The magic of numbers

Randomness and coincidence have fascinated Persi Diaconis as a magician and a mathematician. Jenny Rankine explains.

NOTABLE MATHS PROBLEMS

YANG-MILLS THEORY & MASS GAP

Prove that for any compact simple gauge group G, quantum Yang-Mills theory of R^4 exists and has a mass gap $\Delta>0$.

Simply: Yang-Mills theory is the quantum field theory (QFT) underlying the Standard Model of particle physics. It is a generalization of the Maxwell theory of electromagnetism. As a classical field theory it has solutions which travel at the speed of light, so its quantum version should describe massless particles (gluons). However, the phenomenon of colour confinement permits only bound states of gluons forming massive particles. This is the mass gap. The problem is to establish rigorously the existence of the quantum Yang-Mills theory and a mass gap.

Originators: Chinese physicist Yang Chen-Ning Franklin, and American physicist Robert Mills tried to extend the concept of gauge theory for an Abelian group in 1954. It didn't work as predicted; the idea was shelved until the 1960s when symmetry-breaking in massless theories was initiated by J. Goldstone, Y. Nambu and G. Jona-Lasinio, with particles acquiring mass via Yang-Mills.

Incentive: $US 1million, one of seven Millennium Prize Problems of the Clay Mathematics Institute.

Exploration: This active field includes attempts using lattice gauge theory and four-dimensional supersymmetric theories. A solution will require fundamental new ideas and would demonstrate key physics of the forces binding protons and neutrons into nuclei.

Interesting aspect: Has led to new discoveries in quantum groups and new QFT mathematical structures like the Jones polynomial.

Diaconis exhibits the multiple paths one can take to mathematics. He left home and high school in New York at 14 to apprentice himself to travelling sleight-of-hand magician Dai Vernon, and by 17 was making a living on his own as a magician.

The pair hunted down crooked gamblers, also skilful at sleight of hand. Diaconis was seduced by their conversations about calculating odds and bought a book on probability, only to find he couldn't understand the mathematics.

So he enrolled at night school to learn calculus and made his living as a magician by day. In his mid-20s, he was accepted into Harvard University as a graduate student, despite his lack of high school qualifications and minimal maths skills, partly on the basis of two original card tricks he had submitted to the Scientific American puzzle page.

Now a Professor of Statistics and Mathematics at Stanford University in California, he has specialised in the mathematics of randomness, including card shuffling and coin tossing, and used statistical techniques to test and debunk professional psychics.

Diaconis is still one of about 100 people in the world who can riffle shuffle a deck eight times in under a minute and return the cards exactly to where they started. He also taught himself at 13 to toss a coin so that it always landed as heads.

He and Dave Bayer proved in 1992 that it takes seven riffle shuffles (splitting the pack and interleaving both sets of cards) to randomize a deck, and about four for games like blackjack where suits don't matter. As a result Nevada changed its laws for casino games and the American Contract Bridge League its rules to require seven shuffles.

The maths of shuffling turned out to apply to Markov Chain Monte Carlo simulations, a class of sampling algorithms that use randomness to solve problems in biology, chemistry, physics and linguistics. The methods Diaconis developed to recognize when a dealer can safely stop shuffling cards also tell when a computer can stop running a simulation.

At packed public meetings at Albany and Palmerston North in January, he demonstrated the magic of numbers to academics, magic fans and magicians. He shuffled a pack of cards, wound a rubber band around them and tossed them to someone in the front row.

They tossed it randomly to the back and then he asked five different people to cut the cards and toss them back to him. He asked those people to concentrate on their card (cue laughter from the academics), and then asked those who had cut a red card to stand. He then correctly named all the cards the people had seen.

It wasn't a trick, just simple maths. The deck had been arranged so that every possible combination occurred only once, and the pattern of red cards enabled him to calculate how the deck had been repatterned as it was cut.

He is also fascinated by "problems that are simple enough to say in English but that are hard to do", and is investigating the maths of the carry numbers used in addition. For example, to add eight and five, we write three down and carry one. "It turns out that carries sit in an esoteric corner of group theory called group cohomology, how groups fit together," he says.

At this year's NZIMA maths conference in Hanmer Springs, Diaconis gave three talks about carries, shuffling and their relationship. "New Zealand is terrific at group theory; I posted questions and people suggested things I could try. I made real progress talking with New Zealand mathematicians."

