2.4 Computing probabilities in finite sample spaces

Enough theory! Now, let's write down some concrete probability spaces and actually calculate something. To do this, we will make some extra assumptions that reduce our general notion of a probability space down to a special case where we can use the most basic of tools to compute probabilities: counting.

Assumption 1 (A1): finite sample space

Assume the sample space is finite:

$$S = \{s_1, s_2, ..., s_n\}.$$

So we can list out all of the possible outcomes one after another, and eventually we're finished. This already describes many real-world phenomena of interest, not to mention our standby examples: coins, dice, playing cards. To completely specify the probability space, it suffices to assign an individual probability to each of the individual outcomes:

$$\begin{aligned} s_1 &\longleftrightarrow P(\{s_1\}) \\ s_2 &\longleftrightarrow P(\{s_2\}) \\ &\vdots \\ s_i &\longleftrightarrow P(\{s_i\}) \\ &\vdots \\ s_n &\longleftrightarrow P(\{s_n\}). \end{aligned}$$

Naturally, we require $P(\{s_i\}) \geq 0$ and $\sum_{i=1}^{n} P(\{s_i\}) = 1$ so that the axioms are satisfied. Also, $P(\{s_i\})$ is pretty clunky notation. As needed, we might substitute $P(s_i)$, $P(i)$ or p_i. Anyhow, that covers the individual outcomes, but what about $P(A)$ for more general events A?

Because the entire sample space is a finite set, any event $A \subseteq S$ will also be a finite set with $m \leq n$ elements, generically denoted:

$$A = \{s_1^*, s_2^*, ..., s_m^*\}.$$

Because sets are collections of unique objects, the s_i^* are all distinct. As such, we can rewrite A as a disjoint union of **singleton sets** (sets that just have one element in them):

$$A = \{s_1^*, s_2^*, ..., s_m^*\} = \bigcup_{j=1}^{m} \{s_j^*\}.$$

Then countable additivity tells us that

$$P(A) = P(\cup_{j=1}^{m} \{s_j^*\}) = \sum_{j=1}^{m} P(\{s_j^*\}).$$

So the bottom line is this: when you have a finite sample space, the probability of any event A is just the sum of the individual probabilities of the outcomes in A. In other words, the problem of computing probabilities collapses to an adding problem.

Outcomes	Probabilities
$s_1 = 1$	$P(\{s_1\}) = 3/6 = 1/2$
$s_2 = 2$	$P(\{s_2\}) = 1/6$
$s_3 = 3$	$P(\{s_3\}) = 1/6$
$s_4 = 4$	$P(\{s_4\}) = 1/6$

Table 2.1: Probability space for a loaded, four-sided die

Example 2.2. Imagine you have a *loaded*, four-sided die with probabilities summarized in Table 2.1. The event that you roll an odd number is $A = \{1, 3\} = \{1\} \cup \{3\}$, and so the probability of this event is

$$\begin{aligned} P(A) = P(\{1\} \cup \{3\}) &= P(\{1\}) + P(\{3\}) \\ &= \frac{3}{6} + \frac{1}{6} \\ &= \frac{4}{6} \\ &= \frac{2}{3}. \end{aligned}$$

Not exactly earth-shattering, but the point I mean to emphasize is the fact that "add up the individual probabilities" is a consequence of two things in particular: finite sample space and countable additivity. At this stage in the course where we are building everything up from first principles, every step, no matter how small, needs to be clearly justified.

Assumption 2 (A2): equally likely outcomes

If, in addition to a finite sample space, we assume "equally-likely outcomes" (ELO), things simplify further. This assumption just means that every outcome has the same probability. But since these must add up to one, there is only one value that probability could be:

$$P(\{s_i\}) = \frac{1}{n} \quad \forall i = 1, 2, ..., n.$$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Table 2.2: Sums of two six-sided die rolls.

