Math 130A — Probability I

Neil Donaldson

Spring 2022

1 Combinatorial Analysis

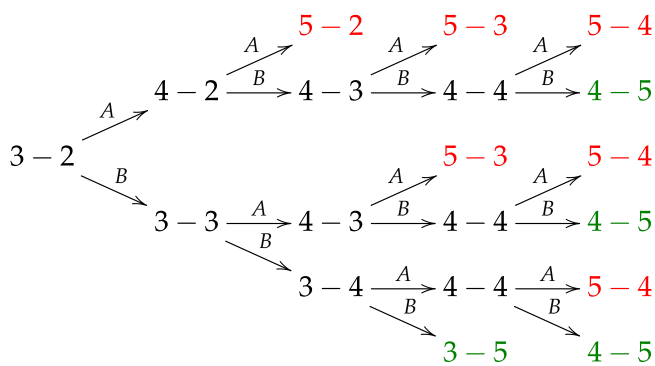
1.1 A Little History & A Lot Of Counting

Modern probability evolved from basic counting and gambling problems. For thousands of years we have asked questions regarding the expected outcome of experiments/activities (How much grain will a field produce? How much can we expect to sell it for?). Indeed modern probability was essentially founded in 1654 as a result of a simple gambling problem discussed by Blaise Pascal and Pierre de Fermat, a version of which we now discuss.

Example 1.1. Two players stake \$100 in a first-to-five coin flip with the winner claiming the full \$200. The game is forced to end early after only 6 rounds, with player A ahead 3–2. How should the stakes be split in a fair manner?

It would be unreasonable for each player to claim their original \$100, since player A is clearly ahead and therefore *more likely* to win. But *how likely*? This problem lead to the idea of an *expected outcome* and to a rigorous measurement and computation of probabilities.

To attack the problem we consider what might have happened if the game continued uninterrupted.



The red outcomes indicate a win for player A and the green for player B. A first thought might be that player A wins 6 out of 10 outcomes and so has a 60% chance of winning the game given the initial position. This doesn't account for the fact that the final outcomes are not equally likely: any outcome in the third column has a $\frac{1}{2^2}$ chance of happening (player A wins 5–2 in 25% of cases), in the fourth $\frac{1}{2^3}$ and in the fifth $\frac{1}{2^4}$. The correct observation is that player A has a

$$\frac{1}{2^2} + \frac{2}{2^3} + \frac{3}{2^4} = \frac{11}{16}$$

chance of winning the game and should therefore claim $200 \cdot \frac{11}{16} = \$137.50$.

The value $\frac{11}{16}$ represents the *expectation* player A has of victory given the game state 3–2. Otherwise said, if the 'finish the game' experiment were run 16 times, we'd expect player A to win *approximately* 11 times. The expectation, or average outcome, is the most important piece of data we could find; the application of probability is geared towards the calculation and estimation of this. By considering probability *distributions*, we also intend to give rigorous meaning to the word 'approximately.'

Even our simple problem is a little tricky, and becomes trickier still when generalized: if player A needs a points to win and player B needs b points, how should the stakes be split? The problem at least illustrates a general approach: *count* the possible outcomes. Here is a simpler problem.

Example 1.2. The first four cards dealt from a standard pack of 52 cards are hearts. There remain 48 cards, of which 9 are hearts. The probability that the next card is a heart is therefore $\frac{9}{48} = 18.75\%$.

Representing the probability as a fraction provides a natural interpretation:

1. From the 48 remaining cards, deal the next card and record whether we have a heart. Return the dealt card to the pack and shuffle.
2. Repeat step 1 a *very large* number of times.

Over the course of the experiment, we *expect* roughly 9 out of every 48 trials to result in a heart.

Note again how *counting* is at the core of the problem: we count how many cards remain and, of these, how many are hearts. Now extend the example.

