### 1.4 Forward Kolmogorov equation

Let us now consider evolving probabilities for a generic situation where the states are ordered along a line, such as in the previous examples with population size $n=0,1,2 \cdots, N$. The general form of the Master equation is

$$
\begin{equation*}
\frac{d p_{n}}{d t}=+\sum_{m \neq n} R_{n m} p_{m}-\sum_{m \neq n} R_{m n} p_{n} \tag{1.42}
\end{equation*}
$$

In many relevant circumstances the number of states is large, and the probability varies smoothly from one value of $n$ to the next. In such cases it is reasonable to replace the discrete index $n$ with a continuous variable $x$, the probabilities $p_{n}(t)$ with a probability density $p(x, t)$, and the rates $R_{m n}$ with a rate function $R\left(x^{\prime}, x\right)$. The rate function $R$ depends on two variables $x$ and $x^{\prime}$, denoting respectively the start and end positions for a transition along the line. We have the option of redefining the two arguments of this function, and it is useful to reparameterize it as $R\left(x^{\prime}-x, x\right)$ indicating the rate at which, starting from the position $x$, a transition is made to a position $\Delta x=x^{\prime}-x$ away. As in the case of mutations, there is usually a preference for changes that are local, i.e. with rates that decay rapidly when the separation $x^{\prime}-x$ becomes large.


These transformations and relabelings,

$$
\begin{equation*}
n \rightarrow x, p_{n}(t) \rightarrow p(x, t), R_{m n} \rightarrow R\left(x^{\prime}-x, x\right) \tag{1.43}
\end{equation*}
$$

enable us to transform Eq. (1.42) to the continuous integral equation

$$
\begin{equation*}
\frac{\partial}{\partial t} p(x, t)=+\int^{*} d x^{\prime} R\left(x-x^{\prime}, x^{\prime}\right) p\left(x^{\prime}, t\right)-\int^{*} d x^{\prime} R\left(x^{\prime}-x, x\right) p(x, t) \tag{1.44}
\end{equation*}
$$

Some care is necessary in replacing the sums with integrals, as the summations in Eq. (1.42) exclude the term with $m=n$. To treat this restriction in the continuum limit, we focus on an interval $y$ around any point $x$, and consider the change in probability due to incoming flux from $x-y$ and the outgoing flux to $x+y$, thus arriving at

$$
\begin{equation*}
\frac{\partial}{\partial t} p(x, t)=\int d y[R(y, x-y) p(x-y)-R(y, x) p(x)] . \tag{1.45}
\end{equation*}
$$

Note that the contribution for $y=0$ is now clearly zero. The flux difference for small $y$ is now estimated by a Taylor expansion of the first term in the square bracket, but only with respect to the location of the incoming flux, treating the argument pertaining to the separation of the two points as fixed, i.e.

$$
\begin{equation*}
R(y, x-y) p(x-y)=R(y, x) p(x)-y \frac{\partial}{\partial x}(R(y, x) p(x))+\frac{y^{2}}{2} \frac{\partial^{2}}{\partial x^{2}}(R(y, x) p(x))+\cdots . \tag{1.46}
\end{equation*}
$$

While formally correct, the above expansion is useful only in cases where typical values of $y$ are small (i.e. when only almost local transitions occur). Keeping terms up to the second order, Eq. (1.45) can be rewritten as

$$
\begin{equation*}
\frac{\partial}{\partial t} p(x, t)=-\int d y y \frac{\partial}{\partial x}(R(y, x) p(x))+\frac{1}{2} \int d y y^{2} \frac{\partial^{2}}{\partial x^{2}}(R(y, x) p(x)) \tag{1.47}
\end{equation*}
$$

The integrals over $y$ can be taken inside the derivatives with respect to $x$,

$$
\begin{equation*}
\frac{\partial}{\partial t} p(x, t)=-\frac{\partial}{\partial x}\left[p(x)\left(\int d y y R(y, x)\right)\right]+\frac{1}{2} \frac{\partial^{2}}{\partial x^{2}}\left[p(x)\left(\int d y y^{2} R(y, x)\right)\right] \tag{1.48}
\end{equation*}
$$

after which we obtain

$$
\begin{equation*}
\frac{\partial p(x, t)}{\partial t}=-\frac{\partial}{\partial x}[v(x) p(x, t)]+\frac{\partial^{2}}{\partial x^{2}}[D(x) p(x, t)] \tag{1.49}
\end{equation*}
$$

