Turning points: Fisher's random idea

Abstract

In this second article in our series on key moments in the history of statistics, **Simon Raper** invites us to a field in England, where R. A. Fisher urged scientists to stop trying to control nature and embrace randomness instead

naumoid/Bigstock.com

A freshly ploughed field is a dull setting for a revolution. Yet it was in the flat, muddy farmland of Hertfordshire, England, that R. A. Fisher, father of modern statistics, did something unexpected and momentous. Instead of bringing order and control to a scientific experiment, he added chaos.

The field was part of an agricultural research station on the grounds of Rothamsted Manor.[1] The experiment, which we think took place sometime between 1918 and 1924, was designed to test the effect of different fertiliser mixes on the growth of potatoes. It sounds prosaic, but Fisher's innovation really was revolutionary.

For his colleagues at Rothamsted faced a problem that dogged any experiment outside of the lab: how to deal with the many factors that might affect the outcome of an experiment. Besides the fertiliser, the crop yield could be influenced by the weather, the fertility of the soil, the water drainage, the possibility of disease, and so on. Researchers could control for effects over time (the weather or disease) by running two experiments simultaneously, one with and one without the fertiliser. But this would require separate plots of land, thus introducing differences due to drainage, weeds, and soil fertility. If they tried to counter this by running the two experiments in sequence on the same plot of land, then effects over time would creep back in.

The response of the researchers at Rothamsted was to invent their way out of the dilemma with ever more elaborate designs. It was obvious to Fisher, however, that whatever pattern they created (and they had some intriguing ones, including sandwich and checkerboard designs) it was at least conceivable that there was some natural phenomenon that also followed that pattern.

Fisher's solution was both utterly left field and ferociously logical: design something without a pattern. In other words, allocate the fertiliser randomly to the plots. Since there is no systematic pattern in randomness, there is nothing that might, by chance, be correlated with some unforeseen phenomenon.

It sounds simple, maybe even obvious. But at the time, Fisher's use of randomisation was viewed as inconvenient, prone to human error and completely unnecessary. This is a puzzle. What could Fisher see that his contemporaries could not?

Nature on the rack

To understand this initial reaction to Fisher's idea, we need to expose a misconception about scientific practice that has a history as long as science itself. This is the notion that complete control over the conditions of an experiment is not just a useful starting point but rather a necessary condition for establishing scientific knowledge.

We can trace the beginnings of this misconception back to Francis Bacon, the Elizabethan scholar whom many consider the originator of the empirical method of scientific discovery. Although he probably did not say, as is commonly claimed, that we should put “Nature to the rack”, the metaphor works well for the tradition he initiated. For one thing, it emphasises the scientist's assumed position of dominance over nature, and the one-question-at-a-time interrogation style that many assume, wrongly, to be a hallmark of good scientific practice. For another, the gradual tightening of the wheels of the rack reminds us of what the philosopher John Stuart Mill called “the method of concomitant variations”, where one variable at a time is slowly and precisely adjusted while we watch for a corresponding change in the experimental outcome – correspondence indicating causality. We should note that Fisher was, for reasons we will later explore, very much against the metaphor of the scientist as interrogator:

> No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or ideally one question, at a time. The writer is convinced that this view is wholly mistaken. Nature, he suggests, will best respond to a logical and carefully thought out questionnaire; indeed, if we ask her a single question, she will often refuse to answer until some other topic has been discussed.[2]

Bacon himself advocated tables of discovery, in which the effect of a phenomenon is isolated by listing those occasions on which the phenomenon is present and comparing them with occasions that are equivalent but for the fact that the phenomenon is absent. As a method, it lacked power and precision, but it was the beginning of something greater. By the time we reach the Victorian era – a golden age of experimentation, with scientist-showmen such as Humphrey Davy and Michael Faraday performing in front of rapt audiences – the notion of the scientific experiment had matured considerably, along with the apparatus needed to carry it out. To run an experiment meant meticulously identifying all possible influencing factors, pinning each of them down, or else eliminating them one by one, until all that remained was the cause in question, and its effect.

A paradigm case would be the work of the chemist Louis

Pasteur. In a famous experiment he disproved the then popular theory of the spontaneous generation of life, which held that life could spring into existence out of the inanimate. He did this by systematically ruling out, in ingenious ways, every possible means by which microbial life might enter a flask of broth. He then demonstrated that it was only when these conditions were relaxed that life again appeared. Furthermore, he repeated the experiment over and over again, under different conditions, until he was convinced that no other factors could be responsible for the absence of life.