But if that is the case, then we get the following for any $A \subseteq S$:

$$\begin{aligned} P(A) &= \sum_{j=1}^{m} P(\{s_j^*\}) \\ &= \sum_{j=1}^{m} \frac{1}{n} \\ &= \sum_{j=1}^{m} \frac{1}{n} \cdot 1 \\ &= \frac{1}{n} \sum_{j=1}^{m} 1 \\ &= \frac{1}{n} \underbrace{(1 + 1 + \dots + 1)}_{m \text{ times}} \\ &= \frac{m}{n} \\ &= \frac{\text{\# of outcomes in } A}{\text{\# of outcomes in } S}. \end{aligned}$$

So the problem of computing the probability of an event collapses to a counting problem, and the interpretation of this is that “the more ways that an event can occur, the more likely it is.”

Example 2.3. Imagine you have a fair (unloaded) six-sided die. So $S = \{1, 2, 3, 4, 5, 6\}$, which contains six outcomes, and so each outcome has probability 1/6 of occurring. The event that you roll an odd number is $A = \{1, 3, 5\}$, which has three outcomes in it, so $P(A) = 3/6 = 1/2$.

Example 2.4. Imagine we are casting two *fair*, six-sided die and then adding up the numbers on the two faces. The sample space is given in Table 2.2. Since both die are fair, each cell in the table is equally-likely, so if we want to compute the probability of A = “sum is even," we just have to count the total number of outcomes that result in an even sum and divide this by the total number of outcomes. So

$$\begin{aligned} \#(A) &= 18 \\ \#(S) &= 36 \\ P(A) &= 18/36 = 1/2. \end{aligned}$$

Summary

If we have...

- (**A1: finite sample space**) $S = \{s_1, s_2, ..., s_n\}$;
- (**A2: equally-likely outcomes**) $P(\{s_i\}) = 1/n$ for all $i = 1, 2, ..., n$,

then for any event $A \subseteq S$, we have

$$P(A) = \frac{\#(A)}{\#(S)}, \tag{2.4}$$

where $\#(A)$ will be our generic notation for the total number of elements in the set A. So, to compute probabilities in this special case of finite S and ELO, we have to be able to count. How?

2.4.1 The counting principle

When we assume a finite sample space and equally-likely outcomes, the problem of computing probabilities collapses to a counting problem. We will study various techniques for counting the number of elements in potentially elaborately specified sets, but the backbone of all of it will be a very simple principle:

Theorem 2.5. (**The counting principle I**) Suppose each element in your set is uniquely identified by specifying two traits. Suppose there are m options for the first trait, and for each of choice of first trait, there are n options for the second trait. Then the total number of elements in the set is $m \cdot n$.

This is hardly earth-shattering. To visualize the situation, let $\{a_1, a_2, ..., a_m\}$ be the possibilities for trait 1, let $\{b_1, b_2, ..., b_n\}$ be the possibilities for trait 2, and consider arraying things in a matrix:

$$\begin{array}{ccccc} & b_1 & b_2 & \cdots & b_n \\ a_1 & \star & \star & \cdots & \star \\ a_2 & \star & \star & \cdots & \star \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ a_m & \star & \star & \cdots & \star \end{array}$$

If you wanted to count the total number of distinct combinations (a_i, b_j) of two traits, you see that it is just $m \cdot n$.

Example 2.5. Recall Example 2.4 and Table 2.2. We had a pair of dice, each with six faces. To specify the outcome, you need two numbers: the number for the first roll (1-6) and the number for the second roll (1-6). So the total number of outcomes is $6 \times 6 = 36$.

Example 2.6. Consider a standard deck of playing cards without the jokers. To uniquely identify a card, you need to determine its suit (diamond, spade, heart, or club) and its rank (2, 3, 4, ..., 10, Jack, Queen, King, or Ace). There are 4 suits and 13 ranks, so a total number of $4 \times 13 = 52$ cards in the deck.

Remark 2.7 (Traits are not symmetric)**.** Consider the possible letter grades on a midterm and a final. There are 5 letter grades (A,B,C,D,F) possible for each test, so $5 \times 5 = 25$ total pairs of two grades. Note! (B,F) means B on the midterm, F on the final. This is very different from (F,B), which means F on the midterm and B on the final. Thus in the counting principle, the two traits are not "symmetric"; they have *distinct* meanings.

