Fermi Estimates: from Harry Potter to ET

David Wakeham

May 30, 2019

Abstract

Some notes on Fermi estimates and methods for doing them, presented at the UBC Physics Circle in May, 2019. We'll look at various applications, including global computer storage, the length of the Harry Potter novels, and the probability of intelligent aliens in the galaxy. *Note:* This is a rough draft only; I hope to polish these and add pictures and additional exercises at some point. For any comments, suggestions, or corrections, please get in touch at `<david.a.wakeham@gmail.com>`.

Contents

1 Introduction

I'm going to be talking about *Fermi problems*: what they are, methods for solving them, and if there's time, a fun application to the search for extraterrestrial intelligence, i.e. aliens we can talk to. Who knows what a Fermi problem is? Has anyone done one before? A Fermi problem is an *order of magnitude estimate*. Roughly speaking, it means to guess an answer to the nearest power of 10.

1.1 Why Fermi problems?

Fermi problems are great because the nearest power of 10 is a very forgiving notion of correctness, and because they're so forgiving, we can accurately estimate many things which seem impossible at first sight. This makes Fermi estimates a great party trick. You can impress your friends and family. But they're also extremely useful! I'm going to quickly talk about some things you can do with them.

- First of all, if you're doing some sort of fiddly, delicate calculation, you may want a way to double-check your result. Fermi estimates are perfect for this! They're simple, they're robust, you can often literally do them on the back of an envelope, so they're a good thing to compare to a difficult calculation. If the fiddly calculation is very different from the Fermi estimate, maybe something's gone wrong there.
- You can also use them to double-check claims that other people make. If you read a statistic in the newspaper or online which strikes you as fishy, you can do some quick googling and cobble together a Fermi estimate and see if the statistic is plausible. So Fermi problems can help protect you from fake news!
- Sometimes, though, you're not comparing against something more precise, you just want a ballpark figure. For instance, you can use Fermi approximation to settle the question of whether there are more stars in the sky, or grains of sand on the earth's beaches. I won't spoil the answer, because you'll get a chance to try this problem yourself. But the point is that you just need to know roughly how big they are to answer the question. In fact, this idea of ballparking numbers is so useful outside of physics that Fermi problems actually get taught in business degrees! You can Fermi estimate costs, which is really useful in day-to-day life.

Of course, impressing friends and family is fun, and I don't want to discourage that! That's a legitimate application as well.

1.2 Powers of 10

Before starting on methods for Fermi estimates, let's talk a bit about orders of magnitude, and how this relates to scientific notation, the usual way we write estimates and measurements. In scientific notation, we write answers as significant figures with a decimal after the first digit, then a power of 10 to shift the decimal point if we need to. For instance, the population of the world is around 7.7 billion, or

$$7.7 \times 10^9 \text{ people.}$$

This has two significant figures, 7 and 7. Say we want to approximate by keeping only a *single* significant figure. We're going to throw a digit away and round 7.7 up to 8, so the population becomes

$$8 \times 10^9 \text{ people,}$$

or 8 billion people. It may seem like that's as approximate as you can get, but there is another level of approximation: throw away significant figures altogether! We just keep the order of magnitude. Since 8 is closer to 10 than 1, we round up to get

$$10 \times 10^9 \text{ people} = 10^{10} \text{ people}.$$

To the nearest order of magnitude, there are 10 billion in the world.

The two rounding steps (from 7.7 to 8, and from 8 to 10) are subtly different. We all know how to round a decimal to the nearest whole number: 0.5 rounds up to 1, and anything less rounds down to 0. But to round something to the nearest power of 10 means to round it to something of the form 10^n, for a whole number n. We really just want to round the index up or down! For instance, $10^{0.5}$ should round up to $10^1 = 10$, and $10^{0.4}$ rounds down to $10^0 = 1$. This sounds reasonable, but when you plug in $10^{0.5}$ on a calculator, you get a bit of surprise: it's around 3.2! So in powers of 10, 3.2 is closer to 10 than to 1! This will be important for us. But it's initially counterintuitive.

