

## Chapter 2

# Technical Background

In this chapter, I will review some of the fundamental ideas and mathematical tools employed in the analysis of stochastic systems. Throughout, the discussion will be concerned with Markov stochastic processes. This class of processes will be introduced more formally in the following section, but for now they are simply described as random processes without memory. We will see that the probability of a system existing in a certain state may be described by a PDF which evolves in time according to an equation known as the master equation. The master equation is a set of partial difference equations and is in general completely intractable. It will be shown that the situation can be greatly simplified however if the state variable is assumed to be continuous.

Formally the approximation introduced is akin to the Kramers-Moyal expansion, which allows the master equation to be approximated by a Fokker-Planck equation (FPE). The FPE is a partial differential equation (PDE) in as many variables as there are distinct types of individual in the original system. In most cases it is far more tractable than the original master equation. Some useful manipulations of the FPE are therefore introduced in Sect. 2.5.

While useful as a calculational tool, the FPE formalism is not very intuitive. Since intuition will play an important role in developing the fast-variable approximation schema in Chaps. 3 and 4, an alternative but equivalent formalism is introduced, that of the stochastic differential equation (SDE). Naïvely, these can be interpreted as ODEs of type Eq. (1.1) but with the addition of some small random noise.

The final sections of this chapter will then concentrate on a review of some essential tools from deterministic dynamical systems theory and their generalisation to stochastic systems. The fundamental problems underlying fast-variable elimination in stochastic systems can then be addressed.

In order to make the discussion less abstract, I will make reference throughout to the concrete example of a Moran model. While a neutral Moran model was discussed in the introduction, for various pedagogical reasons the Moran model with mutation will serve as a preferable illustrative example.

## 2.1 Stochastic Processes and the Markov Assumption

To begin, let us consider a single stochastic variable  $X$  which evolves in time. The measured value of  $X$  at some point in time is the non-negative integer  $n$ . This defines the state of the one-dimensional system. As stated in the introduction, in a stochastic process the state of the system at some time is not necessarily known with certainty. Therefore rather than consider the time evolution of the state of the system, we must consider the time evolution of the PDF of states in the system.

Say the probability of being in state  $n$  at time  $t$  is given by the PDF  $P_1(n, t)$ . The probability of following a specific trajectory in discrete time, beginning at a point  $n_1$  at  $t_1$ , moving to point  $n_2$  at  $t_2$  and so on over  $k$  steps until point  $n_k$  at  $t_k$  is a joint PDF which can be denoted

$$P_k(n_k, t_k; n_{k-1}, t_{k-1}; \dots; n_1, t_1). \quad (2.1)$$

The process can be broken down by introducing the conditional probability distribution,  $P_{k|l}(n_{k+l}, t_{k+l}; \dots; n_{l+1}, t_{l+1} | n_l, t_l; \dots; n_1, t_1)$ . This is the probability of following the trajectory  $n_{l+1}$  to  $n_{k+l}$  given that the trajectory  $n_1$  to  $n_l$  was first followed. The two distributions are related via Bayes' theorem [49], which states

$$\begin{aligned} P_{k|l}(n_{k+l}, t_{k+l}; \dots; n_{l+1}, t_{l+1} | n_l, t_l; \dots; n_1, t_1) \\ = \frac{P_{k+l}(n_{k+l}, t_{k+l}; \dots; n_l, t_l; \dots; n_1, t_1)}{P_l(n_l, t_l; \dots; n_1, t_1)}. \end{aligned} \quad (2.2)$$

This allows the PDF to be separated into a hierarchy.

A Markov process is one in which the probability of moving from state  $n_{k-1}$  to state  $n_k$  depends only on the state of the system at  $t_{k-1}$  and not on the state of the system at any previous times. This amounts to the conditional PDF having the following property [37],

$$P_{1|k-1}(n_k, t_k | n_{k-1}, t_{k-1}; \dots; n_1, t_1) = P_{1|1}(n_k, t_k | n_{k-1}, t_{k-1}). \quad (2.3)$$

This is the mathematical meaning of the Markov processes being memoryless. Substituting this into Eq. (2.2) and rearranging, we arrive at

$$P_2(n_k, t_k; n_{k-1}, t_{k-1}) = P_{1|1}(n_k, t_k | n_{k-1}, t_{k-1}) P_1(n_{k-1}, t_{k-1}). \quad (2.4)$$

The Markov process is fully determined by the functions  $P_{1|1}(n_k, t_k | n_{k-1}, t_{k-1})$  and  $P_1(n_k, t_k)$ , since, for any  $l > 2$ , one can write

$$P_l(n_l, t_l; \dots; n_1, t_1) = P_1(n_1, t_1) \prod_{k=2}^l P_{1|1}(n_k, t_k | n_{k-1}, t_{k-1}). \quad (2.5)$$

The function  $P_{1|1}(n_k, t_k | n_{k-1}, t_{k-1})$  will be referred to as the transition probability.

The final step in completing our initial outlook on Markov processes is in noting two summation relations. The first is rather obvious and is simply

$$P_1(n_2, t_2) = \sum_{n_1} P_{1|1}(n_2, t_2 | n_1, t_1) P_1(n_1, t_1). \quad (2.6)$$

The second relation comes from first taking Eq. (2.5) with  $l = 3$  and summing over  $n_2$ ,

$$P_2(n_3, t_3; n_1, t_1) = P_1(n_1, t_1) \sum_{n_2} P_{1|1}(n_3, t_3 | n_2, t_2) P_{1|1}(n_2, t_2 | n_1, t_1). \quad (2.7)$$

Dividing both sides by  $P_1(n_1, t_1)$ , one can then substitute for the left-hand side using Eq. (2.2) to obtain

$$P_{1|1}(n_3, t_3 | n_1, t_1) = \sum_{n_2} P_{1|1}(n_3, t_3 | n_2, t_2) P_{1|1}(n_2, t_2 | n_1, t_1). \quad (2.8)$$

This equation is a discrete form of the Chapman-Kolmogorov equation [38].

As stated earlier, functions  $P_1$  and  $P_{1|1}$  fully determine the Markov stochastic process. However, these functions must obey Eqs. (2.6) and (2.8) in order to describe a Markov process. The fact that Markov processes obey these convenient relationships makes their analysis far more manageable than for their non-Markovian counterparts. In part, this amenability explains their prevalence in stochastic modelling. There are however many examples of processes that are non-Markovian. For instance, successive draws of coins without replacement from a purse containing a finite number of coins in different denominations is a non-Markovian process; selecting a particular denomination at some time decreases the probability of again picking that denomination at each successive time. In this case, the Markovian assumption would clearly be a very bad modelling choice. However, it is Markov processes which we will be concerned with for the rest of the thesis.

## 2.2 The Master Equation and the Limit of Continuous Time

In the previous section, the fundamental properties of a subset of stochastic processes, Markov processes, were introduced. Let us now restrict our attention slightly further to processes which are also homogeneous. A process is homogeneous if the transition probability between any two given states at any two times is only dependent on the time interval between those times [49]. That is

$$P_{1|1}(n', t + \Delta t | n, t) = W_{\Delta t}(n' | n). \quad (2.9)$$

We now wish to simplify the Chapman-Kolmogorov equation, (2.8), by looking at its infinitesimal time-evolution. This will allow the system's time-evolution to be described by a set of differential equations. Let the system initially be at some state  $n_0$  at  $t_0$ . We wish to know the conditional probability that the system is in state  $n$  some small time  $\Delta t$  later. This is given by the Chapman-Kolmogorov equation, Eq. (2.8), which in this new notation reads

$$P_{1|1}(n, t + \Delta t | n_0, t_0) = \sum_{n'} P_{1|1}(n, t + \Delta t | n', t) P_{1|1}(n', t | n_0, t_0). \quad (2.10)$$

If the process is homogeneous we can substitute in Eq. (2.9) to arrive at

$$P_{1|1}(n, t + \Delta t | n_0, t_0) = \sum_{n'} W_{\Delta t}(n | n') P_{1|1}(n', t | n_0, t_0). \quad (2.11)$$

The assumption is now made [15] that for small  $\Delta t$  the transition probability  $W_{\Delta t}(n | n')$  can be written

$$W_{\Delta t}(n | n') = \begin{cases} T(n | n') \Delta t + \mathcal{O}(\Delta t)^2 & \text{if } n \neq n' \\ 1 - \sum_{n \neq n''} T(n | n'') \Delta t + \mathcal{O}(\Delta t)^2 & \text{if } n = n'. \end{cases} \quad (2.12)$$

Essentially this assumes that in a small time interval, the probability of changing state is to first order proportional to the time interval. Meanwhile the probability of not changing state is much larger, so that as the time interval tends to zero there is no probability that the system can change state. The function  $T(n | n')$  is only defined for  $n \neq n'$  and is called the probability transition rate, or alternatively just the transition rate. As Eq. (2.12) suggests,  $T(n | n')$  can be interpreted as the probability per unit time of a transition from state  $n'$  to a state  $n$ . Choosing to define the transition probability in this manner ensures that, at least to order  $\Delta t$ , the transition probability remains appropriately normalised;  $\sum_n W_{\Delta t}(n | n') = 1$ .

Substituting Eq. (2.12) into Eq. (2.11) and rearranging, one finds

$$\frac{P_{1|1}(n, t + \Delta t | n_0, t_0) - P_{1|1}(n, t | n_0, t_0)}{\Delta t} = \quad (2.13)$$

$$\sum_{n \neq n'} [T(n | n') P_{1|1}(n', t | n_0, t_0) - T(n' | n) P_{1|1}(n, t | n_0, t_0)] + \mathcal{O}(\Delta t). \quad (2.14)$$

Letting  $\Delta t \rightarrow 0$ , the master equation is obtained [15],

$$\frac{dP(n, t)}{dt} = \sum_{n' \neq n} [T(n | n') P(n', t) - T(n' | n) P(n, t)], \quad (2.15)$$

where for notational convenience the subscript for the conditional distribution has been suppressed, and an initial condition  $n_0$  at  $t_0$  is implicitly assumed. This notation will be used in the remainder of the thesis. The interpretation of the master equation is intuitively clear: the probability that the system is in state  $n$  increases with the probability that the system moves into it from one of the surrounding states  $n'$ , but decreases with the probability that the system is already in state  $n$  but transitions to another state. The transition rates fully determine the model.

Thus far no discussion on how a stochastic system might be modelled in this formalism has been presented. Let us rectify this with the aid of a specific example; a Moran model with mutation. The only processes in the neutral model that change the state of the system are a birth of type  $A$  followed by a death of type  $B$  or conversely a birth of type  $B$  followed by a death of type  $A$  (see Eq. (1.9)). Adding mutation there are a further two reactions which can occur, mutation from  $B$  to  $A$  and mutation from  $A$  to  $B$ .<sup>1</sup> For simplicity, we will here assume that these are independent from birth/death events.

