1

Self-organisation and emergence

Mario Nicodemi, Yu-Xi Chau, Christopher Oates,
Anas Rana and Leigh Robinson

Abstract

Many examples exist of systems made of a large number of comparatively simple elementary constituents which exhibit interesting and surprising collective emergent behaviours. They are encountered in a variety of disciplines ranging from physics to biology and, of course, economics and social sciences. We all experience, for instance, the variety of complex behaviours emerging in social groups. In a similar sense, in biology, the whole spectrum of activities of higher organisms results from the interactions of their cells and, at a different scale, the behaviour of cells from the interactions of their genes and molecular components. Those, in turn, are formed, as all the incredible variety of natural systems, from the spontaneous assembling, in large numbers, of just a few kinds of elementary particles (e.g., protons, electrons).

To stress the contrast between the comparative simplicity of constituents and the complexity of their spontaneous collective behaviour, these systems are sometimes referred to as "*complex systems*". They involve a number of interacting elements, often exposed to the effects of chance, so the hypothesis has emerged that their behaviour might be understood, and predicted, in a statistical sense. Such a perspective has been exploited in statistical physics, as much as the later idea of "universality". That is the discovery that general mathematical laws might govern the collective behaviour of seemingly different systems, irrespective of the minute details of their components, as we look at them at different scales, like in Chinese boxes. While the single component must be studied on its own, these discoveries offer the hope that we might understand different classes of complex systems from their simpler examples.

A univocal definition of "complexity" can be elusive, but the above criteria hopefully draw a line to distinguish, in a more technical sense, "complex" from the much broader category of "complicated" systems. Here we introduce some of the basic mathematical tools employed to describe their emergent behaviours. We discuss some basic concepts and several applications (e.g., Brownian motion in physics, asset pricing in finance) of the theory of stochastic processes, which is presented more generally in Chapter 3. We also consider some more advanced topics such as statistical mechanics and its applications to define the emergent properties in interacting systems. The foundations of statistical mechanics are discussed in more detail in Chapter 4. Finally, we introduce more recent topics such as self-organised criticality and network theory.

The course was taught by Mario Nicodemi. The notes that form this chapter were written by Yu-Xi Chau, Christopher Oates, Anas Rana and Leigh Robinson, four students of the Complexity Science Doctoral Training Centre of the University of Warwick who attended the lectures in 2009.

1.1 Random walks

1.1.1 Introduction

Put simply, a **random walk** is a mathematical formalisation of a path a "particle" traces out after taking a sequence of random steps. The idea of a random walk is central to the modelling of a wide range of phenomena, including financial modelling, the diffusion of gases, genetic drift, conformation of polymers, and a large number of other applications where the phenomenon in question evolves by a random process in time.

Various different types of random walk exist but can be grouped into broad categories depending on what the random walker is said to "walk on", and how the time evolution is defined. For example, a random walker may be defined on a graph that evolves in discrete time, moving from one node to another in one discrete time step, or just as well defined would be a random walker that moved in continuous time along the whole real line, $\mathbb{R}$. We shall give no further thought to these kinds of random walks and restrict our discussion to ones that occur along the integers, $\mathbb{Z}$ in discrete time steps.

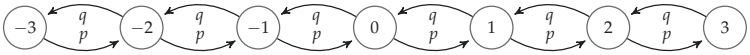

Figure 1.1 A discrete time walk on the integers, $\mathbb{Z}$. Probabilities p and q show how likely that transition is from state to state.

To understand perhaps the simplest example of a random walk we imagine a particle that can inhabit one of the integer points on the number line. At time 0 the particle starts from a specific point and moves in one time step to its next position in the following way: we flip a coin with the result governing how the particle moves. If the coin comes up heads then the particle moves *one* position to the right while if it comes up tails then the particle moves *one* place to the left. If we make n such coin tosses then what will be final position of the particle? Obviously being a random process we can't predict exactly where it will end but we can say a good deal about the distribution of possible outcomes.

