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 2.1. 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 $A \subseteq S$:

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

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 2.2. 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 2.3. 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 2.6. (**The counting principle I**) If Experiment 1 can result in m possible outcomes, and Experiment 2 can result in n possible outcomes, then the total number of outcomes for both experiments jointly is $m \cdot n$.

This is hardly earth-shattering. To visualize the situation, let $\{a_1, a_2, ..., a_m\}$ be the outcomes of Experiment 1, let $\{b_1, b_2, ..., b_n\}$ be the outcome of Experiment 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 (a_i, b_j) pairs that can be formed from the outcomes of the two experiments, you see that it is just $m \cdot n$.

Example 2.4. Recall Example 2.3 and Table 2.2. We had a pair of dice, each with six faces. So the total number of outcomes is $6 \times 6 = 36$. Each die is like an *experiment* in the sense of the counting principle.

Example 2.5. Consider a standard deck of playing cards without the jokers. There are four suits: diamonds, spades, hearts, clubs. There are thirteen ranks: Ace, 2, 3, 4, ..., 10, Jack, King, Queen. Thinking of these as *experiments*, the counting principle says that the total number of ways you can mix-and-match the ranks and suits is $4 \times 13 = 52$, which we know is the total number of cards in the deck.

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

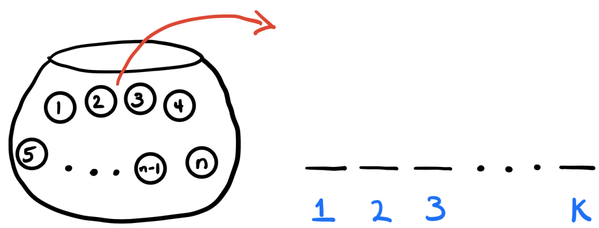

Figure 2.5

Theorem 2.7. (**The counting principle II**) If p experiments are performed, each with n_i outcomes, then the total number of outcomes across all p experiments is $n_1 \cdot n_2 \cdot ... \cdot n_p$.

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

Example 2.6. How many length-n binary strings are there? You can think of each digit or bit in the string as an "experiment" with two possible outcomes, 0 or 1:

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

Then the counting principle tells us that the total number of ways that those outcomes can be combined to form a length-n string is

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

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 $\#(S) = 2^3 = 8$.

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 $k \leq n$ 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:

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

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

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.

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 2.7. 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?

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

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

Remark 2.7. 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 n 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 $n!/(n-n)! = n!/0! = n!/1 = n!$, so $n!$ is the total number of ways that you can re-order or *permute* n 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. Think of those two features as experiments in the sense of 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:

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

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

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

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**:

Theorem 2.8. If $x,\ y \in \mathbb{R}$ and $n \in \mathbb{N}$, then

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

With replacement; order does not matter

Left as an exercise for the reader. ☺