In order to model this process in the formalism described above, it is assumed that the probability of an event in the population occurring is proportional to the probability of picking the individuals involved in instigating the event randomly from the population. The transition probability rate is then equal to this product multiplied by a rate constant, which controls the rate at which the interactions occur. Denoting the rate of birth/death as  $b$ , the rate of mutation from  $B$  to  $A$  as  $\omega_1$  and the rate of mutation from  $A$  to  $B$  as  $\omega_2$ , the probability transition rates are [32]

$$\begin{aligned} T(n+1|n) &= \overbrace{b \left(\frac{n}{N}\right) \left(\frac{N-n}{N-1}\right)}^{\text{A born/B dies}} + \overbrace{(1-b)\omega_1 \left(\frac{N-n}{N}\right)}^{\text{B mutates to A}}, \\ T(n-1|n) &= \overbrace{b \left(\frac{N-n}{N}\right) \left(\frac{n}{N-1}\right)}^{\text{B born/A dies}} + \overbrace{(1-b)\omega_2 \left(\frac{n}{N}\right)}^{\text{A mutates to B}}, \end{aligned} \quad (2.16)$$

with all other transitions set to zero. Birth/death events are proportional to the product of frequencies of  $A$  and  $B$ , since each event involves both types. Mutation events only involve one type and are therefore linear in the relevant frequencies. Setting  $\omega_1 = \omega_2 = 0$ ,  $b = 1$ , one recovers the neutral Moran model, Eq. (1.9). Substituting Eq. (2.16) into Eq. (2.15), the master equation reads

$$\begin{aligned} \frac{dP(n, t)}{dt} &= b \left[ \frac{n+1}{N} \frac{(N-n-1)}{(N-1)} P(n+1, t) + \frac{n-1}{N} \frac{(N-n+1)}{(N-1)} P(n-1, t) \right. \\ &\quad \left. - \left( 2 \frac{n}{N} \frac{(N-n)}{(N-1)} \right) P(n, t) \right] \end{aligned}$$

---

<sup>1</sup>We choose to add mutation because, as discussed in Chap. 1, the neutral model has no deterministic dynamics. This makes it a poor model with which to illustrate general results.

$$\begin{aligned}
& + (1 - b) \left[ \omega_2 \frac{n+1}{N} P(n+1, t) + \omega_1 \frac{N-n+1}{N} P(n-1, t) \right. \\
& \left. - \left( \omega_2 \frac{n}{N} + \omega_1 \frac{N-n}{N} \right) P(n, t) \right],
\end{aligned} \tag{2.17}$$

which together with the initial frequency of type  $A$  individuals,  $n_0$ , fully describes the time evolution of the PDF.<sup>2</sup> Models constructed in this way are referred to as individual based models (IBMs) since they take account of explicit interactions between individuals.

Despite this simple description, the master equation is very rarely analytically tractable. Though there are methods of making analytic progress for a restricted set of problems (usually with some exploitable symmetry), in general obtaining a solution is a formidable task. The situation is further complicated if the system under consideration is multidimensional. While the derivation of the master equation can be easily extended to a multivariate system [49] described by the state vector  $\mathbf{n}$ ,

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_{\mathbf{n}' \neq \mathbf{n}} [T(\mathbf{n}|\mathbf{n}')P(\mathbf{n}', t) - T(\mathbf{n}'|\mathbf{n})P(\mathbf{n}, t)], \tag{2.18}$$

solutions to the equation become even harder to obtain.

There are ways to make progress however. We will consider two avenues in turn: an analytic approximation (the Fokker-Planck equation) and a method of simulating particular realisations of a process obeying the master equation. For pedagogical reasons, we shall consider the method of simulation first.

## 2.3 The Gillespie Algorithm

A particular realisation of a process obeying the master equation may be simulated using the Gillespie algorithm [17]. A collection of such realisations can then be used to estimate properties of the underlying distribution,  $P(\mathbf{n}, t)$ . While the algorithm itself is not essential to the work in this thesis, it does provide a nice insight into the behaviour of systems obeying the master equation. As such it is useful to review the salient points of the algorithm here.

We begin by describing a system comprised of  $m$  different species, so that the state of the system is given by an  $m$ -dimensional state vector  $\mathbf{n}$ . The system is set to  $\mathbf{n}_0$  at  $t_0$  so that the initial distribution is simply a delta peak  $p(\mathbf{n}, t_0) = \delta(\mathbf{n} - \mathbf{n}_0)$ . We suppose that there are  $u$  different types of reaction which may occur given the system is in this state. In order to make the following discussion more clear, it is useful to introduce the stoichiometric matrix,  $\nu$ . The stoichiometric matrix is an  $m$  by  $u$  matrix which gives a concise way of stating which species were transformed

---

<sup>2</sup>Note that while we could have specified the birth/death events and mutation events by separate transition rates in Eq. (2.16), this would have led to no change in Eq. (2.17).

in a given reaction. Each element  $\nu_{i\mu}$  ( $i = 1 \dots m$ ,  $\mu = 1 \dots u$ ) gives the change in number of the  $i$ th species due to the  $\mu$ th reaction. The  $u$  reactions then take the system from state  $\mathbf{n}$  to  $\mathbf{n}' = \mathbf{n} + \boldsymbol{\nu}_\mu$ , where the vector  $\boldsymbol{\nu}_\mu$  is the  $\mu$ th column of the matrix  $\boldsymbol{\nu}$ . In this notation, for a given state  $\mathbf{n}$  at time  $t$ , we can describe all the transitions by  $T_\mu(\mathbf{n} + \boldsymbol{\nu}_\mu | \mathbf{n})$ . The master equation may be rewritten in this new notation as

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_{\mu=1}^u [T_\mu(\mathbf{n} | \mathbf{n} - \boldsymbol{\nu}_\mu) P(\mathbf{n} - \boldsymbol{\nu}_\mu, t) - T(\mathbf{n} + \boldsymbol{\nu}_\mu | \mathbf{n}) P(\mathbf{n}, t)]. \quad (2.19)$$

In the Moran model with mutation, we simply have  $m = 1$  and  $u = 2$ , while the stoichiometry matrix is

$$\boldsymbol{\nu} = (1, -1) \quad (2.20)$$

since the system is described by a single species variable and the transitions either increase or decrease this number by one.

We wish to simulate a realisation of the stochastic process which obeys the master equation. For each reaction realised, this requires evaluating two stochastic quantities. Not only do we need to know which of the  $u$  reactions occurs, we also need to know when it occurs. In order to do this in a way consistent with the probability distribution evolving according to Eq. (2.19), we first move to considering an equivalent form of the Eq. (2.19) in terms of a new distribution  $P(\tau, \mu | \mathbf{n}, t)$ . This distribution is defined such that  $P(\tau, \mu | \mathbf{n}, t) \Delta\tau$  is the probability that the next reaction is the  $\mu$ th and that it occurs in the time interval  $[t + \tau, t + \tau + \Delta\tau]$ . In other words, it is the probability that no reaction occurs between  $t$  and  $t + \tau$  and that when the reaction occurs it is the  $\mu$ th reaction in the interval  $[t + \tau, t + \tau + \Delta\tau]$ . We now note that the conditional distribution,  $P(\mathbf{n}, \tau | \mathbf{n}, t)$ , is the probability that no reaction has occurred between  $t$  and  $\tau$ . Since, given some state  $\mathbf{n}$ , the probability that the system transitions to state  $\mathbf{n}'$  in time  $\Delta\tau$  is given by the transition probability  $W_{\Delta\tau}(\mathbf{n} + \boldsymbol{\nu}_\mu | \mathbf{n}) \approx T_\mu(\mathbf{n} + \boldsymbol{\nu}_\mu | \mathbf{n}) \Delta\tau$  (see Eq. (2.12)),  $P(\tau, \mu | \mathbf{n}, t) \Delta\tau$  and  $P(\mathbf{n}, \tau | \mathbf{n}, t)$  can be related by

$$P(\tau, \mu | \mathbf{n}, t) \Delta\tau = P(\mathbf{n}, \tau | \mathbf{n}, t) T_\mu(\mathbf{n} + \boldsymbol{\nu}_\mu | \mathbf{n}) \Delta\tau. \quad (2.21)$$

To calculate  $P(\mathbf{n}, \tau | \mathbf{n}, t)$  we return to the master equation, Eq. (2.19), and fix the state of the system to  $\mathbf{n}$ . The resulting equation is

$$\frac{dP(\mathbf{n}, \tau | \mathbf{n}, t)}{d\tau} = -P(\mathbf{n}, \tau | \mathbf{n}, t) \sum_{\mu=1}^u T_\mu(\mathbf{n} + \boldsymbol{\nu}_\mu | \mathbf{n}). \quad (2.22)$$

Given the system is already in state  $\mathbf{n}$ , the master equation can thus be solved for  $P(\mathbf{n}, \tau | \mathbf{n}, t)$  to show

$$P(\mathbf{n}, \tau | \mathbf{n}, t) = \exp \left( - \sum_{\mu=1}^u T_\mu(\mathbf{n} + \boldsymbol{\nu}_\mu | \mathbf{n}) \tau \right); \quad (2.23)$$

the time until the next reaction occurs,  $\tau$ , is exponentially distributed. Substituting this into Eq. (2.21), one obtains

$$P(\tau, \mu | \mathbf{n}, t) \Delta\tau = \exp \left[ - \sum_{\mu=1}^u T_{\mu}(\mathbf{n} + \boldsymbol{\nu}_{\mu} | \mathbf{n}) \tau \right] T_{\mu}(\mathbf{n} + \boldsymbol{\nu}_{\mu} | \mathbf{n}) \Delta\tau. \quad (2.24)$$

Since this can be equivalently expressed

$$P(\tau, \mu | \mathbf{n}, t) = \left( \sum_{\mu=1}^u T_{\mu}(\mathbf{n} + \boldsymbol{\nu}_{\mu} | \mathbf{n}) \exp \left[ - \sum_{\mu=1}^u T_{\mu}(\mathbf{n} + \boldsymbol{\nu}_{\mu} | \mathbf{n}) \tau \right] \right) \left( \frac{T_{\mu}(\mathbf{n} + \boldsymbol{\nu}_{\mu} | \mathbf{n})}{\sum_{\mu=1}^u T_{\mu}(\mathbf{n} + \boldsymbol{\nu}_{\mu} | \mathbf{n})} \right),$$

we see clearly that the time until the next reaction and the specific reaction which takes place are statistically independent. The time until the next reaction and the next reaction itself can therefore be separately specified by two distinct random numbers drawn from the above distributions [16]. This allows us to proceed with an algorithm to evaluate a trajectory for the master equation:

1. Calculate the transition rates based on the initial conditions
2. Use these to determine the distributions of next reaction times and reactions
3. Draw a random time until the next reaction and a type of reaction to occur from these distributions
4. Change the time and state of the system accordingly
5. Recalculate the transition rates based on the new state of the system
6. Repeat.

While this simulation procedure gives a useful interpretation of the evolution of the system at short times, it gives no insight into the statistical behaviour of the system. Further, though one can look at the statistics of many realisations of the Gillespie algorithm to infer the properties of  $P(\mathbf{n}, t)$ , this does not provide as deep an understanding of the system as an analytical treatment. While the Gillespie algorithm will prove useful to compare against analytic results, for a deeper understanding we must resort to solving an approximation of the master equation.

## 2.4 The Expansion of the Master Equation

So far a general formalism for modelling a class of stochastic systems has been presented. It has been stated that the governing equation, the master equation, is in the vast majority of cases very difficult to solve. While simulating the master equation has been shown to be relatively straightforward, we are still no further in making the analytic progress which we initially sought. It will now be shown that the problem can be simplified significantly by using an approximation which resembles the Kramers-Moyal expansion.



The Kramers-Moyal expansion is one of a set of schema which approximate the master equation by assuming a continuous state space. The van-Kampen system size expansion, or linear noise approximation (LNA), is another such scheme [49]. Technically, the approximation used here is not the Kramers-Moyal expansion, which features no explicit small parameter [15]. The approximation presented is instead a hybrid of the two methods, since it follows the methodology of the Kramers-Moyal expansion but using an explicit small parameter determined from the system size.