A simple way to perform the rounding with powers of 10 is to first take a logarithm in base 10, which plucks out the index, round this to the nearest whole number, and use this as the index. In a sense, we just want to measure things on a logarithmic ruler, with a line drawn at every whole number index n. Let's do a quick example. To round the world population to the nearest power of 10, we take a logarithm of 8 and find that

$$\log_{10} 8 \approx 0.9.$$

That clearly rounds up to 1, so the nearest power of 10 is $10^1 = 10$. Hence, 10 billion people.

In practice, we don't always round to a power of 10 when we do Fermi estimates. For instance, we might have 8 billion as an order of magnitude estimate for the world's population. That seems very confusing at first, because first of all, I've told you that we're guessing things to the nearest power of 10, and seconly, in scientific notation, this guess would have to be correct to the nearest billion.

But we want a looser notion of correctness, allowing for half a power of 10 on either side, in the same way that we would allow for an error of half a billion on either side in scientific notation. But since $10^{0.5} \approx 3$, this means that if our Fermi estimate is 8 billion, we only expect it to be right *up to a factor of 3 biger or smaller*. You can think of this as just shifting marks on the logarithmic ruler, so we have a mark centred at 8 billion, with the next mark up at 80 billion, and the next mark down at 800 million. We're saying that the real number should be closer to 8 billion than these two other marks.

We use the $\sim$ symbol to indicate order of magnitude estimates. For instance

$$\text{world population} \sim 8 \times 10^9 \text{ people}$$

means we only expect this to be true up to a factor of 3 bigger or smaller. But if we instead want 8 to be a significant figure, with a much smaller error of half a billion on either side, we use an equals sign:

$$\text{world population} = 8 \times 10^9 \text{ people}.$$

This is a subtle but important difference!

2 Fermi problems

2.1 Comparison

Now that we know what an order-of-magnitude estimate is, we can start to do some estimates ourselves. Let's start with a simple question: how many people are in this room, to the nearest order of magnitude? The answer is clearly 10, but there are a couple of different ways to get here.

The first is just to do a head count: either counting individuals, or clumps of 2 or 5 or whatever works for you. Counting here is doable, but for an order-of-magnitude estimate it's unnecessary, and won't generalise to harder problems where you can't count things or you don't want to count things. A different method, which is less precise but does generalise well, is to *compare to things we already know*. To estimate the number of people in this room, I can compare to the size of the room. It looks around half full. An average classroom can comfortably fit 30 people, which I know because I've been in classes and taught classes, so there's probably 10-15 people here.

> **Exercise 1**. Estimate the population of Australia. Since you can't go out and count, compare to something you know. (You can answer either with a power or 10, or if you're game, with a digit followed by a power of 10. Either way, you're correct if you're within a factor of 3.)
>
> **Solution**. What did you estimate, and how did you do it? One option is to compare to the population of Canada. You probably know that the population of Canada is around 35 million, but Australia is a bit smaller, so a reasonable guess is 30 million. The actual answer is about 25 million. You're correct if your answer is between 8 million and 80 million.

Thus far, we've compared to a single data point, one thing that we knew, like the size of a classroom or the population of Canada. But what if something is quite different from the numbers we know? In some cases, there's a nice trick, where we can take an *average* of something bigger and something smaller to get an estimate. But the subtlety is that it's an average on a logarithmic ruler!

Let me illustrate with an example, and estimate the distance to the moon. What we're going to do is pick a distance which is smaller, and a distance which is larger, and average the two. But not just any upper and lower estimate will work. A good strategy is to make your lower estimate as big as possible, while still being below the thing you're trying to estimate, and similarly, make your upper as estimate as small as possible, while still being bigger than the target. On the logarithmic ruler, we want to sandwich the estimate on both sides so it doesn't have much room to move around. Think of the upper and lower guess as two slices of bread we're pushing together, and the target as a tasty slice of cheese in the middle we want to trap.