We have introduced

$$
\begin{equation*}
v(x) \equiv \int d y y R(y, x)=\frac{\langle\Delta(x)\rangle}{\Delta t} \tag{1.50}
\end{equation*}
$$

and

$$
\begin{equation*}
D(x) \equiv \frac{1}{2} \int d y y^{2} R(y, x)=\frac{1}{2} \frac{\left\langle\Delta(x)^{2}\right\rangle}{\Delta t} . \tag{1.51}
\end{equation*}
$$

Equation (1.49) is a prototypical description of drift and diffusion which appears in many contexts. The drift term $v(x)$ expresses the rate (velocity) with which transitions change (on average) the position from $x$. Given the probabilistic nature of the process, there are variations in the rate of change of position captured by the position dependent diffusion coefficient $D(x) .{ }^{4}$ The drift-diffusion equation is known as the forward Kolmogorov equation in the context of populations. As a description of random walks it appeared earlier in physics literature as the Fokker-Planck equation.

In the context of population dynamics, it is convenient to introduce the variable $x=n / N$, such that in the continuum limit $x$ is limited to the interval $[0,1]$. The rates in Eq. (1.23) change $n$ by $\pm 1$, and hence
$v(x)=\frac{\langle\Delta n\rangle}{N}=\frac{R_{n+1, n} \times(+1)+R_{n-1, n} \times(-1)}{N}=\frac{1}{N}\left[\mu_{1}(N-n)-\mu_{2} n\right]=\mu_{1}(1-x)-\mu_{2} x$,
while

$$
\begin{equation*}
D(x)=\frac{\left\langle\Delta n^{2}\right\rangle}{2 N^{2}}=\frac{R_{n+1, n}+R_{n-1, n}}{2 N^{2}}=\frac{1}{2 N^{2}}\left[\mu_{1}(N-n)+\mu_{2} n\right]=\frac{\mu_{1}(1-x)+\mu_{2} x}{2 N} . \tag{1.52}
\end{equation*}
$$

[^0]
### 1.4.1 Binomial selection

Consider a population with two forms of an allele, say $A_{1}$ and $A_{2}$ corresponding to blue or brown eye colors. The probability for a spontaneous mutation to occur that changes the allele for eye color is extremely small, and effectively $\mu_{1}=\mu_{2}=0$ in Eq. (1.24). Yet the proportions of the two alleles in the population does change from generation to generation. One reason is that some individuals do not reproduce and leave no descendants, while others reproduce many times and have multiple descendants. This is itself a stochastic process and the major source of rapid changes in allele proportions. In principle this effect also leads to variations in population size. In practice, and to simplify computations, it is typically assumed that the size of the population is fixed.

In the model of binomial selection, the process or reproduction from one generation to the next is assumed to be as follows: Let us assume that in a population of $N$ alleles, $N_{1}=n$ are $A_{1}$, and $N-n$ are $A_{2}$. The population at the next generation may have $m$ individuals with allele $A_{1}$, and the probability for such a transition is

$$
\begin{equation*}
\Pi_{m n}=\left(\frac{n}{N}\right)^{m}\left(1-\frac{n}{N}\right)^{N-m}\binom{N}{m} \tag{1.54}
\end{equation*}
$$

The process leading to such probability is like reaching into a bag with $n$ balls of blue color and $N-m$ balls of brown color, recording the color of the selected ball and throwing it back to the bag. After repeating such selection $N$ times, the probability that the blue color is recorded $m$ times is given by the above binomial distribution. (The probability of getting a blue ball in each trial is simply $n / N$, and $1-n / N$ for brown.) Clearly some balls can be picked up multiple times (multiple descendants), while some balls are never picked (no offspring).

