

Bill Foote

Counting the Ways



Contents

Counting the Ways	3
Plausibility, probability and information	3
Some Surprise	4
How many ways?	4
Back to data	6

Counting the Ways

Plausibility, probability and information

According to Aristotle, if two claims are well-founded, their truth values can be ascertained. A plate is either green, or it is not. When thrown, the plate will land without a chip or it will break. If I claim that when I throw the plate, it will land with out a chip and you disagree, I can simply throw the plate to find out who was correct. One part of this is a mind game, a thought experiment with a potential outcome. The other is the reality of actually throwing the plate and observing its status on landing. We will eventually call the mind-game a hypothesis and the reality a datum. Actually both are data: the hypothesis is unobserved data, and the so-called observations are observed data.

It is in disagreement that logical deduction might (plausibly) break down. There is no guarantee that the plate will break, or, for that matter, that it will chip. We must simply experiment with plate(s), green, blue or otherwise, to support the claim (or not). These claims arise in everyday life. For example, despite my poor performance in plate throwing in the past, there is no cogent reason to believe that it is absolutely, positively false that the plate I throw would land without a chip. There is a degree of acceptance, of plausibility, in one side of the claim, and on the other as well. Certainly it is not as false as the claim that $2 + 2 = 5$ in base 10 arithmetic, or the patently spurious claim that true is false or my cat is a dog.

Claims about things that are neither definitely true nor definitely false arise in matters both mundane and consequential: producing weather reports, catching the bus, predicting the outcomes of elections, interpreting experimental vaccine results, and betting on sports games, throwing the plate, to name just a few.

So we would benefit from a method of comparing claims in these situations – which atmospheric model produces better predictions? What is the best source for predicting elections? Should I blow three times on my lucky dice, or is this all just a figment of my denial based imagination?

Some Surprise

The goal, in all the cases above, is to guess about something that we don't or can't know directly, like the future, or the fundamental structure of the economy, or reasons why customer preferences change, on the basis of things we do know, like the present and the past, or the results of an existing or past experiment.

Mostly we guess. Some us try to systematically consider and attempt to support with evidence the guess. Lacking precision, and sometimes even accuracy, we try to avoid bad surprises. Goods ones are often welcome. If I use a government weather application on my smart phone it might not surprise me to see rain pelting down at 1630 (this afternoon). After all the app indicated as much. In advance of the rain I brought in chair pads and anything else that might get ruined with rain. An airline pilot knows all the defects of her aircraft. That knowledge saves lives.

Our mortal inferences, clever or dumb as they are, must have a surprise somewhere between *totally expected*, or zero surprises and thus certain 100% of the ways to make the statement, and *totally surprising* and 0% chance of anticipation. We will generally be making statements like: *it will probably rain tomorrow*, or *nine times out of ten, the team with a better defense wins*. This motivates us to express our surprise in terms of plausibility and we hanker for more precision with probability.

How many ways?

Let's use a simple example. We have four voters in an upcoming election. They may be red or blue voters. Three of us go out and talk to three voters at random, that is, indiscriminately. One of us happens to come upon a blue voter, another of us, independently, happens to find a red voter, and the other separately finds a blue voter. This is the very definition of a random sample. Each of the finders does not know what the other is doing, all three do know that there are four voters out there and they happened to have independently talked to two blue and one red voter. How many red voters and how many blue voters are there?

Here are all of the possible conjectures we can make for $blue = \square$ and $red = \square$ voters.

Reading this we see that there are 4 voters and 5 different voter compositions ranging from all red to all blue. Our sample is 2 blue and 1 red voter, so we

TABLE 0.1 voter conjectures

1	2	3	4
□	□	□	□
□	□	□	□
□	□	□	□
□	□	□	□
□	□	□	□

TABLE 0.2 ways voter conjectures turn out

1	2	3	4	proportion	ways	plausibility
□	□	□	□	0.00	$0 \times 4 \times 0 = 0$	0.00
□	□	□	□	0.25	$1 \times 3 \times 1 = 3$	0.15
□	□	□	□	0.50	$2 \times 2 \times 2 = 8$	0.40
□	□	□	□	0.75	$3 \times 1 \times 3 = 9$	0.45
□	□	□	□	1.00	$4 \times 0 \times 4 = 0$	0.00

can very safely eliminate the first and fifth conjectures from our analysis, but for the moment just keep them for completeness sake.