From the laboratory to the field

The still familiar image of the laboratory scientist, patiently ruling out all other possible causes, must have been prominent in the minds of Fisher's co-workers, directing their efforts as they tried in vain to devise experiments to isolate the effect of the fertilisers. But Fisher understood that outside the laboratory, in the field, where many factors lie out of sight and beyond our control, this was a doomed enterprise: "[I]t would be impossible to present an exhaustive list of such possible differences appropriate to any one kind of experiment, because the uncontrolled causes which may influence the result are always strictly speaking innumerable."[3] And, of course, this objection extends far beyond the domain of agriculture, to all human, biological, and environmental sciences – for the simple reason that they too concern phenomena that lie outside the controlled environment of the laboratory.

> **At first, Fisher's use of randomisation was viewed as inconvenient, prone to error and completely unnecessary**

Why, then, was Fisher more open than his contemporaries to the idea of giving up some of this control? One possible answer is that the idea of the perfect laboratory experiment coevolved with the Cartesian-Newtonian picture of an ordered, deterministic (and therefore non-random) universe that was at least, in principle, fully observable. By contrast, the work of statisticians in the nineteenth and early twentieth century had been gradually chipping away at this idea. Initially, statisticians viewed randomness as only apparent: a consequence of the imperfect measurement of a world that was essentially ordered. Then, towards the middle of the nineteenth century, nature itself began to be thought of as something that exhibited randomness but which could still be understood in terms of statistical distributions whose parameters, at least, were fixed and knowable. Finally, with Fisher himself, a quasi-Platonic view of statistical knowledge is adopted: we can only ever estimate the parameters of distributions; their true values are unknowable. In other words, the trajectory is one of increasing humility, ending in an awareness that randomness is an ever-present and irreducible part of our world.

Fisher was, in fact, perfectly placed to bring randomness into the fold. As the academic Nancy Hall points out,[4] part of the reason why Fisher may have been so open to the use of randomisation is that he pictured random sampling in terms of the most physical of the branches of mathematics: geometry. It is usual and intuitive to think of observations as points plotted in a space: a data set of five observations, for example, could be visualised as five points on a number line. But there is another, less intuitive, way of looking at things. Imagine that instead of five points, we have a space with five dimensions – one for each of the observations. Now the whole data set can be represented by a single point whose location marks all five values at once. What Fisher showed was that, in such a space, any random sample is necessarily a point lying on the surface of a sphere (or, strictly speaking, a hypersphere) whose properties are determined by the mean and variance of the sample. Thought of in this way, randomness begins to feel like something that belongs inside the ordered world of mathematics.

A final possible explanation for Fisher's more open attitude to randomisation is that the sciences that, in his day, made the greatest use of statistics were those that hinged upon mechanisms of randomness. Quantum physics, a field in which Fisher did graduate work, is one example. Genetics is another (although here the efforts by statisticians, including, very prominently, Fisher, to assert control through the advocacy of eugenics make up one of the most shameful episodes in the subject's history).

Guessing games

Fisher was not the first to introduce randomisation into the design of an experiment. That honour goes to the American philosopher and mathematician C. S. Peirce. Peirce had doubts about a claim by the German psychologist, Gustav Fechner, that there exists a threshold point (in psychological terminology, a *limen)* beyond which a human being can no longer tell the difference between the weights of two objects. Peirce suspected that rather than ending abruptly at a threshold, our ability simply decreases in proportion to the weight difference. To test this, he arranged for a subject to be presented with pairs of weights in succession and asked to state which, if any, was the heavier. He worried, however, that if the weights were presented in a particular sequence – say, heaviest first – the subject might suspect a pattern and use this suspicion to work out the correct answer. His solution was to change the order of the weights at random, and for this purpose a pack of playing cards was used, with the colour of the suit determining the choice of weight. (The same device – drawing from a shuffled deck of cards – was later used by Fisher to decide which plots would receive fertiliser.) Peirce, incidentally, was proven right.