To illustrate the idea, the procedure will be applied to the specific example of the Moran model with mutation (introduced in Sect. 2.2) before a more general treatment is provided. We begin by taking the master equation (2.17), and introduce a new variable  $x$  such that  $x = n/N$ . The master equation then becomes

$$\begin{aligned} \frac{\partial p(x, t)}{\partial t} = & b \left[ \left( x + \frac{1}{N} \right) \frac{N}{N-1} \left( 1 - x - \frac{1}{N} \right) p(x + 1/N, t) \right. \\ & + \left( x - \frac{1}{N} \right) \frac{N}{N-1} \left( 1 - x + \frac{1}{N} \right) p(x - 1/N, t) \\ & \left. - \left( 2x \frac{N}{N-1} (1-x) \right) p(x, t) \right] \\ & + (1-b) \left[ \omega_2 \left( x + \frac{1}{N} \right) p(x + 1/N, t) \right. \\ & \left. + \omega_1 \left( 1 - x + \frac{1}{N} \right) p(x - 1/N, t) - (\omega_2 x + \omega_1 (1-x)) p(x, t) \right], \end{aligned} \quad (2.25)$$

where we note  $p(x, t)$  is a new continuous distribution. The recurrent factors of  $1/N$  in this equation give us a clue as how to proceed. If  $N$  is large, a Taylor expansion of  $p(x, t)$  about  $x$  can be conducted. Assuming that the mutation rate is small (of order  $N^{-1}$ ) and collecting terms order by order in  $1/N$ , one arrives at a one-dimensional example of the Fokker-Planck equation;

$$\begin{aligned} \frac{\partial p(x, t)}{\partial t} = & -\frac{(1-b)}{N} \frac{\partial}{\partial x} [(\omega_1 - (\omega_1 + \omega_2)x) p(x, t)] \\ & + \frac{b}{2N^2} \frac{\partial^2}{\partial x^2} [2x(1-x)p(x, t)] + \mathcal{O}(N^3). \end{aligned} \quad (2.26)$$

This PDE in two variables,  $x$  and  $t$ , is far more amenable to analysis than the  $N + 1$  partial difference equations comprising the master equation. Its physical interpretation is perhaps most evident when read as a convection-diffusion equation; the term preceding  $p(x, t)$  in the first spatial derivative governs its ‘bulk advection’, while that in the second derivative describes diffusion. For this reason they are referred to

as the drift and diffusion terms respectively.<sup>3</sup> We can also write a continuity equation for this system by introducing the probability flux or probability current,  $J(x, t)$ ;

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial J(x, t)}{\partial x}. \quad (2.27)$$

Say we want to calculate how the mean value of  $x$ ,  $\langle x \rangle = \int x p(x, t) dx$ , evolves in time. We can derive an equation for its time-evolution by multiplying Eq. (2.26) by  $x$  and integrating over all  $x$ . The expression can be further simplified by noting that  $J(x, t)$  must be zero at the reflecting boundaries  $x = 0$  and  $x = 1$  (we will discuss meaning of this term and other boundary conditions in Sect. 2.5) and that in this particular model the diffusion term is also zero at the boundaries. Letting  $\tau = t/N$ , the resulting equation is then

$$\frac{d\langle x \rangle}{d\tau} = (1 - b) [\omega_1 - (\omega_1 + \omega_2)\langle x \rangle]. \quad (2.28)$$

The time-evolution of the mean is governed entirely by the drift term. This behaviour can also be obtained by rescaling time in Eq. (2.26) such that  $\tau = t/N$ , taking the limit  $N \rightarrow \infty$  and noting that  $p(x', t) = \delta(x(\tau), x')$  is a solution of the resulting equation. We will call this the macroscopic behaviour of the system, or alternatively the deterministic limit. Note that for the neutral model, with  $\omega_1 = \omega_2 = 0$ , there are no deterministic dynamics, as stated in Eq. (1.10).

The idea is generalised as follows. Let us postulate that there is some large parameter  $N$  which is both inversely proportional to the reaction rates and some measure of the typical size of the system. In the case of the Moran model this is specifically the size of the system, though in general it could be the typical size (or volume) of the system which governs the interaction rate, or even the inverse of a naturally small reaction rate. A new set of variables  $\mathbf{x} = \mathbf{n}/N$  is introduced. If  $N$  is a measure of system size, the new variables can be naturally interpreted as some measure of the concentration of each species in the population.

Introducing the functions  $f_\mu(\mathbf{x}) = T_\mu(N\mathbf{x} + \boldsymbol{\nu}_\mu | N\mathbf{x})$ , the master equation Eq. (2.19) can be reexpressed

$$\frac{dp(\mathbf{x}, t)}{dt} = \sum_{\mu=1}^u \left[ f_\mu(\mathbf{x} - \boldsymbol{\nu}_\mu N^{-1}) p(\mathbf{x} - \boldsymbol{\nu}_\mu N^{-1}, t) - f_\mu(\mathbf{x}) p(\mathbf{x}, t) \right], \quad (2.29)$$

where we have again changed from a distribution  $P(\mathbf{n}, t)$  to  $p(\mathbf{x}, t)$ . In this form it is clear that if  $N$  is large, one may proceed in the same way as in the Moran model

---

<sup>3</sup>An unfortunate clash of nomenclature appears here between the physics and biology communities. In population genetics, genetic drift is the process by which the composition of a population is changed by noise. Drift in the context of population genetics therefore refers to the noisy component of a system's behaviour, rather than the deterministic component, as in physics.

example by assuming  $\mathbf{x}$  is continuous and implementing a Taylor expansion about  $\mathbf{x}$ . This is the essence of the master equation expansion.

While the expansion contains results in decreasing orders of  $N$ , it can be shown that the resulting expression only leads to a strictly positive PDF if truncated at order  $N^{-2}$ , or alternatively keeping the infinite series of terms. This is the Pawula theorem [38]. Since in going to infinite precision the expression is as complicated as the master equation, we truncate the expansion at  $\mathcal{O}(N^{-2})$  to achieve our desired simplification. The resulting equation is a multivariate FPE [38] which takes the form

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = -\frac{1}{N} \sum_{i=1}^m \frac{\partial}{\partial x_i} [A_i(\mathbf{x}) p(\mathbf{x}, t)] + \frac{1}{2N^2} \sum_{i,j=1}^m \frac{\partial^2}{\partial x_i \partial x_j} [B_{ij}(\mathbf{x}) p(\mathbf{x}, t)], \quad (2.30)$$

where  $A(\mathbf{x})$  is now the drift vector, while  $B(\mathbf{x})$  is the diffusion matrix. Their forms are governed by the stoichiometry matrix and the transition rates [33] such that

$$A_i(\mathbf{x}) = \lim_{N \rightarrow \infty} \sum_{\mu=1}^u \nu_{i\mu} f_{\mu}(\mathbf{x}) \quad (2.31)$$

and

$$B_{ij}(\mathbf{x}) = \lim_{N \rightarrow \infty} \sum_{\mu=1}^u \nu_{i\mu} \nu_{j\mu} f_{\mu}(\mathbf{x}). \quad (2.32)$$

Once again, rescaling time such that  $\tau = t/N$  and taking the limit  $N \rightarrow \infty$ , the resulting equation admits  $P(\mathbf{x}', t) = \delta(\mathbf{x}(\tau), \mathbf{x}')$  as a solution such that

$$\frac{d\mathbf{x}}{d\tau} = A(\mathbf{x}), \quad (2.33)$$

describes the deterministic, macroscopic dynamics. We note that it is also possible to obtain an analogous macroscopic description directly from the master equation (2.29) in terms of the mean value  $\langle \mathbf{x} \rangle$ . However, for a general nonlinear system, this requires making the further assumption that the decomposition  $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$  can be made in order to close the resulting equations. This is only generally true if the distributions for  $x_i$  are delta functions, or as we have seen to be equivalent, if the system size  $N$  is infinite.

Since it is possible to calculate  $A(\mathbf{x})$  and  $B(\mathbf{x})$  entirely from the stoichiometry matrix and reaction rates, models are sometimes expressed in the notation of chemical reactions, in terms of reactants and products (in fact the term ‘stoichiometric matrix’ is borrowed from chemistry). For an arbitrary  $m$ -dimensional IBM, whose dynamics are fully described by a set of  $u$  reaction rates, the model can be expressed in chemical reaction notation as

$$\sum_{i=1}^m a_{\mu i} X_i \xrightarrow{r_{\mu}} \sum_{i=1}^m b_{\mu i} X_i, \quad \forall \mu = 1, \dots, u, \quad (2.34)$$

where  $a_{\mu i}$  and  $b_{\mu i}$  respectively specify the reactants and products of the  $\mu$ th reaction, and  $r_{\mu}$  are the reaction rate constants. The elements of the stoichiometric matrix are then given by  $\nu_{i\mu} = b_{\mu i} - a_{\mu i}$ , while the reaction rate constants  $r_{\mu}$  are related to the transition rates by an equation of the form

$$T_{\mu}(\mathbf{n}'|\mathbf{n}) \propto r_{\mu} \prod_{i=1}^m a_{\mu i} \frac{n_i}{N}. \quad (2.35)$$

Implicit in this notation is the assumption that the probability of a reaction occurring is proportional to the product of the reactant concentrations. This is also known as the law of mass action. Most often this is the case, however situations may arise in which we wish to incorporate additional state dependence. An example of such a system will be given in Sect. 2.9.1. For clarity we will not describe such reactions in this notation. Additionally we have introduced a measure of system size  $N$ . The determination of what this parameter should be is highly dependent on the system under consideration.

In expanding the master equation, we have moved from the microscopic description of an IBM involving as many equations as there are states (the master equation) to one in which there are only as many variables as there are species (the FPE). The crux of the approximation is in changing to a new set of variables  $\mathbf{x} = \mathbf{n}/N$  which are approximately continuous (the diffusion approximation) before applying a Taylor expansion to the master equation and neglecting terms of order  $N^{-3}$ . When modelling populations, the parameter  $N$  is usually identified as the size of the population. In the limit  $N \rightarrow \infty$ , we have seen that the FPE describes a deterministic dynamic which we have termed the macroscopic dynamic. In this spirit, since the FPE lies between the master equation and deterministic equation in terms of detail, it is often referred to as a mesoscopic description.

The Fokker-Planck equation is clearly more amenable to analysis than the master equation with which we began. Of particular interest is the one-dimensional Fokker-Planck equation, from which many properties of interest can be calculated analytically.

## 2.5 The Fokker-Planck Equation and Some Useful Manipulations

In this section we will discuss the way in which the one-dimensional FPE can be used to obtain three particular quantities of interest; the stationary probability distribution, the first passage probability and the first passage time. However, let us first discuss some particular boundary conditions of the one-dimensional FPE which will come in useful later.

We begin by imagining the system existing on an interval  $[a_1, a_2]$  in state space, from which it cannot leave. The FPE (2.26) describes such a system, with  $x$  lying on the interval  $[0, 1]$ . In this case there should be zero flow of probability across the boundary. Therefore the probability current must be zero when evaluated at the boundaries;  $J(a_1, t) = J(a_2, t) = 0$ . Such boundaries are called reflecting.

Now let us consider a system in which there exist states where there are no dynamics. Such states are called absorbing. In order to account for this, we define the barriers as existing outside of the interval, so that when the system reaches the boundary it is removed. The probability of being at either of the boundaries is then zero,  $p(a_1, t) = p(a_2, t) = 0$ . If the barriers  $a_1$  and  $a_2$  of the system are absorbing, it is clear that given an infinite amount of time, the probability of the system remaining on the interval will tend to zero as probability ‘leaks out’ of the interval. An example of such a system is the neutral Moran model with transition rates (1.9) discussed in the introduction. Finally, if the boundaries are at infinity, we expect  $p(x, t)$ , along with  $J(x, t)$ , to vanish at  $x = \pm\infty$ .