Remark 2.8 (Options for trait 2 can depend on trait 1)**.** Consider the number of strings of two letters which don't repeat, i.e. AB, AC, AD,..., BA, BC, BD,... and so on. If the first letter (i.e. value for trait 1) is A, then the second letter can be B,C,...,Z but not A. If the first letter is B, then the second letter can be A, C, ...,Z but not B. Thus the possible options for trait 2 change *depending on trait 1*. But the counting principle still applies: for each of the 26 options for trait 1, the *number* of options for trait 2 is the same: 25. Thus there are 26×25 elements in this set.

That's two traits, but there's nothing special about two:

Theorem 2.6. (**The counting principle II**) Suppose each element in your set is uniquely identified by specifying p traits. Suppose there are n_1 options for the first trait. For each choice of the first trait, there are n_2 options for the second trait. For each choice of the first and second trait, there are n_3 options for the third. And so on, up to the last trait: for each choice of the first $p-1$ traits, there are n_p options for the pth trait. Then the total number of elements in the set is $n_1 \cdot n_2 \cdot \ldots \cdot n_p$.

Armed with the statement of the counting principle for two traits, you can prove the general result using *mathematical induction*. If you're feeling froggy, give it a shot!

Example 2.7. How many length-n binary strings are there? You can think of each digit or bit in the string as trait, which can take the value 0 or 1:

$$\underbrace{\underline{0 \text{ or } 1}}_{\text{Digit 1}} \quad \underbrace{\underline{0 \text{ or } 1}}_{\text{Digit 2}} \quad \cdots \quad \underbrace{\underline{0 \text{ or } 1}}_{\text{Digit } n}$$

Then the counting principle tells us that the total number of combinations of n traits is

$$\underbrace{2 \times 2 \times \ldots \times 2}_{n \text{ times}} = 2^n.$$

This observation will be generally useful. Consider for instance the random phenomenon of flipping a coin three times in a row. The sample space S is the set of triples like HHT or THT. How many outcomes are there? In this case, it isn't so bad just to list them out and count by hand:

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

So we see that $\#(S) = 8$, which could be a handy bit of information when it comes down to computing probabilities (assuming the coin is fair). But if we were flipping the coin more than three times, this would get tedious. A sequence of H/T flips is really just a binary string, so without enumerating the sample space, we could have calculated that $\#(S) = 2^3 = 8$.

Example 2.8. Consider strings of three letters, none of which repeat, e.g ABC, ABD, and so on. Analogously to Remark 2.8, there are 26 options for the first letter, 25 for the second, and 24 for the third. So in total, we have $26 \times 25 \times 24$ options.

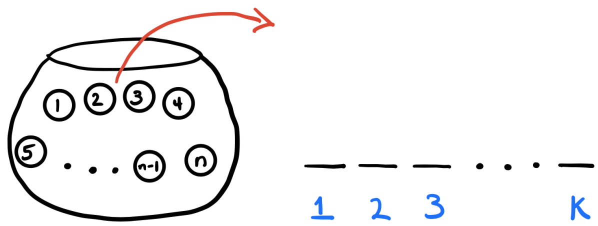

Figure 2.5

	order matters	order does not matter
without replacement	$P_{n,k} = \frac{n!}{(n-k)!}$	$C_{n,k} = \underbrace{\binom{n}{k} = \frac{n!}{k!(n-k)!}}_{\text{“}n\text{ choose }k\text{”}}$
with replacement	n^k	$\binom{n+k-1}{k}$

Table 2.3: How many ways can you select k from n?

2.4.2 Selecting k from n

Counting the number of elements in a set will often take the form of a "selecting k from n" counting problem. Imagine you have an urn with n balls in it. Your task is to withdraw balls from the urn in order to fill $k \le n$ empty slots, as in Figure 2.5.

The question we wish to answer is "how many ways can I fill the k slots?" Before you can answer that question, there are two issues you have to sort out:

1. (**with or without replacement?**) after I draw a ball from the urn to fill a slot, do I put it back in the urn so that I can potentially draw it again to fill another slot? Or after I draw a ball, is it "out of play" forevermore?

2. (**order does or does not matter?**) Say $k = 2$. I could draw (Ball 1, Ball 3) or (Ball 3, Ball 1). Do I count these as two separate outcomes because the order is different, or do I consider them the same outcome because the contents are identical? So, do I ignore or acknowledge the ordering of the draws when I count?