A. Barringron Brown, © Gonville & Claus College. Coloured by Science Photo Library

Another early use of randomisation was in parapsychology, as practised in the late nineteenth century, specifically in studies of telepathy, where the researchers had similar concerns to Peirce about the subject's ability to spot patterns. So why treat Fisher's use of randomisation as a turning point? One answer is that applying randomisation in the layout of experiments was a bolder, more counterintuitive move than using it simply to create a barrier to knowledge: after all, the latter use must have already been familiar in guessing games and magic tricks. But the other, more significant, reason is that randomisation for Fisher was more than just a useful trick; it was an integral part of his statistical system and one that opened up possibilities for statistical modelling in ways that were completely new.

> **Fisher's methods were a success, and they spread from agriculture to medicine to the social sciences**

A significant step forward

As we have seen, it was the strategy of scientists before Fisher to seek to eliminate unwanted factors through the design of their experiments. Not only did Fisher side-step this requirement through his clever use of randomisation, he did something even bolder: he brought the anonymous mass of "strictly speaking innumerable" uncontrolled factors into his mathematical models, treating them as statistical noise – a step that would not have been possible without randomisation.

The modelling of unwanted factors as noise, with statistical properties such as variance, was enormously important since it allowed Fisher to ask questions including “How much variation would we expect to see in crop yields regardless of the fertiliser?” and “How probable is it that we would see large differences between the treated and untreated plots if, in actual fact, the fertiliser was making no difference?”

The latter is, of course, the bare bones of Fisher's significance test: his controversial but still enormously influential mechanism of statistical inference. For example, in Fisher's analysis of variance, the F-test (named after Fisher) compares the amount of variation between groups (the difference between the average crop yield of the treated plots and the average of the untreated plots) and the amount of variation within the groups (the statistical noise). With a few refinements, we can work out the probability of seeing particular ratios of group-to-noise variation under a null hypothesis of no effect from the fertiliser. If the observed ratio is highly improbable then this counts against the null hypothesis.

But notice how this procedure falls apart if we take away randomisation. For one thing, our estimates of the statistical properties of the noise will be wrong since factors that should be part of it risk being absorbed into the estimates of the treatment effects. For another, our test of the null hypothesis is broken: it was formed on the assumption that, as long as the null hypothesis holds, the only thing that can possibly generate a difference between the treated and untreated plots is random allocation. Without randomisation, the difference could be caused by some other factor, spuriously correlated with the treatment allocation. The danger is that we may conclude that there is a treatment effect when there is not. As Fisher put it: "[F]or the systematic arrangement of our plots may have ... features in common with the systematic variation of fertility, and thus the test of significance is wholly vitiated."[5]

The scope of what Fisher achieved should now be clear: his combination of statistical modelling and randomised design gave us, for better or for worse, a method of making inferences without needing to specify, and control for, every possible influencing factor, thus broadening the application of the scientific method to huge swathes of knowledge lying outside the controlled conditions of the lab. If we have any remaining doubts about the benefits of randomisation then we need only look at cases where it is, by the nature of things, quite impossible. Consider, for example, econometric modelling: if we could randomly allocate the effects of, say, a fiscal policy change to the days in a year, a great many statistical problems in economics would be instantly solved!

Randomisation as a tool

Despite initial doubts about the usefulness of randomisation, Fisher's methods were a success. They allowed him to draw firm conclusions from relatively small amounts of data and thus put the development of human-made fertilisers onto a path of continuous improvement. Researchers and scientists were soon writing to him for advice on how to lay out their own trials, some even sending their own staff to Rothamsted for training. Between 1925 and 1930, Fisher's designs for randomised experiments increased in complexity to cope with a wider range of research questions. The models and designs were novel in another respect: they allowed Fisher to test several hypotheses at once, dealing a blow to another great shibboleth of Victorian science: the idea that the scientist only asks one question at a time. This was the implication contained in the earlier quote: "Nature ... will best respond to a logical and carefully thought out questionnaire; indeed, if we ask her a single question, she will often refuse to answer."

The application of Fisher's methods spread from agriculture to medicine to the social sciences and to business, and in all these fields the impact has been enormous. But we can trace another line of influence, one that cuts through the twentieth century and is possibly only now coming to full fruition. When Fisher used a shuffled deck of cards to allocate treatments to plots, he was employing an early, crude, random sequence generator. This marks the beginning, in earnest, of the use of randomness as a tool. Before this, it was merely a problem to be solved. Without this step, perhaps there would not have been a route through to the tools of twenty-first-century statistics: bootstrapping (see page 8), Monte Carlo simulation and Bayesian estimation methods – all of which have, at their heart, an idea that took root nearly a century ago in a field in England.

References