1.1.2 One-dimensional discrete random walk

To try and answer such questions we need to introduce some formalism. We define independent random variables, X_i, that can take the values -1 and 1, with $P(X_i = 1) = p$ and $P(X_i = -1) = 1 - p = q$. The X_i represent the direction of the ith step of our random walk. Pictorially we can represent this arrangement as shown in Fig. 1.1. To see how such a system behaves statistically we calculate the first and second moments and variance of X_i as follows:

$$\langle X_i \rangle = \sum kP(X_i = k)_k = 1 \times P(X_i = 1) - 1 \times P(X_i = -1) \quad (1.1)$$

$$= p - (1 - p) = 2p - 1. \quad (1.2)$$

With the second moment and variance given by

$$\langle X_i^2 \rangle = p + (1 - p) = 1, \quad (1.3)$$

$$Var[X_i] = \langle X_i^2 \rangle - (\langle X_i \rangle)^2 = 1 - (2p - 1)^2 \quad (1.4)$$

$$= 1 - 4p^2 + 4p - 1 = 4p(1 - p). \quad (1.5)$$

We now define a new random variable, Z_n, as the sum of n such X_i variables and this defines the distribution of the value of the random

walk after n steps:

$$Z_n = \sum_{i=1}^{n} X_i, n > 0. \tag{1.6}$$

We can now look at some of the statistics of Z_n, in particular the average position, $\langle Z_n \rangle$, and the variance: $Var[Z_n]$ of this position:

$$\langle Z_n \rangle = \langle \sum_{i=1}^{n} X_i \rangle = \sum_{i=1}^{n} \langle X_i \rangle = n(2p-1), n > 0. \tag{1.7}$$

Since by definition each of the X_i's are independent the second moment can be easily calculated,

$$\langle Z_n^2 \rangle = \langle (\sum_{i=1}^{n} X_i)^2 \rangle = \sum_{i=1}^{n} \langle X_i^2 \rangle + \sum_{i=1}^{n} \sum_{j=1 i \neq j}^{n} \langle X_i X_j \rangle \tag{1.8}$$

$$= n + n(n-1)(2p-1)^2. \tag{1.9}$$

Hence,

$$Var[Z_n] = \langle Z_n^2 \rangle - (\langle Z_n \rangle)^2 = 4np(1-p). \tag{1.10}$$

In particular, notice that

$$Var[Z_n] \propto n, \tag{1.11}$$

which gives us the result that the variance increases as we walk for more steps. This has important consequences for finance as we shall see later.

Unbiased random walk

So far we have been considering a random walk with general transition probabilities, p and q. The special case where $p = q = 1/2$ is called *unbiased* – since each decision is equiprobable. For these random walks the statistical properties collapse to

$$\langle Z_n \rangle = 0, \tag{1.12}$$

$$Var[Z_n] = \langle Z_n^2 \rangle = n. \tag{1.13}$$

From (1.13) we note that the root-mean-square of Z_n is simply $\sqrt{n}$, which hints that the average absolute distance moved after n steps, $E[\mid Z_n \mid] = O(\sqrt{n})$. This is indeed the case, but will not be proven here. Trajectories for a collection of unbiased random walkers are shown in Fig. 1.2.

A further interesting property of unbiased random walkers is the notion of recurrence. Imagine we choose any point, $i \in \mathbb{Z}$. How many times would you expect the random walker to cross this point if the walker

could travel forever? Perhaps surprisingly the answer is that the walker will cross any selected point an infinite number or times. From a financial perspective this property has been termed *gambler's ruin*, since if you are a gambler betting in a casino on a *fair* game then your current wealth will evolve as a unbiased random walk and as such must eventually cross 0. At this time you have lost all your money and the "walk" cannot continue. All the casino has to do to force this win is to have substantially more money than you (to absorb the periods where you are winning) and entice you to keep on playing! It seems there is wisdom in the adage "quit while you are ahead".

