Lecture notes for Modelling in Biology II: Stochastic processes and networks

Thomas Ouldridge

November 27, 2018

Contents

Chapter 1

Note to students

Dear all,

Hopefully you will find these notes a helpful and comprehensive guide to the second half of the Modelling in Biology course. I've tried to include a reasonable level of theoretical background, and then demonstrate how these ideas and techniques can be applied to a range of biologically related problems. Although this text is far from a mathematically detailed introduction to stochastic processes and graph theory, I have endeavoured to provide derivations of many of the results. In particular, I've tried to demonstrate analysis techniques that apply widely – hopefully this text will therefore serve as a useful reference for future work and research, rather than simply as something to use to pass an exam. Many of the results are presented as exercises to be completed by the student (solutions are provided at the end of the notes). Since simply reading these notes will only get you so far, and there are no GTA classes associated with this course, I strongly recommend that you attempt these problems yourself before consulting the solutions.

Please note that in this third edition, I have re-worked the notation slightly for greater internal consistency (and consistency with the third year course on probability and statistics), with particular emphasis on distinguishing random varables/processes from sample values/trajectories. There may therefore be some minor discrepancies in notation when viewing the coursework and exam from 2016-2017.

I have deliberately kept things simple in many places, and sometimes only alluded to the wider context of the systems and results derived. References are provided in the text to extra resources such as books, notes and webpages. These texts should fill in any technical or conceptual gaps for those that are really interested. Due to time constraints, I haven't cited the many original research papers relevant to the systems and techniques presented here. I should also acknowledge that these notes draw heavily on a number of sources, including Aldo Faisal's lecture notes for a previous version of this course, John Chalker's stochastic processes course, and Uri Alon's book: "An introduction to systems biology: Design principles of biological networks."

Chapter 2

Some Useful Mathematics

2.1 Probabilities

In the first section I will introduce the basic notation and definitions that we will use throughout the course. Hopefully it will all be pretty familiar – if not, please look up your notes from school or the earlier in the course; I've tried to stick reasonably close to the formalism you've been using in the *Probability and Statistics for Bioengineering* module. This is also similar to the formalism in David MacKay's book[1] (from the recommended reading list), which is available for free here: http://www.inference.phy.cam.ac.uk/itprnn/book.pdf. Helpfully, this is also close to A slightly more detailed and rigorous presentation of some of the material here is given in section 2.1-2.2 of that book.

I will then highlight some particularly useful results that will be relevant to the course. Some of the derivations may be new to you, so this will give you a reference when we use them later. Please note that I am by no means attempting a rigorous and thorough mathematical treatment of probability, which gets pretty involved. Fortunately, for many purposes, we can get away with keeping things simple.

2.1.1 The meaning of probability

Colloquially, we use "probability" and "chance" pretty easily, but in surprisingly sophisticated ways. For example, we might wonder about the probability that a political party will win an election, the probability that the next card drawn from the top of a deck will be a king, or the probability that you left your bag on the train rather than back at the office. Superficially, these ideas can seem quite different. You either left your bag on the train or you didn't; the event happened in the past. By contrast, nobody can know for certain who will win an election; the outcome is yet to be determined. The next card in the deck seems to occupy a middle ground: the cards have been shuffled already, so the card has already been selected – but no-one knows which card it is.

In fact, in *all* of the above examples, we're using probability to describe our degree of belief in an outcome given the evidence available: opinion polls, which cards have already been revealed, whether you can remember picking your bag up, *et cetera*. Some people would say that probability is only mathematically properly defined if we're talking about "frequencies of outcomes in random experiments" – such as, for example, the chance of obtaining a head when tossing a coin. But in fact, defining exactly what is meant by "frequency" and "random" is difficult to do in a non-circular manner. Moreover, tossing a coin isn't so different from the examples above. If we had really good

knowledge of the initial conditions, it would be possible to predict the result of a coin flip before the coin landed. So even here, probability is really telling us about our degree of belief in an outcome given the available evidence, just as with the lost bag, the deck of cards and indeed the election mentioned above.

I shall not go further into the subtleties, which relate to deep questions in physics such as the interpretation of quantum mechanics, the meaning of entropy, and the relationship between chaos and randomness. In biological settings, expressing degrees of belief through probabilities allows the use of *Bayesian Inference* in the analysis of experiments.

2.1.2 Basic definitions

For our purposes, we shall simply assume that we are able to express the probability, or degree of belief, that a variable X takes an outcome x from a set of values $\mathscr{A}_X$ in a meaningful way. To start with, we'll consider discrete sets of outcomes such as $\mathscr{A}_X = \{1,2,3,4,5,6\}$ for a six-sided die or $\{\text{train}, \text{office}, \text{elsewhere}\}$ for the places we might have left a bag. It is important to emphasise that $\mathscr{A}_X$ are the possible *outcomes*, not the probabilities themselves. These probabilities are a distinct set (of the same size as $\mathscr{A}_X$), looking something like $\mathscr{P}_X = \{1/6,1/6,1/6,1/6,1/6,1/6\}$ or $\{0.5,0.4,0.1\}$ for the examples above, with each term giving the probability of the corresponding value in $\mathscr{A}_X$.

Note that people typically differentiate between the the variable X, and its outcome in a given realisation, x. This is because there are properties of the variable that do not depend on a specific outcome, such as its average, and this approach allows us to label these quantities sensibly. It is convenient to use the shorthand $P(X = x) = p(x)$ to describe the probability that an outcome is x. Just to make sure this is clear, if we have $\mathscr{A}_X = \{0,1\}$ and $\mathscr{P}_X = \{0.3,0.7\}$, then $p(0) = 0.3$ and $p(1) = 0.7$.

Alternative outcomes

If we're considering alternative outcomes of a single variable X, then it's very easy to ask for the probability that the outcome is either x or x'.

$$P(X = x \text{ or } x') = p(x) + p(x'). \tag{2.1}$$

Note that this expression only holds because we're asking for the probability that a *single* outcome of a *single* variable is either x or x'. The generalisation to a larger subset of $\mathscr{A}_X$, $\mathscr{A}'_X \in \mathscr{A}_X$, is simple: just sum over all the outcomes in the set:

$$P(X = x \in \mathscr{A}'_X) = \sum_{x \in \mathscr{A}'_X} p(x). \tag{2.2}$$

Normalisation

If we sum over the entire set of possible outcomes $\mathscr{A}_X$, then the total probability must be one:

$$P(X = x \in \mathscr{A}_X) = \sum_{x \in \mathscr{A}_X} p(x) = 1. \tag{2.3}$$

Joint probabilities

We are often interested in the combined behaviour of two variables X and Y. These might be two separate instances of the same process (two rolls of the same die, for example) or they might be quite distinct (mRNA copy number and protein copy number at a given point in time, for example). In this case we can talk about the combined outcomes x, y, with the sets $\mathscr{A}_{X,Y}$ and $\mathscr{P}_{X,Y}$ defined as before, but accounting for all possible combinations of outcomes from $\mathscr{A}_X$ and $\mathscr{A}_Y$. $p(x,y)$, the *joint probability*, is the probability of obtaining both $X = x$ *and* $Y = y$. Clearly we can generalise to more than two variables.

Marginal probabilities

If we have a joint distribution $p(x,y)$ for variables X and Y, we can ask for the distribution of X regardless of the value of Y. This process involves summing over all values of Y, and is called *marginalisation*.

$$p(x) = \sum_{y \in \mathscr{A}_Y} p(x,y). \tag{2.4}$$

Similarly, $p(y) = \sum_{x \in \mathscr{A}_X} p(x,y)$.

Conditional probabilities

We can also ask for the probability that X takes the value x given that Y takes the value y. This quantity is the *conditional probability* $p(x|y)$:

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x,y)}{\sum_{x \in \mathscr{A}_X} p(x,y)}. \tag{2.5}$$

Effectively, we normalise the joint probability by the marginal. Similarly, $p(y|x) = p(x,y)/p(x)$. Note that whereas the marginal probability $p(x)$ is a function of x only, $p(x|y)$ is generally a function of both x and y. It immediately follows from Eq. 2.4 that

$$p(x,y) = p(x)p(y|x). \tag{2.6}$$

Thus the probability of an outcome of x, y is the probability of outcome x, regardless of y, multiplied by the probability of an outcome y given an outcome x. We can thus re-write Eq. 2.4 as

$$p(x) = \sum_{y \in \mathscr{A}_Y} p(y)p(x|y). \tag{2.7}$$

The fact that probabilities are normalised to one implies the rule

$$\sum_{x \in \mathscr{A}_X} p(x|y) = 1. \tag{2.8}$$

If this rule did not hold, the total probability of all possible outcomes of X would not be one for at least some values of y.

Independence

The outcome of one variable may or may not be related to another. Whilst it is reasonable to assume that the outcome of one die roll is not influenced by a previous roll, the temperature today is highly informative of the temperature tomorrow. If two variables X and Y are independent, then the conditional probability is the same as the marginal (the outcome y has no bearing on the probability of the outcome x):

$$p(x|y) = p(x),\ p(y|x) = p(y). \tag{2.9}$$

Combining this From Eq. 2.6, we see that for independent processes the joint probability is simply a product of the marginals

$$p(x,y) = p(x)p(y). \tag{2.10}$$

For example, the probability of rolling two consecutive sixes on conventional die rolls is $p(6,6) = p(6)p(6) = 1/36$, as is hopefully intuitive. Note that Eqs. 2.9 and 2.10 only hold when X and Y are independent, otherwise Eqs. 2.5 and 2.6 must be used.

Exercise 2.1.1. This exercise is recommended if you are unfamiliar with conditional and marginal probabilities and independence. I have two binary variables, X and Y. $p(x,y)$ is tabulated in Table 2.1.

$p(x,y)$		x	
		0	1
y	0	1/4	1/8
	1	1/4	3/8

Table 2.1: A tabulation of joint probablities for two variables X and Y.

A. Are X and Y independent?

B. Evaluate $p(x)$ and $p(y)$.

C. Evaluate $p(x|y)$ and $p(y|x)$.

D. You measure X and obtain $x = 1$. What is the probablity that the outcome of Y is $y = 1$?

2.1.3 Expectations of outcomes from probability distributions

Often we're interested in overall properties of the entire probability distribution, rather than single values of joint, conditional or marginal probabilities. These properties are extremely helpful in allowing us to characterise the typical or expected behaviour of a variable, rather than the outcome of one particular realisation.

We can consider a general function $f(X)$ of a variable X. In general the output of this function is uncertain due to the uncertainty in X. However, we can calculate an *expectation* of $f(X)$, which is its mean value, using the probability distribution

$$E(f(X)) = \langle f(X) \rangle = \sum_{x \in \mathcal{A}_X} p(x) f(x). \tag{2.11}$$

Two choices of $f(X)$ are extremely common:

- $f(X) = X$, which simply gives the mean or expaction of X, $E(X) = \langle X \rangle = \sum_{x \in \mathscr{A}_X} x p(x)$.
- $f(X) = X^2$, giving the mean or expectation of X^2, $E(X^2) = \langle X^2 \rangle = \sum_{x \in \mathscr{A}_X} x^2 p(x)$.

It is simple to generalise to multiple variables:

$$E(f(X,Y)) = \langle f(X,Y) \rangle = \sum_{x \in \mathscr{A}_X, y \in \mathscr{A}_Y} p(x,y) f(x,y). \tag{2.12}$$

The quantity $E(XY) = \langle XY \rangle$ is often important. Note that this double sum still holds if $f(x,y)$ is only a function of x or y. For example,

$$E(f(X)) = \langle f(X) \rangle = \sum_{x \in \mathscr{A}_X, y \in \mathscr{A}_Y} p(x,y) f(x) = \sum_{x \in \mathscr{A}_X} p(x) f(x). \tag{2.13}$$

where the marginal $p(x) = \sum_{y \in \mathscr{A}_Y} p(x,y)$ as before.

Variances and covariances

The mean $E(X) = \langle X \rangle$ characterises the average value of X, but gives you no idea of how spread out individual outcomes might be about this average. A useful measure of spread is the variance

$$\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = \sum_{x \in \mathscr{A}_X} x^2 p(x) - \left(\sum_{x \in \mathscr{A}_X} x p(x) \right)^2 = \sum_{x \in \mathscr{A}_X} (x - \langle X \rangle)^2 p(x) \tag{2.14}$$

From the final equality in Eq. 2.14, it is clear that the variance is a sum of positive terms; it is the average squared distance of outcomes from the mean. Therefore the variance must be positive, and its square root is real. This square root is the standard deviation $\sigma(X) = \sqrt{\mathrm{Var}(X)}$. For this reason the variance is often referred to as $\sigma^2(X)$.

The covariance is defined as

$$\mathrm{Cov}(X,Y) = \langle XY \rangle - \langle X \rangle \langle Y \rangle = \sum_{x \in \mathscr{A}_X, y \in \mathscr{A}_Y} xy p(x,y) - \sum_{x \in \mathscr{A}_X} x p(x) \sum_{y \in \mathscr{A}_Y} y p(y) \tag{2.15}$$

The covariance tells us whether fluctuations from the mean of X and Y tend to be correlated ($\mathrm{Cov}(X,Y) > 0$), anti-correlated ($\mathrm{Cov}(X,Y) < 0$) or independent ($\mathrm{Cov}(X,Y) = 0$).

Exercise 2.1.2. This exercise is recommended if you are unfamiliar with means, variances and covariances. Using the definitions provided above,

A. Prove that $\langle aX + bY \rangle = a\langle X \rangle + b\langle Y \rangle$, where a and b are constants.

B. Prove that $\langle X^2 \rangle - \langle X \rangle^2 = \sum_{x \in \mathscr{A}_X} (x - \langle X \rangle)^2 p(x)$.

C. Prove that $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$ if a is a constant.

D. Prove that $\mathrm{Var}(X+Y) = \langle (X+Y)^2 \rangle - \langle X+Y \rangle^2 = \mathrm{Var}(X) + \mathrm{Var}(Y)$ if X and Y are independent.

E. Prove that $\mathrm{Cov}(X,Y) = 0$ if X and Y are independent.

Tip 2.1.1. As proved in Ex. 2.1.2, if Z is a sum of independent random variables X and Y, the mean and *variance* of Z are constructed by summing those of X and Y. Consider now Z_N as the sum of a large number N of independent but identically distributed random variables, each with mean $\langle X\rangle$ and variance $\mathrm{Var}(X)$. Then

$$\langle Z_N\rangle = N\langle X\rangle, \quad \sigma_{Z_N} = \sqrt{N}\sigma_X. \tag{2.16}$$

So although the standard deviation of Z_N grows with the number of variables, it grow *much more slowly* than the mean. Thus, as N becomes large, fluctuations about the mean become less and less relatively important. This is why, as a rule of thumb, small systems show much more evident random fluctuations than larger ones (although, be careful - a large system in which all the variables are strongly correlated will not behave like this).

2.1.4 Continuous probabilities

Sometimes we might be interested in the time at which a biochemical reaction takes place, or the position of a diffusing particle. In this case, the sets of possible outcomes $\mathscr{A}_X$ are continuous sets of real numbers. Essentially everything we've said previously still applies, except that $p(x)$ is now a probability *density*, rather than an absolute probability. This means that $p(x)$ quantifies the probability *per unit of* x, and thus has dimensions (for example, per unit time or per unit distance). When dealing with continuous densities rather than discrete probabilities, sums get replaced by integrals.

The probability that an outcome is in the range $a < x < b$ is an integral between these limits

$$P(a < x < b) = \int_a^b \mathrm{d}x\, p(x). \tag{2.17}$$

When integrating over the full possible range or *support* $\mathscr{A}_X$ of the variable (often 0 to ∞ or $-\infty$ to ∞), we obtain the normalisation condition.

$$\int_{\mathscr{A}_X} \mathrm{d}x\, p(x) = 1. \tag{2.18}$$

Two- or multi-dimensional probability densities can be defined analogously to before. Considering two variables X and Y, marginal probabilities are given by

$$p(x) = \int_{\mathscr{A}_Y} \mathrm{d}y\, p(x,y),\ p(y) = \int_{\mathscr{A}_X} \mathrm{d}x\, p(x,y), \tag{2.19}$$

and conditional probability densities by

$$p(x|y) = \frac{p(x,y)}{p(y)},\ p(y|x) = \frac{p(x,y)}{p(x)}. \tag{2.20}$$

Similarly, expectations of functions $f(X)$ are given by

$$E(f(X)) = \langle f(X)\rangle = \int_{\mathscr{A}_X} \mathrm{d}x\, f(x)p(x) \tag{2.21}$$

and for functions of two random variables

$$E(f(X,Y)) = \langle f(X,Y)\rangle = \int_{\mathscr{A}_X}\int_{\mathscr{A}_Y} \mathrm{d}x\,\mathrm{d}x\, f(x,y)p(x,y). \tag{2.22}$$

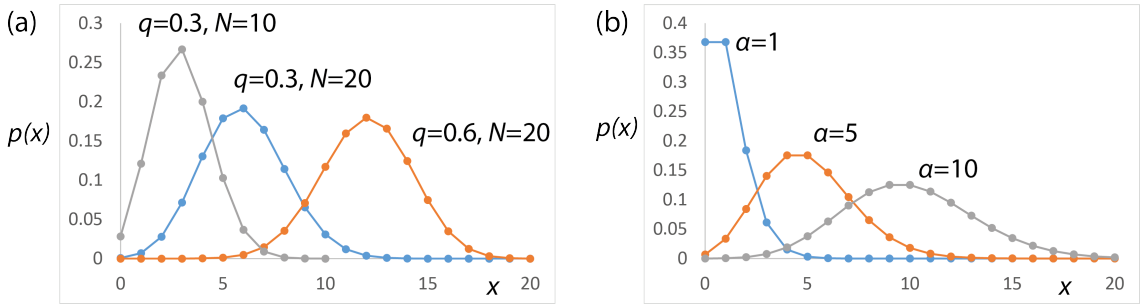


Figure 2.1: Typical binomial (a) and Poisson distributions (b). The lines between points are drawn to guide the eye only. Note that whilst the binomial distributions strictly have $p(x) = 0$ for $x > N$, the same is not true of the Poisson, which has support for $x \to \infty$.

2.2 Common probability distributions and tricks in manipulating them

I don't necessarily expect the content of this section to be familiar to you. Hopefully it will provide a convenient reference for calculations that are relevant to the course, and probabilistic systems in general. By the time of the exam I'll expect you to be able to manipulate these probability distributions given reminders of the relevant techniques, but not reporduce derivations from memory.

2.2.1 Discrete probability distributions

The binomial distribution

The binomial distribution arises when we consider repeated independent "trials" of a process which can have two outcomes, such as heads/tails, success/failure and folded/unfolded. In general, we can represent these outcomes in binary with a "0" or "1". If we have N trials of a process, each with a probability q of giving 1, we obtain the following distribution for X, the number of 1s:

$$p(x) = \frac{N! q^x (1-q)^{N-x}}{x!(N-x)!} \text{ for } 0 \leq x \leq N. \tag{2.23}$$

Equivalently, X can be thought of as the sum of N independent random variables Y_i, each of which take value 1 with probability q and 0 with probability $1-q$. This makes it easy to calculate expectations, without using the distribution directly:

$$\langle X \rangle = \langle \sum_{i=1}^{N} Y_i \rangle = \sum_{i=1}^{N} \langle Y_i \rangle = Nq, \tag{2.24}$$

and

$$\text{Var}(X) = N\text{Var}(Y_i) = Nq(1-q), \tag{2.25}$$

where we have used the result that the variance of the sum of independent variables is the sum of the individual variances (Ex. 2.1.2), and $\langle Y_i^2 \rangle = q$. Typical binomial distributions are plotted in Fig. 2.1 (a).

Where does the distribution in Eq. 2.23 come from? Consider a single specific sequence which contains x 1s and $N-x$ 0s. The probability of obtaining exactly that sequence is simply $q^x(1-$

$q)^{N-x}$. But there are many ways to rearrange the sequence, and still have the same number of x in total. The total probability of obtaining x 1s and $N-x$ 0s in any order is $q^x(1-q)^{N-x}$ multiplied by the number of distinct sequences.

The entries in a sequence of length N can be arranged in $N!$ ways. However, we have only two types of entry – 0 and 1. For every distinct sequence of 0s and 1s, there are $x!$ ways to swap the 1s amongst themselves and not change the overall sequence . Similarly, there are $(N-x)!$ ways to re-order the 0s without influencing the overall sequence. Thus the number of distinct sequences with x 1s and $N-x$ 0s is $N!/x!(N-x)!$. Combining this combinatorial factor with $q^x(1-q)^{N-x}$, we get the desired result (Eq. 2.23).

The Poisson distribution

The Poisson distribution arises in a number of situations

- It is the distribution obtained for the number of entities in a system if those entities are produced at a constant rate, and destroyed at a rate proportional to their total number (an immigration-death process, Sec. 4.3.1).
- It is the distribution of events observed in a given finite period of time if events occur independently and with a fixed average rate (a Poisson point process, Sec. 5.1.4).
- The Poisson distribution with mean α is a good approximation for a binomial distribution with N trials of success probability $q = \alpha/N$, given $N \to \infty$ and α finite.

The distribution is:

$$p(x) = \alpha^x \frac{\exp(-\alpha)}{x!} \text{ for } 0 \leq x. \tag{2.26}$$

The normalisation $\exp(-\alpha)$ follows because $\sum_x \alpha^x/x!$ is the full Taylor expansion of the exponential of α. Poisson distributions are plotted in Fig. 2.1. The secret to finding expectations is "differentiating within the sum", a common approach:

$$\langle X \rangle = \sum_{x=0}^{\infty} x\alpha^x \frac{\exp(-\alpha)}{x!} = \exp(-\alpha) \sum_{x=0}^{\infty} \alpha \frac{\mathrm{d}}{\mathrm{d}\alpha} \alpha^x \frac{1}{x!} = \alpha \exp(-\alpha) \frac{\mathrm{d}}{\mathrm{d}\alpha} \sum_{x=0}^{\infty} \alpha^x \frac{1}{x!} = \alpha. \tag{2.27}$$

Similarly,

$$\langle X^2 \rangle = \sum_{x=0}^{\infty} x^2\alpha^x \frac{\exp(-\alpha)}{x!} = \alpha^2 \exp(-\alpha) \sum_{x=0}^{\infty} \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} \alpha^x \frac{1}{x!} + \langle X \rangle = \alpha^2 + \alpha. \tag{2.28}$$

The above result follows from noting that $\alpha^2 \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} \alpha^x = x(x-1)\alpha^x$. Thus the variance is $\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = \alpha = \langle X \rangle$. The fact that the variance is equal to the mean leads to relatively narrow distributions as α gets large.

2.2.2 Continuous probability distributions

The exponential distribution

One of the most common probability distributions – for example, it's the distribution of waiting times for an event that has a constant probability per unit time of occurring (given that it hasn't occurred already) k.

$$p(x) = k \exp(-kx) \text{ for } 0 \leq x \tag{2.29}$$

Some exponential probability densities are plotted in Fig. 2.2 (a). The secret to finding expectations of powers of x with the exponential distribution is to "differentiate under the integral", a trick that is often very useful.

$$\langle X\rangle = k\int_0^\infty x\exp(-kx)\,\mathrm{d}x = -k\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}k}\exp(-kx)\,\mathrm{d}x = -k\frac{\mathrm{d}}{\mathrm{d}k}\int_0^\infty \exp(-kx)\,\mathrm{d}x = -k\frac{\mathrm{d}}{\mathrm{d}k}(1/k) = 1/k. \tag{2.30}$$

Similalry,

$$\langle X^2\rangle = k\int_0^\infty x^2\exp(-kx)\,\mathrm{d}x = k\int_0^\infty \frac{\mathrm{d}^2}{\mathrm{d}k^2}\exp(-kx)\,\mathrm{d}x = k\frac{\mathrm{d}^2}{\mathrm{d}k^2}(1/k) = 2/k^2. \tag{2.31}$$

Thus $\mathrm{Var}(X) = 1/k^2 = \langle X\rangle^2$. This is quite important – the variance of an exponential distribution is its mean squared, or the standard deviation is the mean. This means that exponential distributions are very broad.

The normal (Gaussian) distribution

Another extremely common distribution, often arising when we define a random variable as the sum over vary many independent (but identically distributed) random variables (this is the Central Limit Theorem – see Ref. 2). The Gaussian distribution is characterised by the mean μ and the standard deviation σ.

$$p(x) = \sqrt{\frac{1}{2\pi\sigma^2}}\exp\left(-(x-\mu)^2)/2\sigma^2\right) \text{ for all } x \tag{2.32}$$

Some Gaussian distributions are plotted in Fig. 2.2 (b). When evaluating integrals involving Gaussians, it is useful to note the following. Define I_n as

$$I_n = \int_{-\infty}^\infty x^n\exp(-\alpha x^2)\mathrm{d}x. \tag{2.33}$$

Then,

$$I_0 = \int_{-\infty}^\infty \exp(-\alpha x^2)\mathrm{d}x. \tag{2.34}$$

This integral is a bit of a pain, but the way to solve it is to square it and then change variables, as below.

$$I_0^2 = \int_{-\infty}^\infty \exp(-\alpha x^2)\mathrm{d}x\int_{-\infty}^\infty \exp(-\alpha y^2)\mathrm{d}y = \int_{-\infty}^\infty\int_{-\infty}^\infty \exp(-\alpha(x^2+y^2))\mathrm{d}x\mathrm{d}y. \tag{2.35}$$

Changing variables to $r = \sqrt{x^2+y^2}$, $\theta = \arctan(y/x)$ (polar coordinates), we find

$$I_0^2 = \int_0^\infty\int_0^{2\pi} r\exp(-\alpha r^2)\mathrm{d}r\mathrm{d}\theta. \tag{2.36}$$

The θ integral is trivial, and the change of variables means that the r integral is feasible since it is of the form $\mathrm{d}f(r)/\mathrm{d}r\times\exp(f(r))$. Thus

$$I_0^2 = -\frac{\pi}{\alpha}\left[\exp(-\alpha r^2)\right]_0^\infty = \frac{\pi}{\alpha}. \tag{2.37}$$

Thus, $I_0 = \sqrt{\pi/\alpha}$. To obtain I_{2n}, simply note that

$$I_{2n} = -\frac{\mathrm{d}}{\mathrm{d}\alpha}I_{2n-2}. \tag{2.38}$$

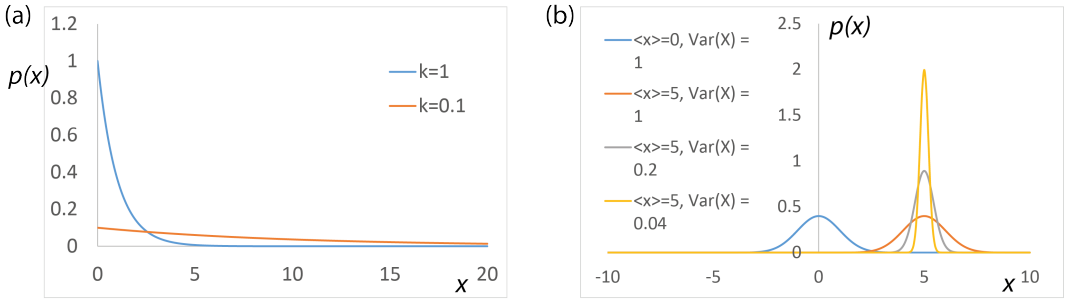


Figure 2.2: Some typical continuous probability densities. (a) Exponential distributions characterised by decay constant k (mean is $1/k$). (b) Gaussian distributions with mean and variance as noted in the legend. The final three, with identical mean but decreasing variance, illustrate a series that eventually culminates in a delta-function

.

Thus, for example, $I_2 = \frac{1}{2}\sqrt{\frac{\pi}{\alpha^3}}$. The integrand of I_1 is the product of an even ($\exp(-\alpha x^2)$) and an odd (x) term, and hence is zero when integrated over all values of x. The same holds for all odd terms I_{2n+1}.

Using these results, we find:

$$\langle X\rangle = \sqrt{\frac{1}{2\pi\sigma^2}}\int_{-\infty}^{\infty} x\exp\left(-(x-\mu)^2)/2\sigma^2\right)\mathrm{d}x = \sqrt{\frac{1}{2\pi\sigma^2}}\int_{-\infty}^{\infty}(y+\mu)\exp\left(-y^2)/2\sigma^2\right)\mathrm{d}y = \mu, \tag{2.39}$$

as expected. Here we have used the results for I_0 and I_1 to evaluate the final step. Similarly,

$$\begin{aligned}\mathrm{Var(X)} = \langle X^2\rangle - \mu^2 &= \sqrt{\tfrac{1}{2\pi\sigma^2}}\textstyle\int_{-\infty}^{\infty}(x^2-\mu^2\exp\left(-(x-\mu)^2)/2\sigma^2\right)\mathrm{d}x \\ &= \sqrt{\tfrac{1}{2\pi\sigma^2}}\textstyle\int_{-\infty}^{\infty}(y^2+2\mu y)\exp\left(-y^2/2\sigma^2\right)\mathrm{d}y = \tfrac{1}{2}\sqrt{\tfrac{1}{2\pi\sigma^2}}\sqrt{\pi(2\sigma^2)^3} = \sigma^2,\end{aligned} \tag{2.40}$$

where we have used the results for I_1 and I_2.

A useful property of Gaussian distributions is that that if $Z = X+Y$, and X and Y are independent random variables with Gaussian distributions, Z also has a Gaussian distribution (with $\langle Z\rangle = \langle X\rangle + \langle Y\rangle$ and $\langle Z^2\rangle = \langle X^2\rangle + \langle Y^2\rangle$.[2]

The Dirac delta function

Hopefully you have met the Dirac delta function before. The delta function isn't a true function, but rather the limit of a family of functions as a parameter is taken to zero. One possible choice of this family (there are many that give the same result) is the family of normalised Gaussian distributions.

$$\delta(x-\mu) = \lim_{\sigma\to 0}\sqrt{\frac{1}{2\pi\sigma^2}}\exp\left(-(x-\mu)^2)/2\sigma^2\right) \text{ for all } x \tag{2.41}$$

The limiting process is illustrated in Fig. 2.2 (b). Hopefully it is clear that the limiting distribution is an infinitely sharp spike centred at $x = \mu$.

