STATISTICS
by Thomas Lumley

Numbers game

From a coin toss to a computer program, generating random numbers is an unpredictable science.

Statistics uses both real and theoretical randomness for a lot of things, from selecting phone numbers in polling and allocating treatments in clinical trials to proving that a set of mathematical assumptions does or doesn't let you distinguish correlation from causation. So what do we think "random" actually means?

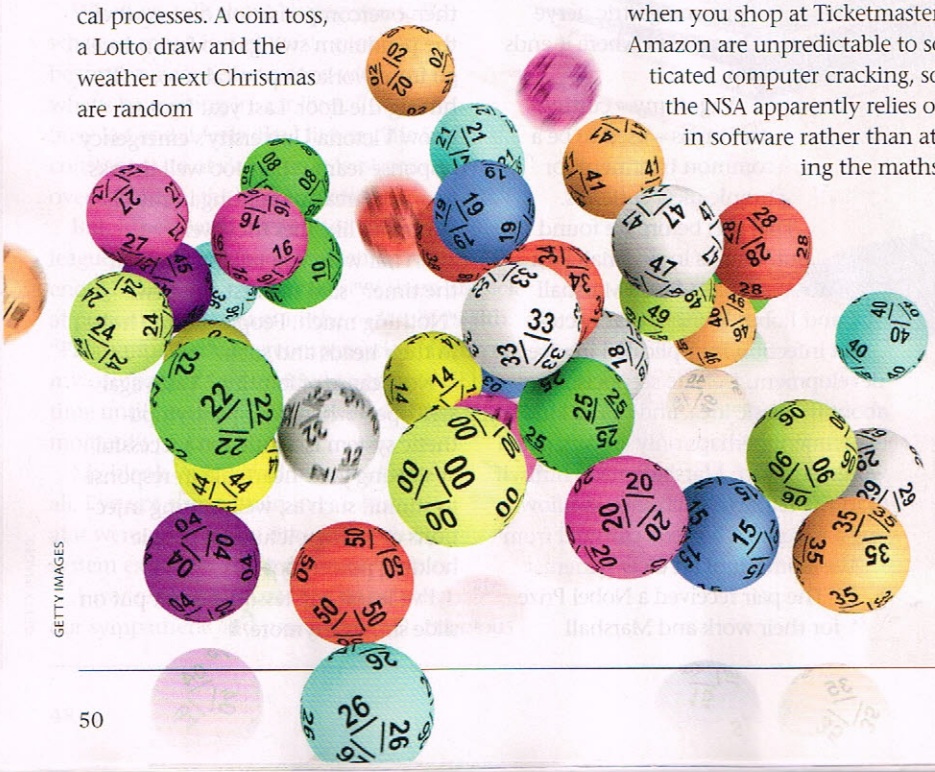
Most of the time, we don't think about it much. It's like electricity: readily available when you need it, without having to think about Lake Tekapo or Otahuhu B. Sometimes, though, you do care where your random numbers are coming from and what that implies.

"Random" basically means "unpredictable". This leaves open the question "Unpredictable by whom?" and that's where the interesting issues arise.

Sometimes we can get randomness from physical processes. A coin toss, a Lotto draw and the weather next Christmas are random because of physical processes that amplify immeasurably tiny differences. One unusual random-number generator used a webcam focused on a lava lamp, exploiting the chaotic physical processes of the rising and falling wax. Some of the same scientists then worked out that you could get even better results without the lamp. Leaving the cap on a digital camera and dialling up the brightness and contrast gives an image that's pure noise, again from chaotic physical processes.

In contrast to the lava lamp, this one's actually usable in practice. Theoretically, processes driven by quantum-mechanical effects are even less predictable. According to the standard interpretations, they're unpredictable given any amount of prior information. As Einstein disapprovingly put it, the theory says even God plays dice.

Computers are designed not to behave unpredictably, so the typical random generators on a computer aren't as solidly unpredictable. Many people won't even call them "random", preferring the term "pseudorandom". They're still good enough for many purposes. The typical random-number generators on a computer are unpredictable if you don't get to see much of the output or if you're not trying to crack them. The properly designed cryptographic random-number generators used when you shop at Ticketmaster or Amazon are unpredictable to sophisticated computer cracking, so even the NSA apparently relies on bugs in software rather than attacking the maths.



GETTY IMAGES

In Lotto, a random-number generator could be enough to stop the draws having exploitable patterns but not enough to allay fears of cheating.

What sort of random numbers you want depends on how much unpredictability you want and how many numbers you need. In a computer-simulation experiment, there's no need to be paranoid: you want numbers that satisfy simple statistical tests of uncorrelatedness and you want them as fast as possible. The same is true in opinion polling, where you just want to make sure every number has the same chance of being called.

In a clinical trial, you want the treatment allocation to be unpredictable to doctors and patients, so using a simple random-number generator inside the trial-enrolment website is sufficient, but using sealed envelopes based on high-quality random numbers is not. It turns out that people cheat.

In Lotto, a cryptographic random-number generator would be enough to stop the draws having exploitable patterns, but not enough to prevent gamblers suspecting that the lottery company was cheating. The big, complicated, physically chaotic machine, the official observers and the televised draws are to make the results obviously unpredictable to Lotto New Zealand.

But if you want to decide who kicks off and which end each team takes in a rugby game, a coin toss is a pretty good random-number generator. It's not perfectly even – according to US statistician Persi Diaconis there's about a 51% chance that the coin lands the same way up that it started – but it's good enough for the purpose. ▮