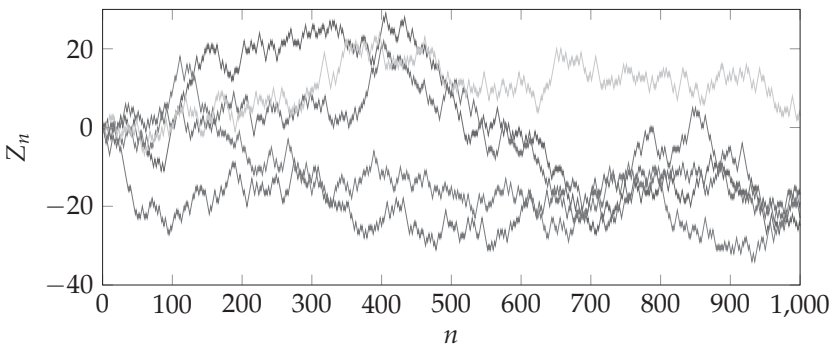

Figure 1.2 Recording the sequence of pairs (n, Z_n) we can plot them on the plane to generate the path of the random walker in time steps. Here we see five unbiased walkers all starting from 0. Note that they remain somewhat centred around 0 – which is what we expect from the results derived in Section 1.1.2

1.1.3 Applications to finance

The simple one-dimensional random walk encountered so far has found great application to finance. Financial forecasting has been labouring under the so-called *random walk hypothesis*, which simply states that stock market prices evolve over time as a random walk.

To see how this works, consider the log price, $S = \log(\text{price})$ of an asset or group of assets, along with ΔS_0, some average price change after a time period of Δt. Economic data suggests that for $\Delta t > 30$ seconds the random walk hypothesis seems valid, as we observe that the correlation between successive prices (the **autocorrelation**) sampled above this threshold is zero.

d	$p(d)$
1	1
2	1
3	0.34
4	0.19
5	0.13
6	0.1

Table 1.1 *Probabilities, $p(d)$ of returning to the origin of a $\mathbb{Z}^d$ random walker initialised at the origin*

Using historical price data you can estimate the value of p and ΔS_0 and so obtain a random walker that behaves *statistically* like the historical data. If you carry out this analysis you will quickly discover that your estimate for p will be approximately 0.5.

1.1.4 Random walks on $\mathbb{Z}^d$ lattices

So far we have restricted our discussion to random walks that occur in one dimension – that is there is only a choice to move left or right along the integers. Formally this is a random walk on the *lattice* defined by $\mathbb{Z}$. An obvious way to extend the concept to higher dimensions is to allow each dimension, d to have its own independent integer random walker, such a random walker is said to walk on the lattice $\mathbb{Z}^d$. An example of a multidimensional random walker is illustrated to two dimensions in Fig. 1.3.

Interestingly not all members of this family of random walkers are recurrent, which may be a little surprising given that they are built from d independent random walks that are recurrent. In fact, for $d > 2$ no such random walks are recurrent. More formally, if we define $p(d)$ as the probability that the random walk on $\mathbb{Z}^d$ starting at 0 will return to 0 some time in the future then we can numerically calculate the probabilities in Table 1.1.

1.1.5 Conformation of polymers

A *polymer* is a large macromolecule composed of repeating structural units typically connected by covalent chemical bonds. Informally we can consider a polymer of length L to be made up of n freely joined con-

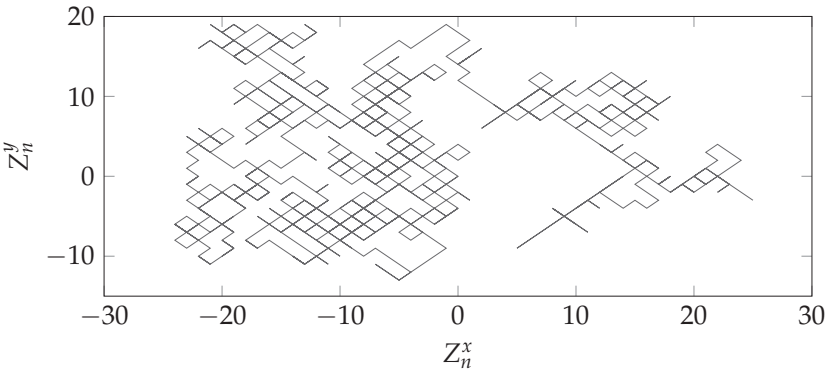

Figure 1.3 An unbiased random walk on $\mathbb{Z}^2$ for $n = 1000$, initialised as the origin.