The biggest number I know smaller than the distance to the moon is the radius of the earth, which is about 6000 km:

$$R = 6 \times 10^3 \text{ km}.$$

My upper estimate is going to be the distance to the sun, since that's the smallest number I know bigger than the distance to the moon. I don't actually know the distance to the sun off the top of my head, but I do know that light takes 8 minutes to arrive from the sun, and that light travels at around 300, 000 km/s. So this translates into a distance

$$D \approx 8 \times 60 \times (3 \times 10^5) \text{ km} \approx 1 \times 10^8 \text{ km}.$$

Now we have our upper and lower estimate, we can take an average. But we don't take the usual arithmetic average where we add our numbers and divide by 2. Since the sun is much further away, this will basically just give us half the distance to the sun, with a negligible contribution from the earth.

We want to find the halfway point in orders of magnitude, in other words, on a *logarithmic* ruler. So we take logs, average those, and equate that average to the log of the distance d to the moon:

$$\log d \sim \frac{1}{2}(\log R + \log D).$$

We can plug this into a calculator and see what we get, but we can actually get rid of the logs using the log laws:

$$\log x + \log y = \log xy, \quad q \log x = \log x^q.$$

So we have

$$\log d \approx \frac{1}{2}(\log R + \log D) = \frac{1}{2}\log(RD) = \log \sqrt{RD}.$$

Undoing the logs, we find that $d \sim \sqrt{RD}$, which is called the *geometric mean* of R and D. Let's calculate it:

$$d \sim \sqrt{RD} \approx \sqrt{(6 \times 10^3)(1 \times 10^8) \text{ km}^2} \approx 8 \times 10^6 \text{ km}.$$

We guess that the moon is about 8 million kilometres away. Drum roll... The actual distance is about 4 million kilometers, so if we we're using scientific notation, we'd be incorrect. But since we're within a factor of 3, and our guess is correct to the nearest order of magnitude! This is pretty cool I think.

In general, the geometric mean gives us a way to average any two order of magnitude estimates. If one way of estimating gives me a number a, and another gives me a number b, and I have no reason to prefer either method, then I should just take the geometric average $\sqrt{ab}$ for my estimate.

Exercise 2. Estimate the population of the US using a geometric mean.

Solution. We've already mentioned a smaller country, namely Canada, with roughly 4×10^7 people. One option for a larger population is China, which has around one billion people. The geometric mean is

$$\sqrt{(4 \times 10^7)(1 \times 10^9)} = \sqrt{4 \times 10^{16}} = 2 \times 10^8,$$

or 200 million. Another option for upper bound is the total world population of 8 billion, which gives an estimate of around 550 million. The actual population of the

US is around 325 million, so both guesses are correct! Finally, let's *average the guesses*:

$$\sqrt{(2 \times 10^8)(5.5 \times 10^8)} \approx 3.36 \times 10^8.$$

This is only off by 10 million or so! This shows that the strategy of averaging guesses obtained by different methods can be very useful.

2.2 Factorisation

For a complicated Fermi estimate, comparison to known things isn't going to cut it. For instance, if we wanted to count the number of grains of sand in the world, what we can compare to? Lower bounds are a dime a dozen, but upper bounds are much harder. This seems like the wrong approach.

Instead, we will need to break the problem down into smaller subproblems, get estimates for each subproblem (I'll call these "subestimates"), then string them back together to get our final guess. "Stringing together" almost always means multiplying, so I'm going to focus on how to break Fermi estimates down into *factors* which we multiply to get our overall result.

This is similar in many ways to factorising whole numbers, so I'll start with a brief recap of how we approach that problem. Suppose we want to factorise a relatively big number, like 153, into its prime number factors. A way to proceed is to identify small factors, divide these out, and continue until we have a prime number left over. Adding the digits gives 1+5+3=9, so it's divisible by 9, with $153/9 = 17$. Since 17 is a prime number, we are done, and 153 can be written as the product of primes

$$153 = 3 \times 3 \times 17.$$

We can represent this with a *factor tree*, where "leaves" of the tree (the nodes at the bottom) are the primes.

Factorising Fermi estimates proceeds in a similar way, but in this analogy, the size of numbers is replaced by the complexity of the estimate. If we start off with a big, complex estimate, we want to break it down into simple estimates - the "primes" of the problem - and multiply them together. A procedure for doing this is to find a single simple estimate that the full estimate is proportional to, and divide it out. Just like the factor of 9! We iterate this process until we are left with a bunch of simple things to estimate. In the same way that 153 has small prime factors (3) and bigger prime factors (17), some of the subestimates will be easier and some will be harder. The way we split 9 into 3×3 also has an analogue: sometimes, a subestimate needs to be split into further subestimates. We can draw "factor trees" for a Fermi estimate as well.

