EIGHT

REDUCTIO AD UNLIKELY

The stickiest philosophical point in a significance test comes right at the beginning, before we run any of the sophisticated algorithms developed by Fisher and honed by his successors. It's right there at the beginning of step 2:

"Suppose the null hypothesis is true."

But what we're trying to prove, in most cases, is that the null hypothesis *isn't* true. The drug works, Shakespeare alliterates, the Torah knows the future. It seems very logically fishy to assume exactly what we're aiming to disprove, as if we're in danger of making a circular argument.

On this point, you can rest easy. Assuming the truth of something we quietly believe to be false is a time-honored method of argument that goes all the way back to Aristotle; it is the proof by contradiction, or reductio ad absurdum. The reductio is a kind of mathematical judo, in which we first affirm what we wish eventually to deny, with the plan of throwing it over our shoulder and defeating it by means of its own force. If a hypothesis implies a falsehood,* then the hypothesis itself must be false. So the plan goes like this.

* Some people will insist on the distinction that the argument is only a reductio if the consequence of the hypothesis is self-contradictory, while if the consequence is merely false the argument is a modus tollens.

- Suppose the hypothesis H is true.
- It follows from H that a certain fact F cannot be the case.
- But F is the case.
- Therefore, H is false.

Say someone exclaims to you that two hundred children were killed by gunfire in the District of Columbia in 2012. That's a hypothesis. But it might be somewhat hard to check (by which I mean that I typed "number of children killed by guns in DC in 2012" into the Google search bar and did not immediately learn the answer). On the other hand, if we assume the hypothesis is correct, then there cannot have been any fewer than two hundred homicides in total in DC in 2012. But there *were* fewer; in fact, there were only eighty-eight. So the exclaimer's hypothesis must have been wrong. There's no circularity here; we've "assumed" the false hypothesis in a kind of tentative, exploratory way, setting up the counterfactual mental world in which H is so and then watching it collapse under pressure from reality.

Put this way, the reductio sounds almost trivial, and in a sense, it is; but maybe it's more accurate to say it's a mental tool we've grown so used to handling that we forget how powerful it is. In fact, it's a simple reductio that drives the Pythagoreans' proof of the irrationality of the square root of 2; the one so awesomely paradigm-busting they had to kill its author; a proof so simple, refined, and compact that I can write it out whole in a page.

Suppose

H: the square root of 2 is a rational number

that is, $\sqrt{2}$ is a fraction m/n where m and n are whole numbers. We might as well write this fraction in *lowest terms*, which means that if there is a common factor between the numerator and denominator, we divide it out of both, leaving the fraction unchanged: no reason to write 10/14 instead of the simpler 5/7. So let's rephrase our hypothesis:

H: the square root of 2 is equal to m/n, where m and n are whole numbers with no factor in common.

In fact, this means we can be sure it's not the case that m and n are both even; for to say both numbers are even is exactly to say both have 2 as a factor. In that case, as in the case of 10/14, we could divide both numerator and denominator by 2 without changing the fraction, which is to say it was not in lowest terms after all. So

F: both m and n are even

is false.

Now since $\sqrt{2} = m/n$, then by squaring both sides we see that $2 = m^2 / n^2$ or, equivalently, that $2n^2 = m^2$. So m^2 is an even number, which means that m itself is even. A number is even just when it can be written as twice another whole number; so we can, and do, write m as 2k for some whole number k. Which means that $2n^2 = (2k)^2 = 4k^2$. Dividing both sides by 2, we find that $n^2 = 2k^2$.

What's the point of all this algebra? Simply to show that n^2 is twice k^2, and therefore an even number. But if n^2 is even, so must n be, just like m is. But that means that F is true! By assuming H we have arrived at a falsehood, even an absurdity; that F is false and true at once. So H must have been wrong. The square root of 2 is *not* a rational number. By assuming it was, we proved that it wasn't. It's a weird trick indeed, but it works.

