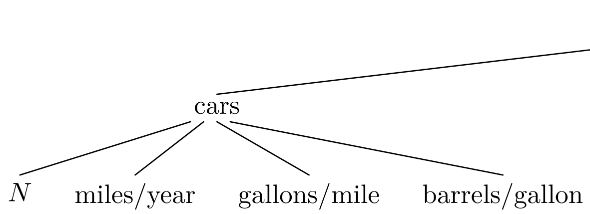
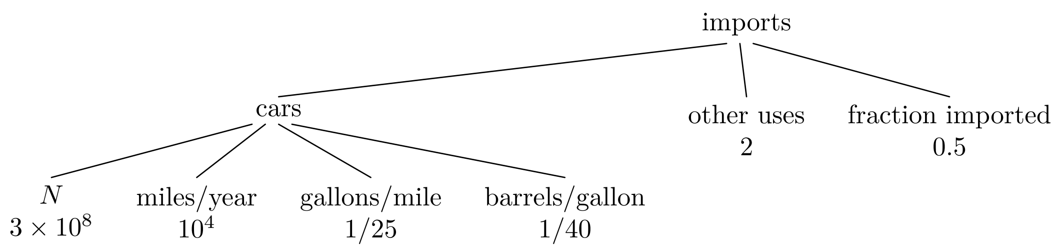
6.055J / 2.038J The Art of Approximation in Science and Engineering
Spring 2008



Now guess values for the unnumbered leaves. There are 3×10^8 people in the United States, and it seems as if even babies own cars. As a guess, then, the number of cars is $N \sim 3\times10^8$. The annual miles per car is maybe 15,000. But the N is maybe a bit large, so let's lower the annual miles estimate to 10,000, which has the additional merit of being easier to handle. A typical mileage would be 25 miles per gallon. Then comes the tricky part: How large is a barrel? One method to estimate it is that a barrel costs about \$100, and a gallon of gasoline costs about \$2.50, so a barrel is roughly 40 gallons. The tree with numbers is:



All the leaves have values, so I can propagate upward to the root. The main operation is multiplication. For the 'cars' node:

$$3\times10^8\,\text{cars}\times\frac{10^4\,\text{miles}}{1\,\text{car–year}}\times\frac{1\,\text{gallon}}{25\,\text{miles}}\times\frac{1\,\text{barrel}}{40\,\text{gallons}}\sim 3\times10^9\,\text{barrels/year}.$$

The two adjustment leaves contribute a factor of $2\times0.5=1$, so the import estimate is

$$3\times10^9\,\text{barrels/year}.$$

For 2006, the true value (from the US Dept of Energy) is 3.7×10^9 barrels/year!

This result, like the pit spacing, is surprisingly accurate. Why? **Section 2.5** explains a random-walk model for it, which suggests that the more you subdivide, the better.

But before discussing that model, try one more example.

2.4 Gold or bills?

2.5 Random walks

The estimates in **Section 2.1** and **Section 2.3** are surprisingly accurate. The true pit spacing in a CDROM varies from 1 μm to 3 μm, according to the so-called *Red Book* where Philips and Sony give the specification of the CDROM; our estimate of 1 μm is not too bad. The true value for the oil imports is only 10% different from our estimate.

Equally important, the estimates are more accurate after doing divide-and-conquer reasoning. My 95% probability interval for oil imports, if I had to guess a value without subdividing the problem, is say from 10^6 b/yr to 10^{12} b/yr. In other words, if someone had claimed that the value is 10 million barrels per year, it would have seemed low, but I wouldn't have bet too much against it. After doing the divide-and-conquer estimate, I'd have been surprised if the true answer were more than a factor of 10 smaller or larger than the estimate.

This section presents a model for guessing in order to explain how divide-and-conquer reasoning can make estimates more accurate. The idea is that when we guess a value far outside our intuitive experience – for example, micron-sized distances or gigabarrels – the error in the exponent will be proportional to the exponent. For example, when guessing a quantity like 10^9 in one gulp, I really mean: 'It could be, say, 10^6 on the low side or, say, 10^{12} on the high side.' And when guessing a quantity like 10^{30} (the mass of the sun in kilograms), I would like to hedge my bets with a region like 10^{20} to 10^{40}. So, in this model any quantity 10^β is really shorthand for

$$10^\beta \to 10^{\beta-\beta/3} \ldots 10^{\beta+\beta/3}.$$

Now further simplify the model: Replace the range of values by its endpoints. So, if we try to guess a quantity whose true value is 10^β, we are equally likely to guess $10^{2\beta/3}$ or $10^{4\beta/3}$. A more realistic model would include 10^β as a likely possibility, but the simplest model is easy to simulate and to reason with (that justification is a fancy way to say that I am lazy).

To see the consequences of the model, I'll compare subdividing and not subdividing by using a numerical example. Suppose that we want to guess a quantity whose true value is 10^{12}. Without subdividing, we might guess 10^8 or 10^{16} (adding or subtracting one-third of the exponent), a wide range.

Compare that range to the range when we subdivide the estimate into 16 equal factors. Each factor is $10^{12/16} = 10^{3/4}$. When guessing each factor, the model says that we would guess $10^{1/2}$ or 10^1 each with $p = 0.5$. Here is an example of choosing 16 such factors randomly from $10^{1/2}$ and 10^1 and multiplying them:

$$10^{0.5} \cdot 10^{0.5} \cdot 10^1 \cdot 10^{0.5} \cdot 10^1 \cdot 10^1 \times 10^{0.5} \cdot 10^1 \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^{0.5} \times 10^{0.5} \cdot 10^{0.5} \cdot 10^1 \cdot 10^{0.5} = 10^{10.5}$$

Here are three other randomly generated examples:

$$10^1 \cdot 10^{0.5} \cdot 10^1 \cdot 10^1 \cdot 10^1 \cdot 10^1 \cdot 10^{0.5} \cdot 10^1 \cdot 10^{0.5} \cdot 10^1 \cdot 10^1 \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^1 \cdot 10^1 \cdot 10^{0.5} = 10^{13.0}$$
$$10^1 \cdot 10^1 \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^1 \cdot 10^1 \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^1 \cdot 10^1 \cdot 10^1 \times 10^{0.5} \cdot 10^{0.5} \cdot 10^{0.5} = 10^{11.5}$$
$$10^{0.5} \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^1 \cdot 10^{0.5} \cdot 10^{0.5} \cdot 10^1 \cdot 10^1 \cdot 10^1 \cdot 10^{0.5} \times 10^{0.5} \cdot 10^{0.5} \cdot 10^1 \cdot 10^{0.5} = 10^{10.5}$$

These estimates are mostly within one factor of 10 from the true answer of 10^{12}, whereas the one-shot estimate might be off by four factors of 10. What has happened is that the errors in the individual pieces are unlikely to point in the same direction. Some pieces will be underestimates, some will be overestimates, and the product of all the pieces is likely to be close to the true value.

This numerical example is our first experience with the random walk. Their crucial feature is that the expected wanderings are significantly smaller than if one walks in a straight line without switching back and forth. How much smaller is a question that we will answer in **Chapter 5** when we introduce special-cases reasoning.