Depending on how you mix-and-match these two settings, you will get a different answer to "how many ways can I fill the k slots?" These are summarized in Table 2.3.

Example 2.9. Say that our jar contains $n = 3$ balls labeled a, b, and c, and we are selecting balls from the jar to fill $k = 2$ slots. How many ways can this be done? It depends on order and replacement. Table 2.4 enumerates all of the possible selections in each of the four cases and counts. We see that the number of possibilities changes in accordance with the formulas in Table 2.3.

So, in a lil' baby example, we are able to literally list out all of the possibilities and count by hand, and the counts we get jive with the formulas in Table 2.3. But where did these formulas come from?

Replace + Order +	Replace - Order +	Replace + Order -	Replace - Order -
aa		aa	
ab	ab	ab/ba	ab/ba
ac	ac	ac/ca	ac/ca
ba	ba		
bb		bb	
bc	bc	bc/cb	bc/cb
ca	ca		
cb	cb		
cc		cc	
$3^2 = 9$	$3! = 6$	$\binom{4}{2} = 6$	$\binom{3}{2} = 3$

Table 2.4: We are filling $k = 2$ spots by drawing from the set $\{a, b, c\}$, which contains $n = 3$ elements. The table enumerates all of the possible selections, depending on whether or not we draw with replacement and whether or not order matters. In each case, the total number of possible selections is different.

With replacement; order matters

This is just the counting principle. Think of each slot as an "experiment." There are n ways I can fill the first slot. But it's with replacement, so when I move onto the second slot, there are still n ways it can be filled. Same with the third slot, and the fourth, etc. So the total number of ways I can fill the k slots is

$$n \times n \times n \times ... \times n = n^k.$$

Without replacement; order matters

This is still the counting principle, but with one wrinkle. When you go to fill the first slot, there are n things you can put there. Then you move on to the second slot. Because we are drawing *without* replacement, there are now only $n - 1$ things you can put there. Then you move on to the third slot. There are only $n - 2$ things you can put there. And so on. Each slot is still an experiment in the sense of the counting principle, but unlike the previous case, each experiment in this case has a different number of options. But at the end of the day, you still multiply them:

$$n \times (n-1) \times (n-2) \times ... \times (n-k+2) \times (n-k+1)$$

That formula is the numerically correct answer, but it looks like hell, and we can clean it up. Recall:

$$n! = \underbrace{n \times (n-1) \times ... \times (n-k+2) \times (n-k+1)}_{\text{our answer from above}} \times \underbrace{(n-k) \times (n-k-1) \times ... \times 2 \times 1}_{(n-k)!}.$$

So a concise way of write the answer in this case is

$$\frac{n!}{(n-k)!}.$$

This number gets a special name, $P_{n,k}$.

Remark 2.9 (Permutations)**.** Say we are drawing without replacement and acknowledging order with $n = k$. So there are just as many objects as slots. This means that by the end of the selection process, all k objects will eventually be drawn, it's just a question of the order. Our formula says that the total number of ways this can be done is $k!/(k-k)! = k!/0! = k!/1 = k!$, so $k!$ is the total number of ways that you can re-order or *permute* k objects. For example, if you have a playlist with 13 songs on it, then there are $13! = 6,227,020,800$ ways you could shuffle it. If Spotify shuffle were truly random, all 13! of these orderings would be equally likely each time you re-press the shuffle button. That is *definitely not* what Spotify does.

Without replacement; order does not matter

Let us momentarily continue thinking about the previous case. If you select k from n without replacement but acknowledging order, a given selection consists of two features:

1. What were the raw contents?
2. How were the contents ordered?

If you know both of those things, then you know what selection was made. Thus these two features are traits, uniquely identifying the selection. We can therefore apply the counting principle: the total number of ways you can select without replacement but acknowledging order is the product of the total number of ways you can select the raw contents and the total number of ways you can re-order the k things you selected. But we already have formulas for most of those counts:

$$\begin{aligned}
\#\begin{pmatrix}\text{without replacement}\\ \text{order matters}\end{pmatrix} &= \#\begin{pmatrix}\text{without replacement}\\ \text{order irrelevant}\end{pmatrix} \times \#\begin{pmatrix}\text{permutations of}\\ \text{the } k \text{ selected objects}\end{pmatrix}\\
\frac{n!}{(n-k)!} &= \#\begin{pmatrix}\text{without replacement}\\ \text{order irrelevant}\end{pmatrix} \times k!.
\end{aligned}$$

