



© 1997–2009, Millennium Mathematics Project, University of Cambridge.

Permission is granted to print and copy this page on paper for non-commercial use. For other uses, including electronic redistribution, please contact us.

September 2004

Features



Coincidence, correlation and chance

by Andrew Stickland



Coincidence?



Coincidence, correlation and chance

Football fans pay attention! How would you like to make hundreds, maybe even thousands, of pounds each week by betting on your favourite team, guaranteed to get the result right every time? Well, now you can. I have devised a computer program, based on complex statistical algorithms, which is 100% accurate every time ... and I am willing to sell one of a limited number of copies to you for the modest sum of £1,000.

It may seem like a lot of money, but just think, you could make it all back in a single bet. I know you'll need convincing after all, who wants to part with a thousand pounds without some sort of guarantee so I'll give you the first prediction for free. In fact, I'll send you one free prediction every week for the next five weeks, and when you see I'm 100% accurate I can guarantee you'll be praying there is still a copy of the program left.

Imagine you received this email from me just before the start of the football season. Would you have taken up my offer and sent me £1,000? Of course not. But what if my first prediction had been correct? Luck, you'd have thought. But what when the second and third predictions were also correct? What about when all five turned out to be correct?

At some point, surely even you would have been convinced. After all, no one's luck is that good. Is it?

So, do you take me up on my offer: yes or no?

Well, if you said yes, I'm afraid you've just thrown away £1,000. Because there is no software, no statistical method. In fact, there isn't even luck. Just quantity. What you don't know is that I sent the same email to a million people at the start of the season. I divided them into three groups and told one group that one team would win, one group that the other team would win, and the third group that the match would be a draw. Once the result was in, I abandoned the two-thirds whose predictions ended up being wrong, and sent new predictions for week 2 only to the third of a million who had the right result in week 1. After the second match, one third of those people had been given two correct predictions, so I focused my attentions on these. After week 3, we're down by a third again, and the same after week 4, but at the end of week 5, there are around 4,000 people who have still only received correct predictions.

But these 4,000 people are seriously impressed, and even if only a few percent decide to take up my offer, I'll still make around £100,000!

But why would you fall for such a scam? Well, for a start, because you don't know about all the other emails and so you have no idea that you are simply one tiny part of a huge operation. Even if your natural caution would have stopped you from sending £1,000 to a complete stranger even such an impressive and believable stranger your own perceptions will have pulled plenty of similar scams on you. What about the times you've been impressed by astounding coincidences, without stopping to think just how many opportunities for coincidence there are in each of our lives? We notice the one extraordinary event, the occasional collision, but not the hundreds of thousands of near misses.

Another reason is that we seem to be constituted to try to find patterns in things. And, after all, there was a fairly impressive pattern here ... perhaps also reinforced by greed and the desire to get something for nothing.

Coincidence, correlation and chance



sometimes it's good to notice things

There is no question but that we have evolved to be noticing animals. Every observation contains an opportunity: maybe a new source of food, early warning of a predator, or just some regularity or change we could exploit. Information, in fact. We notice first, and reason later ... if ever.

So we are biased in favour of seeing something rather than nothing, pattern rather than randomness, certainty rather than uncertainty, existence rather than nonexistence, predator in the bushes rather than random pattern of leaves and shadows, cluster of edible berries rather than random colouring of earth.

In the natural world, individuals who miss opportunities tend to be out-evolved, which is just a nice way of saying that they don't survive. And every single person now alive comes from an unbroken line of survivors. This has led to us being quick to jump to conclusions, quick to spot or even to invent patterns. And when we see a pattern, we intuit a cause; we don't just say, "how extraordinary" and move on, we want to go and look, and find out if there's something there for us.



how unlikely is that?