### 2.5.1 The Stationary Probability Distribution

First consider a system with two reflecting boundaries at  $a_1$  and  $a_2$ . In the limit of long times one might expect the PDF  $p(x, t)$  in Eq. (2.30) to become independent of time. We call this distribution the stationary distribution, and it is defined by

$$p_{\text{st}}(x) = \lim_{t \rightarrow \infty} p(x, t). \quad (2.36)$$

Since  $p_{\text{st}}(x)$  is independent of time, we find it must satisfy the equation [15]

$$-\frac{1}{N} \frac{d}{dx} [A(x)p_{\text{st}}(x)] + \frac{1}{2N^2} \frac{d^2}{dx^2} [B(x)p_{\text{st}}(x)] = 0. \quad (2.37)$$

The solution to this equation is

$$p_{\text{st}}(x) = \exp \left( \int_{a_1}^x dy \frac{2NA(y) - dB(y)/dy}{B(y)} \right) \left[ \int_{a_1}^{a_2} dx \exp \left( \int_{a_1}^x dy \frac{2NA(y) - dB(y)/dy}{B(y)} \right) \right]^{-1},$$

or

$$p_{\text{st}}(x) = \frac{1}{B(x)} \exp \left( 2N \int_{a_1}^x dy \frac{A(y)}{B(y)} \right) \left[ \int_{a_1}^{a_2} dx \frac{1}{B(x)} \exp \left( 2N \int_{a_1}^x dy \frac{A(y)}{B(y)} \right) \right]^{-1} \quad (2.38)$$

where we have appropriately normalised so that  $\int_{a_1}^{a_2} dx p_{\text{st}}(x) = 1$ . This can yield interesting results about the system’s long time behaviour. For instance, we can look at the FPE for the Moran model with mutation, Eq. (2.26), in which there are no

absorbing states for both  $\omega_1$  and  $\omega_2$  greater than zero and  $b$  less than one. Comparing Eq. (2.26) with Eq. (2.30), we see for this system that

$$A(x) = (1 - b)(\omega_1 - (\omega_1 + \omega_2)x), \quad B(x) = 2bx(1 - x). \quad (2.39)$$

Substituting these terms into Eq. (2.38), and noting that  $x$  lies on the interval  $[0, 1]$ , the stationary distribution takes the form

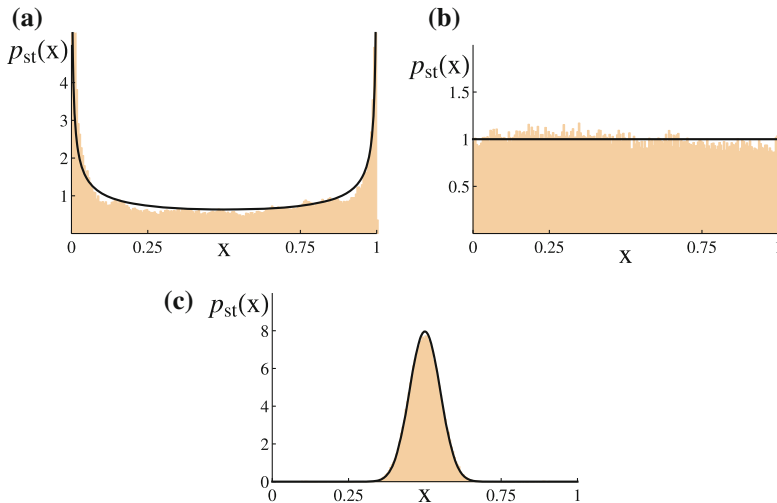
$$p_{\text{st}}(x) = \frac{x^c(1 - x)^d}{\int_0^1 dx x^c(1 - x)^d}, \quad (2.40)$$

with

$$c = N\omega_1(b^{-1} - 1) - 1, \quad d = N\omega_2(b^{-1} - 1) - 1.$$

Some of the different behaviour the system can exhibit is demonstrated in Fig. 2.1.

If the system instead has absorbing boundaries, there is no stationary distribution on the interval  $a_1 < x < a_2$ , since all the stochastic trajectories eventually leave the interval. In such cases, there are other measures of the system's behaviour which are more illuminating.



**Fig. 2.1** Different regimes of Eq. (2.40) with different  $c$  and  $d$  parameters. The histograms are obtained from Gillespie simulations using the model defined by Eq. (2.16). **a** Parameters  $c < 0, d < 0$ . **b** Parameters  $c = 0, d = 0$ . **c** Parameters  $c > 0, d > 0$

### 2.5.2 First Passage Problems

First passage problems are those which ask ‘what is the probability the system reaches a particular final condition?’, or ‘what are the statistics of the time for the system to reach this final condition?’. In order to calculate these first passage properties, we find it useful to work with the backward Fokker-Planck equation (BFPE). This can be expressed in one dimension as [15, 38]

$$-\frac{\partial q(x, t|x_0, t_0)}{\partial t_0} = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} [q(x, t|x_0, t_0)] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} [q(x, t|x_0, t_0)]. \quad (2.41)$$

If the process under consideration is homogeneous, the evolution of the system depends only on the difference between the initial and final time  $t - t_0$ , and the BFPE can be rewritten in terms of the derivative with respect to  $t$ ;

$$\frac{\partial q(x, t|x_0, t_0)}{\partial t} = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} [q(x, t|x_0, t_0)] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} [q(x, t|x_0, t_0)]. \quad (2.42)$$

The key difference between the forward FPE (2.30) and the BFPE, is which set of variables are kept fixed and which vary. In the forward FPE the initial conditions  $x_0$  at  $t_0$  are kept fixed, and one finds for solutions for  $t \geq t_0$ . In the BFPE we keep the final condition  $x$  at  $t$  fixed and calculate for solutions with  $t_0 \leq t$ . Since in the BFPE we fix the final condition, it is clearly more useful when dealing with first passage problems. For simplicity, we restrict ourselves to a one-dimensional homogeneous system.

We wish to know the time until a system first escapes the region between two points,  $x = a_1$  and  $x = a_2$ , given some initial condition  $a_2 > x_0 > a_1$ . This time is clearly a stochastic variable and so it will be described by a PDF indicating the probability of a certain first passage time  $t$  given initial condition  $x_0$ . It is denoted here by  $\mathcal{T}(x_0, t)$ . We begin by defining the probability  $G(x_0, t)$  that, at some time  $t$ , the system is still on the interval;

$$G(x_0, t) = \int_{a_1}^{a_2} dx q(x, t|x_0, t_0), \quad (2.43)$$

where the dependence on the initial time has been suppressed. Integrating Eq. (2.42) over  $x$  between  $a_1$  and  $a_2$ , we find in fact the equation for  $G(x_0, t)$  obeys the same BFPE as  $q(x, t|x_0, 0)$ ;

$$\frac{\partial G(x_0, t)}{\partial t} = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} [G(x_0, t)] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} [G(x_0, t)], \quad (2.44)$$

with initial condition

$$G(x_0, t_0) = 1 \quad \text{if } a_1 < x_0 < a_2, \quad (2.45)$$

$$G(x_0, t_0) = 0 \quad \text{elsewhere,} \quad (2.46)$$

and boundary conditions

$$G(a_1, t) = G(a_2, t) = 0. \quad (2.47)$$

Since  $G(x_0, t)$  is the probability that at time  $t$  the system is still on the interval  $a_1 < x < a_2$ , the quantity  $G(x_0, t) - G(x_0, t + \Delta t)$  is the probability that the system has reached one of the boundaries during  $t$  to  $t + \Delta t$ . This can be related to  $\mathcal{T}(x_0, t)$  quite simply, since

$$\mathcal{T}(x_0, t)\Delta t = G(x_0, t) - G(x_0, t + \Delta t), \quad (2.48)$$

or, rearranging and sending  $\Delta t \rightarrow 0$ ,

$$\mathcal{T}(x_0, t) = -\frac{\partial G(x_0, t)}{\partial t}. \quad (2.49)$$

The mean time for the system to leave the interval, denoted  $T(x_0)$ , can be calculated directly from the distribution  $\mathcal{T}(x_0, t)$  by  $T(x_0) = \int_{t_0}^{\infty} t \mathcal{T}(x_0, t) dt$ . Using the above equality, this can be expressed

$$T(x_0) = -\int_{t_0}^{\infty} t \frac{\partial G(x_0, t)}{\partial t} dt. \quad (2.50)$$

This equation can be integrated by parts; letting  $t_0 = 0$  and assuming that  $tG(x_0, t)$  tends to zero as  $t \rightarrow \infty$ , one arrives at

$$T(x_0) = \int_0^{\infty} G(x_0, t) dt. \quad (2.51)$$

Integrating Eq. (2.44) over time and noting that  $G(x_0, t)$  tends to zero as  $t \rightarrow \infty$ , we then arrive at an equation for the mean time to reach either of the boundaries

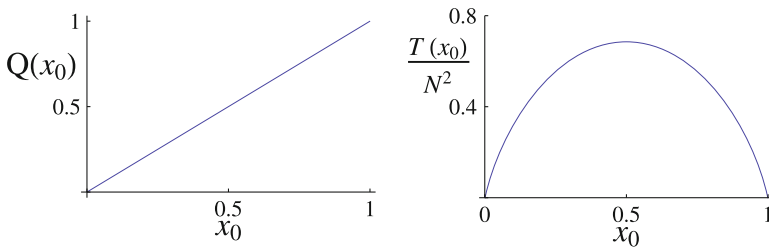
$$-1 = \frac{A(x_0)}{N} \frac{d}{dx_0} T(x_0) + \frac{B(x_0)}{2N^2} \frac{d^2}{dx_0^2} T(x_0) \quad (2.52)$$

with the boundary conditions,

$$T(0) = 0, \quad T(1) = 0. \quad (2.53)$$

In this case the boundary conditions follow by noting that they are the time to reach *either*  $x = a_1$  or  $x = a_2$ , so that at both extremes the system has already fixed on the boundaries.





**Fig. 2.2** Graphical summary of results from the neutral Moran model. *Left panel* probability of fixation of type A,  $Q(x_0)$ , in a neutral system as a function of initial A type concentration,  $x_0$ . *Right panel* time to fixation of either A or B type,  $T(x_0)$ , scaled by the system size squared,  $N^2$ , as a function of the initial concentration of type A

Let us go back to the neutral Moran model described by Eq. (1.9). In this case the points  $x = 0$  and  $x = 1$  are absorbing boundaries from which the system cannot leave. At these points the population is said to have fixated. Using Eq. (2.52) and Eq. (2.53), with  $A(x) = 0$  and  $B(x)$  taken from Eq. (2.39), we can calculate the time this would take to happen. One finds [10]

$$T(x_0) = -N^2 [(1 - x_0) \ln(1 - x_0) + x_0 \ln(x_0)], \quad (2.54)$$

which is plotted in the right panel of Fig. 2.2. In the nomenclature of population genetics this is called the mean unconditional fixation time.

