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 $m \cdot n$. That's two experiments, but there's nothing special about two:

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 2.4. 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:

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

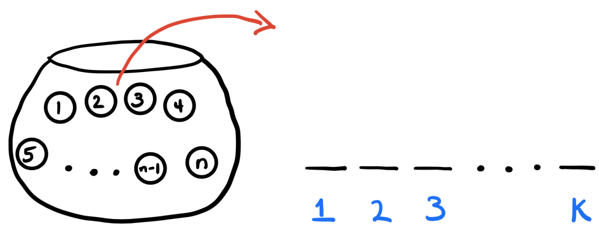

Figure 2.5

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

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

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

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?" This is summarized in Table 2.3.

	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?

With replacement; order matters

See the video for now.

Without replacement; order matters

See the video for now.

Without replacement; order does not matter

See the video for now.

With replacement; order does not matter

Don't worry about it ☺

2.4.3 Worked examples, including the birthday problem

Example 2.5. (**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; 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 $\#(S) = \binom{52}{2}$. Great! Now what about $\#(A)$? To specify a full house, we need to determine four features:

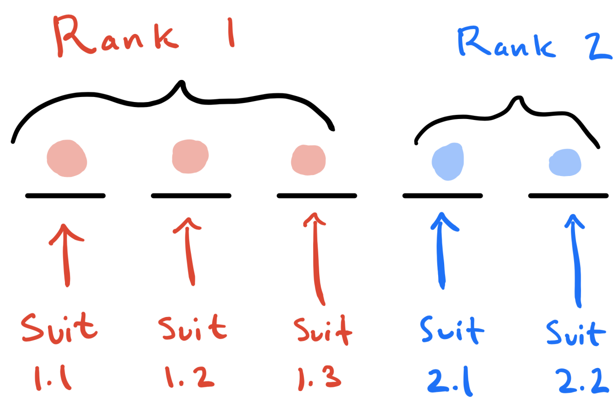

Figure 2.6

- 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 2.6. We can think of each of these features as an "experiment" in the sense of the counting principle. So once we count the total number of ways Rank 1 can be selected, and Rank 2 can be selected, and the Rank 1 suits can be selecting, and the Rank 2 suits can be selected, we just have to multiply them together. 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 2.6. (**Probability that a pair has the same suit**) If a two-card hand is randomly dealt to you from a standard deck of 52 playing cards, what is the probability that you are dealt two cards of the same suit? In words, our sample space and our event are S = "the set of all possible two-card hands" and A = "your two cards are the same suit." Similar to the last example, we are sampling without replacement and order doesn't matter, so $\#(S) = \binom{52}{2} = 1326$. The way to proceed to is realize that we can rewrite the event A as a disjoint union. The event that "your two cards are the same suit" is the same as saying "you get a pair of hearts OR you get a pair of diamonds OR you get a pair of spades OR you get a pair of clubs." A hand could not be both a pair of clubs and a pair of spades, so this is a disjoint union

$$A = \{\text{pair } \heartsuit\} \cup \{\text{pair } \diamondsuit\} \cup \{\text{pair } \spadesuit\} \cup \{\text{pair } \clubsuit\},$$

and the axiom of countable additivity tells us that

$$P(A) = P(\text{pair } \heartsuit) + P(\text{pair } \diamondsuit) + P(\text{pair } \spadesuit) + P(\text{pair } \clubsuit).$$

Furthermore, it pays to recognize that there are the same number of cards of each suit in the deck. Because outcomes are equally likely, no suit is favored, so whatever they are, the summands in the above expression are all equal. So we only have to compute one of them:

$$P(A) = 4 \cdot P(\text{pair } \heartsuit).$$

So, what is the probability of being dealt a pair of hearts? Well, there are 13 hearts in the deck, and we are selecting two of them without replacement and ignoring order. So $\#(\text{pair } \heartsuit) = \binom{13}{2} = 78$, and the probability of interest is

$$P(A) = 4 \cdot \frac{78}{1326} \approx 0.059.$$

Example 2.7. (**the birthday problem**)

See the video for now.

2.5 Conditional probability

2.6 Independent events

2.7 Summary of probability fundamentals