Let's do an example to see how this all works. Here's the question: how many words does an average novel have? Any ideas for factorisation? One factorisation is as follows:

$$\text{words in a novel} = \text{pages in a novel} \times \text{words on a page}.$$

Let's draw the corresponding factors in a tree. How many pages in an average novel? Let's go with 250. What about words in a page? This actually seems a bit hard to do right off, so let's factorise further. Any suggestions? Here's one option:

$$\text{words on a page} = \text{words on a line} \times \text{lines on a page}.$$

This expands this node in the factor tree.

Let's guess the values for words on a line and lines on a page. Words on a line is maybe 10 or 15; lines on a page is maybe 20 or 30. I don't really have a preference between these, so let's take the geometric mean of the upper and lower guess:

$$\text{words on a page} \sim \sqrt{(10 \times 20)(15 \times 30)} = 300.$$

Now we just put it all together:

$$\text{words in a novel} \sim 250 \times 300 = 75,000.$$

How did we do? According to Amazon, the average number of words in a novel is 64,000. We're pretty close!

> **Exercise 3**. (a) Estimate the average length of an English word. (b) Use this to guess the total number of individual letters in all the Harry Potter novels. Draw a factor tree for your guess.
>
> **Solution**. (a) We use the usual strategy of averaging upper and lower bounds. A lower bound is a short word, say 3 letters, and an upper bound is a long word, maybe 10 letters. This gives a geometric mean of $\sqrt{3 \times 10} \approx 5.5$, which is consistent with the various different estimates for average word length in English one finds online.
>
> (b) We'll do something very simple: let's multiply the number of Harry Potter novels by the length of an average novel to get the total number of words. Most of the Harry Potter novels are above average in length, so let's multiply by a factor of 2 for good measure. Finally, we multiply by the average length of an English word, which we just estimated at 5.5. We get
>
> $$\text{words in all Harry Potter novels} \sim 7 \times 75,000 \times 2 \times 5.5 \approx 5.8 \text{ million}.$$
>
> The actual number is 6.3 million, so we are surprisingly close!

2.3 Units

When we made our estimate for the number of words in a novel, you might have noticed that the factors could all be expressed using "units":

$$\frac{\text{words}}{\text{novel}} = \frac{\text{words}}{\text{page}} \times \frac{\text{pages}}{\text{novel}} = \left(\frac{\text{words}}{\text{line}} \times \frac{\text{lines}}{\text{page}}\right) \times \frac{\text{pages}}{\text{novel}}.$$

These are not units in the usual physical sense, but "words", "lines" and "pages" are all things we can sensibly distinguish in the world and count. That's enough to make a unit with! I'm going to call these "general units" as opposed to "physical units", which are associated with physical properties, but you can use the same unit maths.

General units are a nice way to organise your Fermi estimate. For one thing, they let you check that you are stringing together the correct subestimates! In the equation above, lines cancel above and below, and pages cancel above and below, to give words per novel. The units make sense. More importantly, units can help you figure out the right subestimates to use! For instance, starting with

$$\frac{\text{words}}{\text{novel}},$$

we can look for an *intermediate unit*. I 'll talk about generalisations of this method later, but for the moment, we'll think of intermediate units as being like nested containers. A novel is a big container, words are tiny little objects, and we're looking for some kind of container in between.

A natural container is pages, so we arrive at our earlier factorisation:

$$\frac{\text{words}}{\text{novel}} = \frac{\text{words}}{\text{page}} \times \frac{\text{pages}}{\text{novel}}.$$

In our factor tree, adding intermediate units to a node corresponds to expanding that node out. We'll now do an example using the unit method, and try to estimate the total amount of storage space on all computers in the world. I'll write the units, and the factor tree, side by side for comparison. The units of our estimate are

$$\frac{\text{storage}}{\text{world}}.$$

Obviously, "world" is a funny unit, since we only care about our world, not other hypothetical worlds with computers, but it's useful to keep it there anyway. What is a natural intermediate unit? Computers! This contains storage space, and is contained in the world. This leads to

$$\frac{\text{storage}}{\text{world}} = \frac{\text{storage}}{\text{computer}} \times \frac{\text{computers}}{\text{world}}.$$

We now have two simpler estimates: the storage space on a computer, and the number of computers in the world. We can update our tree accordingly.