Example 1.3. The first four cards dealt from a standard pack of 52 cards are hearts. What is the probability that the next *two* cards are *both* hearts?

Again we *count*:

- There are 48 possibilities for the first dealt card; for each, there are 47 possibilities for the second. In total we could deal $48 \cdot 47 = 2256$ distinct (ordered) pairs of cards.
- There are 9 ways for the first dealt card to be a heart; for each card, there remain 8 possible ways to deal a second heart. In total we could deal $9 \cdot 8 = 72$ distinct (ordered) pairs of hearts.
- Putting it together, the probability of dealing two further hearts is $\frac{9 \cdot 8}{48 \cdot 47} = \frac{3}{94} \approx 3.2\%$

We solved the problem by counting how many ways we could do the thing we wanted and dividing by the number of possible outcomes. In both cases we relied on a fundamental principle:

Fundamental Principle of Counting *We conduct k experiments. The first has n_1 possible outcomes. For each outcome of the first experiment, the second has n_2 possible outcomes, etc. Then the number of distinct outcomes of the k experiments is the product $n_1 n_2 \cdots n_k$.*

Example 1.4. You are asked to choose a simple five-character code according to the following rules:

- The first two entries are distinct letters.
- The last three entries are distinct digits $0, \ldots, 9$.

According to the fundamental principle, there are $26 \cdot 25$ ways to choose the letters, and $10 \cdot 9 \cdot 8$ ways to choose the numbers. The total number of possible codes is then

$$26 \cdot 25 \cdot 10 \cdot 9 \cdot 8 = 468,000$$

1.2 Permutations and Combinations

When counting objects we often do not care what *order* they come in.

Definition 1.5. A *permutation* of a set is any ordering of the elements of that set.

Example 1.6. You wish to taste three flavors of ice-cream: Chocolate, Rocky Road & Strawberry. You have a choice of *six* different orders in which to taste all three:

$$CRS,\ CSR,\ RCS,\ RSC,\ SCR,\ SRC$$

Each of the six orderings is a *permutation* of the set of three flavors. This fits with our counting principle: there are $n_1 = 3$ choices for the first flavor, after which there remain $n_2 = 2$ flavors, after which remains only $n_3 = 1$ choice for the final flavor; the number of distinct outcomes (permutations) is therefore $n_1 n_2 n_3 = 3 \cdot 2 \cdot 1 = 6$

The general statement is immediate from the counting principle.

Theorem 1.7. *A set of n elements has $n! = n(n-1)\cdots 2 \cdot 1$ permutations.*

Read the question carefully when several of the elements are, or appear to be, identical.

Examples 1.8.

1. A group of eight children run a race. How many possible outcomes could there be if no two children finish at the same time?

   We have $8! = 40,320$ possible race outcomes.

2. The same eight children have three under-11's and five who are 11 or older. If each of these subsets are ranked only among themselves, how many possible outcomes are there?

   This time we have $3! = 6$ outcomes for the under 11 race and $5! = 120$ for the 11^+: the total number of race outcomes is $3!5! = 720$ (note that the product is in line with the fundamental counting principle!).

3. Finally, the eight children race as teams; under-11's versus 11^+. How many different *team* outcomes are there?

   Within the 8! possible race finishes, we are uninterested in the rankings of the children within the two subsets. The number of possible *team* rankings is therefore

$$\frac{8!}{3!5!} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2} = 8 \cdot 7 = 72$$

To summarize the last example:

Theorem 1.9. *Suppose a collection of n objects has several subsets of identical elements of sizes $n_1, n_2, \ldots, n_r$ respectively. Then the number of permutations of the original collection is $\frac{n!}{n_1! \cdots n_r!}$.*

Example 1.10. A bag contains 10 balls; 5 are red, 4 blue and 1 green. If we remove the balls from the bag one at a time, how many seemingly distinct ways can this be done?

By the theorem, there are $\frac{10!}{5!4!1!} = 1260$ possible color sequences in which we can remove the balls.