Now let's return to the modern world. What if you shuffled a pack of cards thoroughly and turned over the top 10 cards and they were all red? What if you rolled a die 10 times and 6 came up 4 times? What if you flipped

Coincidence, correlation and chance

a coin ten times and turned up 8 heads? What would you expect to get if you flipped again? What if the numbers 1,2,3,4,5,6 came up on the Lottery this Saturday? Or what about six numbers which appear to be random until you happen to remember that these are the same numbers that came up in the previous draw? Does this make them more likely or less?

Or is your head just spinning?

These ideas of randomness, coincidence, uncertainty, what is remarkable and what commonplace, are very slippery and, not surprisingly, we make the most elementary mistakes when we have to deal with them. This is where we're most easily confused and fooled, because we didn't evolve to be good at probability. We didn't evolve to work out the odds on 8 heads out of 10.

If we want to face the challenges of the world we live in now, we can't rely on our intuition, because our intuition is matched to a very different world. We have to be cleverer than that: we have to reason.

Correlation?

Now let's move along to something along a slightly different path and we'll start by spending some time with my neighbours.



First up is Brian, the boy next door, who has just been given a guitar and amp for his birthday. Now each evening I'm treated to a mixture of Brian and Kurt Cobain pounding against my walls at top volume, over and over again as Brian dreams away his musical shortcomings. But I have some bad news for Brian. I've recently discovered a cast-iron statistical correlation between loud music and acne. I've meticulously gone over the figures several times, and there's no question – the more loud rock music you play, the worse your acne is. And with Brian currently desperate to catch the eye of young Jenny from number 37, I'm sure once I've shown him my research we'll have the volume down in no time.

And talking of Jenny, there she is now, buying an ice cream from the local shop. With her family about to go out to Australia for a holiday, I ought to go and warn her that the more ice creams there are sold, the more shark attacks there are. Again, I've done my research quite thoroughly, and the numbers do not lie. Perhaps I should recommend an apple instead!

Finally, let's pop into my local primary school to chat to the head teacher. I want to tell her about research I've uncovered which shows a clear and proven link between literacy levels and hand size in children. Bigger hands make better readers, it seems. With my son starting there in the autumn, maybe now is the time to set up some sort of hand-stretching programme perhaps on Wednesday afternoons, now that PE's been

Correlation?

Coincidence, correlation and chance

scrapped?

These examples may seem bizarre and improbable, but they are not the result of bad statistics. All the information is absolutely correct. Their strangeness comes from our own reasoning. We see two things changing together and our instinct is to assume that they are tied by cause and effect. Unfortunately, our instinct is often wrong. In all these examples a third "confounding" variable is actually the cause of two correlated variables.

It is absolutely true that people who play loud music are more likely to suffer from acne, but only because teenagers make up a big part of both groups. Acne and loud music are certainly correlated. But correlation is not causation. The same thing is true with the sharks and ice cream. The number of shark attacks and ice creams sold both go up during the summer, with the good weather encouraging people both to go in swimming and to eat ice cream. And as for large hands? Older children are bigger, and can read better!

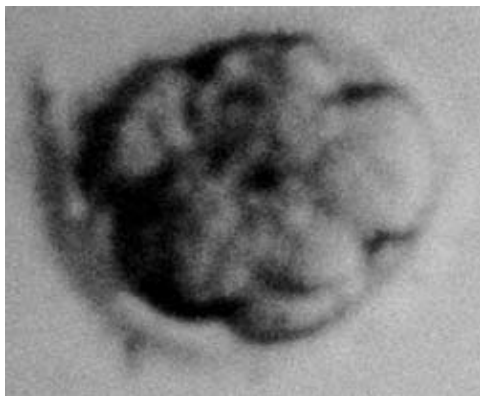
But the situation is not always so clear cut. For example, it was reported recently that mobile phone use can reduce a man's sperm count by almost a third. One plausible theory is that damage caused by electromagnetic radiation makes you infertile. But what about other factors – confounding factors – that could cause both high mobile phone use and low sperm counts?