Let's first figure out storage space. At this point, we have to ask what we mean by "computer", since if we include phones and tablets, for instance, that will change our answer. For simplicity, I'm not going to include these, and we'll just consider laptops and desktop computers. For a modern computer, a 100 GB hard drive is on the small side, while a 1 TB hard drive is on the large side, and the geometric mean is

$$\sqrt{100 \times 1000}\ \text{GB} \approx 300\ \text{GB}.$$

What about computers per world? Perhaps we should break this up again. Can anyone think of a good intermediate unit? The number of people could work:

$$\frac{\text{computers}}{\text{world}} = \frac{\text{computers}}{\text{person}} \times \frac{\text{people}}{\text{world}}.$$

This expands the node in our tree as well. We know the "people per world", i.e. the world population, is around 8 billion.

Computers per person needs a bit more thought. On the one hand, computers are a bit of a luxury, and not everyone has one. In fact, about half the world doesn't have one! On the other hand, there are a lot of additional computers in business, and particularly the tech industry. Think of how much extra storage Google must have! For simplicity, let's just assume these things compensate for each other, and there's about computer per person. Now we have estimates for all the "leaves" of our factor tree, and we can multiply them to get our final result:

$$\begin{aligned}\frac{\text{storage}}{\text{world}} &= \frac{\text{storage}}{\text{computer}} \times \frac{\text{computers}}{\text{person}} \times \frac{\text{people}}{\text{world}} \\ &\sim 300 \text{ GB} \times 1 \times (8 \times 10^9) \\ &= 2.4 \times 10^{12} \text{ GB}.\end{aligned}$$

How did we do? According to statistics aggregator Statista, the total data capacity of world is around 1.5×10^{12} GB, so we're within a factor of 3!

Exercise 4. Using intermediate units, estimate the total number of pets owned in Canada. Draw the corresponding factor tree.

Solution. The estimate has units of pets/Canada. A natural intermediate unit is *people*, with

$$\frac{\text{pets}}{\text{Canada}} = \frac{\text{pets}}{\text{person}} \times \frac{\text{people}}{\text{Canada}}.$$

The second factor is just the population of Canada, around 40 million as we've already discussed. The first factor is a bit harder to estimate, since pets are more often owned associated to *households* rather than individuals. We could introduce households into the factor tree, but I'm going to be lazy, and make a rough guess of pets per person. 1 pet per person is too high, and 1 pet for every 10 people seems too low, so I'm going to guess somewhere in between on the logarithmic scale, or 1 pet for every 3 people. That leads to a guess of

$$\begin{aligned}\frac{\text{pets}}{\text{Canada}} &= \frac{\text{pets}}{\text{person}} \times \frac{\text{people}}{\text{Canada}} \\ &\sim \frac{1}{3} \times 40 \text{ million} \\ &\approx 13 \text{ million.}\end{aligned}$$

According to the Canadian Animal Health Institute, there are around 17 million cats and dogs in Canada, and taking other types of animals into account, a total of around 21 million. So our guess, while a bit low, is still within an order of magnitude.

2.4 Piano tuners

We can be much more creative with intermediate units than simply looking for containers of different sizes. In fact, we're going to need to be to solve more interesting problems! To illustrate, we'll do what is perhaps the most famous Fermi problem of all, posed by the physicist Enrico Fermi, who was, as his name suggests, a master of this sort of problem. So here it is: how many piano tuners are there in the greater Vancouver area? The original problem was "how many piano tuners in Chicago" but we'll do a local version. By "greater Vancouver", I mean the whole metro Vancouver area, including Vancouver itself, but also Richmond, Burnaby, Surrey, etc.

I'm not sure about you, but I know very little about pianos, and even less about piano tuners, but that's the beauty of Fermi problems: we can do the estimate without any detailed knowledge of these things. Somehow, we can cobble together our knowledge of other aspects of the world to get a reasonable guess.