Rearranging, we can solve for the count in the *order irrelevant case*:

$$\#\begin{pmatrix}\text{without replacement}\\ \text{order irrelevant}\end{pmatrix} = \frac{n!}{k!(n-k)!}.$$

As it happens, the number we wrote down is important enough that it gets its own special name and shorthand notation:

Definition 2.2. The **binomial coefficient** is the number

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{2.5}$$

We read $\binom{n}{k}$ as "n choose k." The R command `choose(n, k)` computes it.

The binomial coefficients get their name because of their role in the so-called **binomial theorem**:

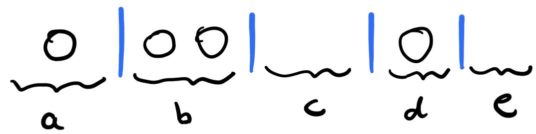

Figure 2.6

Theorem 2.7. If $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$, then

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \tag{2.6}$$

Why is this true? First note that expanding out $(x+y)^n$ means choosing all possible combinations of an x or a y from the first factor, an x or a y from the second factor, and so on. By the counting principle, there are a total of 2^n terms. For example, there are $8 = 2^3$ terms in $(x+y)^3$:

$$\begin{aligned}(x+y)^3 =&(x+y)(x+y)(x+y)\\ =&x \cdot x \cdot x + x \cdot x \cdot y + x \cdot y \cdot x + x \cdot y \cdot y\\ &+ y \cdot x \cdot x + y \cdot x \cdot y + y \cdot y \cdot x + y \cdot y \cdot y.\end{aligned}$$

This is just a generalization of the FOIL method (firsts, outsides, insides, lasts) from algebra, for computing $(a+b)(c+d)$.

We now have to group together like terms. For example, $x \cdot y \cdot y + y \cdot x \cdot y + y \cdot y \cdot x = 3x^2 y$. We therefore see that the coefficient in front of $x^k y^{n-k}$ is the number of terms in the expansion of $(x+y)^n$ which contain k copies of x and $n-k$ copies of y. But there are $\binom{n}{k}$ such terms: just choose k of the n factors to take an x from; the other $n-k$ factors automatically get assigned a y.

With replacement; order does not matter

To get a feel for this scenario, consider $n = 5$ and $k = 4$. Let's label the 5 objects as a,b,c,d,e. Here are a few examples of a choice of four: aaab, aacc, bbdde. As you can see, the different selections differ from one another in how many of each letter there are. In aaab, we have 3 a's, 1 b, 0 c's, 0 d's, and 0 e's. In bbdde, we have 0 a's, 2 b's, 0 c's, 2 d's, and 1 e.

This setting is the same as putting 4 balls in 5 bins: an 'a' bin, a 'b' bin, a 'c' bin, a 'd' bin, and an 'e' bin. The number of balls in each bin tells us how many of each letter to take. For example, Figure 2.6 corresponds to the selection abbd. In Figure 2.6, notice that we separated the 5 bins from one another with 4 blue dividers. Now imagine we have 8 slots in a row. Choose 4 of them to put dividers in, and put balls in the other 4. Each way to put up the 4 dividers corresponds to a different way of putting the 4 balls in 5 bins! See Figure 2.7 for two example arrangements. There are a total of $\binom{8}{4}$ ways to do this.

Now consider the general case of choosing n from k with replacement, where order doesn't matter. We need n bins to count how many of each of the n items we have chosen. Thus we need $n-1$ dividers to delineate the bins from one another. We use k balls to represent the k total items we choose. We have $\binom{n-1+k}{n-1} = \binom{n-1+k}{k}$ ways to arrange the k balls and $n-1$ dividers. This is precisely the number in the bottom right of Table 2.3.