Regarding $R_{m n}$ as the probability to obtain the random variable $m$, given an initial value of $n$, it is easy to deduce from standard properties of the binomial distribution that

$$
\begin{equation*}
\langle m\rangle=N \times \frac{n}{N}=n, \quad \text { i.e } \quad\langle(m-n)\rangle=0 \tag{1.55}
\end{equation*}
$$

while

$$
\begin{equation*}
\left\langle m^{2}\right\rangle_{c}=\left\langle(m-n)^{2}\right\rangle=N \times \frac{n}{N}\left(1-\frac{n}{N}\right) . \tag{1.56}
\end{equation*}
$$

We can construct a continuum evolution equation by setting $x=n / N \in[0,1]$, and replacing $p(n, t+1)-p(n, t) \approx d p(x) / d t$, where $t$ is measured in number of generations. Clearly, from Eq. (1.55), there is no drift

$$
\begin{equation*}
v(x)=\langle(m-n)\rangle=0, \tag{1.57}
\end{equation*}
$$

while the diffusion coefficient is given by

$$
\begin{equation*}
D_{\text {haploid }}(x)=\frac{1}{2 N^{2}}\left\langle(m-n)^{2}\right\rangle=\frac{1}{2 N} x(1-x) . \tag{1.58}
\end{equation*}
$$

Since each allele in the above example is treated as a single individual, the process is similar to reproduction in haploid organisms, where each individual carrier one copy of the
genetic information. A slight modification of binomial selection is also applicable to diploid organisms, where each member carries two copies of an allele. Given two alleles, there are three possible variants (so-called genotypes) of $A_{1} A_{1}, A_{1} A_{2}$, and $A_{2} A_{2}$ in proportions of $x_{11}, x_{12}$, and $x_{22}$ respectively. The offspring from mating of two diploids randomly acquire one copy of the allele from each parent. To mimic a mating event, pick one allele of one individual, another allele from a second individual. Set aside the resulting offspring and return the parents to the initial pool. Repeat the process $N$ times to construct the new generation. For each offspring the probability of selecting allele $A_{1}$ is $x_{11}+x_{12} / 2$, while allele $A_{2}$ is selected with probability $x_{22}+x_{12} / 2$. If the initial population is in Hardy-Weinberg equilibrium, the relative genotype frequencies are related to the proportions of the two alleles simply by $x_{11}=x^{2}, x_{12}=2 x_{1} x_{2}$, and $x_{22}=x_{2}^{2}$. The mating process is thus again equivalent to the process we considered earlier for haploids with $A_{1}$ and $A_{2}$ chosen with probabilities of $x$ and $1-x$ respectively. Since, in a diploid population of $N$ individuals, the number of alleles is $2 N$, the previous result is simply modified to

$$
\begin{equation*}
D_{\text {diploid }}(x)=\frac{1}{4 N} x(1-x) . \tag{1.59}
\end{equation*}
$$

### 1.4.2 Chemical analog \& Selection

Through the reactions in Eq. (1.26), we introduced a chemical mixture that mimicks a mutating population. Consider a system where a reaction between molecules A and B can lead to two outcomes: ${ }^{5}$

$$
\begin{equation*}
\mathrm{A}+\mathrm{B} \rightharpoonup_{c} \mathrm{~A}+\mathrm{A} \quad \text { or } \mathrm{A}+\mathrm{B} \neg^{d} \mathrm{~B}+\mathrm{B} \tag{1.60}
\end{equation*}
$$

at rates $c$ and $d$. In a "mean-field" approximation the number of A molecules changes as

$$
\begin{equation*}
\frac{d N_{A}}{d t}=(c-d) N_{A} N_{B}=(c-d) N_{A}\left(N-N_{A}\right) . \tag{1.61}
\end{equation*}
$$

Equation (1.61) predicts steady states $N_{A}^{*}=0$ for $c<d, N_{A}^{*}=N$ for $c>d$, while any composition is permitted for the symmetric case of $c=d$. As we shall demonstrate, fluctuations modify the latter conclusion.