A related consideration is the number of ways in which r objects can be chosen from a set of n objects. Label the set as follows:

$$\{a_1, \ldots, a_r, b_{r+1}, \ldots, b_n\} = \{a_1, \ldots, a_r\} \cup \{b_{r+1}, \ldots, b_n\}$$

Any choice of r elements amounts to making a permutation and selecting the first r terms. A permutation produces a genuinely distinct subset if and only if it moves at least some element from the first subset to the second and vice versa. The number of subsets of size r therefore equals the number of permutations of the big set which treat the first r and last $n-r$ terms as identically interchangeable. We are therefore in the situation of Theorem 1.9, and we conclude:

Theorem 1.11 (Combinations). *The number of ways to choose r objects from a set of n objects is*

$$\binom{n}{r} := \frac{n!}{r!(n-r)!}$$

The symbol[1] $\binom{n}{r}$ is read 'n choose r' and is known as a *binomial coefficient*; we'll shortly see why. It is worth noting symmetry and the consequence of the convention $0! = 1$;

$$\binom{n}{r} = \binom{n}{n-r}, \qquad \binom{n}{0} = \binom{n}{n} = \frac{n!}{0!n!} = 1$$

Example 1.12. You choose three flavors from the 10 available at the ice-cream shop. Here are several possible questions regarding combinations and permutations.

1. How many flavor combinations are possible if we cannot have two scoops of the same flavor?

   This is simply the binomial coefficient 10 choose 3:

   $$\binom{10}{3} = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2} = 120$$

2. How many flavour combinations are possible if we allow repeat scoops of the same flavor?

   If you allow repeat scoops (e.g. mint, mint, coffee), you must also include $10^2 = 100$ additional outcomes where the first two flavors are identical (10 choices for the two identical flavors and 10 for the third, which can equal the first two!). This results in 220 possible triple-scoop combinations.

3. How many ways can we choose three distinct flavors if we also care about which scoop is on top?

   There are two ways to proceed:

   (a) First choose the top scoop; 10 options. Then choose the remaining two scoops; $\binom{9}{2}$ options. This gives a total of $10\binom{9}{2} = \frac{10\cdot 9\cdot 8}{=}720$ choices.

   (b) Choose the three flavors; $\binom{10}{3}$ options. Then choose the top flavor from these three; 3 options. We therefore have $3\binom{10}{3}$ choices.

   Observe that the two methods produce the same result!

[1] An alternative notation, common on calculators, is ${}^nC_r$. Beware the related 'permutation' symbol ${}^nP_r$ which returns $\frac{n!}{(n-r)!} = r!\binom{n}{r}$, namely the number of ways of choosing an *ordered* subset of r elements from n.

The reasoning in the last example is quite general; for any $k = 1, \ldots, n$, we have

$$n\binom{n-1}{k-1} = k\binom{n}{k}$$

You should think through the reasoning for this *combinatorially*, as we did in the example. It is also very easy to verify *algebraically*:

$$n\binom{n-1}{k-1} = \frac{n \cdot (n-1)!}{(k-1)!\big((n-1)-(k-1)\big)} = \frac{n!}{(k-1)!(n-k)!} = k\binom{n}{k}$$

The discussion of combinations extends to count choices of multiple subsets from a given large set.

Example 1.13. A class of 12 students is to be split into four groups of sizes 6, 3, 2 and 1 respectively. We approach this iteratively:

- There are $\binom{12}{6}$ ways to choose the first group.
- After choosing the first, there are $\binom{12-6}{3}$ ways to choose the second group.
- After choosing the first two groups, there are $\binom{12-6-3}{2}$ ways to choose the third group.
- The final student is chosen by default, and we multiply to count how many ways the groups may be chosen:

$$\begin{aligned}\binom{12}{6}\binom{12-6}{3}\binom{12-6-3}{2} &= \frac{12!}{6!(12-6)!} \cdot \frac{(12-6)!}{3!(12-6-3)!} \cdot \frac{(12-6-3)!}{2!(12-6-3-2)!} \\ &= \frac{12!}{6!3!2!1!} = 54,440\end{aligned}$$

This is an example of a general approach.