What about the probability that the system hits one of these boundaries before the other? Let us return to considering a function of the same form as  $G(x_0, t)$  in Eq. (2.43), but this time introduce two slightly different functions to account for different integration limits;

$$G_{a_1}(x_0, t) = 1 - \int_{a_1}^{\infty} dx q(x, t|x_0, t_0) = \int_{-\infty}^{a_1} dx q(x, t|x_0, t_0), \quad (2.55)$$

$$G_{a_2}(x_0, t) = \int_{a_2}^{\infty} dx q(x, t|x_0, t_0) = 1 - \int_{-\infty}^{a_2} dx q(x, t|x_0, t_0). \quad (2.56)$$

The first of these functions gives the probability that, at time  $t$ , the system is at some point  $x < a_1$  and the second that it is at some point  $x > a_2$ . Of course, this gives us no information about which of these regions the system ended up in *first*. To do this would require a consideration of the trajectories conditioned such that the time to one boundary was less than the time to the other. Our task is significantly simplified however if we force the boundaries  $a_1$  and  $a_2$  to be absorbing. Then once the system hits state  $a_1$  or  $a_2$  it is immediately removed and we do not have to worry about time-ordering. The functions  $G_{a_1}(x_0, t)$  and  $G_{a_2}(x_0, t)$  then tell us respectively whether at time  $t$  the system has hit either  $a_1$  or  $a_2$ . Both functions obey the equations

$$\frac{\partial G_{a_1/a_2}(x_0, t)}{\partial t_0} = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} [G_{a_1/a_2}(x_0, t)] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} [G_{a_1/a_2}(x_0, t)], \quad (2.57)$$

but with different boundary conditions. As  $t \rightarrow \infty$ , the probability of having hit either  $a_1$  or  $a_2$  tends to one. Introducing

$$Q_{a_1/a_2}(x_0) = \lim_{t \rightarrow \infty} G_{a_1/a_2}(x_0, t), \quad (2.58)$$

we see the equation for both functions is

$$0 = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} [Q_{a_1/a_2}(x_0)] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} [Q_{a_1/a_2}(x_0)], \quad (2.59)$$

albeit with different boundary conditions. For  $Q_{a_1}(x_0)$  we have

$$Q_{a_1}(a_1) = 1, \quad Q_{a_1}(a_2) = 0, \quad (2.60)$$

and for  $Q_{a_2}(x_0)$  instead

$$Q_{a_2}(a_1) = 0, \quad Q_{a_2}(a_2) = 1. \quad (2.61)$$

Once again the neutral Moran model, Eq. (1.9), may be used to illustrate the method. We ask the question, what is the probability of the system reaching the point  $x = 1$  given some initial condition  $x_0$ ? Since at  $x = 1$  the system is composed entirely of the  $A$  type individuals, this is called the fixation probability. In the neutral case  $A(x_0) = 0$  and therefore we obtain

$$Q_1(x_0) = x_0, \quad Q_0(x_0) = 1 - x_0. \quad (2.62)$$

The probability of either type fixating is thus simply proportional to their respective initial frequencies in the neutral model (see Fig. 2.2). We note that in population genetics it is most common to simply discuss the probability of fixation of the  $A$  type and for simplicity we will often write this probability  $Q(x_0) \equiv Q_1(x_0)$ .

## 2.6 Stochastic Differential Equations

Stochastic differential equations (SDEs) are perhaps the earliest way in which dynamical stochastic processes were formalised. Whereas the FPE is a deterministic description of the time evolution of a PDF, SDEs are differential equations describing the evolution of a stochastic variable. Say we wished to postulate the form of a general

nonlinear SDE describing a Markov process. Motivated purely by physical reasoning and a desire for simplicity, we choose to express it

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(\mathbf{x}) + g(\mathbf{x})\zeta(t), \quad (2.63)$$

where  $g(\mathbf{x})$  is a matrix. The vector  $\mathbf{A}(\mathbf{x})$  drives the deterministic dynamics, while the  $\zeta(t)$  terms are noise terms arising from some stochastic source. We choose the noise to have zero mean and be delta-correlated in time, with a strength  $g(\mathbf{x})$  which may be proportional to the state of the system:

$$\langle \zeta_i(t) \rangle = 0, \quad (2.64)$$

$$\langle \zeta_i(t)\zeta_j(t') \rangle = \delta_{ij}\delta(t - t'). \quad (2.65)$$

Now let us fix the entire distribution of  $\zeta(t)$  by stating that the noise is Gaussian. Thus each odd moment is zero and each even moment may be calculated from Eq. (2.65). If the noise strength  $g(\mathbf{x})$  is state dependent the noise is said to be multiplicative [49]. Otherwise it is said to be additive. The choice that the noise is delta correlated in time is a reflection of our wish to model a Markov process. We could obviously have chosen the noise to have a non-zero mean, however the result of this would be to simply add some effective deterministic drift to the system, obfuscating matters.

The SDE perspective offers a much more physically intuitive interpretation of stochastic processes than the Fokker-Planck equation; the system evolves in time according to a deterministic contribution  $\mathbf{A}(\mathbf{x})$  plus some noise  $\zeta(t)$ . Ultimately however, it provides an inferior starting point for modelling systems with demographic noise. This inferiority is twofold. First, it is a mesoscopic description which has been arrived at in an ad-hoc manner. While the form of the deterministic contributions could conceivably be derived from the interactions in the system, it is unclear what form the noise should take. Second, the interpretation of Eq. (2.63) is ambiguous. To see this, it is perhaps best to ask the question, how would one integrate Eq. (2.63)? The problem here is essentially where the term  $g(\mathbf{x})$  should be evaluated, since  $\mathbf{x}$  is undefined at the time when a delta function ‘kick’ arrives. The two interpretations that are most common are that of Stratanovich and Itô. Respectively they are [38]

$$\mathbf{x}(t + \Delta t) - \mathbf{x}(t) = \mathbf{A}(\mathbf{x}(t))\Delta t + g\left(\frac{\mathbf{x}(t) + \mathbf{x}(t + \Delta t)}{2}\right) \int_t^{t+\Delta t} \zeta(t')dt' \quad (\text{Stratanovich}),$$

$$\mathbf{x}(t + \Delta t) - \mathbf{x}(t) = \mathbf{A}(\mathbf{x}(t))\Delta t + g(\mathbf{x}(t)) \int_t^{t+\Delta t} \zeta(t')dt' \quad (\text{Itô}).$$

Importantly, not only do these two interpretations describe different stochastic processes, they also each lead to different rules of stochastic calculus. While it can be shown that Eq. (2.63) is equivalent to an FPE, these different rules of calculus lead to different forms of the FPE. If one were to use the SDEs as a starting point in the modelling process, one would not only have to postulate the form of the matrix  $g(\mathbf{x})$ , but also the interpretation. If, conversely, one begins with an FPE, the associated SDEs do not suffer from this ambiguity. The FPE is a well defined PDE which

is equivalent to one set of SDEs in the Stratanovich scheme and another in the Itô scheme. Given an FPE, which of the two schemes we work in is thus a matter of choice. We choose to work in the Itô setting, and therefore ignore the Stratanovich interpretation for the rest of the thesis. In the Itô setting, different rules of calculus apply when conducting a non-linear change of stochastic variables, Itô calculus [15]. However, since only linear transformations are utilised in this thesis (for which the rules of calculus are unchanged) the details of Itô calculus will not be discussed here.

Defining the matrix

$$B(\mathbf{x}) = Ng(\mathbf{x})g^T(\mathbf{x}), \quad (2.66)$$

the SDE (2.63) can be rewritten as

$$\frac{d\mathbf{x}}{d\tau} = \mathbf{A}(\mathbf{x}) + \frac{1}{\sqrt{N}}\boldsymbol{\eta}(\tau), \quad (2.67)$$

with the Gaussian white noise now characterised by

$$\langle \eta_i(\tau) \rangle = 0, \quad (2.68)$$

$$\langle \eta_i(\tau)\eta_j(\tau') \rangle = B_{ij}(\mathbf{x})\delta(\tau - \tau'). \quad (2.69)$$

Interpreting Eq. (2.67) in the Itô sense, it can be shown to be entirely equivalent to the FPE (2.30) with the same drift vector  $\mathbf{A}(\mathbf{x})$  and diffusion matrix  $B(\mathbf{x})$ , and time scaled such that  $\tau = t/N$ . As a result of this alternate form, the diffusion matrix is sometimes also referred to as the noise covariance matrix or noise correlation matrix. In this form some insights about the FPE become very clear. Firstly, in the  $N \rightarrow \infty$  limit, the system follows a deterministic trajectory governed by  $\mathbf{A}(\mathbf{x})$ . Secondly, the size of the noise is governed by  $N^{-1/2}$ . Increasing the system size decreases the relative magnitude of the noise.

We have here shown once again that in the limit of infinite system size the system follows a deterministic trajectory. Further, this trajectory is also the mean path of all stochastic trajectories for large  $N$ . Both these ideas are very intuitive when posed in the SDE formalism. It is for this reason that we will often choose to work in this setting. Since it is clear that the deterministic element of the equation is of great importance in governing the dynamics, we will now review some basic properties of dynamical systems theory, as well as their extensions to certain stochastic problems.

## 2.7 Dynamical Systems Theory

In the deterministic limit, the systems under consideration in this thesis will take the form of a set of ODEs

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(\mathbf{x}). \quad (2.70)$$

We wish to understand the dynamics of such a system. In general, a solution to Eq. (2.70) can only be obtained if the system is linear. In this case the system may be described by

$$\frac{d\mathbf{x}}{dt} = H\mathbf{x}, \quad (2.71)$$

which has solutions

$$\mathbf{x}(t) = \sum_{i=1}^m c_i \mathbf{v}^{(i)} e^{\lambda^{(i)} t}, \quad (2.72)$$

where  $c_i$  are a set of constants determined from the initial conditions of the problem,  $\mathbf{v}^{(i)}$  are the right-eigenvectors of  $H$  and  $\lambda^{(i)}$  their associated eigenvalues. If all the eigenvalues have a negative real part the system will collapse onto the point  $x_i = 0$  as  $t \rightarrow \infty$ . If one of the eigenvalues is positive,  $\mathbf{x}$  will grow exponentially in the direction described by the associated right-eigenvector. If the largest eigenvalue of the system is zero, say  $\lambda^{(1)} = 0$ , the final state of the system will be dependent on the initial conditions via the constant  $c_1$  [46].

If the system under consideration is non-linear, while one may be able to solve the equations in some special instances, it is more than likely we will not be able to make analytic progress. This is particularly true for non-linear systems in many dimensions. We can however answer questions about the potential long-time behaviour of the system.

The fixed points are points at which there are no dynamics. They shall be denoted  $\mathbf{x}^*$  and are the solutions to the equations

$$\mathbf{A}(\mathbf{x}^*) = 0. \quad (2.73)$$

Crucially we want to know what happens to perturbations around these points; do they shrink or do they grow? In other words are the fixed points stable or unstable? To do this we conduct a Taylor expansion in a variable  $\boldsymbol{\xi}$  which perturbs  $\mathbf{x}$  by a small amount  $\epsilon$  about the fixed point such that  $\mathbf{x} = \mathbf{x}^* + \epsilon\boldsymbol{\xi}$ . Conducting the expansion and neglecting nonlinear terms in  $\epsilon$ , we find

$$\dot{\boldsymbol{\xi}} = J\boldsymbol{\xi}, \quad (2.74)$$

where  $J$  is the Jacobian of  $\mathbf{A}$  evaluated on the fixed point  $\mathbf{x}^*$  about which we linearise;  $J_{ij} = (dA_i/dx_j)|_{\mathbf{x}=\mathbf{x}^*}$ . Since Eq. (2.74) has the same form as Eq. (2.71), the solution for  $\boldsymbol{\xi}(t)$  is analogous to Eq. (2.72) and determining the stability of  $\mathbf{x}^*$  therefore comes down to determining the eigenvalues of  $J$ . If all eigenvalues of  $J$  are less than zero, perturbations away from the fixed point will shrink back to the fixed point which is stable. If any eigenvalues are greater than zero, any perturbation from the fixed point will blow up, and so the fixed point is unstable.