$\frac{3}{2}\log_2 n + \theta$ shuffles are necessary and sufficient to mix up n cards

Hybrids and genetic trees

Barbara Holland is enthusiastic about the mathematical possibilities of phylogenetics, the building of evolutionary relationships between different species, with the flood of new genetic information. Jenny Rankine reports.

"The Solexa 2 DNA analyser produces terrifying amounts of data! It really excites me figuring out how to use that data across genomes, and then understanding what may end up being much more complex patterns of evolution than we previously thought."

Phylogenetics is a blend of mathematics, biology and computation. Holland uses stochastic models, graph theory, optimisation and combinatorics in her work at Massey University in Palmerston North and with the Allan Wilson Centre for Molecular Ecology and Evolution.

"It's a tricky business trying to puzzle out what happened 100 to 500 million years ago just using modern DNA sequences. It's slender evidence for such vast timescales." She describes testing mathematical models as "like a fat man in a women's lingerie store; there might an outfit that fits best but that doesn't mean it fits well".

One questionable assumption used by phylogenetic models is that evolution is the same over all parts of the tree. "It's ridiculous," says Holland; "lineages have their own properties". She gives the example of parasites, which use their hosts' functions. This makes many of their genes redundant, reduces selection pressures, and enables them to survive more genetic mutations than other organisms.

Holland and PhD student Liat Shavit changed an existing phylogenetic simulation package so that the same DNA site could accept mutations on only some parts of the evolutionary tree, instead of over the whole tree. A large simulation, faking evolution over and over, found that "it was reasonably difficult to make phylogenetic methods break - it took major deviations before the methods produced inaccurate trees".

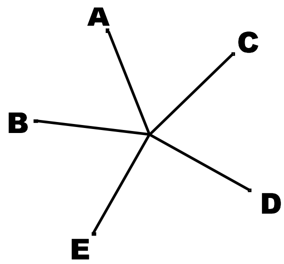
Holland is also fascinated by hybridisation, which is common where species respond rapidly to new environments. Hybrids inherit genetic material from two parent species, making them hard to detect when studying single genes. "Every extra hybridisation doubles the number of evolutionary trees; it's a very challenging problem," says Holland.

However, DNA technology now enables the study of many genes simultaneously. Another problem with current mathematical models is their assumption that any conflict in trees is due to hybridisation.

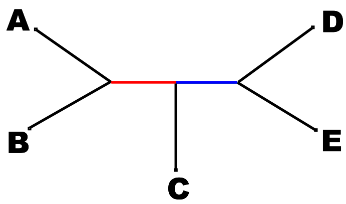
In reality, conflict between possible trees can be caused by estimation errors, missing data, and the random nature of inheritance in populations. Holland has used consensus networks and graph theory to represent these conflicts visually. "Trees are collections of vertices and edges. External vertices correspond to modern species with labels. Internal vertices represent common ancestors. A four-species tree will have two internal vertices and five edges connecting everything. Removing any edge breaks the tree into two parts with two different labels."

"The most common method is majority rule; it's quite restrictive and you end up throwing out a lot of information. Using consensus networks relaxes this rule, showing any edge that appears in more than, say, 33% of the trees. It shows more flexibly where sets of trees agree and disagree." The 33% rule leads to pictures of two-dimensional boxes; a threshold of 25% can be drawn as cubes.

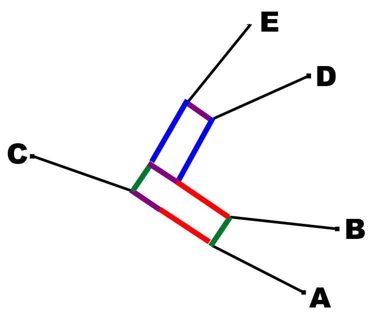
As a mathematician, Holland can make statements like "in animals, our concept of species may make sense, but in plants, it doesn't. However, asking biologists what is a species is like throwing in a hand grenade; they disagree strenuously." Along with Lara Shepherd she's exploring trees for the New Zealand five-finger genus, where she thinks hybridisation has been important, and developing statistical methods and software to show how important hybridisation has been in shaping New Zealand plant and animal life.



(100%) Strict Consensus tree



(>50%) Majority-rule Consensus tree



(≥33%) Consensus network