We'll start as usual by writing the units of our estimate, piano tuners/Vancouver, and introducing an intermediate unit. Any suggestions? An obvious choice is people:

$$\frac{\text{tuners}}{\text{Vancouver}} = \frac{\text{tuners}}{\text{person}} \times \frac{\text{people}}{\text{Vancouver}}.$$

The second factor is the population of greater Vancouver. We could try estimating this with a geometric mean, but instead I'm just going to tell you the answer; it's about 2.5 million. The first factor is piano tuners per person, or in more conventional terms, the percentage of people who are piano tuners. Piano tuners are not like pets; I have basically no background knowledge to draw on, now upper and lower guesses I can reliably average.

So let's look another intermediate unit. Are there any ideas? What is an object that mediates the relationship between people and piano tuners, in other words, that we need to have piano tuners in the first place? Pianos! This leads to the factorisation

$$\frac{\text{tuners}}{\text{Vancouver}} = \frac{\text{tuners}}{\text{piano}} \times \frac{\text{pianos}}{\text{person}} \times \frac{\text{people}}{\text{Vancouver}}.$$

There is a general strategy at work here. The way that a piano tuner is related to a random Vancouverite is that they could tune that Vancouverite's piano! So by just thinking about how piano tuners interact with the bigger group, we can extract an intermediate unit.

> **Exercise 5**. Over to you: have a go at estimating fraction of people with a piano.
>
> **Solution**. Answers will vary, but here is one simple approach. I think maybe 1 in 3 people play an instrument, and piano is a popular choice, so I'm going to guess that 1 in 10 play piano. Just because you play piano doesn't mean you own a piano, and even if you do, you may own an electric piano, since these are cheaper, and electric pianos don't require tuning. I'm going to guess 1 in 2 people who play piano own a piano requiring tuning, so that 1/20 people have a piano which needs tuning.

Let's now focus on the remaining factor, piano tuners per piano. Again, I have no feeling at all for what this number should be, so I can't estimate. But I can think about the relationship between piano tuners and pianos. Presumably, the working week of a piano tuner looks as follows: from 9-5, 5 days a week, they are visiting pianos and tuning them. I'm going to assume that the piano tuning done equals the piano tuning required, or piano tuning supply equals demand.

So the intermediate unit is *tuning time*, so the amount of time a piano tuner actually works over the course of a year for instance, and the amount of tuning an average piano needs over a year. Let's see what that looks like in terms of units:

$$\frac{\text{tuners}}{\text{Vancouver}} = \frac{\text{tuners}}{\text{tuning time per year}} \times \frac{\text{tuning time per year}}{\text{piano}} \times \frac{\text{pianos}}{\text{person}} \times \frac{\text{people}}{\text{Vancouver}}.$$

Assuming that a piano tuner works full time hours and has a couple of weeks break, we can calculate the time they spend tuning per year (which is the inverse of the number in our estimate):

$$\frac{\text{tuning time per year}}{\text{tuner}} \sim 50 \times 5 \times 8 \text{ hours} = 2000 \text{ hours}.$$

Finally, we just estimate the amount of tuning that a piano requires each year. I'm guessing that a typical piano gets tuned once every couple of years, since many people will not get their pianos tuned at all! A piano is a big, delicate instrument, with a profession dedicated to tuning, so I'm guessing that it takes maybe an hour to tune a piano. So that's 30 minutes tuning each year per piano.

Let's plug our numbers in:

$$\begin{aligned}\frac{\text{tuners}}{\text{Vancouver}} &= \frac{\text{tuners}}{\text{tuning time per year}} \times \frac{\text{tuning time per year}}{\text{piano}} \times \frac{\text{pianos}}{\text{person}} \times \frac{\text{people}}{\text{Vancouver}} \\ &= \frac{1}{2000 \text{ hours}} \times (0.5 \text{ hour}) \times \frac{1}{20} \times (2.5 \times 10^6) \\ &\approx 32.\end{aligned}$$

We get an estimate of about 32 piano tuners in the greater Vancouver area. A quick internet search reveals 14 piano tuners in the greater Vancouver area, so our guess looks to be within a factor of 3. In fact, there may be piano tuners I missed, and tuners not on the internet, so this is probably a pretty accurate guess! Ignorance is evidently no barrier to good approximations.