If the system contains noise (such as Eq. (2.67)) then we can still linearise about a fixed point. Setting  $\epsilon = 1/\sqrt{N}$ , the typical fluctuation size, one obtains

$$\dot{\xi} = J\xi + \zeta(t), \quad (2.75)$$

where

$$\langle \zeta(t) \rangle = 0 \quad \left\langle \zeta_i(t) \zeta_j(t') \right\rangle = \delta(t - t') B_{ij}(\mathbf{x}^*). \quad (2.76)$$

This is a linear system with additive noise, since to linear order  $B(\mathbf{x})$  is evaluated at the fixed point. If the fixed point is stable this provides a good approximation for the fluctuations of the system (2.67) in the region of the fixed point.<sup>4</sup> The expression is made even more useful under the assumption that the boundary conditions for this problem lie at  $\pm\infty$ . In this case the associated linear FPE can be shown to describe a Gaussian distribution [30], and thus the entire PDF of the system can be characterised by its first two moments. The equations for these moments are

$$\frac{d}{dt} \langle \xi_k \rangle = \sum_{j=1}^m J_{kj} \langle \xi_j \rangle \quad (2.77)$$

and

$$\frac{d}{dt} \langle \xi_k \xi_l \rangle = \sum_{j=1}^m J_{kj} \langle \xi_j \xi_l \rangle + \sum_{j=1}^m J_{lj} \langle \xi_j \xi_k \rangle + B_{kl}(\mathbf{x}^*). \quad (2.78)$$

The first moment has a solution of the same form as Eq. (2.72). Rather than solve the equation for the second moment, it is more convenient to combine Eqs. (2.77) and (2.78) to give an equation of the covariance matrix  $\Xi$  of the PDF [49];

$$\dot{\Xi} = J\Xi + \Xi J^T + B(\mathbf{x}^*). \quad (2.79)$$

While this equation for  $\Xi$  is the same as that for the second moments, Eq. (2.78), its initial conditions are particularly simple. Given a certain initial condition  $x_{i0}$ , (a delta peak  $p(\mathbf{x}, 0) = \prod_{i=1}^m \delta(x_i - x_{i0})$  in the associated linear FPE),  $\Xi$  vanishes at  $t = 0$ . The solution to Eq. (2.79) can then be shown to be [49]

$$\Xi(t) = \int_0^t e^{(t-t')J} B e^{(t-t')J^T} dt'. \quad (2.80)$$

---

<sup>4</sup>Of course we could have conducted this linearisation in the FPE setting if we desired. We note that if the SDE (2.67) had been derived from a master equation (via the expansion detailed in Sect. 2.4), the resulting linearised FPE is the same as that which would be obtained if the LNA had been applied directly to the master equation.

If all the eigenvalues of  $J$  are negative, the system eventually relaxes to a stationary distribution with zero mean (see Eqs. (2.72) and (2.77)) and a covariance matrix which solves the equation

$$J \Xi + \Xi J + B(\mathbf{x}^*) = 0. \quad (2.81)$$

More generally this equation is known as the Lyapunov equation [22].

## 2.8 Fast-Variable Elimination: The Origins of a Complicated Problem

In Chap. 1, a simple deterministic system featuring a separation of timescales was presented, along with a method of solution. Such a solution goes by many names such as adiabatic elimination, a quasi-steady state approximation or simply fast-variable elimination [19]. In this thesis the term fast-variable elimination will be used. Let us elaborate on the example discussed in the introduction slightly. We may describe a generalisation of a system such as (1.11), in  $m$  variables, as [19]

$$\begin{aligned} \frac{dx_i}{dt} &= f_i(\mathbf{x}, \mathbf{y}), \quad i = 1 \dots r \\ \epsilon_j \frac{dy_j}{dt} &= -y_j + \epsilon_j h_j(\mathbf{x}, \mathbf{y}), \quad j = 1 \dots m - r. \end{aligned} \quad (2.82)$$

If the parameters  $\epsilon_j$  are small and the functions  $f(\mathbf{x}, \mathbf{y})$  and  $h(\mathbf{x}, \mathbf{y})$  are of the same order, we can see that the dynamics in  $\mathbf{y}$  act on a much faster timescale than those in  $\mathbf{x}$ . Now, let us assume that a trajectory of the system in the directions  $y_i$  tends to some finite value. This is somewhat implied by the form of the equations we have chosen. The system reaches this value much faster than the dynamics in the  $x_i$  directions have time to act. We make the approximation that the system reaches its static  $\mathbf{y}$  value instantaneously by setting  $d\mathbf{y}/dt = 0$ . The quasi-steady state value of this system,  $\mathbf{y}_{\text{st}}(\mathbf{x})$ , is termed the slow manifold, and is the solution to the equations

$$-\left[\mathbf{y}_{\text{st}}\right]_j + \epsilon_j h_j(\mathbf{x}, \mathbf{y}_{\text{st}}) = 0. \quad (2.83)$$

The reduced system in  $r$  variables is then given by

$$\frac{dx_i}{dt} = f_i(\mathbf{x}, \mathbf{y}_{\text{st}}), \quad i = 1 \dots r. \quad (2.84)$$

In the course of this thesis, the term ‘slow subspace’ will be used to refer to a subspace in which the deterministic system moves at a slower rate than in the other subspaces. The term centre manifold will be used to refer to a subspace upon which there are strictly no deterministic dynamics. The introduction of the term ‘slow subspace’ is

motivated by the fact that within the mathematical literature the slow manifold has a much more strict definition<sup>5</sup> [1]. The field of deterministic dimensional reduction is a vast one however, and an attempt will not be made to fully adhere to the rigorous definitions detailed in such work. Instead, physical intuition will be used to help guide the analysis and interpretation.

Much of the original work that follows this chapter is concerned with how to appropriately remove fast degrees of freedom from stochastic systems. In the introduction it was merely stated that the generalisation to stochastic systems was not straightforward. Now, with the appropriate mathematical tools and theory in hand, we will briefly see why this is the case, before an outline of some of the existing theory in this area is given.

Since the formulation of the problem has first been discussed in the deterministic setting, an obvious starting point is to consider the stochastic system in terms of SDEs. Indeed this is particularly useful since it is in this setting we will later work. Let us begin, as many authors do, with a consideration of the problem of a particle at position  $x$  in a potential  $U(x)$ , which is subject to noise [15, 42]. A deterministic example of this system, with a quartic potential, was given in the introduction (see Eq. (1.11)). Setting the mass equal to one, this system can be expressed as the SDEs

$$\begin{aligned}\frac{dx}{dt} &= v, \\ \frac{dv}{dt} &= -\beta v - U'(x) + \eta(t),\end{aligned}\tag{2.85}$$

where  $U'(x)$  is the potential gradient,  $\beta$  is the coefficient of friction and  $\eta(t)$  is zero mean Gaussian white noise with correlations  $\langle \eta(t)\eta(t') \rangle = \delta(t - t')\beta D$ . If  $\beta$  is sufficiently large, the SDEs have an analogous form to Eq. (2.82). Mirroring the deterministic approach, one can argue that after a comparatively short time,  $dv/dt = 0$ . Solving for  $v$  we find

$$v = \beta^{-1} (-U'(x) + \eta(t)),\tag{2.86}$$

and therefore the reduced system is described by

$$\frac{dx}{dt} = -\beta^{-1} U'(x) + \beta^{-1} \eta(t).\tag{2.87}$$

The corresponding FPE is given by

$$\frac{\partial p(x, t)}{\partial t} = \frac{1}{\beta} \left\{ -\frac{\partial}{\partial x} [-U'(x)p(x, t)] + D \frac{\partial^2}{\partial x^2} [p(x, t)] \right\},\tag{2.88}$$

---

<sup>5</sup>Formally, the slow manifold is defined by the collection of trajectories which are tangent to the slow right-eigenvector at the fixed point, although in practice there is unlikely to be a closed analytic expression for this surface.



which is known as the Smoluchowski equation. We note that since the noise term is additive, both the Itô and Stratonovich calculus lead to the same FPE. This method will be referred to as direct adiabatic elimination, in the style of [42]. While in this case we have been successful in reducing the dimensionality of the problem, difficulties can be encountered when using this method. Among them are cases where the equation for the time-evolution of the slow variable contains nonlinear contributions. Say for instance, that we replace  $dx/dt = v$  with  $dx/dt = v^2$  in Eq. (2.85). Substituting for the value of the fast-variable  $v$  from Eq. (2.86), one would obtain an equation containing  $\eta(t)^2$ . However it is very hard to attach either a physical or a mathematical meaning to this object since the products of delta functions are ill-defined. Therefore, while the method can be usefully applied in systems where the entire system is linear in the fast-variables, in general other methods of fast-variable elimination must be found. The existing literature on the subject can be coarsely split according to the framework within which the stochastic system is represented; the intuitive SDE formalism or the less mathematically problematic FPE formalism.

The more rigorously derived and well-known Haken slaving principle [20, 40, 41], and several other related methods [29, 50] are developed along similar lines of reasoning as the direct adiabatic elimination method and suffer from the same complications which arise from specifying the slow manifold in a stochastic sense. A notable exception is [31], which deals with a particular model in which there is a true centre manifold (that is, a surface on which there is no deterministic flow), and applies a novel method in which stochastic perturbations away from the manifold are assumed to instantaneously relax along the deterministic trajectories.

More mathematically rigorous work on SDE fast-variable elimination has been conducted for stochastic analogues of normal form coordinate transformations. In general, these seek to determine a nonlinear transformation which simultaneously identifies the fast and slow directions in a controlled way, as well as guaranteeing the absence of ill-defined noise terms. While perhaps the earliest example in the SDE setting was [8], work has been significantly extended in the intervening years [1–3, 44]. However, many of these transformations result in noise convolutions which involve anticipating future unknown noise terms. Further work has proposed the use of additive noise terms to emulate these convolution terms in the limit of long times [6, 23, 43], though these methods are arguably less formal than the theory they rest within. Perhaps the most significant advance in this area therefore came in [39], with the construction of a methodology for stochastic normal form transform that avoids such anticipating memory integrals in many cases, though even here there remain many situations where a long time additive noise substitution must be invoked. A body of work also exists on averaging and homogenisation techniques [4, 36], although both have a more limited range of applicability than stochastic normal forms [39] and the former has been shown in certain cases to be equivalent to a stochastic normal form [43]. One of the biggest drawbacks of the work on normal forms however is that it almost exclusively deals in SDE systems with uncorrelated noise terms, whereas SDEs derived from underlying microscopic IBM models often exhibit strong noise correlations.