For each of the three remaining conjectures we may ask how many ways is the conjecture *consistent* with the collected data. For this task a tree is very helpful. Let's take the first realistic conjecture the □, □, □, □ hypothesis and check if, when we sample all of the four voters, what are all of the ways this conjecture fans out. So here we go.

1. We sampled a □ first. How many □'s are in this version of the composition of voters? only 1.
2. We then sampled independently a □. How many □'s are in this conjecture? Quite a few, 3.
3. Finally we sampled a □ at random. We know there is only one □ in this version of the truth.

So, it is just counting the ways: 1 □ way \times 3 □ ways \times 1 □ way = $1 \times 3 \times 1 = 3$ ways altogether.

When asked, many surmise that the 2 blue and 3 red conjecture is the right one. Are they right? Here is a table of the ways each conjecture pans out. We then in a separate column compute the contribution of each conjecture to the total number of ways across the conjectures, which is $3 + 8 + 9 = 20$ ways. Also each of the conjecture propose a proportion p of the successes, that is, the blue voters in this context.

We cannot help but note that the proportion of ways for each conjecture can range from 0, perhaps to 1, since the proportions add up to 1. The number of ways also expresses the number of true consistencies of the data with the conjecture, an enumeration of the quality of the logical compatibility of conjectures with what we observe.

We might now revise our commonsense surmise that 2 blue and 2 red is the better conjecture. However, if we use the criterion that the conjecture with the most ways consistent with the data is the best choice for a conjecture, then clearly here we would say that there are 3 blues and 1 red. Perhaps we have a better criterion that would choose our equanimous choice of 2 blues and 2 reds? It does not appear to be so.

Ways are the count, the frequencies of logical occurrence of a hypothesis *given* the data. The data includes the knowledge that there are possibly blues and reds, that there are 4 voters, and that we sampled 2 blues and 1 red. The relative frequency of the ways in which our conjectures are consistent with the data is what we will finally call *probability*.

1. The plausibility is a measure between and including 0 and 1.
2. The sum of all plausibilities is 1.
3. The plausibility must be compatible with common sense.

That last requirement, in turn, requires a sketch of why it is important.

- Let's suppose the plausibility of a proposition A (Jaynie is a runner, for example) is changed by another proposition B (Jaynie participated in a marathon).
- Let's mix this up with the plausibility that proposition C (Jaynie is a business analyst, for example) has nothing to do with proposition B.
- Then common sense (as opposed to common non-sense) dictates that the plausibility of A and C together (Jaynie is a business analyst runner) is also increased by proposition B.

We have just quantified the *plausibility* of logical truth values. We have also found a very compelling criterion for the choice of a conjecture given the data and circumstances surrounding our inquiry.

Back to data

The location and scale parameters μ and σ (e.g., mean and standard deviation) we often calculate, and have so since middle school, are not the mean and standard deviation of data x . They are not properties of the physical representation of events called **observed data**. These measures use counts of events,

measures of, perhaps physical properties like heat, size, time, but in the end they are abstractions and summarizations about the events.

These parameters do carry information about the probability distribution of the representation of physical reality in data. To say μ is the mean of the data is to ascribe a mind's eye idea to the physical reality, to invest in physical, objective reality, a property that only exists in the mind, but is in fact **unobserved data**.

This is an example of the *mind projection* or *reification* fallacy much in vogue in the circles of *fake*, or better yet, *chop logic*. In the same way, probabilities only exist in our minds: there is no physical reality that is a probability, just a mental construct that helps us think through potential outcomes. Is information just such a fiction?

Exercising our grip on reality

1. Suppose there are 5 voters of unknown affiliation. What is the probability that 60% are red affiliates?
2. So one of 5 moved out and we are back to 4 voters. What is the most plausible proportion of blue voters if we sample the district and find 2 blue and 3 red voters?