Theorem 1.14 (Multinomial Coefficients). *The number of ways to divide a set of n objects into subsets of sizes $n_1, \ldots, n_r$ (necessarily $n = n_1 + \cdots + n_r$) is given by the multinomial coefficient*

$$\binom{n}{n_1, \ldots, n_r} := \frac{n!}{n_1! \cdots n_r!}$$

Proof. Generalizing the example, we compute

$$\binom{n}{n_1, \ldots, n_r} = \binom{n}{n_1}\binom{n-n_1}{n_2} \cdots \binom{n-n_1-\cdots-n_{r-1}}{n_r}$$

and observe how, except for the first, all factorial terms in the numerator cancel.

Alternatively, we can proceed similarly to how we established Theorem 1.11; this shows why the multinomial coefficient equals the permutation formula in Theorem 1.9. ∎

The standard binomial coefficient is $\binom{n}{k} = \binom{n}{k,n-k}$.

Examples 1.15. 1. $\binom{5}{2,2,1} = \binom{5}{2,1,2} = \binom{5}{1,2,2} = \frac{5!}{2!2!1!} = \frac{5\cdot 4\cdot 3}{2\cdot 2} = 15.$

2. We revisit and modify Example 1.13. In how many ways can we split a class of 12 students into three groups of 4? There are two possible questions here, so additional information is required.

   (a) If the groups are *ordered* amongst each other (a *1*st, *2*nd and *3*rd group), then it matters to which group a student is assigned. The number of possible divisions is then

   $$\binom{12}{4,4,4} = \frac{12!}{4!4!4!} = 34,650$$

   (b) If we don't care about the relative order of the groups, then we divide by 3! to represent permuting them: the result is 5775.

We finish by seeing why the binomial coefficients get their name.

Theorem 1.16 (Binomial/Multinomial Theorem). *For any natural number n,*

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

More generally,

$$(x_1 + \cdots + x_r)^n = \sum \binom{n}{n_1, \ldots, n_r} x_1^{n_1} \cdots x_r^{n_r}$$

where the sum is taken over all choices $(n_1, \ldots, n_r)$ for which $n = n_1 + \cdots + n_r$.

This is often done by induction (you've likely seen this before). Instead we prove directly, using the fact that we know what the binomial/multinomial coefficients *mean*.

Proof. To obtain an $x^k y^{n-k}$ term, we choose k of the $x+y$ factors in

$$(x+y)^n = (x+y)(x+y)\cdots(x+y)$$

to provide an x. By definition, there are 'n choose k' ways to do this!
The multinomial argument is essentially identical but with more subscripts. ∎

Examples 1.17. 1. $(x+y)^4 = x^4 + \binom{4}{3}x^3y + \binom{4}{2}x^2y^2 + \binom{4}{1}xy^3 + y^4 = x^4 + 4x^2y + 6x^2y^2 + 4xy^3 + y^4.$

2. To evaluate $(x+y+z)^3$ we need to consider all triples $n_1 + n_2 + n_3 = 3$ where $0 \le n_j \le 3$. Since a multinomial coefficient is independent of the order of the terms $n_1, \ldots, n_r$, we need only explicitly evaluate three terms,

   $$\binom{3}{3,0,0} = 1, \quad \binom{3}{2,1,0} = \frac{3!}{2!1!} = 3, \quad \binom{3}{1,1,1} = 3! = 6$$

   from which the full expansion is easy:

   $$(x+y+z)^3 = x^3 + y^3 + z^3 + 3(x^2y + xy^2 + x^2z + xz^2 + x^2z + xz^2) + 6xyz$$