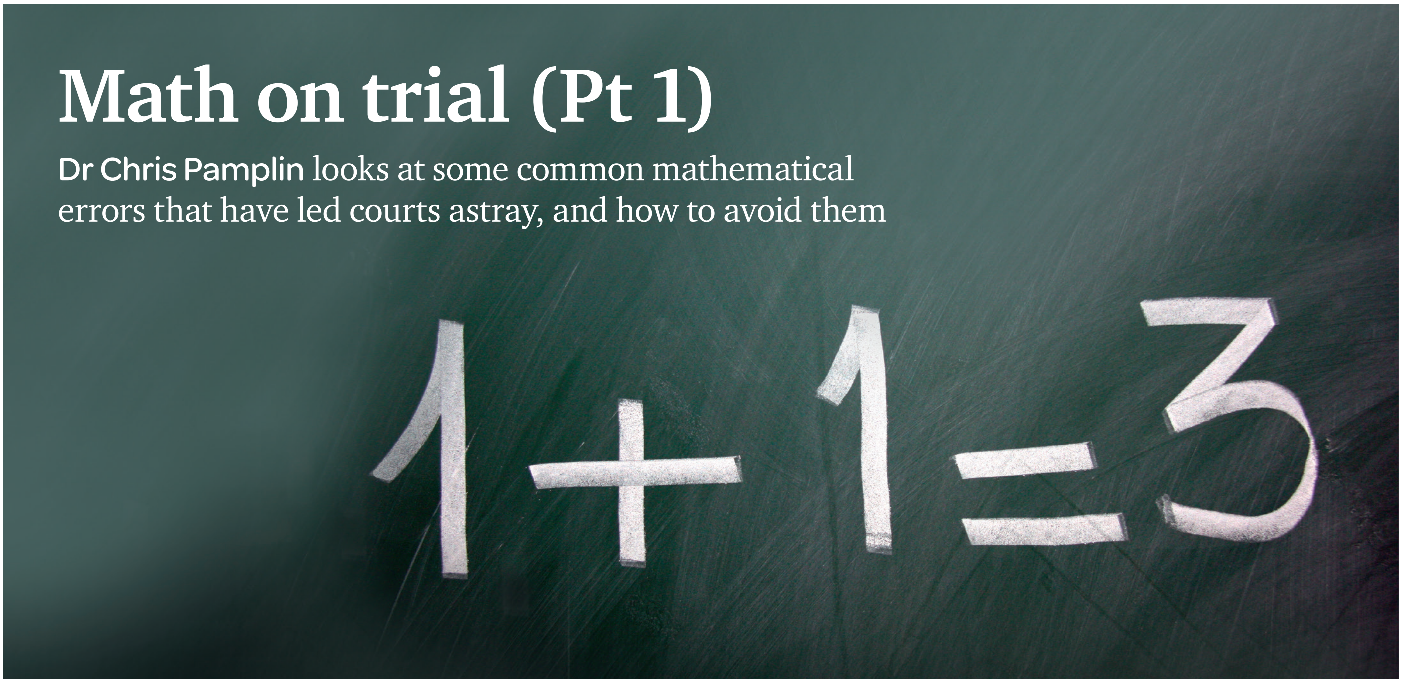
Math on trial (Pt 1)

Dr Chris Pamplin looks at some common mathematical errors that have led courts astray, and how to avoid them



© iStockphoto/virtualphoto

Math on Trial (Schneps, L & Colmez, C, 2013, Basic Books) is an excellent book that catalogues the use—or perhaps that should be misuse—of mathematics in the courtroom. While the publication is well worth reading in its entirety, the purpose here is to summarise the 10 common mathematical errors the authors distil from the legal casebook.

As the authors say, "despite their ubiquity...most of these fallacies are easy to spot". This two-part series offers your very own fallacy-spotting crib sheet.

Error no 1: multiplying non-independent probabilities

Sally Clark was a solicitor who in 1999 was found guilty of the murder of two of her sons. At trial, Professor Sir Roy Meadow, a leading paediatrician, gave evidence for the prosecution. It was his introduction of a published statistic on the likelihood of two cot deaths occurring in one family—given as 1 in 73 million—that is the focus here.

When two events are unrelated, the probability of both events occurring is simply the probability of each event occurring multiplied together. So, if a woman is pregnant with a single child, she has a 1 in 2 chance of having a girl (actually a 50% chance of a girl is only approximately true, but let's not allow biology to distract us). With two children born at different times, the probability that both her children are girls is $(½)^2$, which is 1 in 4. But this simple and common enough calculation is only valid when the two events are independent of each other.

So what's the problem when such an approach is used in cot death cases? Answer: there is no way of knowing that the events are independent. Cot deaths should properly be classified as sudden infant death syndrome (SIDS), and the crucial point is that the death is unexplained. If it transpires that there is an underlying genetic cause, a common environmental cause or, *in extremis,* the mother has been killing her babies, the events are most definitely not independent. In such circumstances, the simple multiplication of probabilities will significantly underestimate the likelihood of both events occurring.

The lesson to learn here is the need to be sure events are truly independent before using the multiplication of their probabilities to work out the chance of all the events happening.

Error no 2: unjustified estimates

The case of Los Angeles resident Janet Collins, in which a fleeting glimpse of an assailant led to wildly unjustified estimates of physical traits in the local population, is used to exemplify error no 2. Put simply, it is the tendency for a number—any number—to add an air of scientific credibility to an argument.

The frequency with which numbers placed in the public domain are plain wrong—whether intentionally, accidentally or through ignorance—is shocking. For example, the authors of *Math on Trial* cite a 2010 Conservative party report that stated under the Labour government 54% of girls in the 10 most disadvantaged areas of England became pregnant under the age of 18. The statement was wrong; the correct figure was 5.4%. The Conservatives swept aside the misplaced decimal as unimportant in the overall conclusion of the report. Clearly unjustified estimates weaken our ability to assess the numbers we are given.

