Advanced Statistical Physics

J.Pearson

August 8, 2009

Abstract

These are a set of notes I have made, based on lectures given by A.McKane at the University of Manchester Sept-Dec '08. Please e-mail me with any comments/corrections: jon@jpoffline.com. These notes may be found at www.jpoffline.com.

Previous courses on statistical mechanics have dealt with systems in equilibrium. This course does not. Perhaps a better title for this is "*A Course in Stochastic Dynamics*". That is, how random variables evolve in time.

Contents

1 Elementary Probability Theory **1**

1.1 Representations 1

1.2 Stochastic Random Variables 3

1.2.1 Continuous Random Variables 4

1.2.2 Generating Function 4

1.2.3 Cumulants 6

1.3 Multi-variable Probability Distribution Functions 6

1.4 Covariance 8

1.5 The Central Limit Theorem 9

1.5.1 Proof 9

1.5.2 Limitations of the Theorem 11

1.6 Time-dependent Systems 11

1.6.1 Stationary Processes: Definition 13

1.6.2 Gaussian Process: Definition 13

2 Markov Processes **15**

2.1 Introduction 15

2.2 Markov Chains 17

2.2.1 Random One-Dimensional Walk 18

2.2.2 Urn Models 20

2.3 Stochastic Matrices 22

2.3.1 Right & Left Eigenvectors 22

2.3.2 Properties of Stochastic Matrices 23

2.3.3 Example: Urn Model 24

2.3.4 General Theory 26

2.3.5 Powers of a Matrix 27

2.4 Examples of Markov Chains 28

2.4.1 The Ehrenfest Urn 30

2.4.2 The Wright-Fisher Model 31
2.5 The Master Equation 33
2.5.1 Derivation 33
2.5.2 Relation to Markov Chains 35
2.6 One Step Processes 35
2.6.1 The Decay Process 36
2.6.2 Linear One-step Processes & Their Solutions 37
2.6.3 The Macroscopic Equation 43
2.6.4 Long Term Behaviour of Non-linear Processes 44
2.6.5 Boundaries 45
2.6.6 Stationary States of the Master Equation 46
2.7 Solution to Master Equation Under Detailed Balance 48
2.8 Summary 59

3 Drift & Diffusion **63**
3.1 Introduction 63
3.1.1 The Moran Model 63
3.2 The Fokker-Planck Equation 66
3.2.1 Computing Jump Moments 68
3.2.2 Simple Diffusion 70
3.2.3 The Moran Model 70
3.2.4 The Moran Model: With Mutation 72
3.3 Properties of the Fokker-Planck Equation 74
3.3.1 Probability Current 75
3.3.2 Boundary Conditions 75
3.3.3 The Adjoint Operator 76
3.3.4 The Stationary State 77
3.3.5 Transformation to Schrodinger-like Equation 78
3.3.6 Transformation to Adjoint Equation 79

3.3.7 Time Independent FPE 80
3.3.8 Example: Brownian Motion in a Potential 83

4 Stochastic Differential Equations **87**

4.1 Brownian Motion Described by the Langevin Equation 87
4.2 The Solution to the Langevin Equation Describing Brownian Motion 88
4.3 Comments on the Langevin Equation 92
4.4 Equivalence to the Fokker-Planck Equation 94

1 Elementary Probability Theory

Suppose we have a total of n events, and that there are m ways of getting a specific event A (say). Then, we say that the probability of A occurring, $P(A)$, is given by

$$P(A) = \frac{m}{n},$$

where m, n may not be "known" (that is, known *a priori*), and must thus be determined by experiment. Also, this only works for an infinite number of trials. Thus, $P(A)$ is actually the limiting value of the relative frequency m/n of A, as the number of trials increases to infinity.

We must make a distinction between the *sample points* and the *sample space*. The former is better described as the "possible outcomes". For example, given two coins, the sample space is

$$S = \{(H, H), (H, T), (T, H), (T, T)\},$$

with each event being a subset of S.