If representing a probability distribution, $\delta(x-\mu)$ tells us that the system is guaranteed to be exactly at $x=\mu$; the probability of finding the particle anywhere else is zero, and the integral of the delta function over any range including μ (no matter how small) is necessarily 1. But the delta function pops up all over the place in science and engineering, not just representing probability. In particular, we often see it in expressions like

$$I = \int_{x_0}^{x_1} \delta(x-\mu) f(x) \mathrm{d}x, \tag{2.42}$$

where $f(x)$ is an arbitrary function. Because $\delta(x-\mu)$ is so sharply peaked, any other function $f(x)$ is effectively constant over the range in which $\delta(x-\mu)$ is non-zero. Consequently,

$$I = f(\mu) \int_{x_0}^{x_1} \delta(x-\mu) \mathrm{d}x = f(\mu) \text{ if } x_1 \leq \mu \leq x_2, \text{ 0 otherwise.} \tag{2.43}$$

The rule of thumb is that, when integrated with another function $f(x)$, the delta function $\delta(x-\mu)$ effectively picks out $f(\mu)$.

Chapter 3

Stochastic Processes

3.1 What is a stochastic process?

The world is full of fluctuations and variability. We see this when we do experiments; you never get exactly the same result twice. Some of this might be down to errors in measurement, some may be due to inherent fluctuations in the system we're measuring, and some may arise due to fluctuations introduced by the environment. Regardless, the net effect is that the output of many processes cannot be perfectly predicted. A famous example of this is the cloned cat "CC", who looks totally different from her genetic donor "Rainbow", despite the identical genomes (she's even a different colour).[3] In some systems, the random noise may simply blur the expected deterministic behaviour; in others, it has fundamental effects that are crucial in understanding behaviour.

There are philosophical arguments to be had about whether much variation is truly random, or is in fact a result of high-order chaos due to deterministic processes that we are in no position to follow accurately. For example, Robert Brown famously observed that pollen particles suspended in water appear to move randomly due to buffeting from the water molecules.[4] We might argue that if we knew the positions and velocities of the water molecules well enough, we could actually predict this motion. But in fact such an undertaking is impossible, for a variety of reasons, including the aforementioned chaotic behaviour. The net result, for all practical purposes, is that the motion of a particle can be extremely well modelled by treating it as subject to a series of random forces, as Einstein famously did.[4,5] This pragmatic approach is at the heart of modelling, and also relates to the pragmatic approach to probability discussed in Chapter 2.

Einstein's model of a suspended particle is an example of a dynamical system in which the rules for time evolution contain a random component. The underlying equations are a *stochastic process*. The word *stochastic* simply means random, and *process* indicates that the system tends to evolve over time. Such a system might have only one evolving variable, or many, and it is important to emphasise that stochasticity, or randomness, does not necessarily imply that there are no biases involved. Although a particle may be equally likely to diffuse in any direction, in most stochastic processes there is a tendency to systematically drift in one direction or another.

3.2 Representing stochastic processes

For simplicity, let us assume that we have a single variable X undergoing a stochastic process. We can think of the sequence of X over time, $X(t)$, as a series of random variables each characterised by its own probability $p(x,t)$. Thus taking X at two different times, $X(t_1)$ and $X(t_2)$, corresponds

to considering two distinct random variables. Typically, of course, these variables are correlated - but nonetheless they are distinct. Thus everything discussed in Chapter 2 for two variables X and Y applies just as well to $X(t_1)$ and $X(t_2)$. We can define a joint probability $p(x_1,t_1;x_2,t_2)$ a the probability that X takes the value x_1 at time t_1, *and* x_2 at time t_2. The marginal probability,

$$p(x_1,t_1) = \sum_{x_2 \in \mathscr{A}_X} p(x_1,t_1;x_2,t_2) \tag{3.1}$$

is simply the probability that X takes the value x_1 at time t_1 regardless of its value at t_2. Similarly, the conditional probability

$$p(x_1,t_1|x_2,t_2) = \frac{p(x_1,t_1;x_2,t_2)}{p(x_2,t_2)} \tag{3.2}$$

is the probability that X takes the value x_1 at time t_1 given that it has x_2 at t_2. It's worth noting that this probability makes sense for both $t_1 > t_2$, in which case it amounts to a probabilistic prediction of the future, and $t_2 > t_1$, in which case it's an inference about the past. $X(t_1)$ and $X(t_2)$ are independent (which is unusual) if and only if the joint probability can be factorised $p(x_1,t_1;x_2,t_2) = p(x_1,t_1)\,p(x_2,t_2)$, as before.

Please note that although $p(x_1,t_1)$ is a function of both x_1 and t_1, the two quantities are playing very different roles. $p(x_1,t_1)$ is *not* a joint probability over x_2 and t_1, and should never be interpreted as "the probability that the $X = x_1$ and the time is t_1." t_1 is simply a label to distinguish the same random variable at different times.

Finally, covariances can be defined as $\mathrm{Cov}(X(t_1),X(t_2)) = \langle X(t_1)X(t_2)\rangle - \langle X(t_1)\rangle\langle X(t_2)\rangle$, with

$$\langle X(t_1)X(t_2)\rangle = \sum_{x_2,x_1 \in \mathscr{A}_X} x_1 x_2 p(x_1,t_1;x_2,t_2)\,. \tag{3.3}$$

If we normalise this covariance by the standard deviations of X at t_1 and t_2, we obtain the *autocorrelation function*

$$R_X(t_1,t_2) = \frac{\mathrm{Cov}(X(t_1),X(t_2))}{\sigma_{X(t_1)}\sigma_{X(t_2)}}. \tag{3.4}$$

The autocorrelation function is perhaps the most important quantity of time-dependent signals. It tells us how much of the variability at time t_2 is a direct reflection of the variability at t_1. Two signals with the same mean and variance, but distinct autocorrelation, are plotted in Fig. 3.1; shorter autocorrelation means that $X(t)$ explores the full range of its possible values much faster.

If we are dealing with a continuous variable X rather than a discrete one, then all probabilities are replaced by probability densities, and sums by integrals, as in Chapter 2. We could also have more than one variable involved our stochastic process; $X(t)$ and $Y(t)$ might evolve together in a coupled fashion. All of the above definitions extend in the natural way.

3.3 Markov processes

In simple terms, a Markov process is a process with no memory, except through the values of the random variable(s) at the current time. The evolution of the system for $t > t_1$ from a given state $X(t_1) = x_1$ is independent of the historic details of how the system reached $X(t_1) = x_1$ over the course of $t < t_1$. This class of stochastic process is by far the simplest to analyse, and fortunately is sufficient to describe many systems of interest.

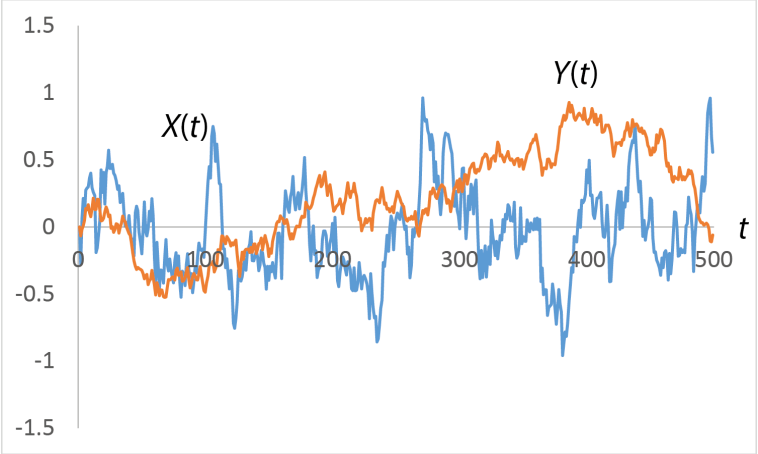


Figure 3.1: Two stochastic processes $X(t)$ and $Y(t)$ with the same mean and variance in the limit of $t \to \infty$ but with an autocorrelation that decays 10 times faster with Δt for $X(t)$ than $Y(t)$.

The mathematical version of the above statement is

$$p(x_n,t_n|x_1,t_1;x_2,t_2;...x_{n-1},t_{n-1}) = p(x_n,t_n|x_{n-1},t_{n-1}) \tag{3.5}$$

for times $t_1 < t_2 ... < t_n$. The conditional probability of the evolution during $t_{n-1} \to t_n$ depends only on the state x_{n-1} at t_{n-1}, and not on the state at earlier times. Markov processes obey the *Chapman-Kolmogorov equation*:

$$p(x_3,t_3|x_1,t_1) = \sum_{x_2 \in \mathscr{A}_x} p(x_3,t_3|x_2,t_2)p(x_2,t_2|x_1,t_1),$$

for $t_3 > t_2 > t_1$, which states that the probability of going from x_1 to x_3 is given by the sum over the probabilities of paths that go through all possible intermediates x_2 at time t_2.

In this course, we will restrict ourselves to the analysis of Markov processes. We will also restrict ourselves to systems in which the underlying dynamics do not change with time; in other words, although $p(x,t)$ may be time-dependent, conditional transition probabilities $p(x_2,t_2|x_1,t_1)$ are not (such Markov processes are *homogenous*). However, the formalism and techniques for solving/simulating Markov processes depend a great deal on whether the variables and time steps are continuous or discrete. In the following sections I will outline the basics of three types of Markov processes, involving:

- Discrete random variables, discrete time.
- Discrete random variables, continuous time.
- Continuous random variables, continuous time.

It is fairly obvious why we might want to use discrete variables; all sorts of biological systems are naturally discrete. In particular, we might be interested in the number of proteins, cells or animals in a given system, or the number of action potentials during a given time window. Many other variables are naturally continuous, such as the size of an organism or the position of a diffusing molecule. It is perhaps less obvious why we might consider discrete time, as we're used to systems

evolving continuously in time. Indeed, in the first half of this course the default approach was to construct differential equations, which by their very nature assume continuous time. However, many systems have natural time scales on which they evolve, or on which we make make observations. For example, we might be interested in a quantity that changes randomly upon cell division, such as the copy number of plasmids; in this case, it makes sense to take about the plasmid copy number discretely as a function of generation, rather than continuously as a function of time. Similarly, grazing behaviour might be best described using a variable that is updated on a day-to-day basis, rather than continuously. Finally, it helps to understand all three perspectives, since it is sometimes convenient to use one perspective to approximate another (we might simulate diffusion using discrete jumps on a grid, for example).

3.4 Discrete random variables, discrete time

The simplest Markov processes have a finite number of discrete states for the variable(s), and involve discrete steps in time. For clarity, I will label these discrete steps with an integer n, so instead of $X(t)$ and $p(x,t)$ we have X_n and $p(x,n)$. These processes are known as *discrete-time Markov chains*.

3.4.1 State vector and transition matrix

Given that the state space is discrete, $p(x,n)$ can be represented by a vector

$$p(x,n) = \begin{pmatrix} p_1(n) \\ p_2(n) \\ \dots \\ p_N(n) \end{pmatrix}. \tag{3.6}$$

Here, $p_i(n)$ is the probability that X takes its i^{th} discrete value at step n, and N is the total number of values that X can take. Note that this is possible even if the outcomes x are not themselves integers – they could be any discrete list, such as $\mathscr{A}_X = \{\text{“blue”}, \text{“green”}, \text{“red”}\}$ that we've simply ordered from 1 to N.

Using this vector notation, the conditional probabilities $p(x_i, n+1|x_j, n)$ can be expressed as an $N \times N$ matrix $T_{ij} = p(x_i, n+1|x_j, n)$, or more explicitly

$$\mathbf{T} = \begin{pmatrix} p(x_1,n+1|x_1,n) & p(x_1,n+1|x_2,n) & \dots & p(x_1,n+1|x_N,n) \\ p(x_2,n+1|x_1,n) & \dots & & \\ \dots & & & \\ p(x_N,n+1|x_1,n) & p(x_N,n+1|x_2,n) & \dots & p(x_N,n+1|x_N,n) \end{pmatrix}. \tag{3.7}$$

$\mathbf{T}$ is the *transition matrix* or *stochastic matrix*. It specifies the evolution of the system, since the conditional probabilities $p(x_i, n+1|x_j, n)$ are the probabilities that a system will be in state x_i after a step, given that the system was in x_j before the step. Indeed, the evolution equation is

$$p(x,n+1) = \mathbf{T}p(x,n), \tag{3.8}$$

or in explicit index notation

$$p_i(n+1) = \sum_{j=1}^{N} T_{ij} p_j(n). \tag{3.9}$$

An example of a transition matrix and a graphical representation of the system it describes are given in Fig. 3.2 .

$$\begin{pmatrix} 0.2 & 0.1 & 0 & 0 \\ 0.3 & 0.7 & 0.2 & 0 \\ 0.5 & 0.2 & 0.5 & 0 \\ 0 & 0 & 0.3 & 1 \end{pmatrix}$$

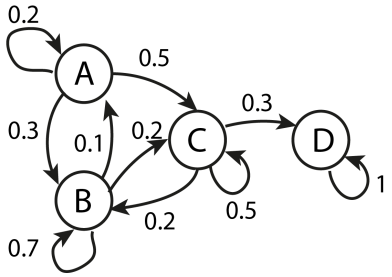


Figure 3.2: A transition matrix that describes the Markov chain illustrated graphically, with four states A, B, C and D. Transition probabilities label the directed arcs between states. The matrix is defined such that A has index 1, B index 2, C index 3 and D index 4. It would be equally valid to take any permutation of this labelling, as long as it was done consistently.

Tip 3.4.1. Note that $T_{ii} = p(x_i, n+1|x_i, n)$, which is the probability that the system starts in state x_i and ends in state x_i, is not generally zero. Each discrete step of the chain in time does not have to correspond to a change of state.

Tip 3.4.2. A single column j in a transition matrix gives the probabilities of ending in each state from the starting point x_j in a single step of the process. A system must end up somewhere after each step, so each column must sum to 1: $\sum_{i=1}^{N} T_{ij} = 1$ for all j. In fact, this is just a statement for the conservation of total probability (Eq. 2.8). Looking at the sum total of each column of a transition matrix that you have constructed is therefore a good sanity check.

Tip 3.4.3. The vector/matrix formalism can be used just as easily if you have more than one discrete variable in your Markov chain. With two coupled variables $X(t)$ and $Y(t)$, you simply need to make a list of the possible combinations of outcomes x and y, and label each combination uniquely.

Exercise 3.4.1. Getting to understand transition matrices. Some days are more productive than others, and I have unproductive days (U), quite productive days (Q) and very productive days (V). When I've been unproductive, it puts me in a bad mood and the next day is unproductive with a 50% probability; otherwise I'm quite productive on the next day. When I've been productive, I tend to rest on my laurels the next day and am only quite productive with an 80% probability (and very productive with a 20% probability). When I've been quite productive, I tend to work harder the next day (very productive with 70% probability) but with a 30% probability I am unproductive.

A. Draw a graphical representation of this Markov chain, with nodes as states and numerically labelled, directed arcs as elements of the transition matrix (as in Fig. 3.2).

B. Write out the full transition matrix, and verify that each column sums to one.

3.4.2 Long time behaviour and stationary distributions

Taking m steps simply corresponds to applying the transition matrix m times.

$$p(x, n+m) = \mathbf{T}^m p(x, n), \qquad p_i(n+1) = \sum_{j=1}^{N} (\mathbf{T}^m)_{ij} p_j(n). \tag{3.10}$$

What sort of behaviour can Markov chains exhibit over many time steps? We're often interested in how many steps the process takes to reach a certain state, or the distribution of states that is reached in the long time limit. Let us define

- A *stationary* or *limiting distribution* $\pi(x)$ is unchanged by the application of the transition matrix $\mathbf{T}$. In terms of its vector components π_i, $\pi_i = T_{ij}\pi_j = (\mathbf{T}^m)_{ij}\pi_j$. Once in the distribution $\pi(x)$, the process will stay there indefinitely.

- The *hitting time* or *first passage time* from i to j is the average time (number of steps) taken to reach state j given a starting point of i.

The stationary distribution

Firstly, we see from the definition $\pi_i = T_{ij}\pi_j = (\mathbf{T}^m)_{ij}\pi_j$ that a stationary distribution is an eignevector of the matrix $\mathbf{T}$ with eigenvalue of 1.

- At least one of these eigenvectors of eigenvalue 1 must exist to conserve probability.

- It's possible to have more than one eigenvector that satisfies this condition.

- It is possible to decompose any probability vector into a sum of eigenvectors: $p_i = \sum_k \alpha_k q_i^k$, where q_i^k is the ith component of the kth eigenvector. We won't go through how to do this explicitly - it's a bit complicated since transition matrices are not symmetric. Please refer to Section 2.4 of Ref. 5 if you are interested. This decomposition is convenient because we can then explore the evolution of each eigenvector separately under multiple applications of $\mathbf{T}$.

- If λ^k is the eigenvalue of the kth eigenvector, then applying $\mathbf{T}$ multiple times to the kth eigenvector gives $(\mathbf{T}^m)_{ij}q_j^k = (\lambda^k)^m q_i^k$. This factor would explode to infinity for $m \to \infty$ (the limit of many steps) if $|\lambda^k| > 1$, which is impossible. Therefore $|\lambda^k| \leq 1$ for all k. But for $|\lambda^k| < 1$, $(\lambda^k)^m q_i^k \to 0$ as $m \to \infty$. Thus only stationary distributions with $\lambda^k = 1$ remain relevant after many steps. If there is only one eigenvector with $\lambda^k = 1$, the process therefore necessarily converges to this distribution in the limit $m \to \infty$.

We therefore have an excellent mechanism for finding the stationary state (and hence long time behaviour) of a Markov chain; find eigenvectors of $\mathbf{T}$ with eigenvalue 1.

Exercise 3.4.2. Finding the stationary state of a simple Markov chain - well worth practising if you're unsure about how to find an eigenvector with a specified eigenvalue.

A. Find the stationary state of the Markov chain discussed in Ex. 3.4.1, using the fact that it is an eigenvector with eigenvalue 1.

B. The other eigenvalues are approximately -0.73 and 0.43. If tell you my productivity on one day, on what sort of time scale does this information become useless in making predictions?

The fundamental matrix and first passage "times"

Before we proceed, it is useful to define the *absorbing states*.

- Some Markov chains have *absorbing states*. Once the process enters this state it will never leave. An example is state D in Fig. 3.2 . Note that a system that is guaranteed to be in an absorbing state is in a stationary state, since it will never leave.

Having identified our absorbing states, we can re-order our states (change the labels i) so that the absorbing states come last. In this case, the transition matrix $\mathbf{T}$ is composed of sub matrices

$$\mathbf{T} = \begin{pmatrix} \mathbf{U} & 0 \\ \mathbf{R} & \mathbb{I} \end{pmatrix}. \tag{3.11}$$

The symbol $\mathbb{I}$ represents an identity, and the lower half of $\mathbf{T}$ has this form because absorbing states are necessarily unaffected by the application of $\mathbf{T}$. With this rearrangement,

$$\mathbf{T}^m = \begin{pmatrix} \mathbf{U}^m & 0 \\ \mathbf{R}+\mathbf{R}\mathbf{U}+\mathbf{R}\mathbf{U}^2...+\mathbf{R}\mathbf{U}^{m-1} & \mathbb{I} \end{pmatrix} = \begin{pmatrix} \mathbf{U}^m & 0 \\ \mathbf{R}\mathbf{W}(m-1) & \mathbb{I} \end{pmatrix}, \tag{3.12}$$

where I have defined $\mathbf{W}(m) = \mathbb{I}+\mathbf{U}+\mathbf{U}^2...+\mathbf{U}^m$ (note that the $\mathbb{I}$ symbol represents an identity of any appropriate dimension - the $\mathbb{I}$ appearing here is typically not the same as that in Eq. 3.11).

What does $\mathbf{W}(m)$ represent? $W_{ij}(0) = \mathbb{I}_{ij}$ is the probability that the process started in j and reached i after zero steps (which is, of course, zero unless $i = j$). $W_{ij}(1)$ is the sum of this probability and the probability that a process that started in state j is in state i after one step. Continuing this reasoning, we see that

$$W_{ij}(m) = \sum_0^m p(x_i, m|x_j, 0). \tag{3.13}$$

In words, $W_{ij}(m)$ is the expected number of times the process is in state i in the first m steps of the Markov chain after starting in j (where both i and j are not absorbing states from the definition of $\mathbf{U}$ in Eq. 3.11). Assuming that at least one absorbing state can be reached from every state in the system, then all processes are eventually absorbed. Since $W_{ij}(m)$ describes only visits to non-absorbing states, $W_{ij}(m)$ will tend to a constant in the limit $m \to \infty$, W_{ij}. This matrix $\mathbf{W}$ is the *fundamental matrix* associated with $\mathbf{U}$, and W_{ij} gives the expected number of times a process starting in j occupies state i prior to absorption, averaged over all possible lengths of trajectories prior to absorption.

The really neat trick is that the limiting $\mathbf{W}$ can be calculated. From its definition, it is easy to see that

$$\mathbf{W}(m) = \mathbb{I}+\mathbf{U}\mathbf{W}(m-1). \tag{3.14}$$

In the limit $m \to \infty$, we have $\mathbf{W}(m) = \mathbf{W}(m-1) = \mathbf{W}$. Thus

$$\mathbf{W} = (\mathbb{I}-\mathbf{U})^{-1}, \tag{3.15}$$

which is simple (if often tedious without matlab) to calculate.

Armed with $\mathbf{W}$ we can calculate the expected number of Markov chain steps before absorption (a first passage "time" t^{abs}) – this is simply the total number of expected steps spent in each non-absorbing state. If we know that the system starts in state j

$$\langle t_j^{\text{abs}} \rangle = \sum_i W_{ij}, \tag{3.16}$$

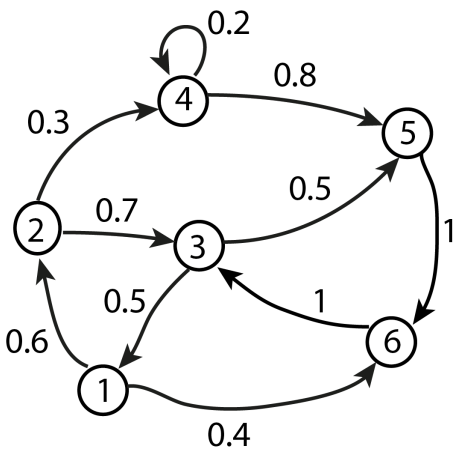


Figure 3.3: Graph representing a discrete-time Markov Chain for the purposes of Ex. 3.4.3.

where the sum runs over the non-absorbing states. If we have an uncertain initial state, distributed with probability $p_j(0)$,

$$\langle t^{\text{abs}} \rangle = \sum_{i,j} W_{ij} p_j(0). \tag{3.17}$$

What about if we've got more than one absorbing state? We might be interested in knowing which one we end up in. Returning to Eq. 3.11, we see that the total probability that a system has transitioned from an initial state j to an absorbing state l during the first m steps is given by the top right of the matrix, $(\mathbf{RW}(m-1))_{lj}$. Note that the index l stands for the lth absorbing state, just as j stands for the jth non-absorbing state. If there are two absorbing states, $l = 1$ or 2, regardless of how many states there are in total. Allowing $m \to \infty$, we obtain $(\mathbf{RW})_{lj}$ as the probability that an initial state of index j reaches the absorbing state of index l after any number of steps. As in Eq. 3.17, we can average $(\mathbf{RW})_{lj}$ over an initial distribution $p_j(0)$ if desired.

Tip 3.4.4. This analysis can also be used to find the first passage time to reach non-absorbing states, or to find out whether one non-absorbing state is reached before another. Simply re-define your process so that the target states are absorbing states (set transition probabilities out of the states in question to 0) and use the above machinery. This works because you're not worried about the dynamics after the system has reached these states for the first time, and so whether they are absorbing or not is irrelevant!

Exercise 3.4.3. Finding first passage times for a discrete-time Markov chain. Highly recommended for gaining familiarity with the fundamental matrix. Consider the Markov chain represented graphically in Fig. 3.3; we are interested in estimating the relative probabilities that a system initiated in state 1 reaches 5 or 6 first, and the average total time it takes to reach one or the other.

A. Write down the appropriate **U** and **R** matrices (hint: there is no need to re-order the states to construct these matrices; if you do, it will be difficult to compare to my answers).

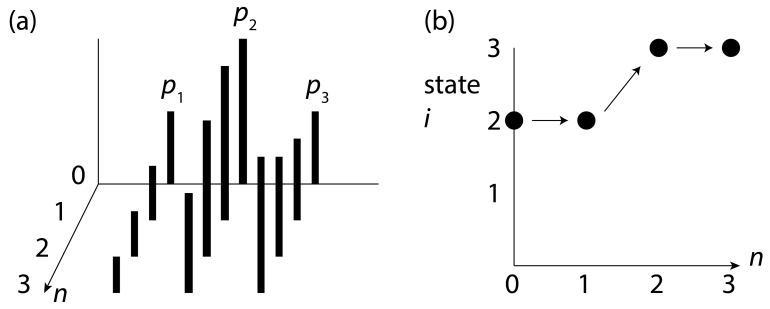


Figure 3.4: A schematic illustration of the difference between evolving a probability distribution over time in (a), and generating a sample trajectory in (b). In both cases, we consider the same 3-state system undergoing 3 Markov chain steps. In (a), we track p_1, p_2 and p_3 as a function of n; in (b), we observe the path of a single representatively sampled trajectory, which in this case is $2 \to 2 \to 3 \to 3$.

B. Verify that

$$\mathbf{W} = \frac{1}{0.632}\begin{pmatrix} 0.8 & 0.28 & 0.4 & 0 \\ 0.48 & 0.8 & 0.24 & 0 \\ 0.336 & 0.56 & 0.8 & 0 \\ 0.18 & 0.3 & 0.09 & 0.79 \end{pmatrix}$$

is the fundamental matrix of $\mathbf{U}$.

C. Calculate the probability that state 5 is reached before state 6, given an initial state $j = 1$.

D. Calculate the expected time (number of discrete MC steps) before either state 5 or 6 is reached, given an initial state $j = 1$.

3.4.3 Simulating a discrete-time Markov chain

It's not always possible to get the answers we want from an analytical treatment. When this happens, we tend to simulate the system. Simulations of a Markov chain are simple and take one of two approaches. We can either

- Evolve the whole probability distribution through time.
- Generate sample trajectories and follow their behaviour.

These two perspectives will crop up repeatedly in this course, so it's worth thinking about them in some detail in this case. The difference between them is illustrated schematically in Fig. 3.4.

Simulating the evolution of the whole probability distribution

Since the intrinsic dynamics is discrete in time, this is conceptually easy. We can simply initalize our system as a vector corresponding to the starting distribution $p_i(0)$, and apply the transition matrix $\mathbf{T}$ m times; the result will be the exact probability distibution after m steps, $p_i(m) = (\mathbf{T}^m)_{ij} p_j(0)$.

Note that there is no need to use random numbers here – the evolution of the probability distribution is deterministic. The randomness comes in when we pick an outcome x_i from that distribution.

Example pseudo-code 3.4.1. Evolving a probability distribution of a discrete-time Markov chain.

1. Define and initialise size of state space d.
2. Define probability vector p of size d, and initialize with values $p_i(0)$.
3. Define transition matrix $\mathbf{T}$ of size $d \times d$ and set its values.
4. Define desired number of Markov chain steps m and set its value.
5. Loop over n from 1 to m.
6. $\quad p = \mathbf{T}p$.
7. Output p as the final probability distribution , with values $p_i(m)$.

Other more sophisticated options include stopping the simulation when a certain state reaches a certain probability of occupancy, or recording how the occupancy of a state changes over time (enabling the calculation of first passage times if this is an absorbing state).

Generating sample trajectories using random numbers

An alternative approach is to generate a string of states that form a representative sample trajectory. Instead of following a whole vector of probabilities, we pick a single starting state randomly from $p_i(0)$. We then randomly chose the next state j with probability $p(x_j, n+1|x_i, n)$, and repeat for m steps. This will give us one possible trajectory of states, sampled from the distribution of trajectories for the process. Combining many independent sample trajectories, generated by distinct random numbers, gives us more statistics. If we're only interested in characterising the steady state, we can (usually) do this by running a single long trajectory and recording the frequency with which states are visited (after an initial transient period).

Unlike the simulation of the whole probability distribution, we need to generate random numbers, as each step involves a random choice. Getting computers (deterministic machines) to generate random numbers is a subtle problem, and in practice we actually rely on high order chaotic processes to do the job. Entertainingly, this means that we're often using high-order chaos to approximate random numbers for our model, which is using random numbers to approximate high-order chaos in the real world. Most programming languages, including Matlab, provide built-in random number generators for you to use; Matlab's is pretty good so you don't need to do anything clever with it. In particular, the instruction X=rand will give you a single random number uniformally distributed in the interval (0,1).

Example pseudo-code 3.4.2. Generating a sample trajectory from a discrete time Markov chain.

1. Define a state variable i and set it equal to an initial value.
2. Define desired number of Markov chain steps n and set its value.
3. Loop over n from 1 to m
4. Identify states l with $p(x_l, n+1|x_i, n) > 0$ and record l in a vector u, and $p(x_l, n+1|x_i, n) > 0$ in a vector v. # This could be a brute force loop over all states, but often only states in a small neighbourhood of i will need to be considered. Don't forget that $l = i$ needs to be considered.

   # Now select reaction outcome and update state.
5. Draw a random number r from an unbiased distribution in the interval (0,1).
6. Define a variable s and set it to zero.
7. Loop over k from 1 to length(v).
8. $s = s + v[k]$
9. If $s > r$
10. $i = u[k]$ (update state to new value).
11. break out of loop over k.
12. Return i as the value of the sampled trajectory after m steps of the chain.

Lines 4 and 5 calculate which states are possible after the next step, and list their indexes and probabilities. Lines 6 to 12 are a method for using a uniform random number r to select one of the options with the correct probability, and assign that state to the updated value of i. It is also possible to pick a random initial state from a distribution $p_i(0)$ using a similar technique.

A further (outer) loop over multiple trajectories would generate an ensemble of trajectories. As with the direct simulation of the evolution of the probability distribution, it is possible to ask more sophisticated questions of the simulation than simply the state i after some fixed chain steps m. Simulations could be stopped when a certain state is reached (and n recorded) to measure first passage times. Additionally, the trajectory – i at each value of n (or a subset) – could be saved to subsequently allow an analysis of the pathways taken through state space.

