Confused at a higher level

Twenty years ago, an analysis of flight duration data recorded automatically from albatross in their ocean environment turned up an intriguing phenomenon. Like many organisms, albatross move erratically as they explore territory, foraging for food, among other things. To a first crude approximation, one might expect bird movements in a relatively homogeneous environment to follow something like a diffusive Brownian motion, and yet the data showed something very different. Albatross seemed to be following a distinct statistical pattern known as a Lévy flight, in which the distribution of flight durations isn't a normal Gaussian, but has broad power-law tails, indicating that albatross frequently take very long flights or excursions into new territory.

Although that finding later proved controversial due to problems in data collection and interpretation, other research soon found what looked like plentiful evidence for Lévy flight dynamics in the movements of many other organisms, including deer, bumblebees, spider monkeys and many marine species. Meanwhile, mathematical results suggested that such behaviour would, in some cases, reflect an optimal foraging strategy. Many organisms, it appeared, were exploring their territory efficiently.

It was a nice story, and yet, as often in science, a little too nice to be the whole story.

Since then, further research has made the story more complicated, not less, and has identified two key issues that early work on movement had overlooked, or didn't emphasize enough. First, the statistical analysis of stochastic processes based on empirical data is perilous, and it is easy to come to false conclusions. Today, animal foraging looks to be consistent with any of a broad class of stochastic processes that generate anomalous diffusion — a deviation from familiar Brownian motion. Second, it's not enough to treat biological data as many early studies did — as if organisms are simple point particles. Fruitful research has to pay much closer attention to the behavioural repertoire of different organisms, which reflect their adaptation to an evolutionary and ecological niche.

On the mathematical side, a major theme of modern statistical physics is the non-ergodic behaviour of many systems in physics and biology, finance or economics. In an ergodic system — a particle undergoing Brownian motion, for example — time averages and ensemble averages are equal. What will happen on one trajectory can be calculated reliably by averaging over the ensemble of probable paths. It works because the dynamics explore phase space in a complete and uniform way. All this breaks down in a non-ergodic system, in which what happens now has a lasting influence over the probabilities of future outcomes.

This issue turns out to be crucial to the interpretation of movement data. Many early analyses of movement time series estimated quantities such as the time-averaged mean square displacement over a time interval, and used this to infer the underlying probability distribution function of an assumed statistical process. This is a dubious step because many simple stochastic processes don't have equal ensemble and time-averaged observables (R. Metzler et al., *Phys. Chem. Chem. Phys.* **16**, 24128–24164; 2014). Indeed, there's no reason even to expect empirically evaluated time averages to converge to reproducible values. This has nothing to do with limits of data, but with the peculiar features of the statistical processes and physical mechanisms driving the dynamics.

Many studies have tracked the locations of animals and analysed the recorded data, treating the organism as a point particle.

In a recent book chapter, physicist Rainer Klages has reviewed some of the unfortunate confusion engendered by a broad failure to recognize these issues (preprint at https://arxiv.org/abs/1804.03738). It's easy to record data, make a histogram and estimate a probability distribution function for the movements — and get results that are entirely mistaken. As Klages points out, really making progress on the statistical characterization of movement data requires identifying the right underlying stochastic process. In other words, empirical studies need more mathematical sophistication.

But previous work on movement has suffered from biological naiveté as well. As Klages notes, many studies have tracked the locations of animals and analysed the recorded data, treating the organism more or less as a point particle lacking differentiated goals, strategies, habits or behavioural predispositions. Yet biologists in recent years have pursued a more informed view, acknowledging that distinct segments of an organism's movements often reflect different phases of behaviour (R. Nathan, *Proc. Natl Acad. Sci. USA* **105**, 19050–19051; 2008). For example, a newly hatched bird may at the outset forage for food, then shift to making excursions designed solely to help it to learn and gain experience of its environment, and then later still migrate to wintering grounds far away. It's a mistake to see all parts of a time series of data as reflecting a single behavioural process.

A good analogy that Nathan invokes is DNA. One might well analyse a DNA strand as a simple linear string of base pairs, and that's legitimate and useful. Yet we know thousands of short segments with specific structures and meanings — genes. DNA analyses that ignore the existence and significance of genes would be seriously impaired, and so too with the analysis of movement data. Biologists refer to 'canonical activity modes' as key behavioural episodes — runs and tumbles for bacteria, or standing, walking and running for larger animals. A lifetime movement track may consist of a patterned sequence of such modes, forming higher-level structures that also carry meaning — migration to a mating territory, for example.

Of course, all this takes place in the context of environmental influences, which may affect movement data just as much as behavioural traits. Klages argues that we'll only begin to deeply understand movement dynamics through studies that include environmental influences, work from an informed biological view of behaviour, and also become more sophisticated about the analysis of stochastic processes. In previous work, frequent miscommunication between theorists and experimentalists has resulted in incorrect or misleading conclusions. In particular, too many studies have settled on the finding of Lévy Flight dynamics when the data is equally consistent with many other non-ergodic statistical processes — continuous random walks, for example, or fractal Brownian motion.

So our understanding of biological movement may, in a paradoxical way, be more confused today than it was a couple decades ago. But, to borrow a phrase from Enrico Fermi, it may be "confused at a higher level", and that's a good thing. ❐

Mark Buchanan

https://doi.org/10.1038/s41567-018-0333-3