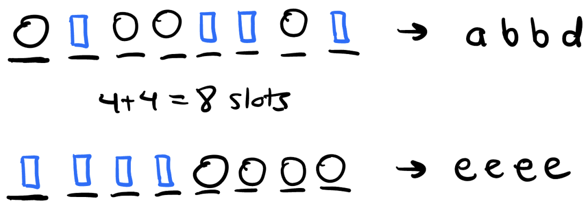

Figure 2.7

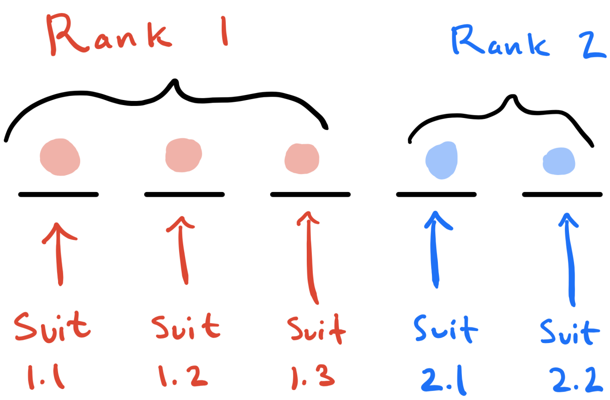

Figure 2.8

2.4.3 Examples

Example 2.10. (**Probability of a full house**) If a five-card hand is randomly dealt to you from a standard deck of 52 playing cards, what is the probability that you are dealt a *full house*: 3 cards of one rank and 2 cards of another rank? In words, our sample space and our event are S = "the set of all possible five-card hands" and A = "your hand is a full house." We have a deck with only 52 cards, and assuming it is adequately shuffled, no card is favored over any other, so we have a finite sample space with equally-likely outcomes. Therefore, we will ultimately apply $P(A) = \#(A)/\#(S)$. To compute $\#(S)$, we have to count all of the ways that five cards could be selected from a deck of 52. So we are "selecting $k = 5$ from $n = 52$." But does order matter, and are we drawing with replacement?

- We are sampling **without replacement**. Once a card is dealt to you, it stays in your hand, it does not go back in the deck to potentially be drawn again;
- **Order does not matter**. The only thing distinguishing one hand from another hand is the raw contents. The order in which the cards were dealt to you is irrelevant.

So, we are selecting $k = 5$ from $n = 52$ without replacement when order does not matter. This means that $\#(S) = \binom{52}{5}$. Great! Now what about $\#(A)$? To specify a full house, we need to determine four traits:

- Rank 1 for three of the cards;
- Rank 2 for two of the cards;

- The suits of the Rank 1 cards;
- The suits of the Rank 2 cards.

This is pictured in Figure 2.8. By the counting principle, we just have to multiply together the number of options for each trait, i.e. the total number of choices for Rank 1, Rank 2, the suits for the Rank 1 cards, and the suits for the Rank 2 cards. So:

- There are 13 ways to choose Rank 1;
- Rank 2 must be distinct from Rank 1, so once Rank 1 is determined, there are 12 ways to choose Rank 2;
- Among the Rank 1 cards, we are drawing without replacement and order is irrelevant, so there are $\binom{4}{3} = 4$ ways to choose the three suits;
- Among the Rank 2 cards, we are drawing without replacement and order is irrelevant, so there are $\binom{4}{2} = 6$ ways to choose the two suits.

Given this, the counting principle tells us that the total number of ways to specify a full house is

$$\#(A) = 13 \cdot 12 \cdot 4 \cdot 6 = 3744$$

So the probability of interest is:

$$P(A) = \frac{\#(A)}{\#(S)} = \frac{3744}{\binom{52}{5}} \approx 0.00144.$$

Example 2.11. (**Probability of a full house, symmetric case**) Now suppose a *four*-card hand is randomly dealt to you from a standard deck of 52 playing cards, and we define “full house" to mean that 2 cards are of one rank, and 2 cards are of the other rank. What is the probability of a full house in this case?