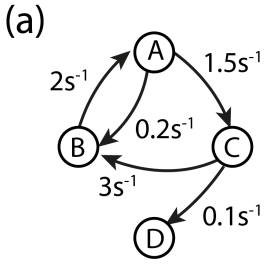
A common approach is to use trajectory-based simulations to sample the stationary distribution (when there is only one for a given system). The idea is to allow the trajectory to run for a long time, so it has time to be representative of the stationary distribution. Then we record the state at each subsequent time point in a histogram, and that histogram gives us the stationary distribution.

Simulating probability distributions vs generating sample trajectories

If simulating the entire probability distribution is conceptually so easy, why would we ever bother with generating sample trajectories?

- Some questions are easier to probe from an ensemble of trajectories. We might ask which pathways a system tends to follow when it moves from one region of state space to another, for example, which is a question about the sequence of states in individual trajectories rather than the evolution of the probability distribution as a whole.

- To evolve the entire probability distribution at each step requires a matrix multiplication, where the matrix $\mathbf{T}$ has the dimension of the number of states in the system. In realistic cases, this state space can easily become very large. For example, we might be interested in the numbers of badgers, foxes and rabbits in a certain region. Even if each is limited to the range



(b)
$$k\begin{pmatrix} -1.7 & 2 & 0 & 0 \\ 0.2 & -2 & 3 & 0 \\ 1.5 & 0 & -3.1 & 0 \\ 0 & 0 & 0.1 & 0 \end{pmatrix}$$

Figure 3.5: A continuous-time, discrete-state Markov process. (a) A graphical reprsentation of the process, showing transitions labelled by rates. Note the absence of $i \to i$ "transitions". (b) The process in (a) represented through a rate matrix, with A indexed by 1, B indexed by 3, C indexed by 3 and D indexed by 4. The constant $k = 1\text{s}^{-1}$.

$(0,100)$, there are still more than a million possible combinations specifying unique states in the system. However, even when state space explodes like this, generating sample trajectories can still be simple. Generally the system will only be able to transition to nearby states in a single step (perhaps $\pm$ one animal of each type, in our case) and the transition probabilities can be calculated using a few universal parameters (birth probabilities, death probabilities) and the details of the current state. In this case, trajectories can be sampled by simply keeping track of the current state, and never worrying about the full state space.

Simulating individual trajectories can be even more advantageous for problems with continuous time or continuous variables, as we shall see.

3.5 Discrete random variables, continuous time

3.5.1 The Master Equation

The first step up in complexity from discrete-time Markov chains is to consider processes in which events can occur in continuous time, but the underlying variables are still discrete. For clarity, I will call such systems discrete-state, continuous-time Markov processes (the term continuous-time Markov chain is also used). Instead of jump probabilities for a single step, our process is now characterised by hopping rates (see the graphical representation of a continuous time Markov process in Fig. 3.5 (a)).

Instead of a difference equation (Eq. 3.8 or 3.9), we now have a set of coupled differential equations in time known as the *Master equation*. The probability distribution $p(x,t)$ can still be represented by a vector with components $p_i(t)$, but now

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = \sum_{j\neq i} \left(K_{ij}p_j(t) - K_{ji}p_i(t)\right). \tag{3.18}$$

Here, K_{ij} is the rate at which a system in state j undergoes transitions to state i. The Master equation is simply a statement that the probability of being in state i is increased by transitions into state i from state $j \neq i$ (the first term on the RHS), and decreased by transitions out of state i into states $j \neq i$ (second term on the RHS).

The quantities K_{ij} look like they form a matrix. Indeed, the Master Equation can be re-written in terms of a true matrix multiplication

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = \sum_j K_{ij} p_j(t) \tag{3.19}$$

if we define $K_{ii} = -\sum_{j\neq i} K_{ji}$. Without index notation, Eq. 3.19 reads $\mathrm{d}p(x,t)/\mathrm{d}t = \mathbf{K}p(x,t)$. The matrix $\mathbf{K}$ is reminiscent of the transition matrix $\mathbf{T}$ that we defined for discrete-time Markov chains – but do not get them confused! I shall call $\mathbf{K}$ the *rate matrix*. The corresponding rate matrix for Markov process in Fig. 3.5 (a) is given in Fig. 3.5 (b).

Some properties of the rate matrix K

- The rate matrix $\mathbf{K}$ has units of 1/time.
- By definition, each column of the rate matrix sums to 0 ($\sum_i K_{ij} = 0$). This is a statement of the conservation of probability; increase in p_j due to K_{ij} is compensated for by decrease in p_i. Note the difference with the transition matrix $\mathbf{T}$ from discrete-time Markov chains, for which each column sums to 1.

Some simple properties of continuous-time Markov processes

- The Markov property means that the rate at which the system undergoes transitions out of state i is constant in time – regardless of how long the system has been in that state. This is why it is possible to write down Eq. 3.19 using simple, history-independent rate constants in the rate matrix $\mathbf{K}$.
- If multiple destination states j exist, the overall rate of leaving i is simply the sum of rates to each individual destination, $k_i^- = \sum_{j\neq i} K_{ji}$.
- The lifetime t_i of state i (the time spent in state i before a jump) follows an exponential distribution $p(t_i) = k_i^- \exp(-k_i^- t_i)$, with an average of $\tau_i = \langle t_i \rangle = 1/k_i^-$. Exponential distributions are discussed in Sec. 2.2.2.
- Due to the Markov property, the additional time spent in state i given that it has already survived until a time t^0, $t_i - t^0$, still follows an exponential distribution: $p(t_i - t^0 | i \,\mathrm{at}\, t^0) = k_i^- \exp(-k_i^-(t_i - t^0))$.
- Regardless of the time at which a transition occurs, the probability that state j is visited immediately after state i is given by $p_{ji} = K_{ji}/\sum_{j\neq i} K_{ji}$. Equivalently, the relative probability of jumping to state j is given by the rate constant K_{ji}.

Exercise 3.5.1. Demonstrating simple properties of continuous-time Markov processes. Helpful if you're not happy taking my claims above at face value! Consider a three-state continuous-Markov process, with states A, B and C. We're only interested in the lifetime of state A, so we'll only analyse transitions out of state A, which have rates k_{BA} and k_{CA} (B and C are absorbing states).

A. Write down the three-state master equation for this process.

B. Show that the differential equation for the probability of being in state A is equivalent to a system with a single transition out of A with a rate $k_{BA} + k_{CA}$.

C. Solve for the probability of state A as a function of time, given that the system starts in state A.

D. Derive the probability density (distribution) of the lifetime of state A (hint: the probability per unit time that the system leaves state A at t is $-\mathrm{d}p(A,t)/\mathrm{d}t$ – as discussed further in Sec. 5.1.4). Show that it matches the distribution claimed above.

E. Adjust your derivation to describe the additional lifetime of A given that it has already survived for time t^0.

F. Derive the probability of B and C over time, and show that the relative probability of transitioning to B or C is given, at each moment in time, by the ratio of rate constants.

3.5.2 Long time behaviour and stationary distributions

Formally, we can integrate Eq. 3.19 to give

$$p_i(t) = \exp(t\mathbf{K})_{ij} p_j(0). \tag{3.20}$$

Unfortunately, $\exp(t\mathbf{K}) = \mathbb{I} + t\mathbf{K} + t^2\mathbf{K}^2/2...$ is the exponential of a matrix, which is not simple to calculate unless we are able to diagonalise $\mathbf{K}$. Diagonalising $\mathbf{K}$ is thus one way to make progress; another common technique is to use generating functions (not part of the course, see Ref. 2 if interested). Stationary distributions, found by setting the LHS of Eq. 3.19 to zero, satisfy

$$0 = \sum_j K_{ij} p_j(t). \tag{3.21}$$

In other words, stationary solutions are eigenvectors of the rate matrix with eigenvalue zero. We shall look at identifying stationary distributions in some special cases using biological examples in Chapter 4.

Continuous-time Markov processes can be used to define discrete-time Markov chains

There are two quite distinct ways to do this. The first is to note that we can chose to observe our system at fixed time intervals Δt; if we do so, we would have an effective discrete-time Markov Chain with a transition matrix $\mathbf{T}(\Delta t)$ given by the exponential of the rate matrix, as implied by Eq. 3.20.

$$p_i(t+\Delta t) = \exp(\Delta t\mathbf{K})_{ij} p_j(t) = \mathbf{T}(\Delta t)_{ij} p_j(t) \tag{3.22}$$

For large Δt, $\mathbf{T}(\Delta t)$ is generally difficult to calculate and connects states i and j with arbitrarily large separation. For small Δt, however, this approach can be useful in designing discrete time step algorithms for approximate simulation of the process via a Taylor expansion;

$$p_i(t+\Delta t) \approx p_i(t) + \Delta t K_{ij} p_j(t). \tag{3.23}$$

The second way of constructing a discrete time Markov chain is conceptually quite different. We can ignore the time spent in each state, and simply ask about the sequence of states visited; we use each jump to define a step of the discrete Markov chain. To construct this type of discrete Markov Chain, simply calculate transition probabilities from i to the next state $j \neq i$ using the relative values of K_{ij}, and set $\mathbf{T}_{ii} = 0$ since, by definition, the next state can never be the same

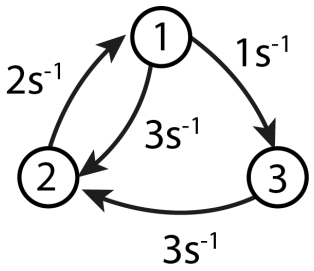


Figure 3.6: A continuous-time, discrete-state Markov process analysed in Ex. 3.5.2.

(although, for completeness, it is convenient to define $\mathbf{T}_{ii} = 1$ for absorbing states only). Once this has been done, we can apply the fundamental matrix techniques outlined in Sec. 3.4.2 to calculate numbers of jumps before absorption and the probabilities of reaching one state before another.

Tip 3.5.1. Note that, since we've forgotten all about time in converting to a discrete-time Markov chain via this second approach, we can't calculate first passage times or times to absorption directly. However, due to the Markov property we know that the expected time spent in each state, each time it is visited, is independent of the the details of the trajectory. Thus having constructed the fundamental matrix $\mathbf{W}$ of our effective Markov chain, and used this to estimate the number of visits to i, we can calculate the expected time simply by multiplying each visit frequency by the average lifetime of the state $\tau_i = 1/k_i^-$:

$$\langle t_j^{\text{abs}} \rangle = \sum_i \mathbf{W}_{ij} \tau_i. \tag{3.24}$$

This is simply Eq. 3.16 augmented with the variable average state lifetime.

Exercise 3.5.2. Using the fundamental matrix of a discrete-time Markov Chain to understand a continuous-time process. Consider the continuous-time Markov process in Fig. 3.5.

A. Write down a discrete time Markov Chain which generates the same sequence of states.

B. Use this chain to calculate the expected first passage time to reach state 3, starting at state 1.

3.5.3 Simulating a continuous-time Markov process

Evolving the whole probability distribution through time

Without diagonalising the matrix $\mathbf{K}$, this can't be done exactly. Instead, we could use an approximate numerical integration (with discrete time steps) of the coupled set of ODEs defined by Eq. 3.19. In the pseuso-code below, I outline a method using simple Euler integration based on Eq. 3.20; more sophisticated numerical integrators, such as those available in matlab, may perform better. It is worth noting, however, that this numerical integrator has the advantage of preserving $\sum_i p_i(t) = 1$ exactly at all time steps. To prevent instabilities, this integrator will require time steps smaller than the smallest value of $1/k_i^-$.

Example pseudo-code 3.5.1. Evolving a probability distribution of a continuous-time Markov process using a discrete time-step approximation.

1. Define and initialize size of state space d.
2. Define probability vector p of size d, and initialize with values $p_i(0)$.
3. Define and initialize the step size Δt.
4. Define and initialize the rate matrix $\mathbf{K}$ of size $d \times d$.
3. Define transition matrix $\mathbf{T}$ and initialize as $\mathbf{T} = \mathbb{I} + \Delta t \mathbf{K}$.
4. Define desired number of time steps m and set its value.
5. Loop over n from 1 to m.
6. $p = \mathbf{T}p$.
7. Output p as the final probability distribution , with values approximating $p_i(m\Delta t)$.

As with Example pseudo-code 3.4.1, it is possible to extract more information than the approximate state after m time steps from such a procedure.

Generating a sample trajectory using random numbers

There is an elegant method for generating exact sample trajectories for continuous-time, discrete-state Markov processes, known as the Gillespie algorithm. Instead of discretising time and generating approximate trajectories, the algorithm is event-based. The idea is to randomly generate the next event type (as in Example pseudo-code 3.4.2) and the next event time using the exact probability distributions specified by $\mathbf{K}$.

Example pseudo-code 3.5.2. Generating a sample trajectory for a continuous-time discrete-state Markov process.

1. Define a state variable i and set it equal to an initial value.
2. Define desired total simulation time T_{fin} and set its value.
3. Define current time t and initialize it to zero.
4. While $t < T_{\text{fin}}$
5. Define total transition rate k, and initialize it to zero.
6. Identify states j with $K_{ij} > 0$ and record j in a vector u and K_{ij} in a vector v. Increment $k = k + K_{ij}$. # This could be a brute-force loop over all states, but often only states in a small neighbourhood of i will need to be considered.

Draw the next reaction time by drawing a random number between 0 and 1, and converting to an exponentially distributed random number. # Often, the language in which you will be programming will allow you to draw an exponentially-distributed random number directly.

7. Draw a random number x from an unbiased distribution in the interval (0,1).
8. Initialize a time step variable dt and set $dt = -\ln(1-x)/k$.

Now select reaction outcome

9. Define a variable i_{new}.
10. Draw a random number r from an unbiased distribution in the interval (0,1).
11. $r = r \times k$.
12. Define a variable s and set it to zero.
13. Loop over l from 1 to length(v).

14. $s = s + v[l]$
15. If $s > r$
16. $i_{\text{new}} = u[l]$ (update state to new value).
17. Break out of loop over l.

If you want to record the states visited by the trajectory, you can record the old state and its lifetime at this point.

18. $t = t + dt$.
19. $i = i_{\text{new}}$.
20. Return i as the value at the end of the trajectory. # Note that this code will simulate for a time slightly larger than T_{fin}. If desired, the simulation can be truncated at exactly $t = T_{\text{fin}}$ with a little extra code.

I often find it helpful to define a function to which I pass the current state of the system, and it then returns the new state and a reaction time (essentially performing steps 5-17). Then all of the simulation management can be done in one script (length of simulation, specifying initial conditions, sampling/recording the trajectory) but the actual mechanics of deciding which state to visit next, and at what time, is kept separate.

As with Example pseudo-code 3.4.2, the external loop can be run multiple times to give a statistical ensemble of trajectories for the time period in question, and more information can be recorded during simulations than simply sampling the states at time T.

Tip 3.5.2. A common mistake is to forget that each step of the Gillespie algorithm takes a different length of time. If you're after an estimation of the time spent in each state in a long simulation (one way to calculate the stationary distribution) don't forget to make a histogram of the total *time* spent in each stae by summing over the actual lifetimes sampled. Otherwise you'll think short-lived states are really common!

3.6 Continuous random variables, continuous time

3.6.1 The Fokker-Planck equation

Finally, we come to systems with naturally continuous degrees of freedom. Now it is not possible to think of only discrete states (at least, not without making assumptions and approximations). $p(x,t)$ can no longer be represented as a finite-dimensional vector, and instead we must consider integration over functions rather than sums over vector components. Thus the Master equation for continuous variables takes the form

$$\frac{\partial p(x,t)}{\partial t} = \int_{\mathscr{A}_X} \mathrm{d}x' K(x|x')p(x',t) - K(x'|x)p(x,t). \tag{3.25}$$

Here, $K(x'|x)$ is a transition probability per unit time, and the result is an analog of Eq. 3.18 for discrete-state systems. In principle, we could imagine that long-range jumps in X are possible – but that's really a problem for purists rather than those of us interested in biologically relevant systems. When only short range jumps are possible, it makes sense to Taylor expand $K(x'|x)$, and after a bit of rearranging (that I will not go through – see Ref. 2 or 5) we arrive at the *Fokker-Planck equation*.

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(A(x)p(x,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(B(x)p(x,t)\right). \tag{3.26}$$

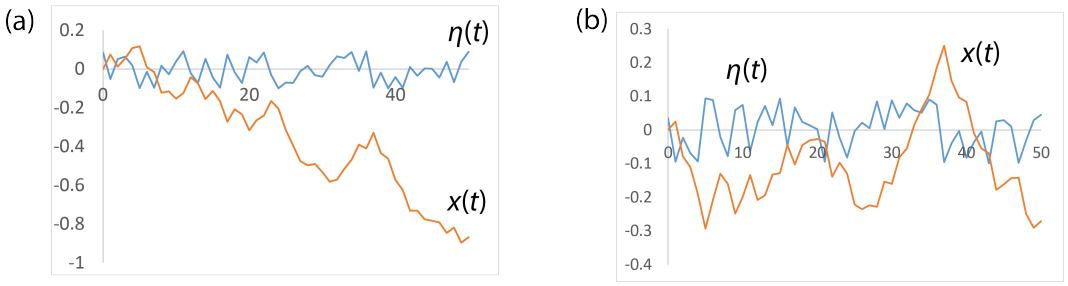


Figure 3.7: Two distinct realisations of trajectories obtained from the same Langevin equation, but with distinct noise histories $\eta(t)$. The functions $x(t)$ in (a) and (b) are solutions to the same Langevin equation with distinct sampled forms for $\eta(t)$, which are also plotted.

Here, $A(x)$ and $B(x)$ are coefficients related to the details of $K(x'|x)$. This linear PDE is generally much more convenient that the integral Master equation (Eq. 3.25), and should look vaguely similar to PDEs that you've seen for processes such as heat flow. Whether or not it is easy to solve depends on the coefficients $A(x)$ and $B(x)$.

The meaning of terms in the Fokker-Planck equation

Here I shall briefly discuss the meaning of the first and second terms in Eq. 3.26.

- The term $A(x)$ is related to the linear term in the expansion of $K(x'|x)$. It is therefore zero if transitions are intrinsically equally likely to be forwards or backwards ($K(x'|x)$ is symmetric about x), making the first term zero as a whole. When $A(x) \neq 0$ the system tends to move asymmetrically in one direction or the other; $-\frac{\partial}{\partial x}(A(x)p(x,t))$ is therefore known as the *drift term*.

- Let $B(x) = 2D$ be a constant (greater than zero). If the drift term $A(x) = 0$, then we see that

$$\frac{\partial p(x,t)}{\partial t} = D\frac{\partial^2}{\partial x^2}p(x,t). \tag{3.27}$$

The physical content of this equation is that probability tends to increase in regions with $\partial^2/\partial x^2 p(x,t) > 0$ (*ie.* troughs of $p(x,t)$) and decrease in regions with $\partial^2/\partial x^2 p(x,t) < 0$ (*ie.* peaks of $p(x,t)$). In fact, Eq. 3.27 is the *diffusion equation*, and it captures the tendency of probabilities to spread out over time. We will explore the diffusion equation in more detail in Chapter 5. In general, $B(x)$ is known as the "diffusion term" and an x-dependence of $B(x)$ implies a non-uniform diffusion coefficient.

3.6.2 The Langevin equation

In Sec. 3.4.3 and 3.5.3, we discussed the approach of generating sample trajectories through simulation for Markov processes with discrete states. For continuous random variables, we can actually write down a stochastic differential equation, the *Langevin equation*, for the process $X(t)$ (I'm not going to prove this – see Ref. 2 for a discussion). For simple systems with constant $B(X) = 2D$, and a non-trivial drift term $A(X)$,

$$\frac{\mathrm{d}X}{\mathrm{d}t} = A(X) + \zeta(t). \tag{3.28}$$

Figure 3.8: The game of snakes and ladders.

Here, the random variable of interest ($X(t)$) evolves according to a deterministic part (the drift $A(X)$) added to a particulalrly simple kind of stochastic process $\zeta(t)$, which provides a "noise" contribution and results in diffusion. $\zeta(t)$ has multiple possible realisations $\eta(t)$, each one corresponding to a different sample trajectory $x(t)$ (see Fig. 3.7). Indeed, we can also write down the differential equation for the evolution of a sample trajectory $x(t)$, if we are given a particular sample of the noise $\eta(t)$ from $\zeta(t)$:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = A(x) + \eta(t). \tag{3.29}$$

Although the outcome $\eta(t)$ of $\zeta(t)$ is variable, we can make simple statements about distributions and expectations of $\zeta(t)$. In particular, for Markov processes with $B(x) = 2D$, we take:

- $P(\zeta(t) = \eta(t)) = p(\eta(t))$ is Gaussian with $\langle\zeta(t)\rangle = \int_{-\infty}^{\infty} \eta(t)p(\eta(t))\mathrm{d}\eta(t) = 0$ at any one moment in time t.
- $\langle\zeta(t)\zeta(t')\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \eta(t)\eta(t')p(\eta(t))p(\eta(t'))\mathrm{d}\eta(t)\mathrm{d}\eta(t') = 2D\delta(t-t')$, setting the variance and autocorrelation of $\zeta(t)$. The $\delta(t-t')$ is a Dirac delta function – see discussion in Sec. 2.2.2. The delta function implies that the noise has no correlation with itself at later times; this type of randomness is known as white noise, and is fundamentally a Markovian property.

The properties of $X(t)$ are generally less simple. However, it is sometimes possible to solve the Langevin equation in terms of $\zeta(t)$; otherwise it is a useful basis for developing a simulation technique. We shall see how solutions in terms of $\zeta(t)$ can be useful in understanding the properties of random systems in Chapter 5.

Tip 3.6.1. If you're finding this hard to understand, think about the game of snakes and ladders (Fig. 3.8). Here, you roll a die and add the score to your old position Y_{n-1} to obtain your new position Y_n. However, if you start a turn on the head of a snake or the bottom of a ladder, then you either slither your way down the slake or climb up the ladder before applying your dice roll

(conventionally, this slithering or climbing is performed at the end of the previous turn, but it makes no difference to the actual game to think about it this way round).

The resultant stochastic process is a discrete analog of a continuous process described by a Langevin equation. The overall position Y_n is the analog of the process $X(t)$; the deterministically added slithering or climbing, along with the average die roll of 3.5, is the analog of $A(X)$; and the deviation of die rolls about 3.5 is the stochastic process corresponding to $\zeta(t)$. A particular sequence of die rolls corresponds to the sample $\eta(t)$, which in turn implies a particular sample trajectory for the position (y_n, the analog of $x(t)$). The fact that all of the die rolls are independent makes the process Markovian, since updates don't depenend on the details of previous updates except through the current position.

For a continuous process, the fluctuations are defined at each point in continuous time, not just at each roll. For the system to be Markovian, the random noise at each point in time must be uncorrelated from all others (just as with the die rolls). But for a continuous process, this effectively means that a random force in the $+x$ direction only lasts for an infinitely short time before we see a random force in the $-x$ direction. It thus sounds like these random forces will cancel out and give us nothing. However, if the variance of $\eta(t)$, $\langle\eta(t)\eta(t)\rangle - \langle\eta(t)\rangle^2$ is infinite, this isn't the case. We have "infinitely strong" forces acting for "infinitely short" times, and the two infinities effectively cancel. $\langle\eta(t)\eta(t')\rangle = 2D\delta(t-t')$ is exactly the mathematical way to quantify these "infinitely strong" forces acting for "infinitely short" times. Overall, during a finite time window Δt, we see a net random kick with a non-zero, but finite, variance (we will use this in simulations as discussed below). Clearly this assumption of δ-correlated noise is a mathematical idealisation, but it turns out to be an incredibly useful and powerful one!

Tip 3.6.2. Note that the construction of the Langevin equation from the Fokker-Planck equation in the way I have outlined becomes more complicated when the random noise is X-dependent. We won't cover the details here: see Ref. 2 if interested.

3.6.3 Simulating a continuous-time, continuous-space Markov process [details not required]

The Fokker-Planck equation (Eq. 3.26) is a partial differential equation for $p(x,t)$. In general, approximate solutions can be obtained using numerical PDE solvers, which typically work by approximating the PDE as a set of discrete difference equations on a grid of points. Matlab is a good tool for solving PDEs; I will not provide pseudo-code here, as it is either too involved or too toolkit-specific for the purposes of this course. Simple approaches for generating trajectories using the Langevin equation, however, are reasonably illuminating. The basic idea is to discretise Eq. 3.28 in units of time δt, and approximate the trajectory evolution during δt (numerical integration). To do so, it is necessary to generate a random increment by sampling from

$$\Delta(\delta t) = \int_{t_0}^{t_0+\delta t} \zeta(t)\mathrm{d}t, \tag{3.30}$$

which is the effect of the random term in incrementing x during time δt. To proceed, we note that $\Delta(\delta t)$ is essentially a sum over uncorrelated Gaussians. A useful property of Gaussian distributions is that the sum of independent Gaussian random variables is also a Gaussian. So $\Delta(\delta t)$ itself is a Gaussian random variable, and the only things we need to find in order to sample from $\Delta(\delta t)$ and

generate random increments are $\langle \Delta(\delta t) \rangle$ and $\mathrm{Var}(\Delta(\delta t))$.

$$\langle \Delta(\delta t) \rangle = \int_{t_0}^{t_0+\delta t} \langle \zeta(t) \rangle \mathrm{d}t = 0, \tag{3.31}$$

since the random noise has zero mean at each time point. For the variance, however, we obtain

$$\mathrm{Var}(\Delta(\delta t)) = \langle \Delta(\delta t)^2 \rangle - 0 = \langle \int_{t_0}^{t_0+\delta t} \zeta(t) \mathrm{d}t \int_{t_0}^{t_0+\delta t} \zeta(t') \mathrm{d}t' \rangle. \tag{3.32}$$

Note that when we square $\Delta(\delta t)$, it is important that we multiply together two separate (but identical) integrals over two separate variables t and t', rather than accidentally using the same dummy variable in each integral – which leads to chaos. This enables us to get the right expression when we combine the two integrands

$$\mathrm{Var}(\Delta(\delta t)) = \int_{t_0}^{t_0+\delta t} \int_{t_0}^{t_0+\delta t} \langle \zeta(t) \zeta(t') \rangle \mathrm{d}t \mathrm{d}t'. \tag{3.33}$$

Using $\langle \zeta(t) \zeta(t') \rangle = 2D\delta(t-t')$, we obtain

$$\mathrm{Var}(\Delta(\delta t)) = \int_{t_0}^{t_0+\delta t} \int_{t_0}^{t_0+\delta t} 2D\delta(t-t') \mathrm{d}t \mathrm{d}t' = \int_{t_0}^{t_0+\delta t} 2D \mathrm{d}t = 2D\delta t. \tag{3.34}$$

The final stages require us to use the properties of the delta function, as discussed in Sec. 2.2.2. Given Eq. 3.34, we are able to propose a simple method for integrating the Langevin equation to find a sample trajectory $x(t)$, based on Euler's method and a particular sequence of random numbers.

Example pseudo-code 3.6.1. Integrating the Langevin equation for a continuous-time, continuous-space Markov process.

1. Define function $A(x)$ that returns the drift term given x.
2. Define and initialise desired integration time T.
3. Define and initialise integration time step δt.
5. Define and initialise the standard deviation of the finite-time increment $\sigma = \sqrt{2D\delta t}$.
4. Define trajectory variable x and initialise as desired.
5. Define current time t and initialise to zero.
6. While $t \leq T$
7. $\quad t = t + \delta t$

Here you need to use a random variable to generate a sample of $\Delta(\delta t)$ over a time step. drawn from a Gaussian distribution with zero mean and standard deviation σ. Most mathematical computing libraries have an appropriate pre-programmed function, although writing one yourself is not too hard.

8. $\quad x = x + A(x)\delta t + rand\_gaussian(0, \sigma)$.
9. Output the final value x.

Once again, an exterior loop would allow the production of many sample trajectories, and it is possible to extract more information than only the final state. The method is only approximate, and improves as $\delta t \to 0$. For serious applications, methods that are less intuitive (but more accurate) should be explored.

Chapter 4

Stochastic processes in biology: Stationary distributions

Stationary distributions are a natural place to start when considering any given stochastic process. Most obviously, many processes reach the stationary distribution on a relatively rapid timescale, and thus it is the probability distribution of direct relevance to our observation. In other cases, comparing the stationary distribution to the current distribution tells us a lot about how we might expect the system to evolve. In this chapter we will focus on some biological example systems in which non-dynamical properties of the stationary distribution are relevant.

It is important to remember that in a stochastic system, being in the stationary distribution does not mean that a system sits in a single, predictable state. In fact, systems will continue to show variability, both between realisations and in one realisation over time. We are often as interested in this variability as the average values obtained in the stationary distribution.

4.1 Detailed balance

There are two fundamentally distinct types of stationary process, those that obey *detailed balance* and those that do not. To understand the difference, it is useful to define the *flux* ϕ_{ji} from a state i to a state j (for convenience, I will use discrete state notation, but everything also applies to the continuous picture):

$$\phi_{ji} = K_{ji}p_i - K_{ij}p_j \tag{4.1}$$

for a continuous-time process, and

$$\phi_{ji} = T_{ji}p_i - T_{ij}p_j \tag{4.2}$$

for a discrete-time Markov chain. The flux ϕ_{ji} is then the *net* probability that a system transitions from i to j per unit time/in a single Markov chain step. In a stationary distribution π with detailed balance, $\phi_{ji} = 0$. In other words, for a distribution with detailed balance, you're just as likely to see a system go from i to j as from j to i when inthe stationary distribution; these two transitions balance each other, and this balance holds for each and every pair of states – which is why it is known as detailed balance. Note that detailed balance is fundamentally *not* the same thing as $K_{ij} = K_{ji}$ (or $T_{ij} = T_{ji}$). If $K_{ij} > K_{ji}$, a stationary distribution with detailed balance distribution will compensate through $\pi_i > \pi_j$.