How about stress? Busy jobs? Exposure to pollution? Age – were the heavy users older? Were they more likely to be smokers? The researchers can't answer these questions; not unless they carry out a carefully designed experiment that controls for all these other variables.

And we must not forget the possibility that this was a random finding in one small study – just 221 men which won't necessarily be replicated by other researchers. If you conduct enough studies, you will inevitably get some interesting findings just by chance.

We all draw conclusions on the basis of what we see. But it is important for us to remember that just because there is a correlation between two facts, there isn't necessarily a cause/effect relationship between them.

Chance?



James, two days after conception

My son, James was conceived by IVF. I have a photo of him when he was just seven cells big and I often look at it and think about how lucky I am to have him. Twenty years ago, the chances of a couple like myself and my wife ever having a child were tiny. Fifty years ago, they were zero.

Coincidence, correlation and chance

But just how lucky were we? Only about 1 IVF treatment in 7 leads to a baby, so does that mean that there was a 1 in 7 chance that we would have James? No, because that 1 in 7 is the chance of an IVF cycle leading to *some* child, not to that particular unique little individual. That chance is so much smaller that it's hard to even get a handle on it, because although James was conceived in a Petri dish, he still shares something with all of us: he is the result of a series of events so unlikely that you'd wonder how anyone was ever born at all.

Take yourself as an example. What if your parents had never met? What if one of them had missed the bus on their way to their first date? What if their eyes hadn't met across a crowded room because someone else was in the way? What if they'd been tired, or had had an argument the night you were conceived and had just gone to sleep instead? And that's all before we get into a biological discussion about the number of potential results arising from each single ... romantic encounter.

But of course people are born all the time. At the moment there are well over 6 billion of us, and each second we're joined by another two-and-a-half newcomers. So how can something which seems so unlikely happen so often? Well, it's all down to probability.

Before you were born, the probability of you being born was almost 0. After you were born, the probability was 1 that is, a certainty. The point is, someone was going to be born, and if it hadn't been you, it would have been a different child. You would never have existed, and that other child would now be reading this article and asking themselves what the chances were that they were ever born.



the venue for this thought experiment

Let's take another, sadly fictitious, example. On Saturday afternoon, you and I each buy a lottery ticket for that night's draw. Each of us has a chance of about one in 14 million of winning. A few hours later, you're tearing up your ticket in disgust and I'm running around and screaming at the top of my voice, several million pounds better off than I was that afternoon. The chance of winning the lottery with a single ticket have not changed, however; it's still 1 in 14 million. But the probability of *me* winning today's draw is now 1 because I won, and the probability of you winning it is now 0 there's no possibility you can win because you don't have the right numbers.

And this distinction, between the chances of a rare thing happening before it happens, and once it's happened, is one way we're easily misled when it comes to our understanding of probability. The thing to remember is

Coincidence, correlation and chance

that rare things happen all the time. People do win the lottery, people do have meteorites land in their lounge, people are struck by lightning twice. These things may not have happened to you, and they may well never happen to you, but that doesn't mean that they never happen at all.

As an experiment, why not think of something unusual that's happened to you this week then imagine that someone had told you beforehand that it was going to happen. Would you have believed them? Would you have put money on it not happening? And if so, how much would you have lost?

About this article



Andrew Stickland with his son, James

Andrew Stickland is a freelance writer and copyeditor. He has written articles for various publications on subjects as diverse as travel and war-gaming, and regularly reviews some of the most accessible mathematical books and films, as well as providing the occasional photo for *Plus*.

The genesis of this article was in three pieces, co-written by Andrew and Helen Joyce, the editor of *Plus*, for the Radio 4 programme, *More or less*. They aired during the summer of 2004, and can still be listened to on the [More or less](#) website. Andrew and Helen would like to thank Michael Blastland, the producer of *More or less*, for his helpful comments and input.



Plus is part of the family of activities in the Millennium Mathematics Project, which also includes the [NRICH](#) and [MOTIVATE](#) sites.