3 Aliens

The piano tuners problem is impressive, but it's also very silly. Unless your baby grand is going flat, you probably don't care how many piano tuners there are in Vancouver. It's a party trick. Here's a much more interesting problem. How many aliens are there in the galaxy? And how many are intelligent? This is perhaps the deepest question we can ask about our place in the universe.

It turns out that we can apply the same machinery of factorising a guess into subestimates and so forth, and end up with an equation for the number of aliens, not so different from the piano tuners. This is actually a simple version of what's called *Drake's equation*. The subestimates we end up with are very hard to do and highly uncertain. We'll leave them to the experts, but if the experts disagree (as they often do), we'll just take a geometric mean of the extreme guesses.

3.1 Counting aliens

We proceed the same way we did earlier: we start by writing the units of our target result, and factorise using intermediate units. The units of our guess are

$$\frac{\text{intelligent aliens}}{\text{galaxy}}.$$

To start off our factorisation, we need an intermediate unit, something simple the number of intelligent aliens is proportional to. Any suggestions? Planets are good:

$$\frac{\text{intelligent aliens}}{\text{galaxy}} = \frac{\text{intelligent aliens}}{\text{planet}} \times \frac{\text{planets}}{\text{galaxy}}.$$

Let's focus on the second factor, the number of planets in the galaxy. This is actually hard to observe directly, since planets are almost invisible to telescopes, but we can see stars, and planets usually orbit stars. We can add stars into the mix:

$$\frac{\text{planets}}{\text{galaxy}} = \frac{\text{planets}}{\text{star}} \times \frac{\text{stars}}{\text{galaxy}}.$$

We'll come back to the numbers later; right now, we'll just focus on getting our factor tree in order.

What about aliens per planet? Here, the idea of containers is really useful. We're going to look at all the planets in the galaxy, and tick off properties that the planets need in order to support intelligent life. That will give us nested sets of planets.

First of all, the planet should be habitable, which means that it should be able to support liquid water, since we think that's essential to life, and it should have a stock of whatever other chemical you need to make complex, self-replicating muck, which is kind of what life is. We can write

$$\frac{\text{habitable planets}}{\text{planet}} = p_H,$$

where p_H is the probability a randomly chosen planet is habitable. In fact, so I don't have to write units everywhere, I'm going to define

$$N = \frac{\text{intelligent aliens}}{\text{galaxy}}, \quad n_S = \frac{\text{stars}}{\text{galaxy}}, \quad n_P = \frac{\text{planets}}{\text{star}}.$$

Now, just because a planet is habitable doesn't mean it develops life! It's possible that lightning has to strike a nutrient-rich pond, or microorganisms need to arrive on an asteroid, or something else equally elaborate. This is a biological question that's hard to answer, but the number of intelligent aliens is going to depend on it. Let's write

$$\frac{\text{planets with life}}{\text{habitable planet}} = p_L,$$

and think of this as the probability that life will arise on a habitable planet. It took the earth a few billion years to go from the basic building blocks of life to tool-wielding monkeys. Life does not automatically mean intelligent life! There's some probability

$$\frac{\text{planets with intelligent life}}{\text{planet with life}} = p_I,$$

which is the probability a planet with life gives rise to a sentient, or self-aware, species, and not just a bunch of weird-looking mushrooms.

Now we just string all of these numbers together to get an estimate for the number of intelligent alien species in the galaxy:

$$N = n_S \times n_P \times p_H \times p_L \times p_I.$$

Exercise 6. Write a "Drake equation" for the number of loonies concealed in couches across Canada. For bonus marks, do the subestimate!

Solution. Many answers are possible! Here's one:

$$N = n_H \times n_C \times p_O \times p_c \times p_L,$$

where N is the number of loonies, n_H is the number of houses in Canada, n_C is the average number of couches per house, and p_O is the probability a couch has an object concealed in it, p_c is the probability that object is a coin, and p_L is the probability the coin is a loonie.