nected elastic rods of length $l = L/n$, where l is the *persistence length* of the polymer. With this $= L/n$, where l is simplification in mind a naive model of how free floating polymers arrange themselves in space is suggested by a three-dimensional random walker of length n. This model predicts that such a polymer would take up a region of space bounded by a radius, $R \propto \sqrt{n}$. Unfortunately, experimentalists disagreed with the value predicted by the model. Instead of seeing $R \propto n^{\upsilon}$ with $\upsilon = 0.5$ they were consistently measuring $\upsilon \approx 0.6$. The discrepancy between the experimental value and the model prediction was resolved by the realisation that polymers obviously cannot self-intersect in the way a random walker can return to previously visited states. A simple modification to create the so-called *self-avoiding* random walker yields a value of $\upsilon \approx 0.6$ – which agrees much better with the experimental value.

The assumption that the polymer is free to float in space is not always valid. DNA for example can self-interact by forming bonds with sites downstream of the molecule forming loop structures (Fig. 1.4). Introducing the possibility of such interactions into the model gives us our first system that exhibits emergent behaviour and phase transitions. Let E_0 be the energy required to break a self-interacting bond, and let T be the temperature of the environment. At higher temperatures, molecules have more energy and it is easier for self-interacting bonds to be broken. We see that there is a "folding point" temperature above which we have our usual free self-avoiding polymer, but below which we see a polymer that is tightly packed together into a looped structure and needs self-interactions to explain its behaviour (Fig. 1.5).

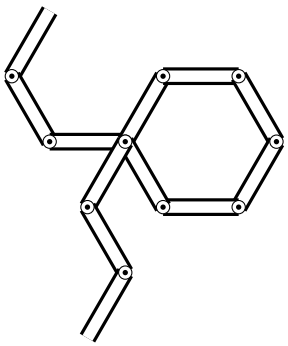

Figure 1.4 Self-interaction of large molecules.

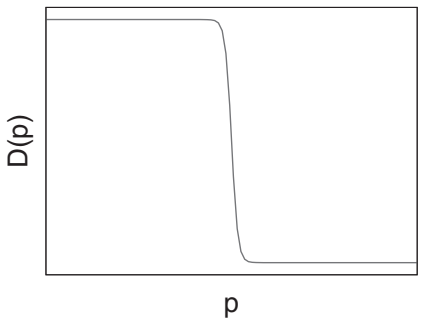

Figure 1.5 Folding point.

This gives us our first important lesson: once we introduce interactions into a model we must be prepared for emergent behaviour to take hold; behaviour that would not be predicted by considering the individual parts of the system alone without considering the interactions. *Interactions are the key.*

1.2 Markov processes

Complex systems often exhibit behaviour which is extremely difficult to quantitatively predict. Whilst initially this might seem problematic, we can turn things to our advantage by constructing models *based* on the apparent randomness. In the previous section we saw how random walks may be used to model simple random processes. In this section we discuss a more general class of models known as stochastic processes and proceed to explore some of the ways in which these models have

been successfully used to describe scientific phenomena. In particular, we will consider Markov processes within cellular biology, where these stochastic models will provide us with a framework from which to derive the dynamics of some prototypical biological systems.

1.2.1 Definitions

To begin with, we extend the random walk model introduced in the previous section to encompass continuous time. Define a **stochastic process** with state space χ to be a collection

$$\{Y(t) \,:\, t \in T\}$$

of χ-valued random variables indexed by $t \in T$. Here t is the generalisation of the number of steps n taken by the random walk and may be considered to represent some measure of time. Notice that the random walks of the previous section correspond to the special case where $T = \{0, 1, 2, \dots\}$. In this chapter we will explore the natural extension to continuous time $T = \{t \in \mathbb{R} \,:\, t \geq 0\}$.