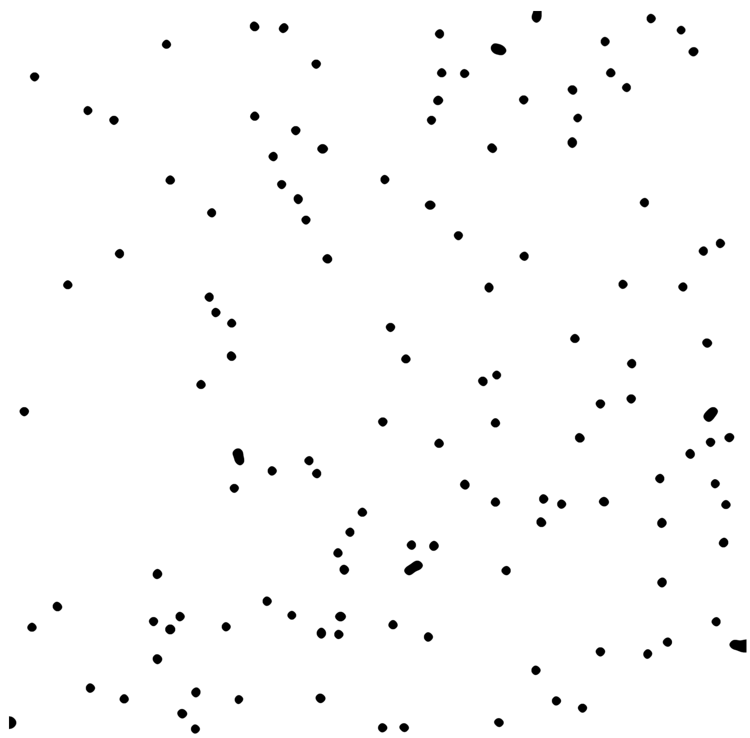
You can think of the null hypothesis significance test as a sort of fuzzy version of the reductio:

- Suppose the null hypothesis H is true.
- It follows from H that a certain outcome O is very improbable (say, less than Fisher's 0.05 threshold).
- But O was actually observed.
- Therefore, H is very improbable.

Not a reductio ad absurdum, in other words, but a reductio ad unlikely.

A classical example comes from the eighteenth-century astronomer and clergyman John Michell, among the first to take a statistical approach to the study of the heavenly bodies. The cluster of dim stars in one corner of the constellation Taurus has been observed by just about every civilization. The Navajo call them Dilyehe, “the sparkling figure”; the Maori call them Matariki, “the eyes of god.” To the ancient Romans they were a bunch of grapes and in Japanese they’re Subaru (in case you ever wondered where the car company’s six-star logo came from). We call them the Pleiades.

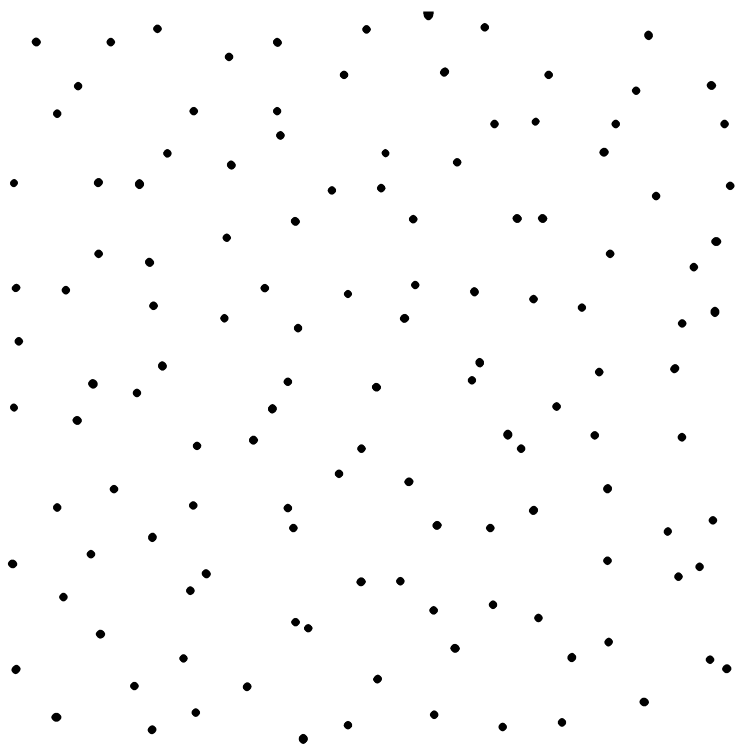
All these centuries of observation and mythmaking couldn’t answer the fundamental scientific question about the Pleiades: is the cluster actually a cluster? Or are the six stars separated by unfathomable distances, but arrayed by chance in almost the exact same direction from Earth? Points of light, placed at random in our frame of vision, look something like this:



You see some clumps, right? That’s to be expected: there will inevitably be some groups of stars that wind up almost on top of one another, simply by happenstance. How can we be sure that’s not what’s going on with the Pleiades? It’s the same phenomenon Gilovich, Vallone, and

Tversky pointed out: a perfectly consistent point guard, who enjoys no hot streaks and suffers no slumps, will nonetheless sometimes nail five shots in a row.