Analogously to Example 2.10, the denominator is $\#(S) = \binom{52}{4}$. But the logic for $\#(A)$ changes. This is because Rank 1 and Rank 2 are *not distinguishable traits*. In Example 2.10, we could indeed distinguish the two: Rank 1 is the one with 3 cards, and Rank 2 is the one with 2 cards. But now we have 2 cards for each of the ranks. So instead of #(Rank 1) × #(Rank 2), we count number of pairs of two ranks. Thus we get

$$\#(A) = \binom{13}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} = 2808,$$

and *not* $13 \cdot 12 \cdot \binom{4}{2} \cdot \binom{4}{2}$. The probability of interest is:

$$P(A) = \frac{\#(A)}{\#(S)} = \frac{2808}{\binom{52}{4}} \approx 0.01037.$$

Example 2.12. (**The birthday problem**) Imagine that k random people convene for a birthday party, and consider two related questions:

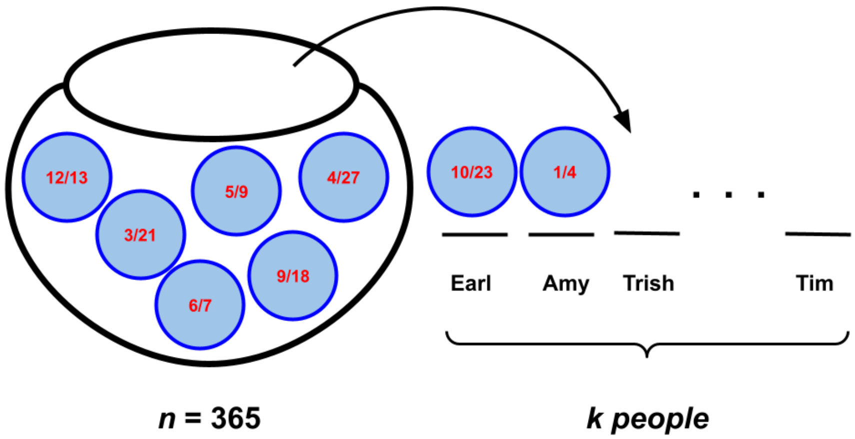

Figure 2.9

- What is the probability that *at least* 2 of the attendees share a birthday?
- How large does the party need to be for the above probability to be at least 50%?

We begin by making a few assumptions:

- There are 365 days in the year (so we ignore leap day);
- A random attendee is equally likely to have any of the 365 days as their birthday (this is empirically false, but oh well);
- The party is not being held at a popular restaurant or a laser tag place or any other such place where you would expect many birthday guys and gals to descend at the same time;
- The attendees are *independent*. If two folks have the same birthday, it is pure coincidence.

We will ultimately be invoking $P(A) = \#(A)/\#(S)$, and our event A is "*at least* 2 of the attendees share a birthday," so what is the sample space S? It is the set of all possible assignments of birthdays to k people. Figure 2.9 visualizes what we have in mind. We have k people (slots) at this party, and their birthdays have been randomly assigned to them as if we were drawing them from the proverbial urn; each ball is one of the possible days of the year. So to compute $\#(S)$, we are selecting k from $n = 365$. But how?

- This is **with replacement** because two people could have been born on the same day. Indeed, if we ruled this possibility out, there would be nothing for us to study. So when a birthday comes out, it goes back in the jar to potentially be drawn again;
- **Order matters**. In Figure 2.9 we have labeled the slots with the names of the unique attendees. This makes each slot distinct, and the order in which the balls (dates) are drawn determines which person receives which date. We consider Trish on 9/3 and Tim on 4/30 a different outcome from Trish on 4/30 and Tim on 9/3, and so order matters here.

Figure 2.10

k	$P(A)$	$\binom{k}{2}$
1	0.0	0
2	0.0027	1
5	0.027	10
10	0.116	45
15	0.2529	105
23	0.5072	253
32	0.753	496
56	0.988	1540
60	0.994	1770
100	0.9999	4950
$\vdots$	$\vdots$	$\vdots$
366	1.0	66795