The difference between a stationary distribution possessing detailed balance and one that does not is illustrated in Fig. 4.1. Note that in both systems, the stationary distribution Π is actually the

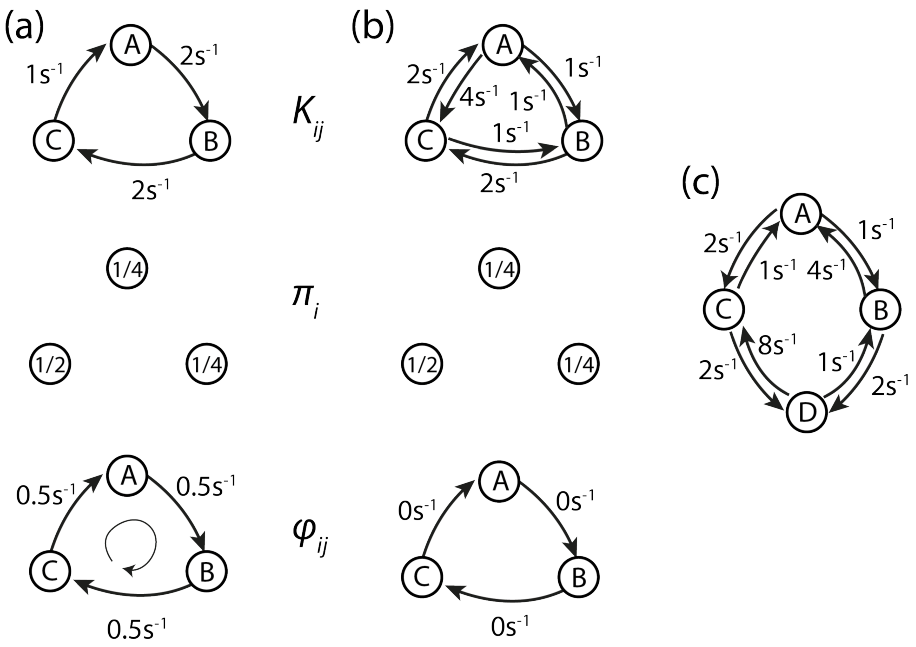


Figure 4.1: (a) and (b) depict two distinct continuous-time Markov processes graphically. Along with the rate matrices K_{ij}, the state occupancies in the stationary distribution (π_i – actually equal for the two systems) and the flux in the stationary distribution (ϕ_{ij}) are shown. The process in (a) does not obey detailed balance, but the process in (b) does. The result is a non-zero net flux ϕ_{ij} in (a), reflecting a tendency for trajectories to move in a clockwise sense around the three-state loop. Conversely, $\phi_{ij} = 0$ in (b). (c) depicts a four-state system that obeys detailed balance, as referenced in Ex. 4.1.1.

same. However, in the system that does not have detailed balance, individual trajectories tend to flow around a loop of states. This behaviour is characteristic in the absence of detailed balance, but is impossible when detailed balance is present.

Since the stationary distribution of a given process is determined by the underlying transition or rate matrices, we can meaningfully say that the matrices themselves (and in fact the process as a whole) either obeys detailed balance or not. This fact is significant; as we shall see in Sec. 4.1.1, we often have good reason to construct stochastic models that obey detailed balance – even if we're not particularly interested in the stationary distribution itself.

Tip 4.1.1. It is easy to see that the system in Fig. 4.1 (a) does not obey detailed balance; all reactions $i \to j$ have no corresponding reaction $j \to i$. Thus $i \rightleftharpoons j$ can never show detailed balance. However, even systems in which every reaction has an inverse may not obey detailed balance. For example, if any one (and only one) of the rate constants in Fig. 4.1 (b) was changed, detailed balance would necessarily be violated.

Tip 4.1.2. The existence of detailed balance makes it very easy to calculate the relative probabilities in the stationary distribution, π_i/π_j. If we know K_{ij} and K_{ji} (or T_{ij} and T_{ji}), then the combination of $\phi_{ij} = 0$ and Eq. 4.1 or 4.2 give

$$\frac{\pi_i}{\pi_j} = \frac{K_{ij}}{K_{ji}} \text{ or } \frac{T_{ij}}{t_{ji}}. \tag{4.3}$$

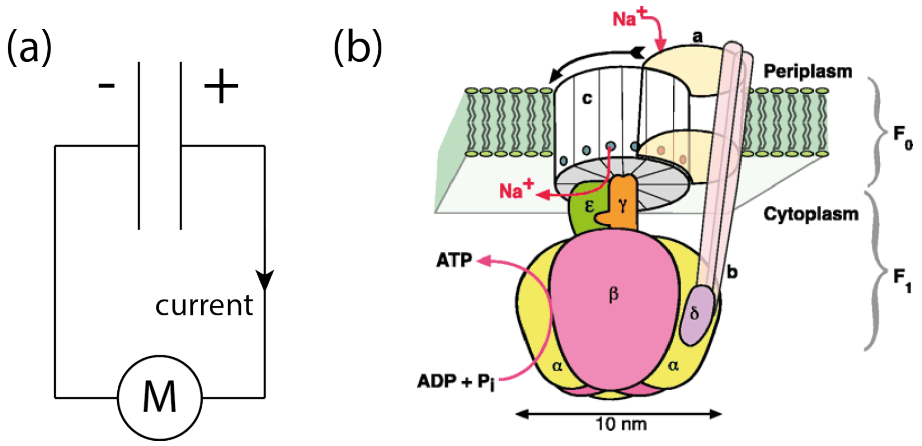


Figure 4.2: (a) A simple circuit with a discharging capacitor connected to a motor. As the capacitor relaxes towards equilibrium, net current flows through the motor and drives a systematic rotation. (b) F0F1-ATPase, used in cells to generate ATP from ion currents through membranes. The membrane acts as a capacitor, which can hold an imbalance of ions on either side due to its low conductivity. The motor acts as a conducting wire, allowing a flow of ions that would lead to equilibration of the membrane capacitor, if the imbalance were not being continuously topped up by the action of ion pumps (not shown). This flow is coupled to motor rotation, which in turn is coupled to conversion of ADP and inorganic phosphate into ATP. Taken from Dimroth *et al.*, Proc. Nat. Acad. Sci 96:4924-4929 (1999).

If it is impossible to directly transition from i to j, we can still calculate π_i/π_j by finding an intermediate state k that allows a connection to be made:

$$\frac{\pi_i}{\pi_j} = \frac{K_{ik}K_{kj}}{K_{jk}K_{ki}} \text{ or } \frac{T_{ik}T_{kj}}{T_{jk}T_{ki}}. \tag{4.4}$$

Clearly this procedure generalises to an arbitrary number of intermediate states. Note that although relative probabilities are easy to calculate in a system with detailed balance, absolute probabilities (which require normalisation) generally are not.

Exercise 4.1.1. Gaining some familiarity with detailed balance.

A. Calculate the relative probabilities of states A and D in Fig. 4.1 (c) in the stationary distribution.

B. Would this answer change if the states and transition rates shown were embedded in a larger state space which still obeyed detailed balance?

4.1.1 Detailed balance, thermal equilibrium and driving

We can imagine a stochastic system that is not externally driven by input of energy in a useful form, such as chemical fuel molecules, electrical energy, or light. The probability distribution of such a system will eventually relax to a stationary distribution known as the *thermal equilibrium distribution*.

Consider, for example, a capacitor. If the two plates are connected by a conductor, charges will tend to flow (in a small enough system, this flow will be noticeably stochastic). If the capacitor isn't constantly topped up (driven) by an external power source, it will eventually discharge and no more (net) current will flow, although in small systems the probability distribution will exhibit noticeable fluctuations in the charge stored. This eventual stationary distribution is thermal equilibrium.

Systems relaxing to equilibrium can be exploited to perform "work". In this case, we could attach an electric motor to the circuit (Fig. 4.1.1 (a)); in the process of relaxing towards equilibrium, the capacitor drives the motor. This setup is in fact a pretty good model of naturally occurring molecular motors such as F1F0 ATPase that are powered by a flow of ions across a membrane (Fig. 4.1.1 (b)). Once the capacitor is at equilibrium, however, it can no longer be used to drive the motor; it no longer stores usable energy in and of itself, a general feature of equilibrated systems. Living cells use ion pumps to make sure the cell membrane "capacitor" remains charged, rather than relaxing to equilibrium (this is a form of driving).

Systems that have reached equilibrium cannot be exploited to perform work or drive another system. If trajectories of equilibrated systems tended to flow around a loop (as in Fig. 4.1 (a)), however, it would in principle be possible to use this flow to drive another process, just as the net flow of electrons can be used to drive an electric motor. Thus:

- Equilibrium distributions necessarily obey detailed balance.
- Stochastic processes that are not driven, and hence eventually relax to equilibrium, have transition/rate matrices consistent with detailed balance.

4.1.2 The equilibrium (Boltzmann) distribution

Thermodynamic equilibrium systems have a well-known stationary distribution. In biology, we're usually interested in systems that are maintained at a constant temperature by their surroundings (with which they can exchange energy in the form of "heat"). In this case, the energy of the system of interest is not constant, and if our "states" are a truly microscopic (atomic-level) description of reality, the stationary distribution is given by

$$\pi_i \propto \exp(-E_i/k_BT), \tag{4.5}$$

in which E_i is the energy of state i, T the temperature and k_B Boltzmann's constant. The proportionality constant is given by normalising via $Z = \sum_i \exp(-E_i/k_BT)$. Typically, however, we don't describe systems at the atomic level of detail – we talk in terms of *macro*states like describing proteins as "active" or "inactive". The probability distribution of macrostates is given by

$$\pi_i \propto \exp(-F_i/k_BT), \tag{4.6}$$

where the *free energy* F_i is given by

$$F_i = -k_BT \ln \sum_{j \in i} \exp(-E_j/k_BT), \tag{4.7}$$

in which the sum over j runs over all microscopic states j in macrostate i. You can think of F_i as just a logarithmic measure of relative probabilities of macrostates inequilibrium, and as a directly measurable quantity in experiment. However, knowing that this F_i comes from Eq. 4.7 is still useful. If we perturb the system, for example by exerting a force, we change the energies of microstates.

If we do this in such a way that the energies of all the microstates in a given macrostate change in energy by the same amount ΔE_j, then $\Delta F_j = \Delta E_j$. We shall exploit this fact in the following sections.

A more complete but still easily readable discussion of the Boltzmann distribution and its underlying basis is given in Ref. 4. More technical discussions can be found in Ref. 6. Note that the free energy introduced here is the Helmholtz free energy, which applies to systems at constant temperature and volume. The Gibbs free energy (G) plays an equivalent role for systems at constant temperature and pressure. We shall not worry about the differences between the two. Indeed, in the common context of modelling solutes (eg. proteins) in solvent that is treated implicitly, they are essentially identical since the solvent maintains an essentially constant volume.[4]

The Boltzmann distribution and detailed balance

Systems that are not driven thus have two useful features: they obey detailed balance, and their stationary state is given by the Boltzmann distribution. Therefore, using Eq. 4.1 and 4.6,

$$\frac{K_{ij}}{K_{ji}} = \frac{\pi_i}{\pi_j} = \exp(-(F_i - F_j)/k_B T). \tag{4.8}$$

Thus free energy differences specify relative rate constants. This is incredibly useful and important when designing models of non-driven systems. If we wish to design a model that respects detailed balance, it is generally easiest to create a model for F_i, and then estimate either K_{ij} or K_{ji}, with the other rate constant following from Eq. 4.8.

Exercise 4.1.2. Designing a Markov process that respects detailed balance. Consider the binding of two competing transcription factors, A and B, to a promoter site. Assume that the free energy changes associated with binding to an empty site are $\Delta F_{A0} = -2k_B T$ and $\Delta F_{B0} = -4k_B T$.

A. What are the probabilities of the three possible states of the system (A-bound, B-bound and empty) in the stationary state?

B. The binding step is assumed to be relatively insensitive to the identity of the transcription factor, occurring at a rate of k_0 in both cases. What is the rate of unbinding of A and B form the promoter, assuming detailed balance?

C. B is observed to directly displace A through a novel competition mechanism, with a rate $k_0/2$. What is necessary for detailed balance to be maintained? Does this change the answer to part A?

4.2 Biological systems in thermodynamic equilibrium

Living systems, of course, never reach equilibrium; being alive means constantly consuming energy-rich molecules to maintain a non-equilibrium state. However, small motifs within living systems can often be described and understood using equilibrium-based modelling; I will give some examples of systems relevant to bioengineering here.

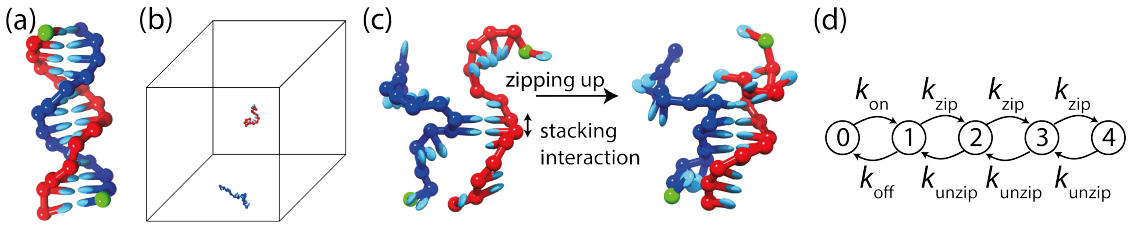


Figure 4.3: DNA, and a simple model of duplex formation. (a) a simple 14-base duplex, as represented by a coarse-grained model. (b) Two complementary strands diluted in a large box. (c) The formation of the first contact between strands brings the remaining base pairs into close proximity; subsequent zipping up is aided by this proximity and the stacking interaction between neighbouring base pairs. (d) A simple continuous-time Markov model for duplex formation; states are labelled by the number of base pairs. Model images taken from Ouldridge *et al.*, Nucl. Acids Res. 41, 8886-8895 (2013).

4.2.1 Heterogeneity of macromolecules

Macromolecules such as proteins and nucleic acids can take on multiple different conformations. Sometimes this is essential to their function, and sometimes it is indicative of a disease state (such a when misfolded proteins aggregate into prions). It is very common to identify discrete macromolecular states and develop Markov models to understand their behaviour. Indeed, Ex 4.1.2 is a very simple example of exactly this approach. Typically, if (implicit) fuel molecules such as ATP are not consumed during the transitions in question, these processes obey detailed balance and often get close to equilibrium.

Cooperativity in nucleic acid binding

DNA (illustrated in Fig. 4.3 (a)) must fulfill two roles; binding of two long strands must be very strong to keep the genetic material intact, but it must be possible to open individual base pairs to read and copy the genetic material. This double-act is achieved through cooperative behaviour. Consider two complementary strands of length N solvated in a box of volume V (Fig. 4.3 (b)). If $N = 1$, each strand is a single base. Unless the box is absolutely tiny, these bases will not be bound for a significant fraction of the time in equilibrium. It takes a large amount of time to find each other by diffusing around the box, but once in contact they are only weakly held together, and will rapidly fall apart. Let's assume that the two bases find each other at a rate $k_{\rm on}$, and detach at a rate $k_{\rm off}$.

Now consider $N > 1$. To form the first base pair, the strands must again find each other. But the task of forming subsequent base pairs once the strands are in contact is much easier: they're already held in close proximity (see Fig. 4.3 (c)). What is more, "stacking" interactions between base pairs (illustrated in Fig. 4.3 (c)) add stability to the structure. We'd therefore expect subsequent base pairs to zip up (Fig. 4.3 (c)) at a high rate compared to $k_{\rm on}$, and fall apart at a low rate compared to $k_{\rm off}$. Indeed, we might define a minimal Markov model for DNA binding using $k_{\rm on}$ and $k_{\rm off}$ for the formation and breaking of the first base pair, and $k_{\rm zip} \gg k_{\rm on}$ and $k_{\rm unzip} < k_{\rm off}$ for all subsequent base pairs. The resultant discrete-state Markov model for $N = 4$ is illustrated in Fig. 4.3 (d).

It is convenient to define

$$\sigma = \frac{k_{\rm on}}{k_{\rm off}} = \exp(-(F_1 - F_0)/k_B T), \tag{4.9}$$

where F_m is the free energy of the m-base-pair state, and

$$t = \frac{k_{\text{zip}}}{k_{\text{unzip}}} = \exp(-(F_{\text{m}} - F_{\text{m}-1})/k_B T) \text{ for } 2 \leq m \leq N, \tag{4.10}$$

In general, we expect $t \gg \sigma$; $t > 1$ implies more base pairs tend to form once the first is in place. The analysis of the stationary distribution of this model is performed in Ex. 4.2.1.

Exercise 4.2.1. Establishing the equilibrium properties of a simple DNA-binding model (Fig. 4.3 (d)).

A. Argue that the equilibrium probability that the number of base pairs M exceeds zero is

$$P(M > 0) = \frac{\sum_{m=1}^{N} \sigma t^{m-1}}{1 + \sum_{m=1}^{N} \sigma t^{m-1}} \tag{4.11}$$

with σ and t defined in Eq. 4.9 and 4.10.

B. Evaluate this fraction, noting the fact that the sums are geometric progressions (you should be familiar with these from school - look them up if necessary!)

C. Argue that $P(M > 0) \approx \frac{\sigma t^{N-1}}{1+\sigma t^{N-1}}$ if t is significantly larger than 1.

We see that if $t^N < \sigma^{-1}$, the two strands are mostly unbound. The stability of the duplex, however, grows exponentially with N. Thus at large N, the bound state is incredibly stable, even if σ is small and t is only marginally greater than 1. This cooperativity explains how very long duplex DNA is stable enough to function as genetic storage material, whilst individual base pairs can be disrupted relatively easily.

4.2.2 Polymer elasticity

Polymers are molecules that consist of repeated units connected to each other in a 1d chain. They are ubiquitous, both in biological and non-biological settings. An important property of polymers is their stiffness; stiffer polymers tend to bend through large angles only on longer length scales (Fig. 4.4). Some polymers, such as typical microtubules in cells, are short compared to the natural length scale of bending. This allows them to provide a fairly rigid structural scaffold. Others, such as genomic DNA, are long enough to be bent easily. In the case of DNA, this is essential in allowing an extremely long polymer to be packed into a cell.

Lattice FJC model

We will consider a very simple lattice model to understand the mechanical properties of flexible polymers in equilibrium. Formally, this model is a freely-jointed chain (FJC) on a lattice (Fig. 4.5). The FJC assumes that a polymer consists of N rigid sections of length b (which may or may not coincide with the constituent monomers). The connection between each section is completely flexible and overlap of monomers is not prohibited in the simplest case (the chain can cross itself or go back on itself). These simplifications mean that all chain configurations are equally likely (equivalently, they have equal free energy). Thus:

- Each segment behaves independently.

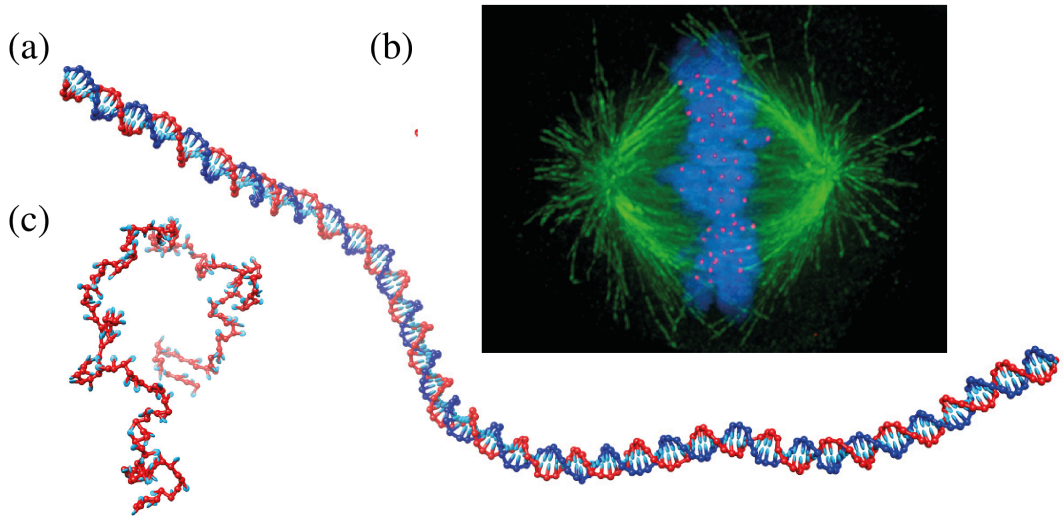


Figure 4.4: Polymers of varying stiffness. (a) A model representation of a DNA duplex, which is stiff on a length scale of less than 150 base pairs (50nm). (b) Microtubules (green) during cell division, showing stiffness on the cellular length scale (the blue region is highly coiled chromosomal DNA). (c) a model of single-stranded poly-T DNA, showing very high flexibility. This system is not far from being a freely-jointed chain at the level of single bases. (a) and (c) taken from Ouldridge *et al.*, J. Chem. Phys. 134:085101 (2011); (b) taken from https://en.wikipedia.org/wiki/Microtubule#/media/File:Kinetochore.jpg

- Every direction is equally favourable for each segment.

Thus each microscopic configuration has the same free energy. We further simplify the FJC model by assuming that each segment is aligned with one of the directions on a cubic lattice of unit length b. Thus, for each segment, there are six equally-probable orientations: $\pm\hat{x}$, $\pm\hat{y}$ and $\pm\hat{z}$.

The end-to-end vector $\mathbf{R} = \sum_{n=1}^{N} \mathbf{R}_n$ is a random variable, where $\mathbf{R}_n$ is the vector of the nth polymer unit.

$$\langle \mathbf{R} \rangle = \sum_{n=1}^{N} \langle \mathbf{R}_n \rangle = 0, \tag{4.12}$$

since for each segment, each component is equally likely to be $\pm b$ in any direction. This does not mean, however, that the typical end-to-end distance is zero; only that there is no bias for one direction over another. Indeed

$$\langle \mathbf{R}^2 \rangle = \sum_{m=1}^{N} \sum_{n=1}^{N} \langle \mathbf{R}_n \cdot \mathbf{R}_m \rangle = \sum_{m=1}^{N} \sum_{n \neq m}^{N} \langle \mathbf{R}_n \cdot \mathbf{R}_m \rangle + \sum_{n=1}^{N} \langle \mathbf{R}_n \cdot \mathbf{R}_n \rangle. \tag{4.13}$$

Since each unit is independent, $\langle \mathbf{R}_n \cdot \mathbf{R}_m \rangle = \langle \mathbf{R}_n \rangle \cdot \langle \mathbf{R}_m \rangle = 0$ if $n \neq m$, and thus the first term on the RHS in Eq. 4.13 is zero. By contrast, $\langle \mathbf{R}_n \cdot \mathbf{R}_n \rangle = \langle \mathbf{R}_n^2 \rangle = b^2$. Thus

$$\langle \mathbf{R}^2 \rangle = Nb^2 \tag{4.14}$$

for the lattice FJC. Eq. 4.14 tells us that the typical end-to-end distance of a flexible polymer in equilibrium scales with only $N^{1/2}$, rather than as N as for a stiff polymer; the flexibility leads to a moderately compact typical structure (Fig. 4.5).

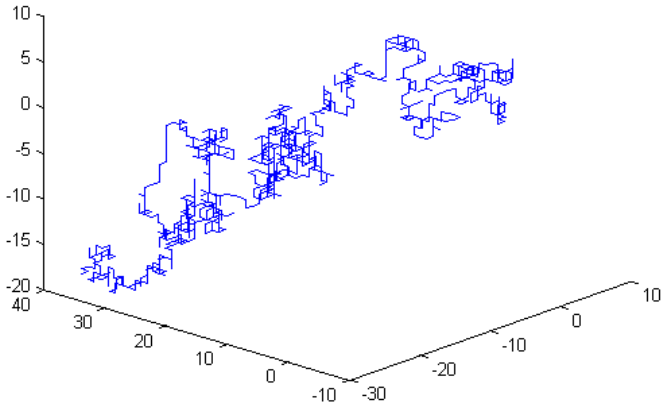


Figure 4.5: A representative configuration of a lattice FJC polymer of 1000 units (each of length 1). the polymer starts at (0,0,0).

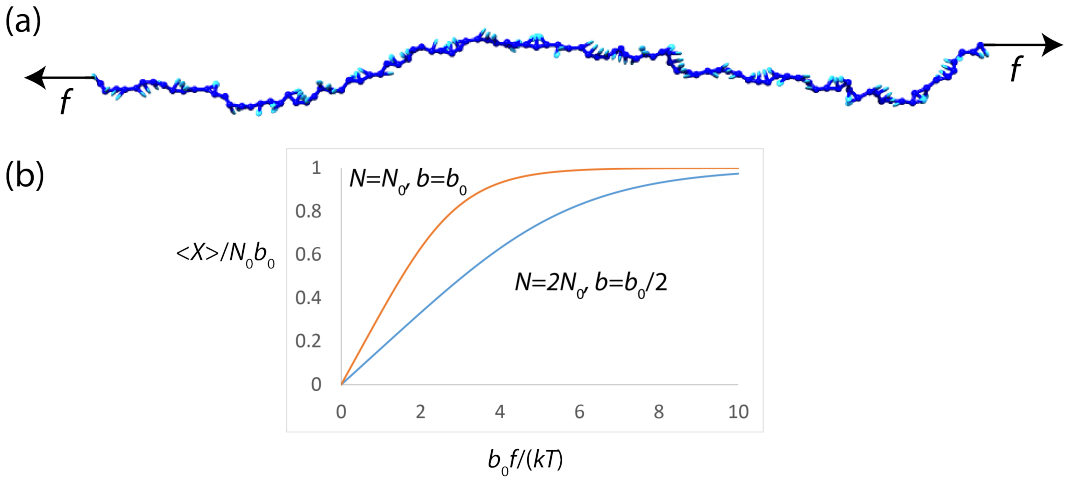


Figure 4.6: Applying force to a polymer. (a) The response of single stranded DNA to an applied tension, as represented by a coarse-grained model. The polymer extends along the force direction, but is not fully aligned. (b) The force-extension response of two lattice polymers, one with $N = N_0$ and $b - b_0$, the second with twice as many units but each of half the length. Note that it is harder to extend the polymer with more units.

Applying a force

We can consider applying a tension to the ends of the polymer, perhaps by optical tweezers or some cellular process. In response, it will extend in the force direction (Fig. 4.6 (a)). If we apply a tension f in the $\hat{x}$ direction, the free energy of each configuration is changed by $\Delta F = -fX = -\sum_{n=1}^{N} fX_n$,where X_n is the x-component of the nth segment in the configuration. As the free energy is simply a sum over terms for individual segments, each segment remains independent and thus

$$\langle X \rangle = \sum_{n=1}^{N} \langle X_n \rangle. \tag{4.15}$$

Unlike the force-free case, $\langle X_n \rangle \neq 0$. In fact,

$$p(x_n) \propto \begin{cases} \exp(bf/k_BT) & \text{for } x_n = b, \\ \exp(-bf/k_BT) & \text{for } x_n = -b, \\ 4 & \text{for } x_n = 0. \end{cases} \tag{4.16}$$

Thus

$$\langle X \rangle = N \frac{b\exp(bf/k_BT) - b\exp(-bf/k_BT)}{4 + \exp(bf/k_BT) + \exp(-bf/k_BT)} = Nb \frac{\sinh(bf/k_BT)}{2 + \cosh(bf/k_BT)} \tag{4.17}$$

The response of the polymer to applied force is plotted in Fig. 4.6 (b) At low forces, the polymer resists stretching like a linear spring. Interestingly, the polymer resists stretching because the extended state has fewer available configurations, not higher energy. This is a classic example of an *entropic force*. Find out more about entropic forces in Ref. 4. At higher forces, the extension saturates at the maximum possible length Nb.

Exercise 4.2.2. Limiting behaviour of extended polymer.

A. By expanding Eq. 4.17 for low force, find the spring constant of an FJC lattice polymer of length N and unit size b.

Tip 4.2.1. It is also possible to view $\mathbf{R}_N = \sum_n^N \mathbf{R}_n$ as a stochastic process, in which the total end-to-end vector of the polymer changes as more units N are considered. The number of units is then the"time" variable. Indeed, this is a lattice-based discrete model of diffusion (see section 5.1.3 for a continuum model), and much of the same intuition applies.

4.2.3 Overdamped motion

Small objects moving in water experience three types of forces:

- Forces due to interactions with other solutes, or a potential $V(X)$.
- A drag force $-\gamma U$, with U the velocity.
- A random noise due to buffeting by the water molecules, which we denote $\gamma\zeta(t)$. We assume this random noise is Gaussian with zero mean and its autocorrelation is proportional to a δ-function (as in Chapter 3). A fundamental theorem known as the *fluctuation-dissipation relation*[2,5], connects friction, temperature and random noise, requiring

$$\langle \zeta(t)\zeta(t') \rangle = \frac{2k_BT}{\gamma}\delta(t-t'). \tag{4.18}$$

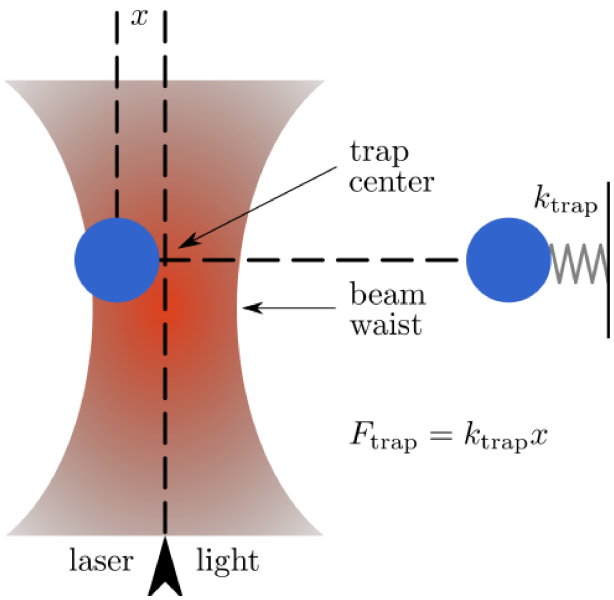


Figure 4.7: A schematic of a dielectric bead restrained in a 1d optical trap at the focus of a laser. The bead experiences a restoring force towards the centre of the trap due to its tendency to refract the light of the beam. Taken from https://commons.wikimedia.org/wiki/File:Optical_trap_principle.svg.

