

6.4 The Birthday Principle

The preceding result is called the Birthday Principle. It can be interpreted this way: if you throw about $\sqrt{N}$ balls into N boxes, then there is about a 50% chance that some box gets two balls.

For example, in 27 years there are about 10,000 days. If we put about $1.177\sqrt{10,000} \approx 118$ people under the age of 28 in a room, then there is a 50% chance that at least two were born on exactly the same day of the same year!

As another example, suppose we have a roomful of people. Each person writes a random number between 1 and a million on a piece of paper. Even if there are only about $1.177\sqrt{1,000,000} = 1177$ people in the room, then there is a 50% chance that two wrote exactly the same number!