As before, let us denote $N_{A}=n, N_{B}=N-N_{A}$, and follow the change in composition after a single reaction. The number of A species may change by $\pm 1$ with rates

$$
\begin{equation*}
R_{n, n+1}=d(n+1)(N-n-1), \quad \text { and } \quad R_{n, n-1}=c(n-1)(N-n+1) \tag{1.62}
\end{equation*}
$$

where the product is over the number of possible pairs of A-B particles that can participate in the reaction. The diagonal terms are again obtained from the normalization condition in

[^1]Eq. (1.14) resulting in the Master equation
$\frac{d p(n, t)}{d t}=d(n+1)(N-n-1) p(n+1)+c(n-1)(N-n+1) p(n-1)-d n(N-n) p(n)-c n(N-n) p(n)$,
for $0<n<N$, and with boundary terms

$$
\begin{equation*}
\frac{d p(0, t)}{d t}=d(N-1) p(1), \quad \text { and } \quad \frac{d p(N, t)}{d t}=c(N-1) p(N-1) \tag{1.64}
\end{equation*}
$$

When the number $N$ is large, it is reasonable to take the continuum limit and construct a Kolmogorov equation for the fraction $x=n / N \in[0,1]$. The rates in Eq. (1.62) change $n$ by $\pm 1$, and hence

$$
\begin{align*}
v(x) & =\frac{\langle\Delta n\rangle}{N}=\frac{R_{n+1, n}-R_{n-1, n}}{N}=\frac{1}{N}[c n(N-n)-d n(N-n)] \\
& =N(c-d) x(1-x) \tag{1.65}
\end{align*}
$$

while

$$
\begin{align*}
D(x) & =\frac{\left\langle\Delta n^{2}\right\rangle}{2 N^{2}}=\frac{R_{n+1, n}+R_{n-1, n}}{2 N^{2}}=\frac{1}{2 N^{2}}[c n(N-n)+d n(N-n)] \\
& =\frac{c+d}{2} x(1-x) \tag{1.66}
\end{align*}
$$

Comparison with Eqs.(1.57) and Eq. (1.59) indicates that the above reaction has the same behavior as binomial selection provided that $c=d=1 /(4 N)$. Indeed the superficial difference in factor of $N$ between the two cases is because in the latter we followed the reactions one at a time (at rate $c=d$ ), while in the former we computed the transition probabilities after a whole generation ( $N$ steps of reproduction and removal). The selection process characterized by Eq.(1.54) treats the two alleles as completely equivalent. In reality one allele may provide some advantage to individuals carrying it. If so, there should be a selection process by which individuals with this allele are more likely to reproduce, on average increasing their population in the next generation. Such advantage then leads to a drift in the appropriate Kolmogorov equation. The process of selection in population genetics was described in a previous section. Indeed the result in Eq. (1.91), can be mathematically reproduced in the binary reaction of Eq. (1.60) with $c \neq d$, given by

$$
\begin{equation*}
c=\frac{1}{4 N}(1+s) \quad \text { and } \quad d=\frac{1}{4 N}(1-s) . \tag{1.67}
\end{equation*}
$$

In the following, we shall employ the nomenclature of population genetics, such that

$$
\begin{equation*}
v(x)=\frac{s}{2} x(1-x), \quad \text { and } \quad D(x)=\frac{1}{4 N} x(1-x) . \tag{1.68}
\end{equation*}
$$

### 1.4.3 Steady states

While it is usually hard to solve the Kolmogorov equation as a function of time, it is relatively easy to find the steady state solution to which the population settles after a long time. Let us denote the steady-state probability distribution by $p^{*}(x)$, which by definition must satisfy

$$
\begin{equation*}
\frac{\partial p^{*}(x)}{\partial t}=0 \tag{1.69}
\end{equation*}
$$

Therefore, setting the right-hand side of Eq. (1.49) to zero, we get

$$
\begin{equation*}
-\frac{\partial}{\partial x}\left[v(x) p^{*}(x)\right]+\frac{\partial^{2}}{\partial x^{2}}\left[D(x) p^{*}(x)\right]=0 \tag{1.70}
\end{equation*}
$$

The most general solution admits steady states in which there is an overall current and the integral over $x$ of the last equation leads to a constant flow in probability. It is not clear how such a circumstance may arise in the context of population genetics, and we shall therefore focus on circumstances where there is no probability current, such that

$$
\begin{equation*}
-v(x) p^{*}(x)+\frac{\partial}{\partial x}\left(D(x) p^{*}(x)\right)=0 . \tag{1.71}
\end{equation*}
$$