Restricting ourselves to 1d, and with only a fixed potential, these forces specify the following stochastic equations of motion for an object of mass m:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = U, \tag{4.19}$$

$$m\frac{\mathrm{d}U}{\mathrm{d}t} = -\gamma U - \frac{\mathrm{d}V(X)}{\mathrm{d}X} + \gamma\zeta(t). \tag{4.20}$$

For small objects, water appears very viscous, and momentum is quickly lost. In this limit, a common approximation is that the $m\mathrm{d}U/\mathrm{d}t$ term is small, so that the RHS of Eq. 4.20 can be set equal to zero and combined with Eq. 4.19 to give the *overdamped Langevin equation* for X only.

$$\frac{\mathrm{d}X}{\mathrm{d}t} = -\frac{1}{\gamma}\frac{\mathrm{d}V(X)}{\mathrm{d}X} + \zeta(t). \tag{4.21}$$

Note that Eq. 4.21 has exactly the same form as Eq. 3.28. Thus it is also possible to describe the small object's motion with the Fokker-Planck equation, using $A(x) = -\frac{1}{\gamma}\frac{\mathrm{d}V(x)}{\mathrm{d}x}$ and $B(x) = 2D = \frac{2k_BT}{\gamma}$:

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{\gamma}\frac{\partial}{\partial x}\left(\frac{\mathrm{d}V(x)}{\mathrm{d}x}p(x,t)\right) + \frac{k_BT}{\gamma}\frac{\partial^2}{\partial x^2}p(x,t). \tag{4.22}$$

We will explore the dynamical properties of overdamped motion in Chapter 5.

A bead in an optical trap

Eq. 4.22 applies to overdamped motion in any potential. A common example is motion in a harmonic trap $V(x) = \kappa/2(x-x_0)^2$. For example, this is the potential exerted by optical tweezers

on a trapped object (see Fig. 4.7), and optical tweezers are one of the key experimental tools in single-molecule studies. For this potential, we obtain

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{\gamma}\frac{\partial}{\partial x}\left(\kappa(x-x_0)p(x,t)\right) + \frac{k_B T}{\gamma}\frac{\partial^2}{\partial x^2}p(x,t) \tag{4.23}$$

It is easily verified that $p(x) \propto \exp(-\kappa(x-x_0)^2/2k_BT)$ is the stationary distribution. Of course, this shouldn't surprise us since a bead in a constant potential is not driven by a supply of fuel and hence should obey the Boltzmann distribution in equilibrium, $p(x) \propto \exp(-V(x)/k_BT)$. Note that the specific choice of noise strength in Eq. 4.18 was necessary for this equality to hold; the fluctuation-dissipation relation is a fundamental feature of systems that can reach equilibrium.

Exercise 4.2.3. Manipulating solutions to the Fokker-Planck equation

A. Prove that $p(x) \propto \exp(-\kappa(x-x_0)^2/2k_BT)$ is a solution of Eq. 4.23 by explicit substitution.

B. Identify the normalisation constant.

4.3 Non-equilibrium stationary distributions

4.3.1 Population processes

Whether the number of badgers in a forest, cells in a petri dish or proteins in a cell, we're often interested in how the population of one or more species varies in a certain environment. If only one species is present, our state space $x \in \mathscr{A}_X$ is the set of non-negative integers, and we can consider a continuous time Markov process within this state space. Typically, we assume that the population changes by ± 1 in a single event, meaning that our process is fully described by an population-dependent increase rate $\lambda(x)$, and a population-dependent decrease rate $\mu(x)$, as illustrated in Fig. 4.8 (a).

Although population process typically do not describe equilibrium systems, if only one species X is considered, and only jumps of ± 1 are allowed, then the process must obey detailed balance. Qualitatively, this is because no loops are possible around which trajectories could tend to flow (Fig. 4.8 (a)). More mathematically, we can see that transitions between $X = 0$ and $X = 1$ must satisfy detailed balance, because an imbalanced flow into $X = 0$ from $X = 1$ cannot be compensated by an imbalance in another transition out of $X = 0$. In turn, this means that transitions between $X = 1$ and $X = 2$ must obey detailed balance, since an imbalance cannot be compensated for elsewhere. This argument applies inductively to all pairs of states $X = x$ and $X = x+1$. With multiple possible species, this result no longer holds, since cycles are now possible.

Tip 4.3.1. In Chapter 3, we discussed discrete- and continuous-time Markov Processes in general, when the transition or rate matrices $\mathbf{T}$ and $\mathbf{K}$ might be arbitrarily complex. It is thus necessary to write down a full $N \times N$ matrix, where N is the size of the state space, to specify the system. In other cases, such as population processes, $\mathbf{T}$ or $\mathbf{K}$ can have a much simpler structure. For example, jumps are typically possible to neighbouring states only, and jump rates are usually a simple function of the current population. In this case, it is possible to specify the entire Master equation (Eq. 3.18) of a process with an infinite number of states in only a single line that holds for all values of $X = x$:

$$\frac{\mathrm{d}p_x(t)}{\mathrm{d}t} = -(\lambda(x)+\mu(x))p_x(t) + \lambda(x-1)p_{x-1}(t) + \mu(x+1)p_{x+1}(t). \tag{4.24}$$

Do not be fooled by this simplicity – we still have an infinite series of coupled equations, since $\mathrm{d}p_x/\mathrm{d}t$ depends on p_{x-1} and p_{x+1}. But we only have to write down a single equation to specify this entire set of coupled ODEs, which can help analysis (and Gillespie simulation - it makes identifying candidate transitions trivial). Note that the equation for $\frac{\mathrm{d}p_0(t)}{\mathrm{d}t}$ includes an apparent $p_{-1}(t)$ term. We simply define $p_{-1}(t) = 0$, and the equation still works for $p_0(t)$.

Birth-death processes

A common model, particularly for colonies of organisms, is the simple birth-death process (Fig. 4.8 (b)). Here, each entity can either reproduce or die with fixed rates λ_0 and μ_0, giving $K_{x+1,x} = \lambda(x) = \lambda_0 x$ and $K_{x-1,x} = \mu(x) = \mu_0 x$. Thus we can write the Master equation as

$$\frac{\mathrm{d}p_x(t)}{\mathrm{d}t} = -(\lambda_0+\mu_0)xp_x(t) + \lambda_0(x-1)p_{x-1}(t) + \mu_0(x+1)p_{x+1}(t). \tag{4.25}$$

An advantage of expressing the Master equation in this way is that we can find the expected evolution of X by multiplying by x and summing over all possible values

$$\frac{\mathrm{d}\langle X\rangle}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{x=0}^{\infty} xp_x(t) = -(\lambda_0+\mu_0)\sum_{x=0}^{\infty} x^2p_x(t) + \lambda_0\sum_{x=0}^{\infty} x(x-1)p_{x-1}(t) + \mu_0\sum_{x=0}^{\infty} x(x+1)p_{x+1}(t). \tag{4.26}$$

which can be re-written

$$\begin{aligned}\frac{\mathrm{d}\langle X(t)\rangle}{\mathrm{d}t} &= -(\lambda_0+\mu_0)\sum_{x=0}^{\infty} x^2p_x(t) + \lambda_0\sum_{x=0}^{\infty} x(x+1)p_x(t) + \mu_0\sum_{x=0}^{\infty} x(x-1)p_x(t)\\ &= -(\lambda_0+\mu_0)\langle X(t)^2\rangle + \lambda_0\langle X(t)^2\rangle + \lambda_0\langle X(t)\rangle + \mu_0\langle X(t)^2\rangle - \mu_0\langle X(t)\rangle\\ &= (\lambda_0-\mu_0)\langle X(t)\rangle. \end{aligned} \tag{4.27}$$

Tip 4.3.2. Note that in order to replace the sums with averages, we had to convert all sums so that they run from $x = 0$ to ∞, but with $p_x(t)$ in the argument. To do this, we change the variable of summation. For example, using $y = x-1$,

$$\sum_{x=0}^{\infty} x(x-1)p_{x-1}(t) = \sum_{y=-1}^{\infty}(y+1)yp_y(t) = \sum_{y=0}^{\infty}(y+1)yp_y(t). \tag{4.28}$$

Here, the final step follows since the $y = -1$ term of the sum is zero. Since y is just a counting (dummy) variable that is being summed over, we are free to rename it x (or anything else for that matter), giving us $\sum_{x=0}^{\infty} x(x+1)p_x(t)$. Thankfully, this trick (combined with the definition $p_{-1}(t) = 0$) works pretty widely, allowing us to reduce the sums to averages in many cases.

Thus the average population obeys the Malthus equation seen on P10 of Dr Stan's notes. The solution of Eq. 4.27 is simple; exponential decay for $\lambda_0 < \mu_0$, exponential growth for $\lambda_0 > \mu_0$, and no change for $\lambda_0 = \mu_0$. The full stochastic system behaviour is very different. We immediately see that $x = 0$ is reachable from all states in the system, and is an absorbing state. Thus the only possible stationary distribution is $\pi_0 = 1$, $\pi_x = 0$ for $x \neq 0$. In fact,

- For $\lambda_0 < \mu_0$, the system performs a stochastic version of the deterministic decay to $X = 0$, and reaches the steady state.

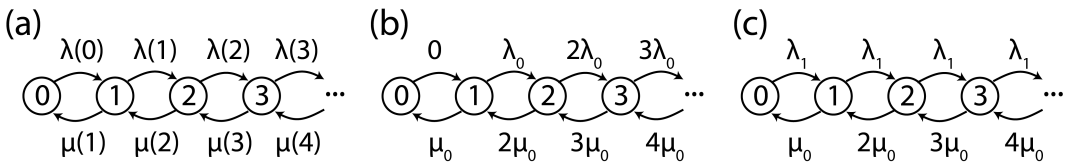


Figure 4.8: Population processes represented as graphs. (a) A generic ad population process with single events only. (b) A birth-death process in which both the rate of increase and decrease of the population are proportional to the current state. (c) An immigration-death model with a constant rate of population increase, but a rate of decrease proportional to the current population.

- For $\lambda_0 = \mu_0$, unlike the deterministic case, the system is not fixed and will perform a "random walk" in X before eventually being absorbed by the $X = 0$ state. Absorption is guaranteed, although the average absorption time is infinite! This reflects the fact that trajectories show wide variance; most trajectories are absorbed relatively quickly, but there is an extremely broad "tail" of long-lived trajectories in the distribution of absorption times.

- For $\lambda_0 > \mu_0$, the deterministic case predicts exponential growth to infinity. For the stochastic system, however, it is possible to reach the absorbing state $X = 0$ through a random fluctuation. However, since births are more likely than deaths, there is no guarantee of absorption in this case. Trajectories that are not absorbed tend towards infinity.

For more details please refer to Ref. 7. The pure birth-death process is therefore somewhat pathological in terms of reaching a non-trivial steady state. One extension would be to make $\mu(X)$ non-linear, to reflect eg. competition effects. This modification will stop trajectories diverging to infinity, but does not prevent guaranteed eventual absorption at $X = 0$.

Immigration-death processes

The alternative immigration-death model assumes $\lambda(x) = \lambda_1$, implying a constant $K_{x+1,x} = \lambda(x) = \lambda_1$ (Fig. 4.8 (c)). Such a growth term might describe a population of red blood cells, for example, since the current generation of red blood cells do not directly produce the next.. We retain $K_{x-1,x} = \mu(x) = \mu_0 x$. The immigration term can be added to a birth-death process to create an immigration-birth-death process, which allows for re-population from $x = 0$, or the immigration-death process can be treated on its own.

Exercise 4.3.1. Describing a population process using the master equation

A. Write down the master equation for the immigration-death process in the form of Eq. 4.24.

B. Find the time-evolution of $\langle X(t) \rangle$.

Unlike the birth-death process, the immigration-death has a non-trivial steady state. Applying detailed balance to the transition between $X = x$ and $X = x+1$, we immediately see that

$$\mu_0(x+1)\pi_{x+1} = \lambda_1 \pi_x \implies \frac{\pi_{x+1}}{\pi_x} = \frac{\lambda_1}{\mu_0(x+1)}, \tag{4.29}$$

Thus

$$\pi_x = \pi_0 \left(\frac{\lambda_1}{\mu_0}\right)^x \frac{1}{x!}. \tag{4.30}$$

This is the Poisson distribution (Sec. 2.2.1), one of the common discrete probability distributions. The normalized version is

$$\pi_x = \left(\frac{\lambda_1}{\mu_0}\right)^x \frac{\exp(-\lambda_1/\mu_0)}{x!}. \tag{4.31}$$

Some typical Poisson distributions are shown in Sec. 2.2.1 – it is particularly worth noting that $\langle X\rangle = \lambda_1/\mu_0$ and $\mathrm{Var}(X) = \lambda_1/\mu_0$. When $\langle X\rangle = \lambda_1/\mu_0$ is large, the standard deviation σ_X is fairly small compared to the mean. An immigration-death process therefore fits our intuition that stochastic fluctuations become relatively less important in large systems.

4.3.2 Biasing stationary distributions through driving

In Sec. 4.2.1, we considered equilibrium molecular systems that obey detailed balance. One way to drive biochemical systems out of equilibrium is to couple the transitions to the consumption of a fuel molecule, such as ATP. If the actual ATP molecules aren't explicitly accounted for in the description of the system, the net result is an apparent violation of detailed balance.

For example, consider the system illustrated in Fig. 4.9 (a), (b). We have a single protein that can be phosphorylated (an inorganic phosphate molecule can bind to one of its amino acid residues). This is a common motif within cells for altering protein function. In the phosphorylated state we label it X^*, and in the unphosphorylated state X. If the only way to become phosphorylated is to bind a phosphate molecule from solution (assumed to have a constant concentration $[P]$), the two-state master equation is given by the graph in Fig. 4.9 (c) (in which we label X as state 1 and X^* as state 2). The transitions are $X \to X^*$ with transition rate $K_{21} = k_{\mathrm{on}}[P]$, and $X^* \to X$ with transition rate $K_{12} = k_{\mathrm{off}}$. The system obeys detailed balance, and will reach equilibrium in the stationary distribution, in which

$$\frac{\pi_2}{\pi_1} = \frac{k_{\mathrm{on}}[P]}{k_{\mathrm{off}}} = \exp(-(F_2 - F_1)/k_B T), \tag{4.32}$$

where F_2 and F_1 are free energies of the phosphorylated and unphosphorylated states.

We can now ask the question, is it possible to alter $\frac{\pi_2}{\pi_1}$ by allowing the protein to participate in further interactions? As we discussed in Sec. 4.1, adding additional states and transitions to the system will have no effect on π_2/π_1 if detailed balance is not broken, since with detailed balance, $\frac{\pi_2}{\pi_1} = \frac{K_{21}}{K_{12}} = \frac{k_{\mathrm{on}}[P]}{k_{\mathrm{off}}}$ necessarily holds. We could, for example, allow X and X^* to bind to a protein Y (Fig. Fig. 4.9 (d)), introducing the states $X-Y$ (labelled 3) and X^*-Y (labelled 4). Regardless of the details of this interaction, and whether $X-Y$ or X^*-Y is more stable, $\frac{\pi_2}{\pi_1}$ is unchanged if detailed balance still holds.

Fuel molecules allow us to break detailed balance from the perspective of X/X^*. Consider the possibility that Y acts as a catalyst in the following reaction, illustrated in Fig. 4.9 (e):

$$Y + X + \mathrm{ATP} \rightleftharpoons Y - X - \mathrm{ADP} - \mathrm{P} \rightleftharpoons Y + X^* + \mathrm{ADP} \tag{4.33}$$

In this reaction X gains a phosphate from ATP, rather than from solution, and Y acts as a catalyst. Catalysts for this type of reaction are widespread in biology, and are called *kinases*. From the perspective of X, both this reaction and the original $X + \mathrm{P} \rightleftharpoons X^*$ involve $X \rightleftharpoons X^*$, but the other inputs and products are different. As a consequence, the overall free energy change for the reaction in Eq. 4.33 need not be the same as $F_2 - F_1$, and the rates need not respect the same detailed balance

ratio. Thus if we do not explicitly keep track of the ATP and ADP molecules, the master equation for X and X^* looks like one that violates detailed balance.

Indeed, a simple model is illustrated in Fig. 4.9 (f), in which the the overall rate for $Y + X +$ ATP$\rightarrow Y - X -$ ATP is proportional to the concentration of Y and ATP, and the overall rate for $Y + X^* +$ ADP$\rightarrow Y - X -$ ATP is proportional to the concentration of Y and ADP. Labelling $Y - X -$ ADP $-$ P with index 3, the rate matrix is

$$\mathbf{K} = \begin{pmatrix} -k_{\rm on}[P] - k_3[ATP][E] & k_{\rm off} & k_4 \\ k_{\rm on}[P] & -k_{\rm off} - k_3[ADP][E] & k_4 \\ k_3[ATP][E] & k_3[ADP][E] & -2k_4 \end{pmatrix}. \tag{4.34}$$

It is straight forward to show (Ex. 4.3.2) that

$$\frac{\pi_2}{\pi_1} = \frac{2k_{\rm on}[P] + k_3[ATP][E]}{2k_{\rm off} + k_3[ADP][E]}. \tag{4.35}$$

Clearly, it is possible to vary π_2/π_1 in this system. In particular, at low $[E]$, the intrinsic exchange of phosphate with solution is kinetically dominant, and $\pi_2/\pi_1 \approx k_{\rm on}[P]/k_{\rm off}$. However, when the concentration of $[E]$ is high, $\pi_2/\pi_1 \approx [ATP]/[ADP]$ which can in principal take any value. Thus by providing the system with a large excess of ATP, it is possible to kinetically pump the protein X into a non-equilibrium steady state with a high yield of X^*. This system is the simplest example of a generic concept known as *kinetic proofreading*. The idea originally arose as a way to explain how replication/transcription/translation could occur with a much lower error than would be expected from the equilibrium specificity of the component molecules (see Sec. 5.1.2, and Ref. 8 for a longer discussion).

Since the system in Fig. 4.9 (f) does not obey detailed balance for $[E] > 0$, there is a net flux ϕ around the loop (clockwise if $[ATP]/[ADP] > k_{\rm on}[P]/k_{\rm off}$). Each clockwise cycle leaves X unchanged, but converts a single ATP into P and ADP. Thus the system constantly uses up the supply of chemical fuel, just like the electric motor in Fig. 4.1.1 uses the potential energy stored in the capacitor. In cellular systems, we generally assume that the concentrations of fuel molecules are maintained (constantly topped up) by mitochondria etc, allowing the X/X^* system to reach a non-equilibrium, driven stationary distribution at constant levels of $[P]$, $[ATP]$ and $[ADP]$, rather than relaxing to equilibrium.

Exercise 4.3.2. Finding the steady state of the system in Eq. 4.34.

A. Write down the simultaneous equations for π_1, π_2 and π_3 implied by Eq. 4.34.

B. If we are only interested in π_2/π_1, it is sufficient to take the equations given by the first two lines of the rate matrix in Eq. 4.34. Use these two equations to eliminate π_3, and hence identify π_2/π_1.

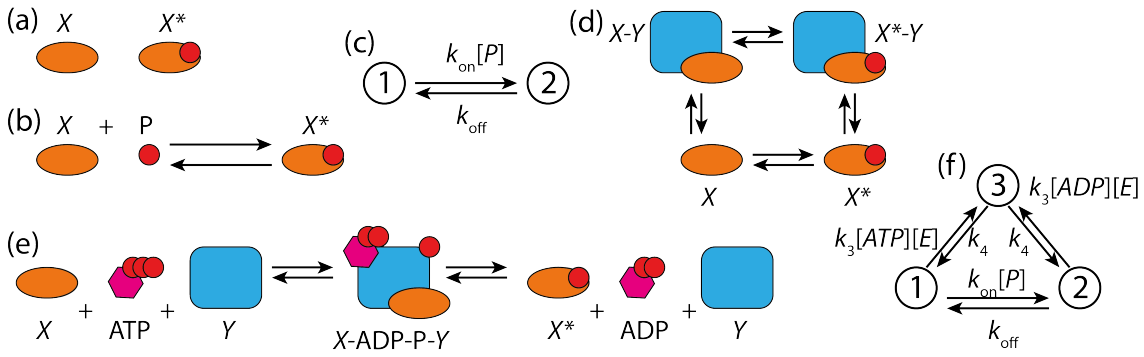


Figure 4.9: Biasing stationary states through non-equilibrium driving. (a) Consider a protein that can be phospohorylated (X^*) or dephosphorylated (X). (b) Transitions between the two states can happen through exchange of phosphate with solution, resulting in the simple Markov process shown in (c). (d) The introduction of extra states and transitions, involving binding to a second protein Y, cannot change the ratio of free X^*/X if the system remains in detailed balance. (e) Detailed balance can be broken by introducing a second catalytic pathway by which $X \rightarrow X^*$, in this case involving the breakdown of ATP. (f) A simple Markov process including both pathways by which X can be converted to X^*. The fuel molecules effectively act as external driving for the X/X^* system, breaking detailed balance (note that the use of k_3 to describe both enzyme-binding processes is a simplification for illustrative purposes).

Chapter 5

Stochastic processes in biology: Stochastic dynamics

5.1 When is dynamics important?

In the previous chapter, we analysed the stationary distributions of various biologically-inspired stochastic systems. We didn't, however, consider much dynamical behaviour, even though we're often interested in systems that are not in a stationary distribution, and hence have a probability distribution that evolves with time. The stationary distribution itself also doesn't tell us much about individual sample trajectories, even when the process is stationary. For example, although we know the equilibrium probability of binding for the DNA model in Sec. 4.2.1, we have not calculated the typical time taken to flip between bound and unbound states. This chapter will deal with biologically-inspired questions of these kinds.

5.1.1 Rare events and transition states

As I have hinted at during this course, we rarely describe processes at the most microscopic (atomistic) level in biology; we typically consider macrostates that group many microscopic configurations together. If we describe the dynamics using a discrete-state Markov process at the level of these macrostates, $X(t)$, we are explicitly assuming that transitions between possible states $x \in \mathscr{A}_X$ are *rare events*. A transition between two macrostates A and B can be described as a rare event if the typical time spent in the grey area between A and B during a transition is short compared to the typical time spent well within A or B between transitions. If this is not the case, it is hard to justify modelling $A \to B$ as an instantaneous process with no intermediates, which is exactly what is done in a discrete state Markov process. For example, it is usually reasonable to describe the binding and unbinding of molecules, such as the DNA strands discussed in Sec. 4.2.1, as rare events. For most of the time the DNA is clearly double- or single-stranded; the intermediate partially-bound states are visited only briefly.

If we have access to a (slightly) more microscopic description than simply $x \in \mathscr{A}_X$, then we can start to make predictions about overall effective transition rates between different X states. For simplicity assume that the coarse-grained states are labelled by $\mathscr{A}_X = \{1,2\}$, and let $Z(t)$ be a stochastic process describing the system at a more microscopic level of detail (see Fig. 5.1). One approach to estimate K_{12}^X and K_{21}^X, the overall effective rates, is known as *transition state theory*. The idea is to identify the microscopic configuration(s) $Z = z_{tr}$ that sit on the dividing line between

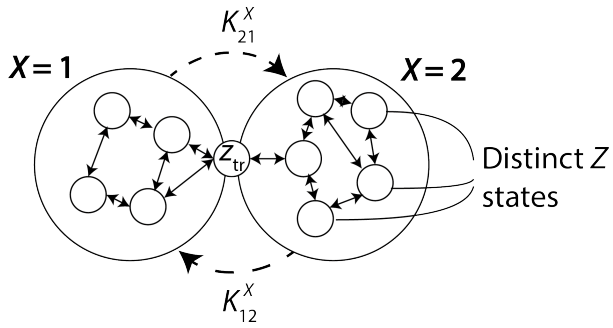


Figure 5.1: A microscopic process $Z(t)$ is grouped into two "macrostates", $X = 1$ and $X = 2$. The overall rates of transition between $X = 1$ and $X = 2$ (K^X_{12} and K^X_{21}) can be estimated by considering the probability that the transition state $Z = z_{tr}$ is occupied.

$X = 1$ and $X = 2$; systems in these *transition states* could go one way or the other. An illustration of this idea is given in Fig. 5.1. In the simplest approach, we assume that

$$K^X_{21} = k_0 \Pi(Z = z_{tr}|X = 1). \tag{5.1}$$

Here, k_0 is a constant with units of rate and $\Pi(Z = z_t|X = 1)$ is the probability of being in the transition state z_t when in a stationary distribution confined to $X = 1$. The basic idea is that the system moves around within macrostate $X = 1$, sampling it fully and reaching a steady state over the microstates of Z within $X = 1$. Very occasionally, during this process, the system reaches the transition state $Z = z_t$, and when it does a transition can occur. Thus the overall rate of transition is proportional to the probability of being in the transition state in the stationary distribution $\Pi(Z = z_t|X = 1)$.

For a single process, Eq. 5.1 isn't fantastically useful. We don't know k_0 *a priori*, and identifying $\Pi(Z = z_t|X = 1)$ is hard. Transition state theory really comes into its own is when we compare similar processes. For example, we might be interested in the relative rates at which a protein unfolds under two different forces, or the rate of dissociation of two DNA duplexes of different lengths. In cases such as these, we can assume k_0 doesn't change much, and look to estimate changes in $\Pi(Z = z_t|X = 1)$ due to changes in conditions. We now explore an example.

Dissociation rates of DNA

We will consider the simple cooperative DNA binding model introduced in Sec. 5.1. In principle, we can calculate the overall rate of escape from a bound macrostate $X = 2$ with non-zero base pairs to an unbound macrostate $X = 1$ with zero base pairs, as a function of duplex length N, using the fundamental matrix method. However, we can get a very quick estimate of the relative rates using transition rate theory. We take the $m_{tr} = 1$ base-pair state from the model in Sec. 4.2.1 as the transition state. The probability of being in this state, given that base pairs are present, is determined by the free energies:

$$\Pi(M = m_{tr}|X = 2) = \frac{\exp(-F_1/k_BT)}{\sum_{m\geq 1}^{N}\exp(-F_m/k_BT)} = \frac{\sigma}{\sum_{m\geq 1}^{N}\sigma t^{m-1}} = \frac{1}{\sum_{m\geq 1}^{N} t^{m-1}}, \tag{5.2}$$

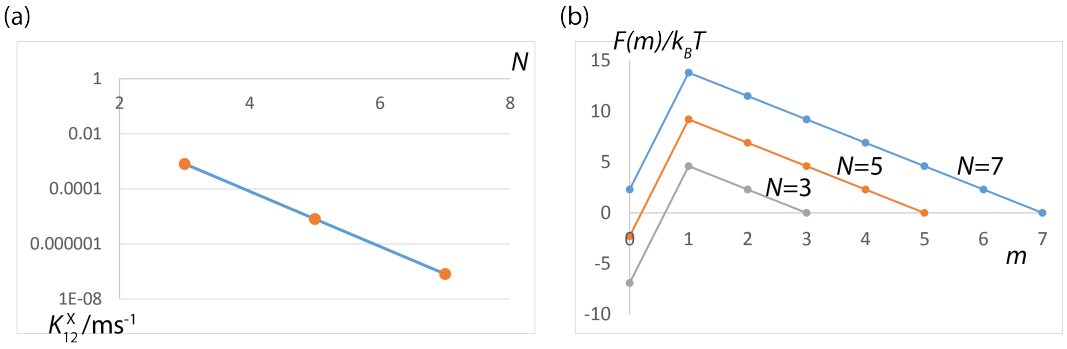


Figure 5.2: (a) Comparison of the predictions of simple transition state theory (solid line) and exact calculations (points) for the transition rate from a bound macrostate to an unbound macrostate, as a function of DNA length N, for a simple DNA model (Sec. 4.2.1). Plot obtained with parameters $k_{\text{zip}} = 1000\,\text{s}^{-1}$, $k_{\text{off}} = k_{\text{unzip}} = 100\,\text{s}^{-1}$ and $k_{\text{on}} = 10^{-3}\,\text{s}^{-1}$; note that k_0 is a free fitting parameter for the transition state theory. (b) Free energy as a function of the number of base pairs m for duplexes of different lengths N in the simple model; there is a clear barrier at $m = 1$ in all cases, justifying the use of transition state theory.

where σ and t are defined in Eq. 4.9 and 4.10. Thus we predict that for two different lengths of DNA N and N',

$$\frac{K_{12}(N)}{K_{12}(N')} = \frac{\sum_{m\geq 1}^{N'} t^{m-1}}{\sum_{m\geq 1}^{N} t^{m-1}} \approx t^{(N'-N)} = \left(\frac{k_{\text{zip}}}{k_{\text{unzip}}}\right)^{(N'-N)}. \tag{5.3}$$

Longer duplexes therefore fall apart exponentially more slowly (if $k_{\text{unzip}} > k_{\text{zip}}$, the process wouldn't be well described by a rare event). This simple prediction is compared to the full calculation of the cooperative DNA binding model in Fig. 5.2 (a), and we see that it does pretty well (of course, the full model here is pretty simple itself!).

Tip 5.1.1. Often, as above, the change in $\Pi(Z = z_{tr}|X = 1)$ between systems can be estimated through the change in free energy of the transition state relative to macrostate $X = 1$. We often think of the transition state as being at the top of a free-energy "barrier" between $X = 1$ and $X = 2$ (see Fig. 5.2 (b) for the DNA example). The time scale of a rare event is the time scale to climb this barrier, which grows exponentially with barrier height.

Exercise 5.1.1. More basic transition state theory.

A. How does the rate of transition from $X = 1$ to $X = 2$ (unbound to bound) depend on DNA length N in this transition state theory description?

Extinction in population processes

As mentioned in Sec. 4.3.1, populations models often have the state with $N = 0$ individuals as an absorbing state. More sophisticated models than simple birth-death (perhaps those with a carrying capacity as introduced on P16 of Dr Stan's notes) can also have a locally stable state with a finite average population $N \sim \bar{N}$. Given enough time, all populations obeying such a model will go extinct; however this process is often extremely slow and the rate at which it occurs can be estimated

using transition state theory, treating the state with $N = 1$ as the transition state and calculating its probability relative to $N \sim \bar{N}$.

5.1.2 Competing outcomes of molecular reactions

When understanding molecular processes, we're often interested in questions such as:

- Given that reactants are bound, what is the probability that the reaction proceeds to completion rather than returning back to the initial reactant state?
- Given multiple possible substrates, what is the probability that an intended substrate is involved in a reaction rather than an unintended one?