Table 2.5

With that, we know from Table 2.3 that $\#(S) = 365^k$. So what is $\#(A)$? Start by registering that "*at least* 2" is spooky language. There are several (too many) ways that *at least two* people could share a birthday. Let's use the complement rule to simplify:

$$P(A) = 1 - P(A^c) = 1 - \frac{\#(A^c)}{\#(S)} = 1 - \frac{\#(A^c)}{365^k}.$$

A^c is "no one shares a birthday," which is much simpler to count. How many ways can we assign birthdays to k people such that no one shares? Well, order continues to matter, but now we are drawing *without* replacement. Once a birthday is assigned, it will not be assigned again, thus ensuring that there are no duplicates. As such, $\#(A^c) = 365!/(365-k)!$, and so

$$P(A) = 1 - P(A^c) = 1 - \frac{365!}{365^k(365-k)!}.$$

Table 2.5 displays this probability for various values of k, the number of partygoers. When $k = 1$, then $P(A) = 0$ because that one person has no one they could match with. By the time $k = 366$ attend, at least one match is guaranteed as a consequence of the *pigeonhole principle*. Imagine 365 people are in attendance, and by some miracle there are no birthday matches; every person has a unique birthday. As soon as a 366th person arrives at the party, their birthday *must* match with someone. All of the possible days of the year have already been exhausted.

So those are the extreme cases of $k = 1$ and $k = 366$. But newcomers tend to be surprised by what happens when $1 < k < 366$. It only takes $k = 23$ for a 50% chance of at least one match, and by the time $k = 60$, it's all but guaranteed. Folks often guess that you need a large number of attendees (on the order of hundreds) for the probability of a match to be high, but what they fail to realize is that every unique pair of people is a fresh opportunity for a match. Table 2.5 shows the number of unique pairs for each k, and while k may be small, the number of pairs gets large pretty quickly, making at least one match more and more likely.

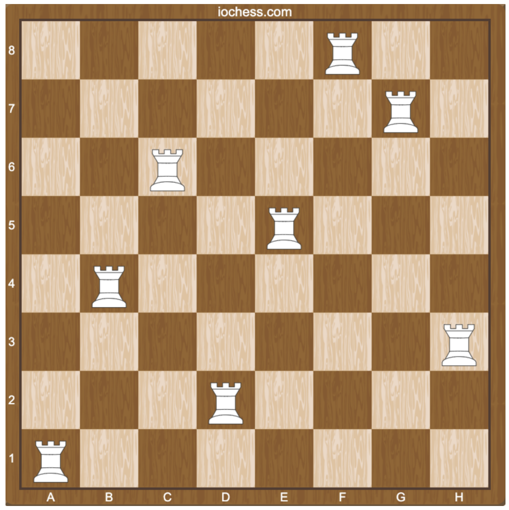

Figure 2.11: Eight safe rooks.

Example 2.13. (**Eight rooks**) Imagine we take eight identical rooks and randomly place them on *distinct* squares on an 8×8 chess board. What is the probability that all of the rooks are "safe" from one another? Figure 2.11 visualizes an example of this. Recall that rooks can move up and down or side-to-side, but not diagonally. In order to be safe from one another, two rooks therefore need to occupy different rows *and* columns of the board. For the eight rooks to be mutually safe, they all must be in different columns and different rows from one another. This is a job for counting. There are a finite number of ways to place the rooks on the board, and all placements are equally likely, so we must count the total number of placements and the total number of *safe* placements. There are $n = 8 \cdot 8 = 64$ squares on the board, and we are selecting $k = 8$ of them at which to place the rooks. The rooks are identical, so order doesn't matter, and we have $\#(S) = \binom{64}{8}$. The event we care about is $A =$ "the rooks are safe from one another," so how many ways can this happen? We know that a safe placement will have every row and column occupied by exactly one rook, so let us imagine building up a safe placement row-wise, starting from the bottom and working upward. There are 8 ways to place a single rook on the bottom row. No other rooks can go there, so we move to the next row. There are only 7 ways to place the next rook in the next row, because we must avoid the column occupied by the first rook. Similarly, there are only 6 ways to place the next rook in the next row, because we must avoid the two columns occupied by the first two rooks. And so on. As a consequence of the counting principle, where we regard each row as an experiment, we see that $\#(A) = 8 \times 7 \times 6 \times \cdots \times 2 \times 1 = 8!$. So $P(A) = 8!/\binom{64}{8} \approx 0.0000091095$.