We can easily rearrange this equation to

$$
\begin{equation*}
\frac{1}{D(x) p^{*}} \frac{\partial}{\partial x}\left(D(x) p^{*}(x)\right)=\frac{\partial}{\partial x} \ln \left(D(x) p^{*}(x)\right)=\frac{v(x)}{D(x)} . \tag{1.72}
\end{equation*}
$$

This equation can be integrated to

$$
\begin{equation*}
\ln D(x) p^{*}(x)=\int^{x} d x^{\prime} \frac{v\left(x^{\prime}\right)}{D\left(x^{\prime}\right)}+\text { constant } \tag{1.73}
\end{equation*}
$$

such that

$$
\begin{equation*}
p^{*}(x) \propto \frac{1}{D(x)} \exp \left[\int^{x} \frac{v\left(x^{\prime}\right)}{D\left(x^{\prime}\right)}\right] \tag{1.74}
\end{equation*}
$$

with the proportionality constant set by boundary conditions.
Let us examine the case of the dynamics of a fixed population, including mutations, and reproduction with selection. Adding the contributions in Eqs. (1.52), (1.53) and (1.68), we have

$$
\begin{equation*}
v(x)=\frac{s}{2} x(1-x)+\mu_{1}(1-x)-\mu_{2} x \tag{1.75}
\end{equation*}
$$

while

$$
\begin{equation*}
D(x)=\frac{1}{4 N} x(1-x)+\frac{\mu_{1}(1-x)+\mu_{2} x}{2 N} \approx \frac{1}{4 N} x(1-x) \tag{1.76}
\end{equation*}
$$

The last approximation of ignoring the contribution from mutations to diffusion is common to population genetics, where mutation rates are much smaller than one. It enables a closed
form solution to the steady state, as

$$
\begin{aligned}
\log D(x) p^{*}(x) & =\int^{x} d x^{\prime} \frac{v\left(x^{\prime}\right)}{D\left(x^{\prime}\right)} \\
& =4 N \int^{x} d x^{\prime}\left[\frac{\mu_{1}}{x^{\prime}}-\frac{\mu_{2}}{1-x^{\prime}}+\frac{s}{2}\right] \\
& =4 N\left[\mu_{1} \ln x+\mu_{2} \ln (1-x)+\frac{s}{2} x\right]+\text { constant }
\end{aligned}
$$

resulting in

$$
\begin{equation*}
p^{*}(x) \propto \frac{1}{x(1-x)} \times x^{4 N \mu_{1}} \times(1-x)^{4 N \mu_{2}} \times e^{2 N s x} . \tag{1.77}
\end{equation*}
$$

In the special case of no selection, $s=0$ and (for convenience) $\mu_{1}=\mu_{2}=\mu$, the steadystate solution (1.77) simplifies to

$$
\begin{equation*}
p^{*}(x) \propto[x(1-x)]^{4 N \mu-1} \tag{1.78}
\end{equation*}
$$

The shape of the solution is determined by the parameter $4 N \mu$. If $4 N \mu>1$, then the distribution has a peak at $x=1 / 2$ and diminishes to the sides. On the other hand, if the population is small and $4 N \mu<1$, then $p^{*}(x)$ has peaks at either extreme - a situation where genetic drift is dominant.

* No relection $(\bar{w}=1)_{j}$ neutial mutations $\left(r_{1}=r_{2}=\mu\right)$



In deriving Eqs. $(1.74,1.77)$, we neglected two important issues: boundary conditions, and normalizability of the probability distributions. These topics will be discussed next.


[^0]:    ${ }^{4}$ The diffusion coefficient is usually associated with the variance, $\left\langle\Delta(x)^{2}\right\rangle_{c} \equiv\left\langle\Delta(x)^{2}\right\rangle-\langle\Delta(x)\rangle^{2}$. However, in the limit of $\Delta t \rightarrow 0$, the squared mean is of second order in $\Delta t$, and can be ignored.

[^1]:    ${ }^{5}$ These reactions mimic an important element of the mating process which stochastically modifies the proportion of alleles in a fixed-size population: The offspring from mating a heterozygote (a diploid organism with different alleles $A_{1}$ and $A_{2}$ ) with a homozygote (say with two copies of allele $A_{1}$ ) may be either heterozygote $\left(A_{1} A_{2}\right)$ or homozygote $\left(A_{1} A_{1}\right)$.