Typically, when thinking about these questions, we construct finite-state, continuous-time Markov processes in which the states represent reaction intermediates. Consider, for example, the scheme in Fig. 5.3 (a), which is a simple model for the process by which mRNA is copied into a protein. In this simple picture, a substrate tRNA molecule s can bind to or unbind from the template mRNA, with rates $k_{\text{on}}[s]$ and k^s_{off}. Once bound, the amino acid carried by the tRNA can be added to the growing polypeptide at a rate k_{cat} (so called because this is a catalytic process).

Also present are alternative tRNA with alternative amino acids that could be incorporated into the growing structure. For simplicity, let us just consider a single competitor i with concentration $[i] = [s]$. Further, let us assume that correct and incorrect molecules are distinguished only by their off rates $k^s_{\text{off}} < k^i_{\text{off}}$ (k_{cat} and k_{on} are the same for both). We can then ask, what is the probability of incorporating the correct rather than the incorrect amino acid into the growing polypeptide?

Consider the process from the perspective of an amino acid currently at the end of the polypeptide chain. It can either have no adjacent amino acid (state labelled E), a correct/incorrect candidate amino acid in the adjacent location but not covalently bound to the chain (s/i), or a correct/incorrect amino acid can be fully bound as part of the chain (S/I). The rate matrix K_{jk} for transitions between these states is depicted graphically in Fig. 5.3 (b). As discussed in Sec. 3.4.2, if we are interested in finding the relative probability of reaching absorbing state S rather than absorbing state I, starting at E, we can do so via the fundamental matrix of a discrete-time process. We can create a discrete time Markov chain from the continuous process by asking for the probability that a transition out of state j will reach state k. The transition matrix T_{jk} for this effective process is illustrated graphically in Fig. 5.3 (c).

To proceed, we split the state space into non-absorbing (E, s, i) and absorbing (S, I) states, and construct the matrices $\mathbf{U}$ and $\mathbf{R}$ as discussed in Sec. 3.4.2. in our case,

$$\mathbf{U} = \begin{pmatrix} 0 & \frac{k^s_{\text{off}}}{k_{\text{cat}}+k^s_{\text{off}}} & \frac{k^i_{\text{off}}}{k_{\text{cat}}+k^i_{\text{off}}} \\ 1/2 & 0 & 0 \\ 1/2 & 0 & 0 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 0 & \frac{k_{\text{cat}}}{k_{\text{cat}}+k^s_{\text{off}}} & 0 \\ 0 & 0 & \frac{k_{\text{cat}}}{k_{\text{cat}}+k^i_{\text{off}}} \end{pmatrix}. \tag{5.4}$$

We now calculate the fundamental matrix $\mathbf{W}$:

$$\mathbf{W} = (\mathbb{I} - \mathbf{U})^{-1} = \frac{1}{1 - \frac{k^s_{\text{off}}}{2(k^s_{\text{off}}+k_{\text{cat}})} - \frac{k^i_{\text{off}}}{2(k^i_{\text{off}}+k_{\text{cat}})}} \begin{pmatrix} 1 & \frac{k^s_{\text{off}}}{k_{\text{cat}}+k^s_{\text{off}}} & \frac{k^i_{\text{off}}}{k_{\text{cat}}+k^i_{\text{off}}} \\ \frac{1}{2} & 1 - \frac{k^i_{\text{off}}}{2(k_{\text{cat}}+k^i_{\text{off}})} & \frac{k^i_{\text{off}}}{2(k_{\text{cat}}+k^i_{\text{off}})} \\ \frac{1}{2} & \frac{k^s_{\text{off}}}{2(k_{\text{cat}}+k^s_{\text{off}})} & 1 - \frac{k^s_{\text{off}}}{2(k_{\text{cat}}+k^s_{\text{off}})} \end{pmatrix} \tag{5.5}$$

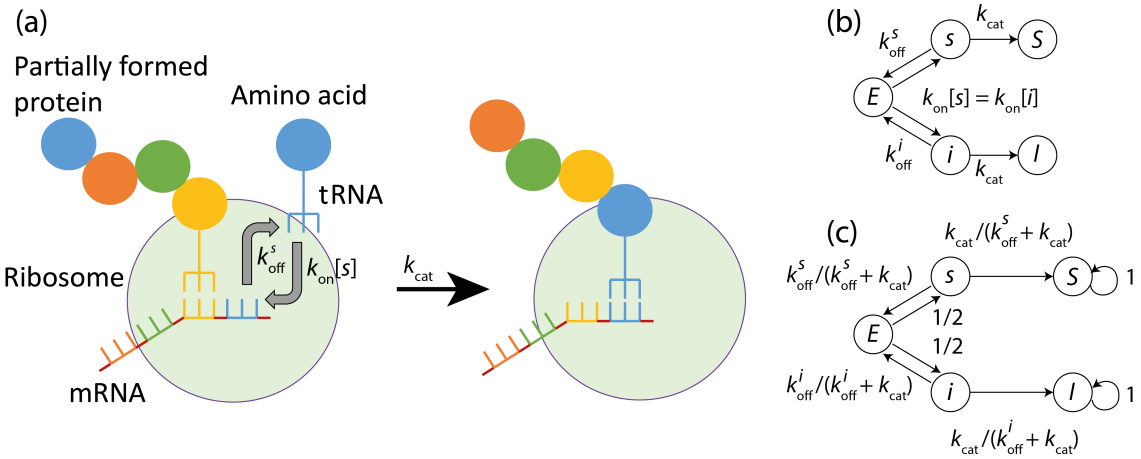


Figure 5.3: (a) A basic model for the addition of amino acids to a polypeptide chain by the ribosome, which is using an mRNA molecule as a template. Monomers can first bind to and unbind from the mRNA site adjacent to the end of the polypeptide (protein) chain; whilst bound in this position, they can be covalently added to the chain by the ribosome at a rate k_{cat}. (b) Graphical representation of the rate matrix for the addition of the next amino acid to a growing chain, as described by this model. Included here is the possibility of an incorrect amino acid i being incorporated. State E corresponds to nothing being bound to the adjacent site to the end of the polypeptide chain; state s corresponds to the correct monomer and i to the incorrect monomer being present, but not covalently bound. States S and I correspond to full covalent binding of the correct/incorrect amino acid to the polypeptide chain. (c) The transition matrix for a discrete time Markov process constructed from the continuous process in (b), using the methodology discussed in Sec. 3.4.2.

and

$$\mathbf{RW} = \frac{1}{1-\frac{k^s_{\text{off}}}{2(k^s_{\text{off}}+k_{\text{cat}})}-\frac{k^i_{\text{off}}}{2(k^i_{\text{off}}+k_{\text{cat}})}}\begin{pmatrix} \frac{k_{\text{cat}}}{2(k_{\text{cat}}+k^s_{\text{off}})} & \frac{k_{\text{cat}}}{k_{\text{cat}}+k^s_{\text{off}}}\left(1-\frac{k^i_{\text{off}}}{2(k_{\text{cat}}+k^i_{\text{off}})}\right) & \frac{k_{\text{cat}}}{k_{\text{cat}}+k^s_{\text{off}}}\frac{k^i_{\text{off}}}{2(k_{\text{cat}}+k^i_{\text{off}})} \\ \frac{k_{\text{cat}}}{2(k_{\text{cat}}+k^i_{\text{off}})} & \frac{k_{\text{cat}}}{k_{\text{cat}}+k^i_{\text{off}}}\frac{k^s_{\text{off}}}{2(k_{\text{cat}}+k^s_{\text{off}})} & \frac{k_{\text{cat}}}{k_{\text{cat}}+k^i_{\text{off}}}\left(1-\frac{k^s_{\text{off}}}{2(k_{\text{cat}}+k^s_{\text{off}})}\right) \end{pmatrix} \tag{5.6}$$

Recalling Sec. 3.4.2, we identify the probability of reaching absorbing state S from state E as $P(S \text{ before } I|E) = (RW)_{11}$ and the probability of reaching I instead as $P(I \text{ before } S|E) = (RW)_{21}$. Thus the ratio of correct to incorrect incorporations is then

$$\frac{P(S \text{ before } I|E)}{P(I \text{ before } S|E)} = \frac{k_{\text{cat}}+k^i_{\text{off}}}{k_{\text{cat}}+k^s_{\text{off}}}. \tag{5.7}$$

We see that in the limit of small k_{cat}, the ratio of specificity is given by the ratio of unbinding rates. For large k_{cat}, any amino acid that binds will get added to the growing protein and all specificity is lost.

In practice, the observed specificity in processes such as translation is higher than can be justified by such simple models and our knowledge of biochemistry, even if we allow $k_{\text{cat}} \to 0$. To explain this behaviour, Hopfield proposed that "proofreading" was carried out by the inclusion of intermediate configurations s^* and i^* between binding states (s/i) and full incorporation into the polypeptide

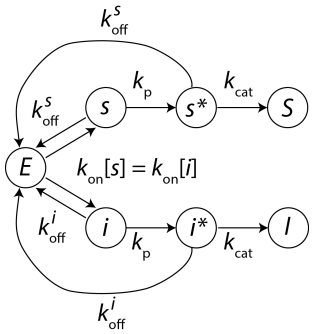


Figure 5.4: A graphical representation of the rate matrix for a process with additional intermediate states s^* and i^* prior to full incorporation into the growing polypeptide.

(S/I). We assume that states s and i are both converted into s^* and i^* at a rate k_p. If fuel consumption is coupled to these transitions, it is reasonable to assume that the reverse process essentially never happens, and thus the overall rate matrix can be described graphically as in Fig. 5.4.

We can once again construct **R**, **U** and **W** matrices to solve for the relative rate at which correct and incorrect amino acids are incorporated into the polypeptide. However, this is even more work than in the previous case. Fortunately, the reaction scheme is so simple that an easier approach can be taken, as outlined in Ex. 5.1.2. The result is

$$\frac{P(S \text{ before } I|E)}{P(I \text{ before } S|E)} = \frac{k_p + k^i_{off}}{k_p + k^s_{off}} \frac{k_{cat} + k^i_{off}}{k_{cat} + k^s_{off}}. \tag{5.8}$$

Effectively, discrimination is performed twice; once from the state s/i, and once from the state s^*/i^*. Proofreading thus leads to an approximate squaring of the ratio $P(S \text{ before } I|E)/P(I \text{ before } S|E)$.

Exercise 5.1.2. Deriving the specificity of proofreading the easy way.

A. Justify that, for both the models in Fig. 5.3 and Fig. 5.4,

$$\frac{P(S \text{ before } I|E)}{P(I \text{ before } S|E)} = \frac{P(s \text{ before } i|E)P(S \text{ before } E|s)}{P(i \text{ before } s|E)P(I \text{ before } E|i)} \tag{5.9}$$

This abuse of notation requires some explanation. $P(S \text{ before } E|s)$ means the probability that a system starting in s will reach S before it reaches E.

B. Using Eq. 5.9, argue that Eq. 5.7 holds for the system in Fig. 5.3.

C. Using Eq. 5.9, argue that Eq. 5.8 holds for the system in Fig. 5.4.

The aim of this whole exercise is really to increase the ratio of correct to incorrect amino acids that are in position to be incorporated into the polypeptide chain. The natural equilibrium ratio is k^i_{off}/k^s_{off}, and by introducing the intermediate state we have improved on this. As discussed in Sec. 4.3.2, driving state ratios out of equilibrium requires the input of fuel – this is why the $s/i \to s^*/i^*$ steps must be coupled to fuel consumption. If they are not, then the proofreading mechanism fails. For more information, see Ref. 8.

5.1.3 Time dependence in overdamped diffusion

In Sec. 4.2.3 we discussed the motion of a small object (such as a protein or colloidal beads) moving in a viscous medium (water), deriving the Fokker-Planck and Langevin equations for overdamped motion in one dimension (Eq. 4.21 and 4.22). We showed that the stationary distribution of a bead in a quadratic potential is a Gaussian, consistent with the expectation from the Boltzmann distribution provided that the fluctuation-dissipation relation connecting noise strength and friction was obeyed. We now return to small objects moving in solution, but this time consider time-dependent properties.

Freely diffusing objects

Firstly, let us consider particles that are not subject to any kind of potential, and are free to diffuse. The original pollen particles observed by Robert Brown would come under this category, or perhaps a protein in a cell (ignoring confinement and crowding effects). In this case, the Fokker-Planck equation (Eq. 4.22) reduces to the simpler diffusion equation

$$\frac{\partial p(x,t)}{\partial t} = \frac{k_B T}{\gamma}\frac{\partial^2}{\partial x^2}p(x,t) = D\frac{\partial^2}{\partial x^2}p(x,t). \tag{5.10}$$

It is possible to solve Eq. 5.10 in the general case, using standard PDE approaches (eg. separation of variables). The general solution is a linear combination of solutions (eigenfunctions) with arbitrary coefficients. These coefficients can be identified for any given situation by matching to boundary conditions, giving the specific solution of relevance.

Since this is not a course on solving PDEs, we won't get bogged down in these details, but will instead jump to a solution of particular interest. It can be verified (Ex. 5.1.3) that

$$p(x,t) = \sqrt{\frac{1}{4\pi D t}}\exp\left(-x^2/4Dt\right) \tag{5.11}$$

is a solution of Eq. 5.10 . Eq. 5.11 is a Gaussian of mean 0 and variance $2Dt$ (see Sec. 2.2.2), and is plotted in Fig. 5.5 for several values of t. As is clear, although the maximum of $p(x,t)$ stays at $x = 0$, the distribution gets broader over time.

This is good to know, but what is the actual meaning of this result? At small t, $p(x,t)$ is tightly peaked meaning that values of $X(t) \approx 0$ are typical. Perhaps the particle is a protein that was produced at $t = 0$ at this point, or perhaps it is a particle whose position we measured at $t = 0$. Over time, however, Brownian motion makes the particle's position increasingly difficult to predict given the knowledge of where it was at $t = 0$. It is this increasing uncertainty that is described by the broadening of the distribution with time, and it reflects the general tendency of diffusing systems to "spread out" over time. This spreading out is actually a relaxation towards equilibrium; with no potential influencing the particle, it should be equally likely to be anywhere. Thus equilibrium predicts a flat distribution in the limit of $t \to \infty$, as does Eq. 5.11.

Exercise 5.1.3. Verifying a solution of the diffusion equation

A. Verify that Eq. 5.11 satisfies the partial differential diffusion equation (Eq. 5.10) by direct substitution.

B. What is the behaviour of the solution as $t = 0$? What sort of function is this? How does this relate to the initial conditions to which this solution corresponds?

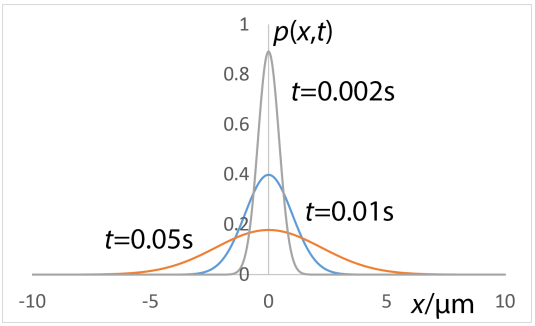


Figure 5.5: Probability density $p(x,t)$ for a system undergoing free diffusion with diffusion constant $D = 10^{-10}\mathrm{m}^2\mathrm{s}^{-1}$, with X initially 0.

Using the properties of Gaussian distributions as discussed in Sec. 2.2.2, it can immediately be seen that

$$\begin{aligned} \langle X(t)\rangle &= 0, \\ \mathrm{Var}(X(t)) = \sigma^2_{X(t)} &= 2Dt. \end{aligned} \tag{5.12}$$

Thus the typical width of the distribution grows with the square root of time: $\sigma_{X(t)} = \sqrt{2Dt}$. Diffusion is therefore a slow way to reach a distant objective in general – the time required scales quadratically with the distance, unlike travelling with a constant velocity. Hopefully Eq. 5.12 is reminiscent of the lattice polymer discussed in Sec. 4.2.2, where the expected end to end distance scales with the square root of the number of monomers. Indeed, the underlying process is very similar – if we made a lattice-based model of diffusion, in which a particle jumped from site to site in an unbiased and memoryless fashion, the distribution of particle trajectories with N hops would be indistinguishable from configurations of a polymer of length N in the lattice polymer model.

Dynamics of a bead in a quadratic trap

In Sec. 4.2.3, we considered the stationary distribution of a bead in a quadratic trapping potential. We will now consider the time-dependent dynamics of individual trajectories, asking how fast the bead "explores" the well. This kind of analysis is simplest using the trajectory-based formalism of the Langevin equation (Eq. 4.21). For a quadratic potential $V(X) = \frac{\kappa}{2}(X)^2$ (let us take the minimum of the potential at $X = 0$ for simplicity), the stochastic process obeys

$$\frac{\mathrm{d}X}{\mathrm{d}t} = -\frac{\kappa X}{\gamma} + \zeta(t), \tag{5.13}$$

where $\zeta(t)$ is a stochastic process with $\langle\zeta(t)\rangle = 0$ and

$$\langle\zeta(t)\zeta(t')\rangle = 2D\delta(t-t'), \tag{5.14}$$

as before. We can formally solve Eq. 5.13 using the integrating factor method (which you should technically remember from school). Eq. 5.13 implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(X\exp\left(\frac{\kappa}{\gamma}t\right)\right) = \zeta(t)\exp\left(\frac{\kappa}{\gamma}t\right). \tag{5.15}$$

Integrating with respect to t gives

$$X(t) = X(0)\exp\left(-\frac{\kappa}{\gamma}t\right) + \exp\left(-\frac{\kappa}{\gamma}t\right)\int_0^t \mathrm{d}t'\,\zeta(t')\exp\left(\frac{\kappa}{\gamma}t'\right). \tag{5.16}$$

This solution, although correct, is not immediately useful, since it involves the integral of a stochastic process. However, since we know the statistical properties of $\zeta(t)$, we can do some averaging. In particular,

$$\langle X(t)\rangle = \langle X(0)\rangle\exp\left(-\frac{\kappa}{\gamma}t\right), \tag{5.17}$$

since $\langle\zeta(t)\rangle$ over all possible noise histories is zero. Thus a particle initially at some displacement will on average fall back toward the centre of the well on a time scale of γ/κ. Individual trajectories, however, will show a great deal of variance. Indeed, we can calculate the autocorrelation function between two points separated by a time t for a system that has reached the stationary equilibrium distribution, $R_X(t) = (-\kappa t/\gamma)$ (evaluated in Ex. 5.1.4). We see that after a time $t \gg \gamma/\kappa$, random noise has destroyed all memory of the state at time 0, since $X(t)$ and $X(0)$ show uncorrelated fluctuations. Combined with Sec. 4.2.3, we now know the stationary (equilibrium) distribution *and* the time scale γ/κ over which a single particle will sample the states in this distribution.

Exercise 5.1.4. Calculating the autocorrelation of a bead in a quadratic well.

A. Calculate $\langle X(t)\rangle$, $\langle X(t)^2\rangle$, $\langle X(0)\rangle$ and $\langle X(0)^2\rangle$ for a particle in the quadratic potential $V(x) = \frac{\kappa}{2}(x)^2$. Assume that the particle is in the stationary distribution at all times (you may find Sec. 4.2.3 and Sec. 2.2.2 helpful).

B. Despite having the same overall distribution at time t and time 0, the particle's position can still be correlated from one moment to the next. Using your answers to part A and Eq. 5.16, calculate $\mathrm{Cov}(X(t),X(0)) = \langle X(t)X(0)\rangle - \langle X(t)\rangle\langle X(0)\rangle$.

C. Evaluate the autocorrelation function $R_X(t) = \mathrm{Cov}(X(t),X(0))/\sqrt{\mathrm{Var}(X(t))\mathrm{Var}(X(0))}$.

5.1.4 Spike trains of nerve impulses: an example of a point process

To a good approximation, the output of a nerve is a series of instantaneous spikes separated by large periods of nothingness (as in Fig. 5.6). The times of these spikes are a realisation of a *point process*, a stochastic process that generates a set of isolated points along some axis - in this case, "time". We can also view the process in terms of the gaps between spikes; if we label the nth time gap as $\Theta(n)$, the whole set of $\Theta(n)$ form a discrete sequence of continuous random variables $\Theta(1),\Theta(2)...$ - a fourth "type" of stochastic process not considered in Chapter 3.

In general, the value of $\Theta(n+1)$ could depend on an arbitrary large set of previous values $\{\Theta(n),\Theta(n-1)...\}$. If so, the series would be non-Markovian. If, however, the value of $\Theta(n+1)$ depends on previous gaps only through $\Theta(n)$, then the series is a Markov process. The simplest possible model would have each $\Theta(n+1)$ independent from all others, including $\Theta(n)$.

A basic example of this simplest case is if spikes arrive independently of the last gap between spikes, and with a constant fixed probability per unit time (or rate) ν. This is known as a *point Poisson process*, and is a very common model in engineering and science. Clearly, real neurons are more sophisticated , but at the very least the point Poisson process gives a baseline against which to compare actual behaviour.

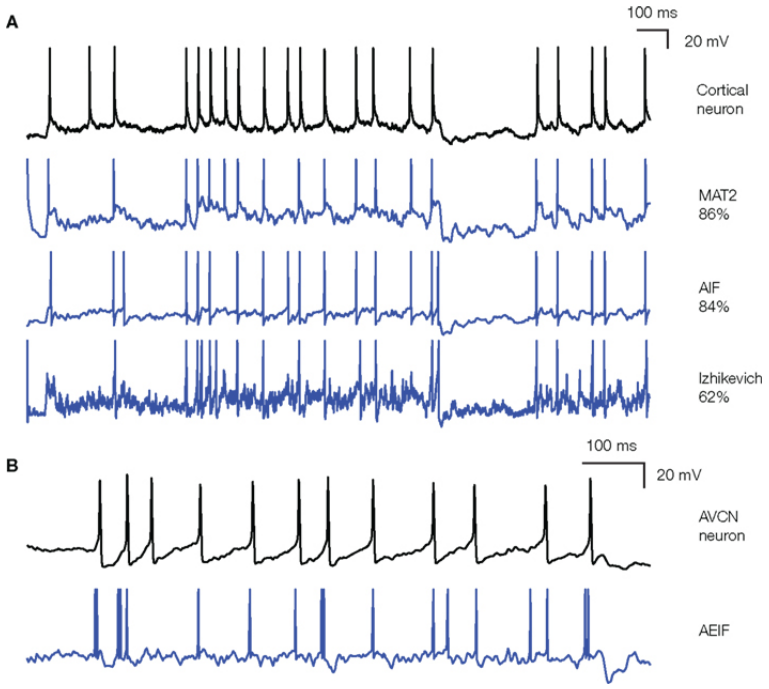


Figure 5.6: Experimental data on the membrane potentials of neurons. Discrete action potential spikes are clearly evident. Taken from Rossant *et al.*, Front. Neurosci. 5:9 (2011).

We can define a probability density of gaps for the point Poisson process, $P(\theta)$ (note: I am using a capital P to avoid confusion in the next section). To calculate this probability density, note that $P(\theta)\delta\theta$, by definition, is the probability that no event occurs within time θ *and* that an event occurs in the subsequent $\delta\theta$, for small $\delta\theta$. To calculate this quantity, let us define a "survival" probability $S(\theta)$, which is the probability that no event occurs for the period θ. In our case, events occur at a constant rate ν. Thus the survival probability follows the equation

$$\frac{\mathrm{d}S(\theta)}{\mathrm{d}\theta} = -\nu S(\theta), \tag{5.18}$$

just like radioactive decay. More formally, consider a two-state Markov process in which the two states are "1: no event has happened" and "2: an event has occurred", with $K_{21} = \nu$ and $K_{12} = 0$. The ODE for $p_1(t)$ of this Markov process is that given by the ODE for S above.

Thus $S(\theta) = \exp(-\nu\theta)$ (which hopefully should surprise no-one). We now need the probability that an event occurs in a small $\delta\theta$, having not occured within θ. But this is simply

$$S(\theta) - S(\theta+\delta\theta) = -\frac{\mathrm{d}S(\theta)}{\mathrm{d}\theta}\delta\theta + \mathscr{O}(\delta\theta^2) \tag{5.19}$$

As we take limit $\delta\theta \to 0$, we obtain

$$P(\theta) = \frac{S(\theta) - S(\theta+\delta\theta)}{\delta\theta} = -\frac{\mathrm{d}S(\theta)}{\mathrm{d}\theta} = \nu\exp(-\nu\theta), \tag{5.20}$$

which is the exponential distribution (Sec. 2.2.2). Again, hopefully this is not a shock to anyone.

Number of events in a fixed window [detailed derivation not expected]

A relevant quantity for many point Poisson processes is the number of events N in a fixed window, Δt. For example, a neuron might trigger a response if and only if the number of pulses in a given time exceeds a threshold. Clearly, N itself is a random variable, since the timing between events is random.

For a point Poisson process with rate ν, the distribution of N can be calculated as a function of ν and Δt (let's call it $p_n(\nu,\Delta t)$). Firstly, it is trivial to calculate the probability of no events - this is just a survival probability:

$$p_0(\nu,\Delta t) = \exp(-\nu\Delta t). \tag{5.21}$$

Now consider the probability of $n+1$ events, $p_{n+1}(\nu,\Delta t)$. By definition, this must be the probability of having a first event after some time $\theta < \Delta t$, $P(\theta)$ from Eq. 5.20, multiplied by the probability of n events in the remaining $\Delta t - \theta$, integrated over all possible first event times θ.

$$p_{n+1}(\nu,\Delta t) = \int_0^{\Delta t} \mathrm{d}\theta P(\theta) p_n(\nu,\Delta t-\theta) = \int_0^{\Delta t} \mathrm{d}\theta \nu \exp(-\nu\theta) p_n(\nu,\Delta t-\theta). \tag{5.22}$$

Now we will guess a solution and check that it works. Consider the Poisson distribution with mean $\nu\Delta t$ (Sec. 2.2.1):

$$p_n^{\mathrm{po}}(\nu\Delta t) = \frac{\exp(-\nu\Delta t)(\nu\Delta t)^n}{n!}. \tag{5.23}$$

It is clear that our guess satisfies $p_0(\nu,\Delta t) = p_0^{\mathrm{po}}(\nu\Delta t)$. We will now argue that if $p_n(\nu,\Delta t) = p_n^{\mathrm{po}}(\nu\Delta t)$, then $p_{n+1}(\nu,\Delta t) = p_{n+1}^{\mathrm{po}}(\nu\Delta t)$ must hold. To see this, we substitute Eq. 5.23 into the RHS of Eq. 5.22, obtaining

$$\begin{aligned} p_{n+1}(\nu,\Delta t) &= \int_0^{\Delta t} \mathrm{d}\theta \nu \exp(-\nu\theta) \frac{\exp(-\nu(\Delta t-\theta))(\nu(\Delta t-\theta))^n}{n!} \\ &= \frac{\nu}{n!}\exp(-\nu(\Delta t)) \int_0^{\Delta t} \mathrm{d}\theta (\nu(\Delta t-\theta))^n \\ &= \frac{(\nu(\Delta t-\theta))^{n+1}}{(n+1)!}\exp(-\nu(\Delta t)) \\ &= p_{n+1}^{\mathrm{po}}(\nu\Delta t). \end{aligned} \tag{5.24}$$

We can now prove that $p_n(\nu,\Delta t) = p_n^{\mathrm{po}}(\nu\Delta t)$ by induction; it is true for $n=0$, and for $n+1$ given if it is true for n, so $p_n(\nu,\Delta t) = p_n^{\mathrm{po}}(\nu\Delta t)$ must hold for all integers $n \geq 0$. To summarise this important conclusion: if independent events occur with a fixed rate ν, then the distribution of number of events N within a time window Δt is the Poisson distribution with mean $\nu\Delta t$.

Chapter 6

Networks

6.1 What is a network?

We're often interested in systems that contain many "entities" interacting with each other in a complex way. These entities might be people in a social network, neurons in the brain or proteins within the cell. The internet is a typical example of a man-made network (Fig. 6.1). We might know everything about these interactions, or we might only know which connections exist, but not how strong they are. We might even only know the distribution of connections in a statistical sense. But in any of these cases, as soon as we start to think about the connections between entities in a complex system, we've entered the world of networks.

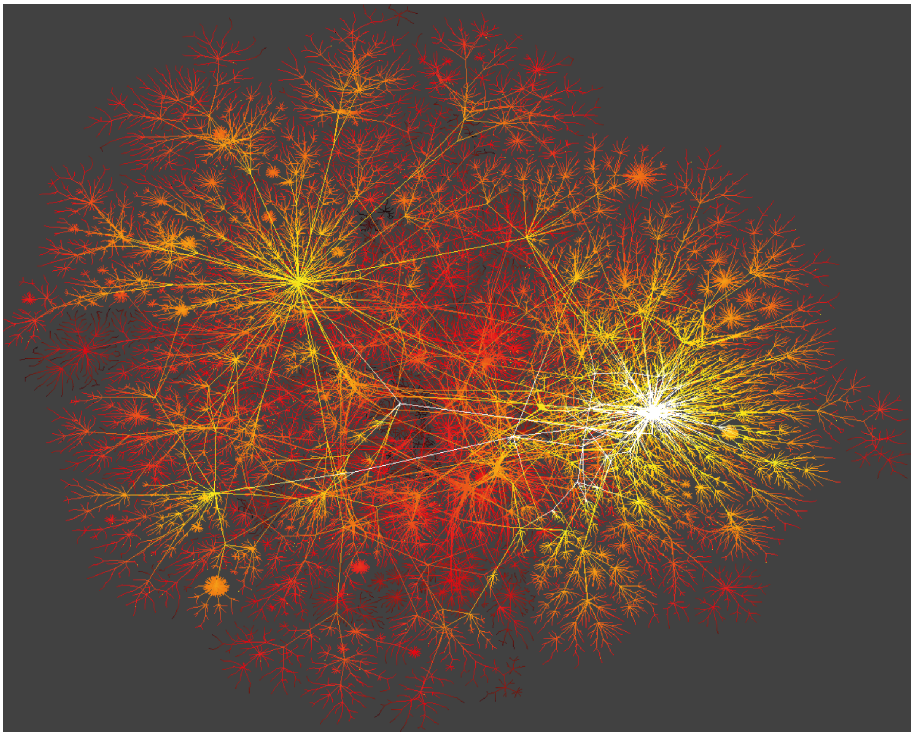


Figure 6.1: Network of internet connectivity. By K. C. Claffy (www.caida.org).