This kind of mathematical error is not new to the English courts. In *R v T* [2010] EWCA 2439 the court of appeal looked at identifying shoe prints by comparison with a database compiled by the Forensic Science Service. The court found the quality of this database to be so poor that any attempt to assess the probability that a given shoe could have made a particular mark based on figures relating to shoe distribution was inherently unreliable.

So lesson 2: always check that estimates are properly grounded in reality.

Error no 3: getting something from nothing

In another US case, Joe Sneed was convicted of murdering his parents on the basis of a probability calculation. Not only did the calculation make error nos 1 *and* 2, but it also introduced a probability estimate based on *not* finding any matches in the sample examined.

How one should treat the findings from a sampling exercise is highly dependent on the distribution of the feature being sought in the overall population. The prosecutor in the *Sneed* case was trying to calculate the probability that two people would share a number of physical traits, known movements and actions around the locality and, crucially, share the same name.

On that last element (a shared name), the prosecutor turned to the phone book. The court examined several telephone directories from the south-western US. The target surname was not found in any of them. With the examined directories containing around 1.2 million names, the prosecutor estimated the frequency of the surname in the general population to be around one per million.

The absence of the surname in the sampled directories gives very little information about the frequency of the name across the whole country. This is because that sampling approach makes the unjustified assumption that surnames are evenly spread across the country. As anyone with a large extended family and an uncommon surname will know, that is simply not true. One cannot draw any conclusion more precise from not finding the name in any of the telephone books than to note the name is not very common.

The lesson to learn here is that if the court is to know how it should treat the result of a sampling exercise, it must know the distribution of the item of interest in the whole population.

Error no 4: double experiments

The court's handling of the case of Meredith Kercher's murder in Italy highlights the next error: the belief that running a test a second time on the same item will provide no more evidence than did the first test.

Suppose a positive blood test gives a 60% probability that a suspected illness is really present. If the first test comes back positive, you can be 60% sure the illness is present. What can be concluded if the test is repeated and another positive result obtained? Are you still 60% certain, or are you even more certain?

Running a test twice can give far more information than one might think. Take two coins, one fair, the other weighted so that it comes up heads 70% of the time. Choose one coin. After the first toss lands heads, the maths tells us we can be 58% certain the coin is weighted. Tossing it again and getting heads now tells us that there is a 66% chance the coin is weighted.

In the Kercher trial, the judge made the error of assuming that a repeat DNA test on tiny samples of DNA taken from the supposed murder weapon would provide no more information than had the first test. The judge said: "The sum of the two results, both unreliable due to not having been obtained by a correct scientific procedure, cannot give a reliable result." That reasoning is to misunderstand the potential for the separate results from two iterations of the same test to add information. As the coin toss example shows, independent runs of a test of moderate reliability can, indeed, in total, give more reliable results.

The lesson here is that probability is a delicate subject that often runs counter to human intuition. Used properly, multiple runs of a test can increase the statistical reliability of an otherwise uncertain test.

Error no 5: the birthday problem

The improvement in DNA analysis in recent times has increased interest in cold cases. The unsolved case of the murder of nurse Diana Sylvester in California in 1972 is one such. A search of a large database of DNA profiles of convicted criminals in California against a degraded DNA specimen from the murder scene returned a single partial DNA match. What should be made of that match?

The prosecution presented data to show that for a partial match on a specific set of nine out of a possible 13 DNA peaks we can expect to find about 1 person matching out of 13 billion. But the defence presented data to show that in a database of 65,000 DNA profiles, more than 100 pairs matched at nine peaks. What's going on? How can a chance of finding a single match out of 13 billion people—a very small chance indeed—be reconciled with finding 100+ pairs of people who match at nine peaks in a database of 65,000 people?

The answer lies in the number of people you need in a room to have a better than evens chance of two of them sharing the same birthday. To the surprise of many, the answer to that question is 23. But that is only if you do not fix the birthday you are seeking. If you fix the shared birthday to, say, 1 January, the room gets much more crowded—you'd need 253 people. This is because you cannot now pair everyone with each other.

What the 1 in 13 billion probability is measuring is pairs of individuals, and the number of pairs of individuals in a population is far higher than the number of individuals. With 65,000 people on a database, the number of pairs is ~1.5 trillion. At a match probability of 1 in 13 billion, you would expect to get 116 pairs of individuals who match.

So lesson 5 is: beware two propositions that sound similar but are actually quite different. **NLJ**

Dr Chris Pamplin is the editor of the UK Register of Expert Witnesses and can be contacted on *nlj@jspubs.com* (*www.jspubs.com*)

Coming up next time...

The next instalment will look at:

- Simpson's Paradox, which occurs when trends mysteriously vanish.
- The conviction of Lucia de Berk as a serial killer of children based on retrospective thinking.
- The power of very large numbers to confound us poor humans.
- The significance of the fact that mathematical models are always simplifications of the real world.
- The fact that unlikely events are not always uncommon.

NewLawJournal

Leading on debate, litigation and dispute resolution

Earn free CPD points at Newlawjournal.co.uk

A New Law Journal subscription provides you with free access to **Newlawjournal.co.uk** – where you will find exclusive content and can earn CPD points.

Visit www.newlawjournal.co.uk today to:

- Access the latest news plus exclusive online content to stay in the know
- Search through the extensive New Law Journal archive
- Earn up to two CPD hours per month by completing the online multiple-choice questionnaire

LexisNexis®

For your login details please contact additionalusers@lexisnexis.co.uk