I estimate $n_H \approx 10^7$ (divide the population by 3), $n_C \approx 1.5$ (most houses have a couch, some have more), $p_O, p_c \approx 0.4$ (decent chance something is in there, and decent chance it is a coin), and finally $p_L \approx 0.2$ (a larger coin you are less likely to lose). Putting it all together, I estimate around

$$N \sim 10^7 \times 1.5 \times 0.4^2 \times 0.2 \approx 500,000.$$

So, there may half a million dollars concealed in couches across Canada!

3.2 Cosmic numbers

There's a huge amount of uncertainty about these numbers, and we'd be kidding ourselves if we thought we could come up with a Fermi estimate which was correct to within a factor of 3. But we'll give it a go anyway; we just need to take our final results with a big grain of salt.

The first three numbers are things that astronomers can study, so have a reasonably good idea of what they are. The numbers of stars n_S is around 300 billion, and the number of planets per star n_P is a bit more uncertain, but something like 1.6. The number of habitable planets depends on what you think habitable means, but around 10% of planets are in what's called the *habitable zone*, meaning they can support liquid water.

On the other hand, some astronomers have argued that there's a whole bunch of other constraints a planet must satisfy in order to be habitable (e.g. it's got to have a circular orbit, it's got to stay away from the bad neighbourhoods of the galaxy, it maybe needs a big moon to help stabilise its axis of rotation, and so on). The claim, anyway, is that the earth is extremely rare, maybe the only truly habitable planet in the galaxy, which corresponds to a probability of 1 in 500 billion. Let's be even-handed, and take the geometric mean of these two guesses for p_H:

$$p_H \sim \sqrt{\frac{1}{500 \times 10^9} \times \frac{1}{10}} \approx \frac{1}{2 \times 10^6}.$$

So we estimate that approximately 1 in every 2 million planets is habitable.

The last numbers are to do with life. First, how often does life arise on a habitable planet? Sadly, we have no data about any planets besides the earth. But the earth tells us something interesting: we can see from the fossil and geological records that, as soon as the conditions were right, life developed immediately! This suggests that the probability is high that habitability leads to to life. It did take a few billion years for these conditions to arise, so let's perhaps conservatively set $p_L = 10\%$.

Before we move onto intelligent aliens, we can already have a guess at the number of planets with not-necessarily-sentient alien life out there, just by dropping the probability p_I from our string:

$$N_{\text{NS}} = n_S \times n_P \times p_H \times p_L \sim (300 \times 10^9) \times 1.6 \times (0.5 \times 10^{-6}) \times 0.1 \approx 25,000.$$

Being fairly even-handed, we get a guess of about 50,000 planets out there which have, or will develop, alien life. Maybe they'll just be weird mushrooms, or bacteria made of energy or something, but it seems overwhelmingly likely that aliens exist.

What about intelligence? What's the probability p_I? Once again, we have nothing but the earth to go on, and like I said earlier, it took billions of years of bacterial soup, and trilobites and dinosaurs and other stuff before monkeys arrived on the scene. A particularly pessimistic way to estimate the probability is to count species. Over the course of history, there have been billions of species on earth, and only one is intelligent. Our lower estimate is 10^{-9}. On the other hand, some people have more optimistically estimated that any planet with life will eventually develop intelligent life. The geometric mean is

$$p_I \sim \sqrt{10^{-9} \times 1} \approx 3 \times 10^{-5}.$$

So if we plug this back in and estimate the number of intelligent species:

$$N = N_{\text{NS}} \times p_I \sim 25,000 \times 3 \times 10^{-5} = 0.75.$$

This is consistent with the fact that there is at least one intelligent species, i.e. us! But there may not be anyone else. If we get lucky, there might be a few other sentient races in the galaxy. But even if they exist, our chances of contacting them are probably very low; that additional probability is factored into the full Drake equation.

We come to the conclusion that the galaxy is probably teeming with alien life, most of which is non-sentient. This answers a question, coincidentally, also posed by Fermi: *where is everybody?* In other words, why haven't we met aliens yet? Or even just spotted their probes? The answer seems to be that sentient is rare and widely distributed. This is called *Fermi's paradox*. So we've given a potential solution to Fermi's paradox with Fermi problems!