A stochastic process can be expressed in terms of its joint density function

$$p((y_1, t_1), (y_2, t_2), \dots, (y_n, t_n)). \tag{1.14}$$

Conversely, given a function $p((y_1, t_1), \dots, (y_n, t_n))$ how can we decide if p defines a stochastic process? It can be shown that the following constraints represent sufficient criteria:

(i) $p((y_1, t_1), \dots, (y_n, t_n)) \geqslant 0 \;\; \forall \;\; n, \; y_j, \; t_j$

(ii) $p(\dots, (y_i, t_i), \dots, (y_j, t_j), \dots) \;\; = \;\; p(\dots, (y_j, t_j), \dots, (y_i, t_i), \dots)$
$\forall \, i \neq j$

(iii) $p((y_1, t_1), \dots, (y_{n-1}, t_{n-1})) = \iint p((y_1, t_1), \dots, (y_n, t_n)) dy_n dt_n$

(iv) $\iint p(y_1, t_1) dy_1 dt_1 = 1.$

Many of the concepts which were introduced in the context of the random walk can be naturally extended to continuous-time stochastic processes. For example, we can define **moments**

$$\langle Y(t_1) \cdots Y(t_n) \rangle = \iint y_1 \cdots y_n p((y_1, t_1), \dots, (y_n, t_n)) dy_1 \dots dy_n \tag{1.15}$$

and **covariances**

$$K(t_1, t_2) = \langle (Y(t_1) - \langle Y(t_1) \rangle)(Y(t_2) - \langle Y(t_2) \rangle) \rangle \,. \tag{1.16}$$

An important concept in stochastic processes is stationarity. Specifically, we say that p^S is a **stationary distribution** if for all $\tau > 0$ we have

$$p^S((y_1, t_1 + \tau), \ldots, (y_n, t_n + \tau)) = p^S((y_1, t_1), \ldots, (y_n, t_n)). \tag{1.17}$$

We can think of a stochastic process given by a stationary distribution p^S as being *in a time-independent steady state* – the probability of an event does not depend on when the event is scheduled to happen.

1.2.2 Markov property

We say that a stochastic process $\{Y(t) : t \in T\}$ is a **Markov process** if $Y(t)$ has the property that

$$p((y_{n+1}, t_{n+1})|(y_n, t_n), \ldots, (y_1, t_1)) = p((y_{n+1}, t_{n+1})|(y_n, t_n)), \tag{1.18}$$

for all n and all pairs $(y_1, t_1), \ldots, (y_{n+1}, t_{n+1})$, where the times $t_1 \leq t_2 \leq \cdots \leq t_n \leq t_{n+1}$ form an increasing sequence. This is commonly known as the **Markov property**. Intuitively it means that the stochastic process is *memoryless* in the sense that the future behaviour (y_{n+1}, t_{n+1}) of the process depends only on the present (y_n, t_n) and not on the past $(y_{n-1}, t_{n-1}), \ldots, (y_1, t_1)$. This is a nice property because it allows us to factorise $p((y_{n+1}, t_{n+1}), \ldots, (y_0, t_0))$ as

$$p((y_{n+1}, t_{n+1})|(y_n, t_n)) \times \cdots \times p((y_1, t_1)|(y_0, t_0)) \times p((y_0, t_0)), \tag{1.19}$$

meaning that the entire process can be completely characterised by simply stating the function $p((y_{n+1}, t_{n+1})|(y_n, t_n))$ and the **initial distribution** $p((y_0, t_0))$. Notice that the random walks of the previous section all obey the Markov property – the distribution of the next state Z_{n+1} depends only on the state Z_n in which you happen to lie.

For the special case where $n = 2$ we have

$$p((y_2, t_2), (y_1, t_1), (y_0, t_0)) = p((y_2, t_2)|(y_1, t_1))p((y_1, t_1)|(y_0, t_0))p((y_0, t_0)), \tag{1.20}$$

and integrating the equation over (y_1, t_1) produces

$$p((y_2, t_2), (y_0, t_0)) = p((y_0, t_0)) \iint p((y_2, t_2)|(y_1, t_1))p((y_1, t_1)|(y_0, t_0))dy_1 dt_1. \tag{1.21}$$