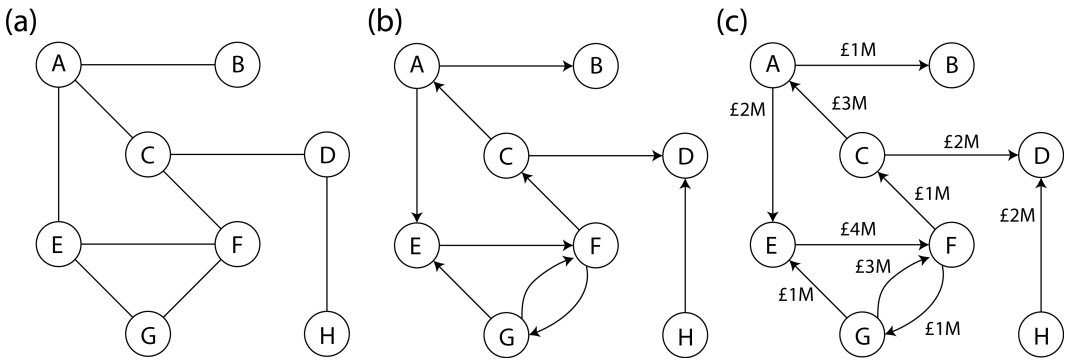


Figure 6.2: Business interactions described by undirected (a), directed (b) and weighted (c) graphs. (a) simply shows connections, (b) indicates the direction of buying/selling (outward edges from a node indicate selling by that node) and (c) indicates the volume of trade by weighting the edges.

6.1.1 Graphs: the mathematical description

A graph is the mathematical description of a network. It consists of a series of *nodes* – these are the "entities" – and *edges* between those nodes that represent the links or connections. These edges can be *directed* (point from one node to another) or not, and sometimes they can have *weights* attached to them.

To illustrate these different possibilities, consider Fig. 6.2, in which I've sketched three different networks involving the same nodes. I've chosen a network of imaginary businesses, some of which trade with each other. Each business is a node. In Fig. 6.2 (a), I've simply connected those businesses that trade with each other. This is an undirected, unweighted graph. In Fig. 6.2 (b), I've used a directed graph to indicate the direction in which trade occurs (an arrow from A to B indicates that A sells goods to B). Note that with a directed graph, it can make sense to draw two edges between a given pair of nodes; one in each direction. In my example, this would imply that both businesses both sell goods to each other. Finally, I could label these directed edges with the total value of goods sold per year, giving me a directed, weighted graph (Fig. 6.2 (c)).

This is pretty much all there is to defining a graph that represents a network. We just need to define a set of nodes, a set of (directed) edges, and possibly a set of weights.

6.1.2 Why are we interested in networks?

Even when we don't know everything about a network, a little information often tells us interesting things

It's unusual to know all the connections and connection strengths in a large, real-life network, such as a human social network or protein interaction network. This means it would be essentially impossible to systematically study the system behaviour, either in a stochastic or deterministic fashion, as all unknown parameters are varied. However, just a few basic facts about the network topology (the way in which nodes are connected) can tell us a lot. For example, if we're interested in the flow of something through a network, we can say a lot just by looking at the statistical properties of the edges between nodes. For example: do nodes have high or low connectivity on average? Do they tend to cluster together in mutually interconnected lumps? Is there always a short path between any two nodes? Different types of networks, which are generated by different underlying

principles/processes, will have very different answers to these questions. Fundamentally different behaviour will then result, regardless of the minute details.

Similarly, if a network has been constructed for a purpose (either by nature or humans), we might ask if its overall structure is optimised for efficiency or robustness (for example). If a certain motifs, like hub nodes with extremely high connectivities, appear more frequently than you might expect by chance, does this mean it is somehow particularly useful? What happens if these hub nodes are somehow crippled or deactivated? Can we learn anything from natural network design when constructing our own?

Networks can show emergent behaviour that is fundamentally different from that of a single node

Most of the discussion in this chapter will focus on what we can learn from typical network topologies about the function and design of a network. But another interesting aspect of thinking in terms of networks is that they allow remarkably simple components to come together and achieve much more sophisticated behaviour. An individual neuron doesn't seem all that exciting, but a whole network of neurons interacting with each other underlies the functionality of the human brain. Recently, the machine learning community has developed (somewhat) similar artificial network architectures that can perform "deep learning". The "deep" here actually refers to the presence of multiple processing layers in an artificial network, and this approach is one of the factors that has driven the great advances in artificial intelligence in recent years. I'm already out of my depth, so I won't say any more.

6.2 Network properties

6.2.1 Basic properties of graphs

Properties of nodes and small sets of nodes

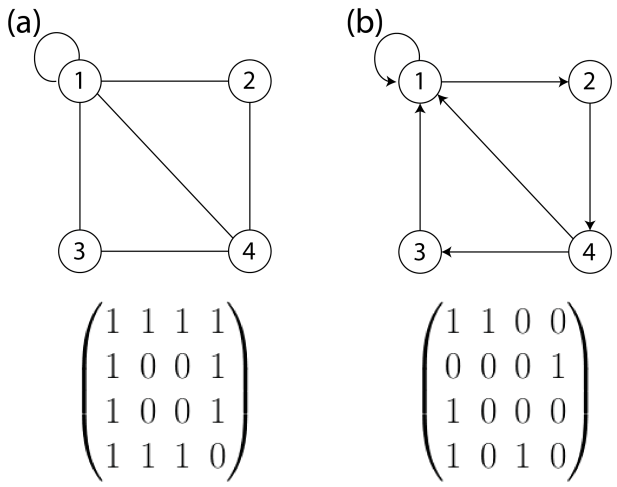
- **Degree**: The degree of a node is the number of edges connected to (adjacent to) a node. If the graph is directional, we can define an *in-degree* and an *out-degree*. Note that undirected self-edges contribute 2 to the degree (the edge connects twice).

- **Adjacency**: Two nodes are adjacent if the are connected by an edge (this edge is also described as being adjacent to both nodes).

- **Shortest path length**: the shortest path L_{ij} is the minimimum path length from node i to j. For an unweighted graph, this length is simply the number of edges that must be travelled. For a weighted graph, the distance is the sum over weights along the links. To find the shortest path between any two nodes, you can apply one of several well-known algorithms (eg. Dijkstra's algorithm[9]).

- **Local clustering coefficient** The local clustering coefficient of a node i is the proportion of the possible connections between the nodes adjacent to i that actually exist.

$$C_i = \frac{2n_i}{k_i(k_i-1)} \tag{6.1}$$



$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Figure 6.3: Simple graphs and their associated adjacency matrices. Note the difference between a directed graph in (b), and an undirected one in (a).

Here, k_i is the number of distinct nodes to which i is connected, and n_i is the number of edges within these k_i nodes (note that this quantity is ill-defined for nodes with 1 or 0 neighbours; we shall simply set the local clustering coefficient to zero in such cases).

Topological properties of the graph as a whole

- **Connectivity**: In a connected graph, there is a path between every pair of vertices (the path can go through multiple intermediate nodes).
- **Average degree**: The average degree $\bar{k}$ is the degree k_i averaged over all nodes in the network.
- **The adjacency matrix**: The adjacency matrix **A** for a graph with N nodes is an $N \times N$ matrix with $A_{ij} = 1$ if there is a connection from i to j in the graph, and zero otherwise (for undirected graphs, $A_{ij} = 1$ if there is an edge between i and j). Examples of adjacency matrices are given in Fig. 6.3.
- **Network diameter**: The network diameter D is the maximum over all pairs i, j in the network of the shortest path length from i to j, L_{ij}. It is therefore the shortest route between the most distant nodes.
- **Global clustering coefficient**: The global clustering coefficient C is the local clustering coefficient averaged over all nodes in the graph.

As well as the averages of quantities such as degree, and clustering coefficient, we are potentially also interested in the distribution of these quantities for large networks. For example, we might consider $p(k)$, the probability that a randomly chosen node has degree k.

Exercise 6.2.1. Topological properties of a simple graph. Consider the graph in Fig. 6.4.

A. Calculate the degree of each node and the average degree.

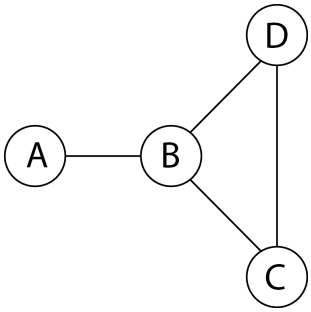


Figure 6.4: A simple network analysed in Ex. 6.2.1.

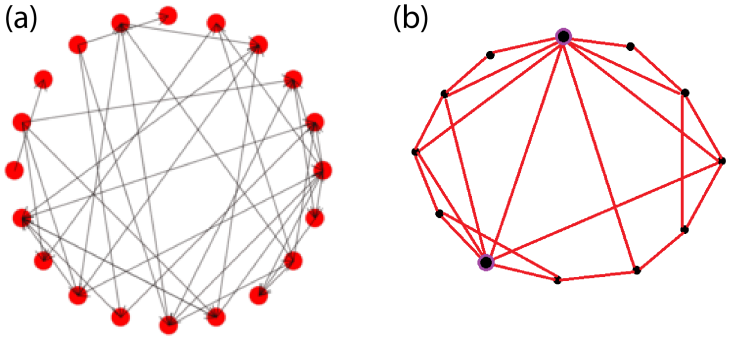


Figure 6.5: Simple examples of (a) an Erdős-Rényi network; and (b) a small world network with notable hubs. Compared to the Erdős-Rényi network, note that the small world network has high clustering and outlier nodes of unusually high connectivity. Taken from https://fr.wikipedia.org/wiki/Theorie_des_graphes#/media/File:Random20-0.1-instance3.png and https://en.wikipedia.org/wiki/Small-world_network#/media/File:Small-world-network-example.png

B. Write down the adjacency matrix.

C. Calculate the shortest path length between each pair of nodes.

D. Calculate the local clustering coefficient of each node.

E. Calculate the diameter of the network.

F. Calculate the global clustering coefficient.

6.2.2 Generic types of network and how to create them

Certain types of networks are observed frequently in real-world data, and/or make good null models against which to compare measured data.

The default null model: Erdős-Rényi networks

Erdős-Rényi networks are simple random networks that make a good starting point for comparisons to real-world data. To construct an Erdős-Rényi network with no self-edges, we specify a desired number of nodes N and a desired average degree $\bar{k}$. We then consider each possible edge between nodes, and include each one with a probability $q = \bar{k}/(N-1)$. The result will be a random graph with N nodes and approximately the desired average degree. A small example is shown in Fig. 6.5 (a).

Tip 6.2.1. If we're constructing an undirected Erdős-Rényi network in which self-edges are allowed, then each edge between distinct nodes is included with a probability $q = \bar{k}/(N)$ and each self-edge is included with a probability $q = \bar{k}/(2N)$ (to account for the fact that it contributes twice to the degree).

Tip 6.2.2. If we're constructing a directed Erdős-Rényi network with self-edges forbidden, then each outward edge from each node is included with a probability $q = \bar{k}_{out}/(N-1)$.

Tip 6.2.3. If we're constructing a directed Erdős-Rényi network with self-edges allowed, then each outward edge from each node (including the self edge) is included with a probability $q = \bar{k}_{out}/(N)$.

Example pseudo-code 6.2.1. Generating a directed Erdős-Rényi with N nodes and average degree approximately equal to $\bar{k}$. In essence, the challenge is to generate the adjacency matrix, which specifies the graph.

1. Declare and define dimension N.
2. Declare a two-dimensional list or matrix A with dimensions $N \times N$.
3. Declare and define a probability of connection $q = \bar{k}/N$

First, we consider generating a directed graph:

4. Loop over i from 1 to N:
5. Loop over j from 1 to N:
6. Define a variable t and set it equal to a random number drawn uniformly from the range (0,1).
7. If $t < q$:
8. $A_{ij} = 1$.
9. Else:
10. $A_{ij} = 0$.

Alternatively, we could consider generating an undirected graph with no self-edges:

4. Loop over i from 1 to N:
5. Loop over j from $i+1$ to N:
6. Define a variable t and set it equal to a random number drawn uniformly from the range (0,1).
7. If $t < q$:
8. $A_{ji} = A_{ij} = 1$.
9. Else:
10. $A_{ji} = A_{ij} = 0$.

Note that matlab has shortcuts for generating random matrices, which may be convenient in practice.

Characteristic features of Erdős-Rényi networks (assume no self-connections):

- For each node, we attempt to create $N-1$ connections, each with a probability $q = \bar{k}/(N-1)$. Thus the expected degree for each node is $\bar{k}$, and the distribution $p(k)$ is binomial (see Sec. 2.2.1).
- The clustering coefficient is simple: each edge has a probability q, so the proportion of the possible connections between the nodes adjacent to any node i that actually exist will, on average, be q. Thus $C = q$.
- The diameter $D \propto \log(N)$. This is relatively small; it is easy to get from one node to another (compare this to a graph that is a regular lattice; distant nodes would take many steps to reach).

Exercise 6.2.2. A simple argument to justify $D \propto \log(N)$ for sparse Erdős-Rényi networks. If the network is sparse, then $\bar{k} \ll N$ (and indeed $k_i \ll N$ for all i).

A. In this sparse limit, approximately how many nodes are connected to a given node i via a shortest path of l edges?

B. Roughly how large must the path length l_{sat} be before most nodes are connected to i by paths of length $l \leq l_{\text{sat}}$? Using this estimate, predict the scaling of diameter D with network size N.

Small-world networks

You have probably heard of the idea of "six degrees of separation" – that it is possible to connect yourself to essentially anyone in the world via a friend of a friend of a friend of a friend of a friend. Whether or not the precise number is six, the phenomenon is real; despite the fact that friendship networks are highly clustered (your friends are friends with your friends), and you don't know that many people in absolute terms, it doesn't take many steps to connect yourself to essentially anyone. The more formal way of phrasing this is that the network diameter D is surprisingly small ($D \propto \log(N)$) given the high global clustering coefficient C and the sparsity of the network ($\bar{k} \ll N$).

Sparse, highly-clustered networks with this small-world property seem to arise all the time in both biological (eg. molecular regulatory networks) and non-biological (eg. the internet) contexts. How do they manage it, when connections are so much more clustered than in an Erdős-Rényi network? One answer is that most small world networks typically have degree distributions that differ markedly from Erdős-Rényi networks with the same average connectivity. $p(k)$ for an Erdős-Rényi network is binomial, giving a fairly sharply peaked distribution with $\text{Var}(k) \sim \bar{k}$ (see Sec. 2.2.1). In many (but not all) small-world network, $p(k)$ is much broader, with substantial probabilities of finding a node with $k \gg \bar{k}$. These highly-connected nodes act as interconnecting hubs, allowing efficient hopping around the network and low D despite high clustering. A typical small world network, showing clear hubs, is illustrated in Fig. 6.5 (b).

Scale-free networks

In the discussion of small world networks, we mentioned that many natural networks have a much higher frequency of nodes with $k \gg \bar{k}$ than would be expected from an Erdős-Rényi network. But

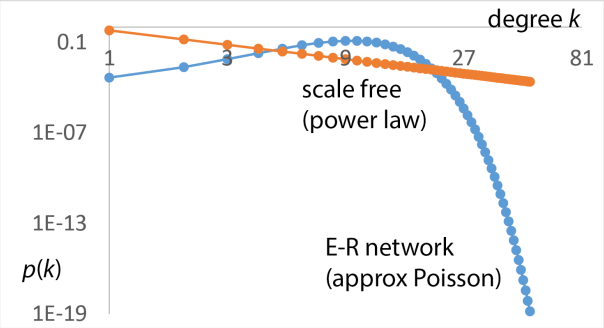


Figure 6.6: Ideal degree distributions for an Erdős-Rényi network and a scale free network, plotted on a log-log scale. Note the peak and rapid falloff at large k for the Erdős-Rényi network, compared to the straight line of the scale free counterpart, which has a power law dependence (specifically, $p(k) \sim k^{-2}$).

what exactly does this mean? Erdős-Rényi networks have a binomial degree distribution with parameters N and $q = \bar{k}/N$. In the limit of large N and finite $\bar{k} \ll N$, as applies to most networks of interest, this binomial can be approximated by a Poisson degree distribution with mean $\bar{k}$ (Sec. 2.2.1). Thus

$$p(k) \approx \frac{\bar{k}^k \exp(-\bar{k})}{k!}. \tag{6.2}$$

This distribution is reasonably tightly peaked around $\bar{k}$, which then sets a natural scale of the distribution. Exploring the behaviour of this function for large k is simpler if we plot $\ln p(k)$ against $\ln k$ – this peaked shape, with increasingly rapid falloff for large k, is typical. When we say that highly connected hubs are more common than expected if the fall of in $\ln p(k)$ against $\ln k$ is slower than this.

Many networks in the wild – including most regulatory and metabolic networks, the internet and social networks, exhibit something known as *scale-free* behaviour. Scale-free networks are characterised by a power law distribution, for which $p(k) \sim k^{\alpha}$ (at least for large k). On a log-log plot, a power law appears as a straight line: $\ln p(k) = \alpha \ln k + \text{const}$ – markedly different from a Poisson or binomial distribution (see Fig. 6.6). In particular, there is no natural peak at $k \approx \bar{k}$, and no increasingly rapid drop-off beyond this point. There is thus no natural "scale" to these networks, which is why they are called"scale-free". The absence of an increasingly rapid drop-off in $\ln p(k)$ at large $\ln k$ for scale-free networks is indicative of an over-representation of highly-connected hubs, when compared to Erdős-Rényi networks. A natural example of a power law degree distribution from a scale-free network is shown in Fig. 6.7.

The *Barabási-Albert* model gives us a mechanism for generating scale-free networks. The basic idea is to have positive feedback, so that highly connected nodes gain even more edges (rich get richer). We start with an initial small network with N_0 nodes, each with at least one edge (we could simply use $N_0 = 2$, with an edge between the two nodes). We then add new nodes one at a time that each connect to m existing nodes. An existing node i is selected for connection with a probability proportional to its current degree, k_i. Heavily-linked nodes therefore tend to accumulate edges at a faster rate, giving a much slower drop-off in $\ln p(k)$ at large $\ln k$ than for Erdős-Rényi networks. A graph generated using the Barabási-Albert model, with clear hubs, is shown in Fig. 6.8. It can be shown that $\alpha = -3$ for the Barabási-Albert model, but deriving that result is beyond the scope of

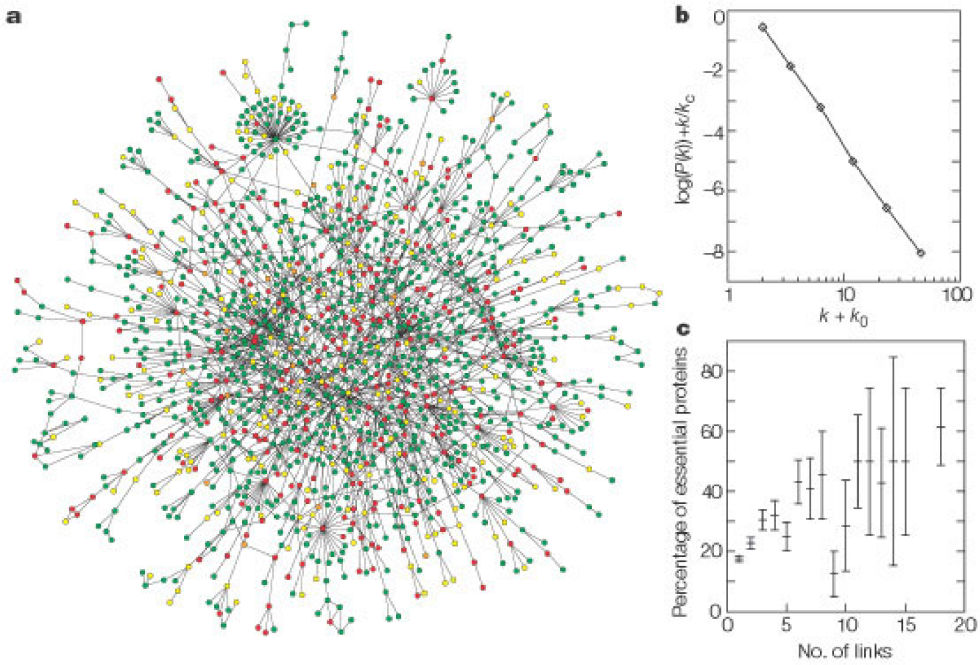


Figure 6.7: A protein-protein interaction network, taken from H. Jeong *et al.*, Nature 411, 41-42 (2001). (b) Shows the power law dependence of degree distribution.

this course.

Might the method of artificially generating scale-free networks using the Barabási-Albert model, a form of *preferential attachment*, tell us something about why scale-free networks occur so often in practice? On the world wide web, for example, it makes sense that new pages link more frequently to pre-existing pages with high degree. It also makes some sense for protein networks, since a common mechanism for evolving a new protein is to duplicate an existing one, which then drifts away from the parent via mutation. When the copy is made, all those proteins linked to the original are also linked to the duplicate, and there is a good chance that they will stay connected as the protein and its parent differentiate. Thus highly connected proteins are likely to acquire more edges, simply because they are connected to more proteins that might be duplicated.

Tip 6.2.4. Although the *Barabási-Albert* approach provides great insight, it isn't a perfect model of typical real networks. As is clear from Fig. 6.8, the clustering coefficient C isn't very high; there is no strong tendency for two adjacent nodes to have a third node adjacent to both.

Tip 6.2.5. Many networks are not just clustered, but *modular*. A module is a tightly connected group of nodes that functions effectively as a unit, and may be closely analogous to other modules performing related tasks. Typically, a small subset of nodes within the module act as inputs and outputs to couple to the rest of the network. The assumption that cellular networks are essentially modular allows the kind of engineering performed by synthetic biologists, in which small circuits

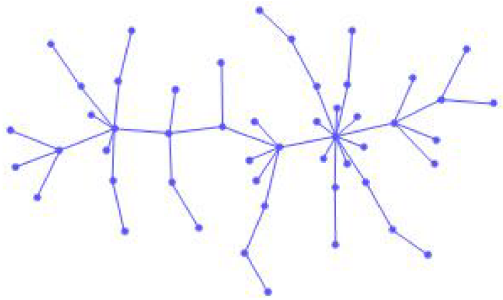


Figure 6.8: A scale-free network generated using the Barabási-Albert approach. Note the presence of highly connected hubs, but also the absence of clustering. Taken from https://commons.wikimedia.org/wiki/File:Barabasi_Albert_generated_network.jpg

are treated largely in isolation, and components are taken out of one context and plugged into another.

6.3 Networks in biology

6.3.1 Spread of disease over a population

A simple model of an epidemic might assume that the population is split into diseased, healthy and recovered fractions, and the fractions are updated over time using either deterministic or stochastic equations. This might be sufficient for some purposes, but it inherently assumes that the population is "well-mixed", by which we mean anyone can catch the disease from anyone else. In practice, this is a very crude description - we're much more likely to contract diseases from people we know well and see often. A more realistic model would be to include the underlying social network in our model; infections are propagated along edges between nodes only (perhaps with a propensity proportional to a weight of connection). One might expect that the network topology has a large effect on the progress of the epidemic – a question we will explore in the practical.

The spread of disease is just one characteristic example of a (stochastic) process that is influenced by an embedded network structure. Another would be evolution, in which we could assign a node to each genotype that produces a functional phenotype, and connect genotypes that are related by a single point mutation with an edge. Clearly, stochastic evolution within this network will be highly influenced by its topology.

6.3.2 Identifying over-represented – and potentially functional – motifs within networks

Consider the network shown in Fig. 6.9, which is a partial gene regulation network from *E. coli*. Nodes correspond to proteins encoded by the *E. coli* genome, and edges correspond to transcriptional regulation. These edges are directed, and an edge from i to j implies that the presence of i either increases or decreases the production of j. Self edges are possible; proteins can promote or suppress their own production.

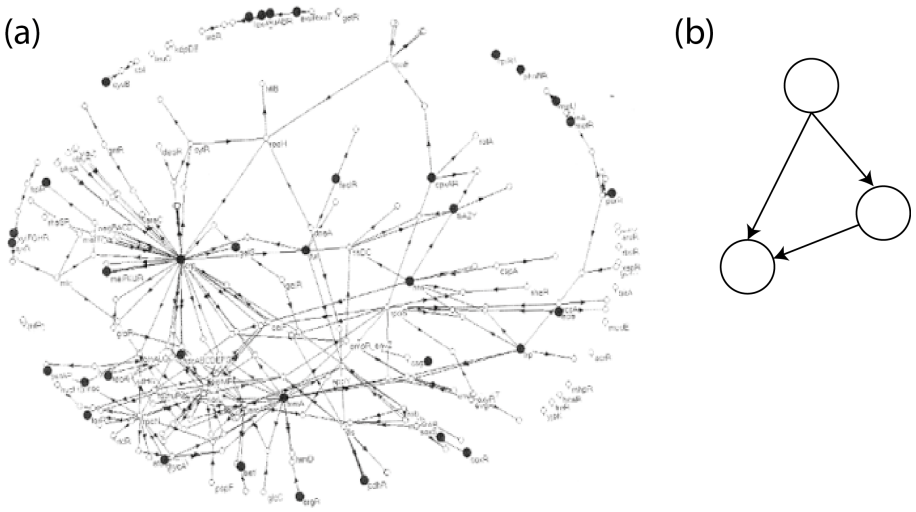


Figure 6.9: (a) Partial map of transcription regulation in *E. coli*, taken from Urii Alon: An introduction to systems biology.[8] Edges indicate transcriptional regulation (either suppression or activation); the direction of the edges indicates the direction of influence. Self edges (autoregulating transcription factors) are indicated by nodes represented in bold. (b) A feed-forward loop subgraph.

The *E. coli* network has evolved to meet certain functional requirements. Are some subgraph structures (small regions of connected nodes) particularly helpful in achieving a range of useful behaviours? Useful subgraphs are given the name of "motifs", and we'd expect them to appear frequently within the overall network. To find candidate motifs, we therefore need to hunt for over-represented subgraphs. We thus need a null model for the purposes of comparison, to identify when a subgraph is over-represented compared to our expectations. Directed Erdős-Rényi networks are a good place to start. Simple examples of how this might work in practice are discussed below. This discussion borrows heavily from Ref. 8.

Autoregulation

Autoregulation is the simplest possible candidate motif; one node with a self edge. In the *E. coli* network shown, there are $N = 420$ nodes with $E = 520$ edges, including $N_s = 40$ self-edges. Is this more or less than we'd expect in an Erdős-Rényi network with the same number of nodes and edges?

In a directed Erdős-Rényi network with N nodes, there are N^2 possible distinct edges (including self-edges). Thus, any given edge exists with a probability E/N^2. The number of self-edges therefore follows a binomial distribution (Sec. 2.2.1) with N trials and probability $q = E/N^2$ of success, giving

$$\langle N_s \rangle_{ER} = Nq = E/N, \qquad \mathrm{Var}(N_s)_{ER} = Nq(1-q) \approx E/N, \qquad \sigma_{N_s} = \sqrt{E/N}. \tag{6.3}$$

For $N = 420$ and $E = 520$, $\langle N_s \rangle_{ER} = 1.2$ and $\sigma_{N_s} = 1.1$. It is thus very difficult to see how $N_s = 40$, as observed in the real network, could have happened by chance; some underlying factor must encourage the existence of autoregulation relative to a random network. One explanation is that autoregulation is an advantageous motif that has been selected repeatedly by evolution.

Feed-forward loop

A more complex candidate motif is the three-node feed-forward loop, illustrated in Fig. 6.9 (b). This particular subgraph is observed 42 times in the network in Fig. 6.9 (a) – how does this compare to an Erdős-Rényi network? We analyse this candidate motif in Ex. 6.3.1.

Exercise 6.3.1. Identifying the expected number of feed-forward loops in an Erdős-Rényi network. Let us first ask for the expected number of V-shaped patterns of edges facing outwards from the same node (ie., the structure at the head of a feed-forward loop; Fig.6.9). To simplify the analysis, assume that the number of self edges in an Erdős-Rényi network with large N can be neglected.

A. If a node has an out-degree of k_{out}, how many different pairs of outward-facing edges can be identified?

B. Show that the expectation of the number of pairs of outwards facing edges for each node is $\frac{\bar{k}_{\text{out}}^2}{2}(1-1/N) \approx \bar{k}_{\text{out}}^2/2$. Here, $\bar{k}_{\text{out}}$ is the average out degree of a node. You may wish to use the fact that the number of out edges on each node follows a binomial distribution with parameters N and $q = \bar{k}_{\text{out}}/N$ (Sec. 2.2.1).

C. How many outwards-facing V patterns are therefore expected in the whole network?

D. What is the probability that any given one of these outwards-facing V-motifs have an additional edge joining the nodes at the external tips of the V together? Using this result, argue that the expected number of feed-forward motifs is $\langle N_{FFL}\rangle_{ER} \approx \bar{k}_{\text{out}}^3$, independent of N at fixed $\bar{k}_{\text{out}}$.

For our parameters, $\langle N_{FFL}\rangle_{ER} = \bar{k}_{\text{out}}^3 \approx 1.2^3 \approx 1.7$. Although we haven't worked it out, a reasonable guess for $\text{Var}(N_{FFL})_{ER}$ is $\text{Var}(N_{FFL})_{ER} \approx \langle N_{FFL}\rangle_{ER} \approx 1.7$ (we'd probably expect the total number of feed-forward loops to be roughly binomial- or Poisson-distributed). Thus, again, it is difficult to account for the number of instances of this subgraph without some underling factor, such as its potential as an advantageous motif.

Tip 6.3.1. Not every subgraph tested will turn out to be over-represented; other three-node motifs do not actually appear to be within this network.

Tip 6.3.2. Beware over-interpretation of results like this. The high prevalences of autoregulation and feed-forward loops strongly suggest that these subgraphs are widely-useful "motifs". However, other explanations are possible - perhaps these subgraphs are common just because they are easy to produce in the evolutionary dynamics, rather than because they provide a large fitness benefit. In fact, in these cases, we have good evidence that autoregulation and feed-forward loops are actually functionally useful. They provide important dynamic capabilities, as outlined further by Alon[8].