In fact, if there were no big visible clusters of stars, as in this picture:



that itself would be evidence that some nonrandom process was at work. The second picture might look "more random" to the naked eye, but it is not; it testifies that the points have a built-in disinclination to crowd.

So the mere appearance of an apparent cluster shouldn't convince us that the stars in question are actually clumped together in space. On the other hand, a group of stars in the sky might be so tightly packed as to demand that one doubt it could have happened by chance. Michell showed that, were visible stars randomly strewn around in space, the chance that six would array themselves so neatly as to present a Pleiades-like cluster to our eyes was small indeed; about 1 in 500,000, by his computation. But there they are above us, the tightly packed bunch of grapes. Only a fool, Michell concluded, could believe it had happened by chance.

Fisher wrote approvingly of Michell's work, making explicit the analogy he saw there between Michell's argument and the classical reductio:

"The force with which such a conclusion is supported is logically that of a simple disjunction: Either an exceptionally rare chance has occurred, or the theory of random distribution is not true."

The argument is compelling, and its conclusion correct; the Pleiades are indeed no optical coincidence, but a real cluster—of several hundred adolescent stars, not just the six visible to the eye. The fact that we see many very tight clusters of stars like the Pleiades, much tighter than would be likely to exist by chance, is good evidence that the stars are not placed randomly, but rather are clumped by some real physical phenomenon out there in the void.

But here's the bad news: the reductio ad unlikely, unlike its Aristotelian ancestor, is not logically sound in general. It leads us into its own absurdities. Joseph Berkson, the longtime head of the medical statistics division at the Mayo Clinic, who cultivated (and loudly broadcast) a vigorous skepticism about methodology he thought shaky, offered a famous example demonstrating the pitfalls of the method. Suppose you have a group of fifty experimental subjects, who you hypothesize (H) are human beings. You observe (O) that one of them is an albino. Now, albinism is extremely rare, affecting no more than one in twenty thousand people. So given that H is correct, the chance you'd find an albino among your fifty subjects is quite small, less than 1 in 400,* or 0.0025. So the p-value, the probability of observing O given H, is much lower than .05.

We are inexorably led to conclude, with a high degree of statistical confidence, that H is incorrect: the subjects in the sample are not human beings.

It's tempting to think of "very improbable" as meaning "essentially impossible," and, from there, to utter the word "essentially" more and

* As a good rule of thumb, you can figure that each of the fifty subjects contributes a 1/20,000 chance of finding an albino in the sample, yielding 1/400; this isn't exactly right, but is usually close enough in cases like this one, where the result is very close to 0.

more quietly in our mind's voice until we stop paying attention to it.* But impossible and improbable are not the same—not even close. Impossible things never happen. But improbable things happen a lot. That means we're on quivery logical footing when we try to make inferences from an improbable observation, as reductio ad unlikely asks us to. That time in North Carolina when the lottery combo 4, 21, 23, 34, 39 came up twice in a week raised a lot of questions; was something wrong with the game? But each combination of numbers is exactly as likely to come up as any other. For the numbers to show 4, 21, 23, 34, 39 on Tuesday and 16, 17, 18, 22, 39 on Thursday is precisely as improbable as what actually took place—there's just one chance in 300 billion or so of getting those two draws on those two days. In fact, *any* particular outcome of the Tuesday and Thursday lottery draws is a one in 300 billion shot. If you're committed to the view that a highly improbable outcome should lead you to question the fairness of the game, you're going to be the person shooting off an angry e-mail to the lottery commissioner every Thursday of your life, no matter which numbered balls drop out of the cage.

Don't be that person.

PRIME CLUSTERS AND THE STRUCTURE OF STRUCTURELESSNESS

Michell's critical insight, that clusters of stars might appear to our eye even if stars were randomly distributed around our field of vision, doesn't apply only to the celestial sphere. This phenomenon was the hinge for the pilot episode of the math/cop drama *Numb3rs*.† A series of grisly attacks, marked by pins on the wall map at HQ, showed no clusters; ergo, a single cunning serial killer intentionally leaving space between victims, not an unconnected burst of psychos, was at work. It was somewhat contrived as a police story, but mathematically it was perfectly correct.

* Indeed, it's a general principle of rhetoric that when someone says "X is essentially Y," they generally mean "X is not Y, but it would be simpler for me if X were Y, so it'd be great if you could just go ahead and pretend X is Y, sound good?"

† Disclosure: I used to read *Numb3rs* scripts in advance to check their mathematical accuracy and provide comments. Only one line I suggested ever made it on the air: "trying to find a projection of affine three-space onto the sphere subject to some open constraints."