Other authors have focused on master or Fokker-Planck type equations as the basis for fast-variable elimination. The main advantage to working in this formalism is that one avoids encountering the ill-defined noise terms which can present themselves in the SDE setting. Much of the work in this formalism is heavily influenced by the work of Zwanzig [53, 54], who proposed that a reduction of dimension could be achieved through the application of projection operators, as illustrated by Gardiner [14], and developed by others [47]. While other variants in the FPE setting have been constructed [7, 28, 38], the projection operator technique remains the most popular. The essential idea is to define a projection operator used to separate the system into two spaces, those of the fast and slow directions. It is most instructive to give a cursory description of the method with reference to the system described by Eq.(2.85). The FPE for this system is known as Kramers equation [15]

$$\begin{aligned} \frac{\partial p(x, v, t)}{\partial t} = & -v \frac{\partial}{\partial x} [p(x, v, t)] + U'(x) \frac{\partial}{\partial v} [p(x, v, t)] \\ & + \beta \frac{\partial}{\partial v} [(v) p(x, v, t)] + D\beta \frac{\partial^2}{\partial v^2} [p(x, v, t)], \end{aligned} \quad (2.89)$$

which we can express in the more simple notation

$$\frac{\partial p(x, v, t)}{\partial t} = (L_1 + \beta L_2) p(x, v, t), \quad (2.90)$$

where the operators  $L_1$  and  $L_2$  can be read off from the previous equation. The aim is to find an equation for just  $x$  by solving an equation for  $v$  as a function of  $x$  in the limit of large  $\beta$ . Essentially it is assumed that an approximate solution to Eq.(2.90) will be formed from the marginal distribution  $p(x, t) = \int p(x, v, t) dv$  and the stationary distribution in the limit of large  $\beta$ ,  $L_2 p_{\text{eq}}(v) = 0$ . This idea is formalised by introducing a projection operator,  $R(x, v)$ , such that when applied to a function  $f(x, v)$  one obtains

$$(Rf)(x, v) = p_{\text{eq}}(v) \int f(x, v) dv. \quad (2.91)$$

Applying  $R(x, v)$  to  $p(x, v, t)$  one then obtains

$$(RP)(x, v) = p_{\text{eq}}(v) p(x, t), \quad (2.92)$$

where  $p(x, t)$  is the marginal distribution. Applying  $R(x, v)$  to Eq.(2.90) and then separately applying the operator  $(1 - R)$  one obtains two separate PDEs for the system. The aim is then to solve the second PDE in such a way as to obtain an equation for  $p(x, t)$ . In this particular case it can be shown that the reduced FPE does indeed converge to Eq.(2.88) in the limit of infinite  $\beta$  [42]. However, the procedure is by no means straightforward; the resulting reduced FPE has a non-Markovian character which can only be eliminated by a careful consideration of the limiting

behaviour with respect to  $\beta$ . A correct treatment only begins to become clear by either introducing a highly non-linear change of variables [42] or by utilising a Laplace transform. It is perhaps testament to the difficulty of working in this formalism, that the first attempt to derive the Smoluchowski equation as the limiting form of the Kramers equation was given (albeit incorrectly) by Brinkman in 1956, while the first correct treatment was given by Stratanovich in 1963 [45] and again independently by Wilemski in 1976 [51] and Titulaer in 1978 (the reader is referred to [15] for a complete discussion). The complexity of working in this methodology increases massively as one encounters systems with multiplicative noise and correlated noise. Additionally, many systems exist in which the large or small factor is not factorisable in as convenient form as Eq. (2.90), leading to further complications. Unfortunately, the details of these methods are cumbersome, especially in cases where the fast degrees of freedom are distinct from the natural basis of the problem.

The difficulties encountered in two of the main areas of research into fast-variable elimination have been discussed. In the SDE formalism one has to work very hard to avoid introducing ill-defined noise terms and systematically eliminate non-Markovian effects. In the FPE formalism, the mathematical manipulations are better defined, but one loses all sense of physical intuition so that progress becomes difficult. In both settings, the reduced system often ends up taking a very complicated form for most cases of interest. Moreover, it could be argued that in both streams of the literature a fundamental and practical question is often avoided; what advantage is gained by replacing a moderately complicated system in many variables with a more complicated system in fewer variables?

## 2.9 The Moran Model

So far this chapter has been primarily concerned with the more technical aspects of stochastic theory and fast-variable elimination. Chapters 4 and 5 will focus on the application of fast-variable elimination to a model inspired by the Moran model. A slightly more in depth discussion of the Moran model and its variants in thus called for, along with an historical overview of the relevant areas of mathematical population genetics.

It is often said that there are four main processes which drive the evolution of a population: genetic drift, mutation, selection, and migration [21]. A model which incorporates stochastic effects address the first of these issues, as was first illustrated by Fisher and Wright who considered simple stochastic processes in systems where the population size,  $N$ , was finite—the Wright-Fisher model [13, 52]. Subsequent work tended to follow their original approach, assuming discrete generations and discrete state variables (corresponding to the number of individuals in the population of a particular type) [10]. However as more details are added into these models, their discrete-time nature can make an analytic treatment difficult [11].

Progress can be made however if one considers an analogous model in continuous time with individual birth-death events, the Moran model. In the Introduction, a verbal description of the neutral Moran model was presented, while throughout this chapter, the specific example of a Moran model with mutation (see Eq. (2.16)) was used to illustrate some of the methods and ideas of stochastic theory. In Sect. 2.4 we saw what makes this model more amenable to analysis; it is formulated as a master equation with small ‘jumps’ from one state to another, allowing the expansion of the master equation to be conducted, which results in an FPE for the system dynamics. This approximation, although originally suggested by Fisher [12], was popularised by Kimura [9, 26], and proved to be a powerful tool and the starting point for many studies of more complex processes in population genetics [24, 25, 27].

We recall that the neutral Moran model is a model for pure genetic drift described by probability transition rates (1.9). As we have defined it, it consists of two types of individual, labelled  $A$  and  $B$ . In the nomenclature of population genetics these types are called alleles, which are loosely traits which an individual can carry. If one wished to model the Mendellian genetics of a human population, each individual could be modelled as having two alleles, reflecting the two sets of genetic data available from two chromosome sets [18]. Throughout this thesis we will deal exclusively with models in which each individual carries only one allele. These are referred to as haploid (as opposed to diploid) models. Mutation can be added to the Moran model, as in Eq. (2.16).<sup>6</sup> We now turn our attention to the third evolutionary process on our list, selection.

Before proceeding, we note that the designations  $A$  and  $B$  for the allele types may appear slightly confusing when presented simultaneously with the drift and diffusion terms of the FPE,  $A(\mathbf{x})$  and  $B(\mathbf{x})$ . This notation has been chosen for consistency with the population genetics literature in the first instance and the physics literature in the second. In order to avoid confusion, the drift and diffusion terms will always be presented with their functional dependence.

### 2.9.1 The Moran Model with Selection

Selection is the process by which the constitution of a population changes according to some bias. The fitness of an allele within the population is a measure of this bias. Of course, this leaves the term fitness as something of an umbrella label, embodying many potential mechanisms, some of which will be explored in Sect. 6.3. Throughout this thesis, when dealing with the Moran model, frequency-independent (or absolute) fitness will be considered, that is fitness that is independent of the constitution of the population [34].

Within the scope of the Moran model, matters are complicated slightly by the rather artificial nature of the model. In order to maintain a fixed system size  $N$ ,

---

<sup>6</sup>The way in which mutation is incorporated is of course by no means unique. Mutation could for instance be coupled to birth and death. Some of these ideas are explored in [5].

birth and death events are coupled. Therefore increasing the birth rate necessarily increases the death rate. Let us introduce the parameters  $b_A$  as the birth rate of allele  $A$  and  $b_B$  as the birth rate of allele  $B$ . The transitions rates are then

$$T(n+1|n) = b_A \frac{n}{N} \frac{(N-n)}{N-1}, \quad (2.93)$$

$$T(n-1|n) = b_B \frac{(N-n)}{N} \frac{n}{N-1}. \quad (2.94)$$

We now ask, what is the total birth/death rate of the population? Here we must bear in mind that in our formulation of the master equation, we have only taken (and need only take) account transitions which change the state of the system. However there are of course two other processes implicitly assumed to be taking place; allele  $A$  reproducing to replace another allele  $A$  and likewise for allele  $B$ . The time at which *any* reaction occurs is of course a stochastic variable, but as was shown in Sect. 2.3, it is exponentially distributed and characterised by the sum of all possible reactions (see Eq. (2.23)). The sum of all transitions in this instance can be expressed as

$$\begin{aligned} \sum_{\mu=1}^4 T_{\mu}(n + \nu_{\mu}|n) &= \frac{1}{N(N-1)} \left\{ b_A n [(n-1) + (N-n)] \right. \\ &\quad \left. + b_B (N-n) [(N-n-1) + n] \right\}, \\ &= \frac{1}{N} [b_A n + b_B (N-n)]. \end{aligned} \quad (2.95)$$

The birth/death rate changes according to the constitution of the population. If all the population is comprised of individuals carrying allele  $A$ ,  $n = N$  and the mean birth/death rate is  $b_A$  while if the entire population is  $B$  the mean birth/death rate is  $b_B$ . While at first glance this seems like a reasonable result, it suggests that the ‘fittest type’ would reproduce quickly, but also die quickly. Essentially this is an unnatural pathology of a model which requires that the system size is fixed. We can avoid this behaviour if instead we ask that  $b_A$  and  $b_B$  are weighted such that the mean birth rate of the population is fixed to be one;

$$b_A = \frac{Nw_A}{nw_A + (N-n)w_B} \quad b_B = \frac{Nw_B}{nw_A + (N-n)w_B}. \quad (2.96)$$

In this way the fitter type in some sense has a ‘greater share’ of a fixed birth rate.

It is in this fashion we choose to work, with the parameters  $w_A$  and  $w_B$  introduced to represent the fitness weightings of alleles  $A$  and  $B$  respectively [35]. The transition rates are then

$$T(n+1|n) = \frac{nw_A}{nw_A + (N-n)w_B} \frac{(N-n)}{N-1}, \quad (2.97)$$

$$T(n-1|n) = \frac{(N-n)w_B}{nw_A + (N-n)w_B} \frac{n}{N-1}. \quad (2.98)$$

The appearance of  $n$  in the denominator complicates the expansion of the master equation slightly. This is usually addressed by rewriting the fitness parameters  $w_A = 1+s$  and  $w_B = 1$ , and expanding in powers of  $s$  under the very reasonable assumption that  $s$  is small. Positive  $s$  indicates a small fitness advantage for individuals with allele  $A$ . The expansion is carried out for a more general case in Appendix C. One can obtain the results required in this instance by setting  $\mathcal{D} = 1$  in the appendix.

An expansion of the master equation leads to the FPE (2.30) with  $m = 1$  and  $A(x)$  and  $B(x)$  given respectively by

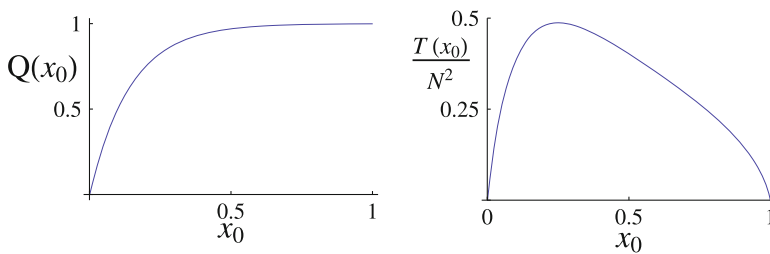
$$A(x) = sx(1-x) \quad \text{and} \quad B(x) = 2x(1-x), \quad (2.99)$$

for small  $s$ . Here we have included only the lowest order contribution in  $s$  to  $A(x)$ , and omitted the order  $s$  correction in  $B$  altogether, since it will be negligible compared with  $2x(1-x)$ . We note that in the case  $s = 0$  we obtain the neutral model. In the same manner as in Sect. 2.5.2, we find equations Eqs. (2.59) and (2.52), with  $A(x)$  and  $B(x)$  given by Eq. (2.99), for the fixation probability of type  $A$ ,  $Q(x_0) = Q_{a_2}(x_0)$ , and the unconditional fixation time  $T(x_0)$  as a function of  $x_0$ , the initial concentration of allele  $A$ .

In this case, solving Eq. (2.59), the familiar expression for the probability of fixation,  $Q(x_0)$ , is found [11]:

$$Q(x_0) = \frac{1 - \exp(-Nsx_0)}{1 - \exp(-Ns)}. \quad (2.100)$$

While the mean time to fixation can also be obtained analytically (see Eqs. (5.12) and (5.13) with  $M$  set to  $N$  and  $\sigma$  set to  $s$ , and Appendix E), it is sufficient to determine it numerically. Illustrative plots are shown in Fig. 2.3.