Chapter 7

Solutions to Exercises

7.1 Chapter 2

Exercise 2.1.1

A. X and Y are not independent. There are many ways to see this, including noting that $P(Y = 0|X = 0) \neq P(Y = 0|X = 1)$, for example.

B.
$$p(x) = \begin{cases} 1/2 & \text{for } x = 0 \\ 1/2 & \text{for } x = 1 \end{cases} \qquad p(y) = \begin{cases} 3/8 & \text{for } x = 0 \\ 5/8 & \text{for } x = 1 \end{cases}.$$

C.
$$p(x|y) = \begin{cases} 2/3 & \text{for } x = 0, y = 0 \\ 1/3 & \text{for } x = 1, y = 0 \\ 2/5 & \text{for } x = 0, y = 1 \\ 3/5 & \text{for } x = 1, y = 1 \end{cases} \qquad p(y|x) = \begin{cases} 1/2 & \text{for } y = 0, x = 0 \\ 1/2 & \text{for } y = 1, x = 0 \\ 1/4 & \text{for } y = 0, x = 1 \\ 3/4 & \text{for } y = 1, x = 1 \end{cases}.$$

D. $p(y = 1|x = 1) = 3/4$.

Exercise 2.1.2

A.
$$\langle aX + bY \rangle = \sum_{x \in \mathscr{A}_X} \sum_{y \in \mathscr{A}_Y} (ax + by) p(x,y) = a \sum_{x \in \mathscr{A}_X} \sum_{y \in \mathscr{A}_Y} x p(x,y) + b \sum_{y \in \mathscr{A}_Y} \sum_{y \in \mathscr{A}_X} y p(x,y). = a\langle X \rangle + b\langle Y \rangle$$

B.
$$\langle X^2 \rangle - \langle X \rangle^2 = \langle X^2 \rangle - 2\langle X \rangle^2 + \langle X \rangle^2 = \sum_{x \in \mathscr{A}_X} x^2 p(x) - 2\langle X \rangle \sum_{x \in \mathscr{A}_X} x p(x) + \langle X \rangle^2 \sum_{x \in \mathscr{A}_X} p(x)$$

In this first line I have simply added and subtracted $\langle X \rangle^2$, then re-written some of the expectations explicity. I have also used the normalisation identity $\sum_{x \in \mathscr{A}_X} p(x) = 1$ (Eq. 2.3). The

trick now is to realise that $\langle X\rangle^2$ and $\langle X\rangle$ do not depend on single outcomes x. Therefore they can be moved inside the sums, giving

$$\langle X^2\rangle - \langle X\rangle^2 = \sum_{x\in\mathscr{A}_X} \left(x^2 - 2x\langle X\rangle + \langle X\rangle^2\right) p(x) = \sum_{x\in\mathscr{A}_X} \left(x - \langle X\rangle\right)^2 p(x),$$

as required. It is actually slightly easier to work backwards from the final expression to show that it is equal to the initial one.

C.

$$\mathrm{Var}(aX) = \sum_{x\in\mathscr{A}_X} a^2x^2p(x) - \left(\sum_{x\in\mathscr{A}_X} axp(x)\right)^2 = a^2\sum_{x\in\mathscr{A}_X} x^2p(x) - a^2\left(\sum_{x\in\mathscr{A}_X} xp(x)\right)^2 = a^2\mathrm{Var}(X).$$

D.

$$\mathrm{Var}(X+Y) = \sum_{x\in\mathscr{A}_X}\sum_{y\in\mathscr{A}_Y} (x+y)^2p(x,y) - \left(\sum_{x\in\mathscr{A}_X}\sum_{y\in\mathscr{A}_Y} (x+y)p(x,y)\right)^2.$$

If X and Y are independent, $p(x,y) = p(x)p(y)$ (Eq. 2.10). Thus, using $\sum_{x\in\mathscr{A}_X} p(x) = 1$ (Eq. 2.3) and its equivalent for y, we can expand and simplify:

$$\mathrm{Var}(X+Y) = \sum_{x\in\mathscr{A}_X} x^2p(x) + \sum_{y\in\mathscr{A}_Y} y^2p(y) + 2\sum_{x\in\mathscr{A}_X} xp(x)\sum_{y\in\mathscr{A}_Y} yp(y) - \left(\sum_{x\in\mathscr{A}_X} xp(x) + \sum_{y\in\mathscr{A}_Y} yp(y)\right)^2.$$

Expanding the bracket and cancelling terms gives

$$\mathrm{Var}(X+Y) = \sum_{x\in\mathscr{A}_X} x^2p(x) - \left(\sum_{x\in\mathscr{A}_X} xp(x)\right)^2 + \sum_{y\in\mathscr{A}_Y} y^2p(y) - \left(\sum_{y\in\mathscr{A}_Y} yp(y)\right)^2 = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

E.

$$\mathrm{Cov}(X,Y) = \sum_{x\in\mathscr{A}_X, y\in\mathscr{A}_Y} xyp(x,y) - \sum_{x\in\mathscr{A}_X} xP(x)\sum_{y\in\mathscr{A}_Y} yp(y).$$

If X and Y are independent, $p(x,y) = p(x)p(y)$ and hence the first term can be re-written to be equivalent to the second.

$$\mathrm{Cov}(X,Y) = \sum_{x\in\mathscr{A}_X, y\in\mathscr{A}_Y} xyp(x)p(y) - \sum_{x\in\mathscr{A}_X} xp(x)\sum_{y\in\mathscr{A}_Y} yp(y) = \sum_{x\in\mathscr{A}_X} xp(x)\sum_{y\in\mathscr{A}_Y} yp(y) - \sum_{x\in\mathscr{A}_X} xp(x)\sum_{y\in\mathscr{A}_Y} yp(y) = 0.$$

7.2 Chapter 3

Exercise 3.4.1

A. Please refer to Fig. 7.1.

B. Indexing U as 1, Q as 2 and V as 3,

$$\begin{pmatrix} 0.5 & 0.3 & 0 \\ 0.5 & 0 & 0.8 \\ 0 & 0.7 & 0.2 \end{pmatrix}$$

Each column sums to one by inspection.

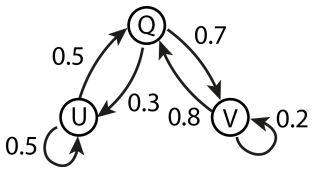


Figure 7.1: Graphical representation of discrete-time Markov Chain in Ex. 3.4.1

Exercise 3.4.2

A. We seek the following:

$$\begin{pmatrix} 0.5 & 0.3 & 0 \\ 0.5 & 0 & 0.8 \\ 0 & 0.7 & 0.2 \end{pmatrix} \begin{pmatrix} \pi(U) \\ \pi(Q) \\ \pi(V) \end{pmatrix} = \begin{pmatrix} \pi(U) \\ \pi(Q) \\ \pi(V) \end{pmatrix},$$

with $\pi(U)+\pi(Q)+\pi(V)=1$. Considering the top line of the matrix equation, we have $5\pi(U)+3\pi(Q)=10\pi(U)$, or $3\pi(Q)=5\pi(U)$. From the third line, we have $7\pi(Q)+2\pi(V)=10\pi(V)$, or $7\pi(Q)=8\pi(V)$. Thus $\pi(U):\pi(Q):\pi(V)$ has the ratio $3/5:1:7/8$, giving a steady state of

$$\begin{pmatrix} \pi(U) \\ \pi(Q) \\ \pi(V) \end{pmatrix} = \begin{pmatrix} 24/99 \\ 40/99 \\ 35/99 \end{pmatrix},$$

B. At $n=0$, all three eigenvectors contribute to the initial state. Over time, those with $|\lambda|<1$ decay, becoming less important – the system then relaxes to the stationary distribution, and the initial knowledge has no predictive power. The slowest decaying eigenvector (apart from the stationary state) has $|\lambda|\approx 0.73$. Thus after two days, the deviation from the stationary distribution due to this presence of this component in $P(x,n)$ will be about half as big as at the start, and after 10 days it will only contribute 10% of its initial value on the first day. The other component, with $|\lambda|\approx 0.43$, decays even faster.

Exercise 3.4.3

A.

$$\mathbf{U} = \begin{pmatrix} 0 & 0 & 0.5 & 0 \\ 0.6 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 \\ 0 & 0.3 & 0 & 0.2 \end{pmatrix} \quad \mathbf{R} = \begin{pmatrix} 0 & 0 & 0.5 & 0.8 \\ 0.4 & 0 & 0 & 0 \end{pmatrix}$$

B. Simply perform the matirx multiplication $\mathbf{W}(\mathbb{I}-\mathbf{U})$ or $(\mathbb{I}-\mathbf{U})\mathbf{W}$, obtaining $\mathbb{I}$.

C. The probability that state 5 is reached is given by $\mathbf{RW}_{11}=0.494$. The probability that state 5 is reached is given by $\mathbf{RW}_{21}=0.506$. The two add up to 1, as they must, since all trajectories must end in one absorbing state or the other.

D. $\langle t_1^{\text{abs}}\rangle = \sum_i W_{i1} = 2.84$ discrete time steps.

Exercise 3.5.1

A.
$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} P(X=A,t) \\ P(X=B,t) \\ P(X=C,t) \end{pmatrix} = \begin{pmatrix} -(k_{BA}+k_{CA}) & 0 & 0 \\ k_{BA} & 0 & 0 \\ k_{CA} & 0 & 0 \end{pmatrix}\begin{pmatrix} P(X=A,t) \\ P(X=B,t) \\ P(X=C,t) \end{pmatrix}$$

B. Follows directly from the first line of the master equation; $\mathrm{d}/\mathrm{d}tP(X=A,t) = -(k_{BA}+k_{CA})P(X=A,t) = -k_A^- P(X=A,t)$, with $k_A^- = k_{BA}+k_{CA}$.

C. Simple first order ODE, solved by $P(X=A,t) = D\exp(-(k_{BA}+k_{CA})t)$, with D constant. Use initial condition of $P(X=A,0)=1$ to obtain $D=1$ and $P(X=A,t) = \exp(-(k_{BA}+k_{CA})t)$.

D. Probability density of leaving A at time t_A is $p(t_A) = -\mathrm{d}/\mathrm{d}tP(X=A,t) = (k_{BA}+k_{CA})\exp(-(k_{BA}+k_{CA})t) = k_A^-\exp(-k_A^- t)$, as claimed.

E. The solution to the differential equation is the same, and the initial condition is $P(X=A,t)=1$ at $t=t^0$, giving $P(X=A,t) = \exp(-k_A^-(t-t^0))$. Thus the probability density of leaving state A after an additional time $t_A - t^0$ is $p(t_A - t^0 | A \text{ at } t^0) = k_A^-\exp(-k_A^-(t-t^0))$, as claimed.

F.
$$\frac{\mathrm{d}P(X=B,t)}{\mathrm{d}t} = k_{BA}P(X=A,t) \Longrightarrow P(X=B,t) = -\frac{k_{BA}}{k_A^-}\left[\exp(-k_A^- t)\right]_0^t = \frac{k_{BA}}{k_A^-}\left(1-\exp(-k_A^- t)\right).$$

Similarly,
$$P(X=C,t) = \frac{k_{CA}}{k_A^-}\left(1-\exp(-k_A^- t)\right).$$

It is clear that $P(X=B,t)/P(X=C,t) = k_{BA}/k_{CA}$ for all t, so the ratio of transitions to B and C at each moment in time is k_{BA}/k_{CA}, as claimed.

Exercise 3.5.2

A. The discrete time Markov chain with the transition matrix
$$\mathbf{T} = \begin{pmatrix} 0 & 1 & 0 \\ 0.75 & 0 & 1 \\ 0.25 & 0 & 0 \end{pmatrix}$$

B. To calculate the average first passage time from state 1 to state 3, turn state 3 into an absorbing state and construct $\mathbf{U}$.
$$\mathbf{U} = \begin{pmatrix} 0 & 1 \\ 0.75 & 0 \end{pmatrix}$$

Now calculate fundamental matrix
$$\mathbf{W} = (\mathbb{I}-\mathbf{U})^{-1} = \begin{pmatrix} 4 & 4 \\ 3 & 4 \end{pmatrix}.$$

Expected first passage time $\langle t_1^{\text{abs}}\rangle = W_{21}\tau_2 + W_1 1\tau_1$. Using $\tau_1 = 0.25\,\text{s}$ and $\tau_2 = 0.5\,\text{s}$ gives $\langle t_1^{\text{abs}}\rangle = 2.5\,\text{s}$

7.3 Chapter 4

Exercise 4.1.1

A. First, label states $A = 1$, $B = 2$... etc. Then we can chose either $\frac{\pi_1}{\pi_4} = \frac{K_{13}}{K_{31}}\frac{K_{34}}{K_{43}}$ or $\frac{\pi_1}{\pi_4} = \frac{K_{12}}{K_{21}}\frac{K_{24}}{K_{42}}$. In both cases, we find $\frac{\pi_1}{\pi_4} = 2$.

B. The answer would not change. If detailed balance is satisfied for the whole network, then it is necessarily possible to evaluate $\frac{\pi_1}{\pi_4}$ in the same manner as above.

Exercise 4.1.2

A. First, label states A-bound as 1, B-bound as 2 and empty as 0. Then, from the free energies, $\pi_1/\pi_0 = \exp(2)$ and $\pi_2/\pi_0 = \exp(4)$. Thus, enforcing normalisation,

$$\pi_0 = \frac{1}{1+\exp(2)+\exp(4)} \approx 0.016, \quad \pi_1 = \frac{\exp(2)}{1+\exp(2)+\exp(4)} \approx 0.117, \quad \pi_2 = \frac{\exp(4)}{1+\exp(2)+\exp(4)} \approx 0.867.$$

B. Unbinding rate of A is $k_0\exp(-2) \approx 0.135k_0$. Unbinding rate of B is $k_0\exp(-4) \approx 0.0183k_0$.

C. Detailed balance implies that A must also be able to directly displace B, with a rate of $k_0\exp(-2)/2 \approx 0.0677k_0$. Since the system still possesses detailed balance even with this extra transition, the answer to A is unchanged.

Exercise 4.2.1

A. The equilibrium respects detailed balance. Thus $\pi_1/\pi_0 = \exp(-(F_1 - F_0)/k_BT) = \sigma$, and for $m > 1$, $\pi_m/\pi_{m-1} = \exp(-(F_m - F_{m-1})/k_BT) = t$. Thus the probability of a state with $m > 0$ base pairs, relative to the 0-base-pair state, is $\pi_m/\pi_0 = \sigma t^{m-1}$. Consequently, the probability of having any number of base pairs $M > 0$ relative to $M = 0$ is $\sum_{m=1}^{N} \sigma t^{m-1}$. The absolute probability in Eq. 4.11 follows immediately from this relative probability.

B. $\sum_{m=1}^{N} \sigma t^{m-1}$ is a geometric progression with N terms, initial term σ and common ratio t. Using the standard result,

$$\sum_{m=1}^{N} \sigma t^{m-1} = \frac{\sigma(t^N - 1)}{t - 1}.$$

Substituting this into Eq. 4.11, we find

$$P(M > 0) = \frac{\sigma(t^N - 1)}{t - 1 + \sigma(t^N - 1)}.$$

C. If $t \gg 1$, then $t - 1 \approx t$ and $t^N - 1 \approx t$. Thus

$$P(M > 0) \approx \frac{\sigma(t^N)}{t + \sigma(t^N)} = \frac{\sigma(t^{N-1})}{1 + \sigma(t^{N-1})},$$

as required.

Exercise 4.2.2

A. From the definition of $\sinh(x) = 1/2(\exp(x) - \exp(-x)$ and $\cosh(x) = 1/2(\exp(x) + \exp(-x)$, at small x

$$\sinh(x) \approx x + x^3/6, \quad 2 + \cosh(x) \approx 3 + x^2/2.$$

Thus Eq. 4.17 can be expanded as

$$\langle X \rangle \approx \frac{Nb^2 f}{3k_B T}.$$

This is equivalent to a spring with a constant $\kappa = 3k_B T/Nb^2$.

Exercise 4.2.3

A. Substitute $p(x) = C\exp(-\kappa(x-x_0)^2/2k_B T)$, where C is a constant, and $V(x) = \kappa/2(x-x_0)^2$, into RHS of Eq. 4.23:

$$\begin{aligned} \text{RHS} &= \frac{1}{\gamma}\frac{\partial}{\partial x}\left(\kappa(x-x_0)C\exp(-\kappa(x-x_0)^2/2k_B T)\right) + \frac{k_B T}{\gamma}\frac{\partial^2}{\partial x^2}C\exp(-\kappa(x-x_0)^2/2k_B T) \\ &= \frac{\kappa}{\gamma}C\exp(-\kappa(x-x_0)^2/2k_B T) - \frac{\kappa^2}{\gamma k_B T}(x-x_0)^2 C\exp(-\kappa(x-x_0)^2/2k_B T \\ &\quad + \frac{\kappa^2}{\gamma k_B T}(x-x_0)^2 C\exp(-\kappa(x-x_0)^2/2k_B T - \frac{\kappa}{\gamma}C\exp(-\kappa(x-x_0)^2/2k_B T \\ &= 0, \end{aligned} \tag{7.1}$$

proving that $p(x) \propto \exp(-\kappa(x-x_0)^2/2k_B T)$ is indeed the stationary distribution.

B. To normalise, we require

$$1 = \int_{-\infty}^{+\infty} \mathrm{d}x C\exp(-\kappa(x-x_0)^2/2k_B T).$$

A simple change of variables to $y = x - x_0$ renders the integral equivalent to I_0 in Eq. 2.34, with $\alpha = \kappa/2k_B T$.

$$1 = \int_{-\infty}^{+\infty} \mathrm{d}y C\exp(-\alpha y^2) = CI_0 = C\sqrt{\frac{\pi}{\alpha}}.$$

Thus the normalisation constant $C = \sqrt{\frac{\kappa}{2\pi k_B T}}$.

Exercise 4.3.1

A. For each x, the system can leave that state through immigration with a transition rate λ_1, or by death with a transition rate $\mu_0 x$. The system can enter x by immigration from $x-1$ with a transition rate λ_1, or by death from $x+1$ with a transition rate $\mu_0(x+1)$. Thus the Master equation is

$$\frac{\mathrm{d}p_x(t)}{\mathrm{d}t} = -(\lambda_1 + \mu_0 x)p_x(t) + \lambda_1 p_{x-1}(t) + \mu_0(x+1)p_{x+1}(t).$$

B. Multiplying through by x and summing over x,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle X(t)\rangle &= -\sum_{x=0}^{\infty}(\lambda_1 x+\mu_0 x^2)p_x(t)+\lambda_1\sum_{x=0}^{\infty}xp_{x-1}(t)+\mu_0\sum_{x=0}^{\infty}x(x+1)p_{x+1}(t).\\ &= \lambda_1\langle X(t)\rangle-\mu_0\langle X^2(t)\rangle+\lambda_1\sum_{x=0}^{\infty}(x+1)p_x(t)+\mu_0\sum_{x=0}^{\infty}(x-1)xp_x(t)\\ &= \lambda_1\langle X(t)\rangle-\mu_0\langle X^2(t)\rangle+\lambda_1(\langle X(t)\rangle+1)+\mu_0(\langle X^2(t)\rangle-\langle X\rangle)\\ &= \lambda_1-\mu_0\langle X(t)\rangle. \end{aligned} \tag{7.2}$$

Exercise 4.3.2

A. The simultaneous equations are

$$\begin{aligned}-(k_{\mathrm{on}}[P]+k_3[ATP][E])\pi_1+k_{\mathrm{off}}\pi_2+k_4\pi_3 &= 0\\ k_{\mathrm{on}}[P]\pi_1-(k_{\mathrm{off}}+k_3[ADP][E])\pi_2+k_4\pi_3 &= 0\\ \pi_1+\pi_2+\pi_3 &= 1\end{aligned}$$

The first two of the above simultaneous equations follow from using the first two lines of **K** to identify an eigenvector with eigenvalue 0. The third equation is the normalisation of total probability. Note that taking the third line of **K** gives no new information relative to the first two lines when looking for an eigenvector with a specific eigenvalue, but it is valid to use it as an alternative to one of the first two lines.

B. We can subtract the second of the above equations from the first, giving

$$\pi_1(2k_{\mathrm{on}}[P]+k_3[ATP][E])=\pi_2(k_{\mathrm{off}}+k_3[ADP][E]),$$

and thus

$$\frac{\pi_2}{\pi_1}=\frac{2k_{\mathrm{on}}[P]+k_3[ATP][E]}{2k_{\mathrm{off}}+k_3[ADP][E]}$$

as required.

7.4 Chapter 5

Exercise 5.1.1

A. In this model, the free energy difference between $M=0$ and $M=1$, the transition state, is N-independent ($F_1=F_0=-k_BT\ln\sigma$). Therefore the binding rate, $k_{B\to A}$, is predicted to be N-independent in this picture.

Exercise 5.1.2

A. Each time the system visits state E, it can either transition to s or i. From s, it will either eventually make it to S or return to E; similarly, from i it will either eventually make it to I or return to E. If the system returns to E, the process is repeated, each time with the same relative probabilities of the various outcomes. Thus to identify the relative probabilities of reaching S or I after any number of abortive attempts that return to E, it is enough to calculate the relative probabilities of success the first time the system leaves state E.

The probability that the system leaves state E and reaches S before returning to E is simply $P(s \text{ before } i|E)P(S \text{ before } E|s)$. Similarly, the probability that it leaves E and reaches I before returning to E is simply $P(i \text{ before } s|E)P(I \text{ before } E|i)$, giving the relative probability quoted in Eq. 5.9.

B. For the simple system in Fig. 5.3,

- $P(s \text{ before } i|E) = 1/2$.
- $P(S \text{ before } E|s) = \frac{k_{\text{cat}}}{k_{\text{cat}}+k^s_{\text{off}}}$.
- $P(i \text{ before } s|E) = 1/2$.
- $P(I \text{ before } E|i) = \frac{k_{\text{cat}}}{k_{\text{cat}}+k^i_{\text{off}}}$.

Thus, as claimed,

$$\frac{P(S \text{ before } I|E)}{P(I \text{ before } S|E)} = \frac{k_{\text{cat}}+k^i_{\text{off}}}{k_{\text{cat}}+k^s_{\text{off}}}$$

since all other terms cancel between numerator and denominator.

C. For the more complex system in Fig. 5.4, the only difference is that $P(S \text{ before } E|s)$ and $P(I \text{ before } E|i)$ are both the products of two simpler probabilities:

$$P(S \text{ before } E|s) = P(s^* \text{ before } E|s)P(S \text{ before } E|s^*) = \frac{k_{\text{cat}}}{k_{\text{cat}}+k^s_{\text{off}}}\frac{k_p}{k_p+k^s_{\text{off}}}, \tag{7.3}$$

and

$$P(I \text{ before } E|i) = P(i^* \text{ before } E|i)P(I \text{ before } E|i^*) = \frac{k_{\text{cat}}}{k_{\text{cat}}+k^i_{\text{off}}}\frac{k_p}{k_p+k^i_{\text{off}}}. \tag{7.4}$$

Thus for the more complex system,

$$\frac{P(S \text{ before } I|E)}{P(I \text{ before } S|E)} = \frac{k_{\text{cat}}+k^i_{\text{off}}}{k_{\text{cat}}+k^s_{\text{off}}}\frac{k_{\text{p}}+k^i_{\text{off}}}{k_{\text{p}}+k^s_{\text{off}}} \tag{7.5}$$

as required.

Exercise 5.1.3

A. Substitute

$$p(x,t) = \sqrt{\frac{1}{4\pi Dt}}\exp\left(-x^2/4Dt\right)$$

into the LHS of Eq. 5.10. We obtain

$$\frac{\partial p(x,t)}{\partial t} = \frac{x^2}{4Dt^2}p(x,t) - \frac{1}{2t}p(x,t).$$

Similarly, substituting into the RHS of Eq. 5.10 gives

$$\frac{\partial^2 p(x,t)}{\partial x^2} = -D\frac{\partial}{\partial x}\frac{x}{2Dt}p(x,t) = -\frac{1}{2t}p(x,t) + \frac{x^2}{4Dt^2}p(x,t).$$

We immediately see that the RHS = LHS, proving that Eq. 5.11 indeed solves Eq. 5.10.

B. As $t \to 0$ we have an infinitely narrow Gaussian centred on $x = 0$. This is a delta function. It corresponds to an initial condition of x being guaranteed to be found at exactly $x = 0$ at $t = 0$.

Exercise 5.1.4

A. In the stationary distribution, $X(t)$ and $X(0)$ are both Gaussian distributed (see Sec. 4.2.3):

$$p(x(t)) = p(x(0)) = \sqrt{\frac{\kappa}{2\pi k_B T}} \exp(-\kappa x^2/2k_B T).$$

Using the analysis of Gaussian distributions presented in Sec. 2.2.2, $\langle X(t)\rangle = \langle X(0)\rangle = 0$ and $\langle X^2(t)\rangle = \langle X^2(0)\rangle = k_B T/\kappa$.

B.

$$\mathrm{Cov}(X(t),X(0)) = \langle X(t)X(0)\rangle - \langle X(t)\rangle\langle X(0)\rangle = \langle X(t)X(0)\rangle,$$

where I have used the fact that the average displacement is zero in the equilibrium stationary distribution. To find $\langle X(t)X(0)\rangle$, multiply Eq. 5.16 by $X(0)$ and take the expectation. Since $\langle \zeta(t)\rangle = 0$, this leaves only

$$\mathrm{Cov}(X(t),X(0)) = \langle X(t)X(0)\rangle = \langle X(0)^2\rangle \exp\left(-\frac{\kappa}{\gamma}t\right) = \frac{k_B T}{\kappa}\exp\left(-\frac{\kappa}{\gamma}t\right).$$

C. The autocorrelation function is defined by

$$R_X(t) = \frac{\mathrm{Cov}(X(t),X(0))}{\sqrt{\mathrm{Var}(X(t))\mathrm{Var}(X(0))}} = \exp\left(-\frac{\kappa}{\gamma}t\right),$$

as is required.

7.5 Chapter 6

Exercise 6.2.1

A. Degrees of A: 1. B: 3. C: 2. D: 2. Average: 2.

B.

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

C. AB: 1. AC: 2. AC: 2. BC: 1. BD: 1. CD: 1.

D. $C_A = 0, C_B = 1/3, C_C = 1.$ $C_D = 1.$

E. $D = 2.$

F. $C = 7/12.$

Exercise 6.2.2

A. For a sparse Erdős-Rényi network, the clustering coefficient is low. Thus most edges from nodes adjacent to a node i go to other nodes not adjacent to i. Therefore the average number of nodes that are connected by a minimum length of one edge is $\bar{k}$, two edges is approximately $\bar{k}^2$, three edges is approximately $\bar{k}^3$, etc. Thus approximately $\bar{k}^l$ nodes are connected to any node i by a minimum path of l edges.

B. This scaling clearly breaks down when $\bar{k}^l \sim N$, when the number of new nodes starts to run out. Thus for path lengths $l_{\text{sat}} \sim \ln N / \ln \bar{k}$, most nodes are connected by a path of length $l \leq l_{\text{sat}}$. This length l_{sat} will set the scale for connections across the network, including the network diameter, and hence D scales as $\ln N / \ln \bar{k}$.

Exercise 6.3.1

A. The number of ways to pair k_{out} edges is $\frac{k_{\text{out}}(k_{\text{out}}-1)}{2}$.

B. Let us treat K_{out} as a stochastic variable; $p(k_{out})$ is then a binomial distribution with parameters N and $q = \bar{k}_{\text{out}}/N$. We need to average $\frac{K_{\text{out}}(K_{\text{out}}-1)}{2}$ over this distribution. Rearranging the expectation,

$$\langle K_{\text{out}}(K_{\text{out}}-1)/2 \rangle = \frac{1}{2}\left(\text{Var}(K_{\text{out}}) + \langle K_{\text{out}} \rangle^2 - \langle K_{\text{out}} \rangle\right)$$

Using the results for a binomial from Sec. 2.2.1, we obtain

$$\langle K_{\text{out}}(K_{\text{out}}-1)/2 \rangle = \frac{1}{2}\left(N(\bar{k}_{\text{out}}/N)(1-\bar{k}_{\text{out}}/N) + \bar{k}_{\text{out}}^2 - \bar{k}_{\text{out}}\right) \approx \bar{k}_{\text{out}}^2/2.$$

In the final step I have ignored a a term $\bar{k}_{\text{out}}^2/2N$ that is relatively small for large N.

C. In the whole network there are N nodes, each with an average number of outwards-facing Vs of $\bar{k}_{\text{out}}^2/2$. Thus, for the whole network, the average number of outward facing V patterns is $N\bar{k}_{\text{out}}^2/2$.

D. There are two possible ways to complete the feed-forward motif with an extra edge between the final two nodes; each has a probability $\bar{k}_{\text{out}}/N$ of existing, and in a sparse network the probability of both existing is minimal. Thus the probability of either one or the other is approximately $2\bar{k}_{\text{out}}/N$ for each V-shaped motif, giving an expected number of feed-forward loops of

$$\langle N_{FFL} \rangle_{ER} = 2\bar{k}_{\text{out}}/N \times N\bar{k}_{\text{out}}^2/2 = \bar{k}_{\text{out}}^3,$$

as required.

Bibliography

[1] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2003.

[2] N. G. Van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, Amsterdam, 2007.

[3] https://en.wikipedia.org/wiki/CC_(cat).

[4] P. Nelson. *Biological physics: Energy, information life*. W. H. Freeman, NY, NY., 2003.

[5] https://www-thphys.physics.ox.ac.uk/people/JohnChalker/theory/lecture notes.pdf.

[6] K. Huang. *Statistical Mechanics, Second Edition*. John Wiley & Sons, Inc., New York, 1987.

[7] E. Renshaw. *Stochastic Population Processes: Analysis, Approximations, Simulations*. Oxford Universtiy Press, Oxford, UK, 2011.

[8] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Networks*. CRC press, Boca Raton, FL, 2007.

[9] https://en.wikipedia.org/wiki/Dijkstra%27s_algorithm.