**Fig. 2.3** Probability of fixation of allele  $X$ ,  $Q(x_0)$ , and mean time to fixation of either allele,  $T(x_0)$ , in a system where allele  $A$  has a selective advantage  $s$  over allele  $B$ , as a function of initial  $A$  concentration. While  $Q(x_0)$  is obtained from Eq. (2.100),  $T(x_0)$  has been obtained numerically

## 2.10 Technical Background Overview

In this chapter, the fundamental mathematical tools which are necessary to understand the following research have been discussed. Let us briefly review how they relate to the modelling process.

We began by reviewing the definition of Markov process, which we verbally described as one which has no memory. Clearly however many real-world systems are non-Markovian. In Sect. 2.1 a simple example of such a system was given, however there are many other examples; when considering reproduction in a population, it could cogently be argued that there is a correlation in time between reproduction events, that a single reproduction event at time  $t$  makes another at time  $t + \Delta t$  less likely. However, if this time is sufficiently small it can be argued that the effect is negligible (effectively another fast-timescale approximation). It may be true that “Non-Markov is the rule, Markov is the exception” [48], but the benefits of simplicity and tractability garnered from treating the process as Markov often outweigh those gained from a more detailed treatment.

In Sect. 2.2, we went on to treat a Markov model in continuous time. This required defining a set of time-independent probability transition rates describing the probability per unit time of moving state. While the master equation provides a total description of the system, we noted that it was in most cases very difficult to solve. Two methods of making progress were then proposed. In Sect. 2.3, the Gillespie algorithm was introduced. While this provides stochastic realisations whose statistics obey the master equation, it does not provide any meaningful understanding of an underlying model. It therefore best serves as a comparative tool for analytic approximations and solutions of the master equation. To make analytic progress a system-size expansion of the master equation was detailed, an approximation which allows the system to be described as an FPE in as many variables as there are species in the system. The equivalent SDE form of the FPE was then given in Sect. 2.6, since this form will often be used in this thesis for intuitive reasons.

In the penultimate section, Sect. 2.8, we arrived at the crux of the thesis; the removal of fast-variables in stochastic systems. Here the origins of the problem in a mathematical sense were discussed, along with the limitations of existing work. In particular, attention was drawn to one of the fundamental problems many existing methods fail to address; a simplification of a problem via a reduction in the number of variables which comes at the cost of its over-complication in other regards, is often little simplification at all.

Throughout this chapter, reference was made to the Moran model and one of its many variants, the Moran model with mutation. This is because a model of its ilk is the subject of much of the thesis. In order to place this work in context, the Moran model was discussed from a more historical standpoint in Sect. 2.9. In this section an additional variant of the model was also introduced which includes the effect of selection.

In what follows, I will introduce two methods of simplifying stochastic problems with a separation of timescales. Both procedures are mathematically explicit,

straightforward to apply, and address the effect of correlated noise terms. In Chap. 3 the first of these approximations will be introduced, the conditioning method. To begin I develop the method with the aid of a simple illustrative example with an ecological interpretation in Sect. 3.2, before providing a general formulation. The results from this example demonstrate the success of the approximation scheme even in regimes where the fixed point is weakly unstable. In Sect. 3.4 the method is applied to an epidemiological model with seasonal forcing. This model has been identified as suffering from the technical numerical difficulties associated with a large separation between eigenvalues. It will be shown how the method may be used in tandem with the LNA to provide a very good approximation to results coming from stochastic simulations.

In Chap. 4 the second of the methods will be introduced, which shall be referred to as the projection matrix method. In particular, I investigate the method applied to a generalisation of the Moran model which incorporates migration between a number of well-mixed populations. Birth and death events are later moderated by a weak selection pressure. The technique allows the equations for the system (in as many variables as there are islands) to be reduced so that they resemble those for a single island, which are amenable to analysis. Once again, the idea behind the method is simple, its application is systematic, and the results are in very good agreement with simulations of the full model for a range of parameter values.

Following the reduction of the metapopulation Moran model, Chap. 5 is devoted to analysing the reduced model, and comparing it to the results obtained from stochastic simulation. A rich array of behaviour is found, all of which is well predicted by the reduced system.

Chapter 6 contains work which extends and connects that in the previous chapters. This includes incorporating mutation into the Moran model with migration, drawing links between the conditioning method and the projection matrix method and looking at the possibility of relaxing the fixed system size assumption inherent in the Moran model.

## References

1. L. Arnold, *Random Dynamical Systems*, Springer Monographs in Mathematics (Springer, Berlin, 2003)
2. L. Arnold, P. Imkeller, Normal forms for stochastic differential equations. *Probab. Theory Relat. Fields* **110**, 559–588 (1998)
3. L. Arnold, X. Kedai, Simultaneous normal form and center manifold reduction for random differential equations, in *Equadiff 91 : International Conference on Differential Equations*, ed. by C. Perello, C. Sim, J. Sola-Morales (1991)
4. N. Berglund, B. Gentz, Geometric singular perturbation theory for stochastic differential equations. *J. Differ. Equ.* **191**, 1–54 (2003)
5. R.A. Blythe, A.J. McKane, Stochastic models of evolution in genetics, ecology and linguistics. *J. Stat. Mech.* **P07018** (2007)
6. X. Chao, A.J. Roberts, On the low dimensional modelling of Stratanovich stochastic differential equations. *Phys. A* **225**, 62–80 (1996)



7. M.N. Contou-Carrere, V. Sotiropoulos, Y.N. Kaznessis, P. Daoutidis, Model reduction of multi-scale chemical Langevin equations. *Syst. Control Lett.* **60**, 75–86 (2011)
8. P.H. Coughlet, C. Elphick, E. Tirapegui, Normal form of Hopf bifurcation with noise. *Phys. Lett. A* **111**, 277–282 (1985)
9. J.F. Crow, M. Kimura, Some genetic problems in natural populations, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, University of California Press, Berkeley (1956), pp. 1–22
10. W.J. Ewens, *Population Genetics* (Wilmer Brother Limited, Birkenhead, 1969)
11. W.J. Ewens, *Mathematical Population Genetics*, 2nd edn. (Springer, Berlin, 2004)
12. R.A. Fisher, On the dominance ratio. *Proc. Roy. Soc. Edinb.* **42**, 321–341 (1922)
13. R.A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930)
14. C.W. Gardiner, Adiabatic elimination in stochastic systems. I. Formulation of methods and application to few-variable systems. *Phys. Rev. A* **29**, 2814–2822 (1984)
15. C.W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2009)
16. D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976)
17. D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977)
18. A.J.F. Griffiths, S.R. Wessler, R.C. Lewontin, S.B. Carroll, *Introduction to Genetic Analysis*, 9th edn. (W. H. Freeman and Company, New York, 2007)
19. H. Haken, *Synergetics* (Springer, Berlin, 1983)
20. H. Haken, A. Wunderlin, Slaving principle for stochastic differential equations with additive and multiplicative noise and for discrete noisy maps. *Z. Phys. B* **47**, 179–187 (1982)
21. D.L. Hartl, A.G. Clark, *Principles of Population Genetics*, 4th edn. Sinauer Associates Inc., Sunderland, Mass. (2007)
22. R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, 1991)
23. R.Z. Khasminskii, A limit theorem for the solutions of differential equations with random right-hand sides. *Theor. Probab. Appl.* **11**, 390–406 (1966)
24. M. Kimura, Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harb. Symp. Quant. Biol.* **20**, 33–53 (1955)
25. M. Kimura, Diffusion models in population genetics. *J. Appl. Probab.* **1**, 177–232 (1964)
26. M. Kimura, *Population Genetics, Molecular Evolution and the Neutral Theory* (Chicago University Press, Chicago, 1994)
27. M. Kimura, G.H. Weiss, The stepping stone model of population structure and the decrease in genetic correlation with distance. *Genetics* **49**, 561–576 (1964)
28. E. Knobloch, K.A. Wiesenfeld, Bifurcations in fluctuating systems: the center manifold approach. *J. Stat. Phys.* **33**, 611–637 (1983)
29. Y. Lan, T.C. Elston, G.A. Papoian, Elimination of fast variables in chemical Langevin equations. *J. Chem. Phys.* **129**, 214115 (2008)
30. M. Lax, Fluctuations from the nonequilibrium steady state. *Rev. Mod. Phys.* **32**, 25–64 (1960)
31. Y.T. Lin, H. Kim, C.R. Doering, Features of fast living: on the weak selection for longevity in degenerate birth-death processes. *J. Stat. Phys.* **148**, 646–662 (2012)
32. T. Maruyama, *Stochastic Problems in Population Genetics* (Springer, Berlin, 1977)
33. A.J. McKane, T. Biancalani, T. Rogers, Stochastic pattern formation and spontaneous polarisation: the linear noise approximation and beyond. *Bull. Math. Biol.* **76**, 895–921 (2014)
34. R.E. Michod, *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality* (Princeton University Press, New Jersey, 2000)
35. M.A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, 2006)
36. G.A. Pavliotis, A.M. Stuart, *Averaging and Homogenization*, vol. 53, Multiscale Methods (Springer, New York, 2008)
37. L.E. Reichl, *A Modern Course in Statistical Physics* (Wiley VCH, New York, 1998)
38. H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1989)

39. A.J. Roberts, Normal form transforms separate slow and fast modes in stochastic dynamics systems. *Phys. A* **387**, 12–38 (2008)
40. G. Schöner, H. Haken, The slaving principle for Stratanovich stochastic differential equations. *Z. Phys. B* **63**, 493–504 (1986)
41. G. Schöner, H. Haken, A systematic elimination procedure for Itô stochastic differential equations and the adiabatic approximation. *Z. Phys. B* **68**, 89–103 (1987)
42. R. Serra, M. Andretta, M. Compiani, G. Zanarini, *Introduction to the Physics of Complex Systems* (Pergamon Press, Oxford, 1986)
43. N. Sri Namachchivaya, Equivalence of stochastic averaging and stochastic normal forms. *J. Appl. Mech.* **57**, 1011–1017 (1990)
44. N. Sri Namachchivaya, Y.K. Lin, Method of stochastic normal forms. *Int. J. Nonlinear Mech.* **26**, 931–943 (1991)
45. R.L. Stratanovich, *Introduction to the Theory of Random Noise* (Gordon and Breach, New York, 1963)
46. S.H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Westview Press, New York, 2001)
47. P. Thomas, R. Grima, A.V. Straube, Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators. *Phys. Rev. E* **86**, 041110 (2012)
48. N.G. van Kampen, Remarks on non-Markov processes. *Braz. J. Phys.* **28**, 90–96 (1998)
49. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007)
50. W. Wang, A.J. Roberts, Slow manifold and averaging for slow-fast stochastic differential system. *J. Math. Anal. Appl.* **398**, 822–839 (2013)
51. G. Wilemski, On the derivation of Smoluchowski equations with corrections in the classical theory of Brownian motion. *J. Stat. Phys.* **14**, 153–170 (1976)
52. S. Wright, Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931)
53. R. Zwanzig, Ensemble method in the theory of irreversibility. *J. Chem. Phys.* **33**, 1338–1341 (1960)
54. R. Zwanzig, Memory effects in irreversible thermodynamics. *Phys. Rev.* **124**, 983–992 (1961)

Fast Variables in Stochastic Population Dynamics

Constable, G.W.A.

2015, XVII, 163 p. 40 illus., 31 illus. in color., Hardcover

ISBN: 978-3-319-21217-3