1.1 Representations

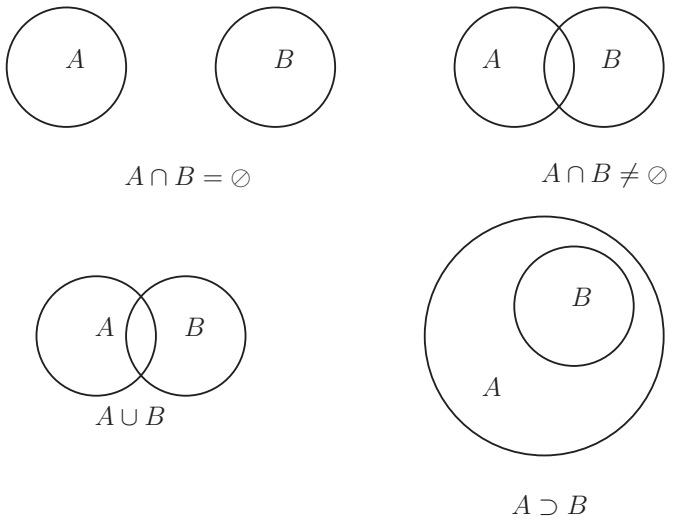

Figure 1.1: Some standard representations in elementary probability theory.

We use the following symbols:

- $A \cap B$ to denote what the "cross-over" of the sets A and B are; i.e. only that which is common to both A and B.

- $A \cup B$ to denote the union: everything within both A and B.
- $A \supset B$ to denote the subset.

So, for example, the statement

$$A \cup B = \oslash$$

reads "the union of A with B is the empty set". That is, A and B posses no common elements. Similarly,

$$A \cap B \neq \oslash$$

denotes that the two sets have common elements. The statement

$$A \supset B$$

reads "B is a subset of A". That is, all elements of A lie within B.

We see that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

is the probability of an event anywhere within the sets A and B. We subtract the 'cross-over' term as we are counting that region twice. If we have a situation where $A \cup B = \oslash$, then events A and B are *mutually exclusive*. In this case, we have that

$$P(A \cup B) = P(A) + P(B);$$

So that we read $P(A \cup B)$ as "the probability of event A or B occurring".

We have that, for some total number of events n, the number of events which fall under the class $A \cup B$ is $n_{A \cup B}$. Then, by our previous definition of the probability being the limiting case,

$$P(A \cup B) = \lim_{n \to \infty} \frac{n_{A \cup B}}{n}.$$

Consider the setup where some set m is within some set k. That is, $m \supset k$. Then, the probability of event m occurring, given that event k has occurred is denoted

$$P(m|k).$$

We can get an intriguing result from this. Consider that the probability $P(m|k)$ is the probability of m occurring from a space where k has already occurred, in a space of n total possibilities. Then, we see that

$$P(m|k) = \frac{P(m)}{P(k)}, \tag{1.1}$$

where each may be expressed as

$$\begin{aligned} \frac{P(m)}{P(k)} &= \lim_{n \to \infty} \frac{n_m/n}{n_k/n} \\ &= \lim_{n \to \infty} \frac{n_m}{n_k}. \end{aligned}$$

Now, consider that the union of two sets A and B is denoted M; $A \cup B = M$. Then, by (1.1), we see that the probability of M occurring given that B has already occurred, is given by

$$\begin{aligned} P(M|B) &= P(A \cup B|B) \\ &= \frac{P(A \cup B)}{P(B)}. \end{aligned}$$

From this, we have **Bayes' theorem**:

$$P(A \cup B) = P(A|B)P(B), \tag{1.2}$$

which reads "the probability of both A and B occurring is the product of the probabilities that B occurs, with the probability that A occurs given that B occurred".

We have that

$$P(A \cap B) = P(A)P(B)$$

only if A and B are independent. That is, there is no overlap between the two sets.

1.2 Stochastic Random Variables

Think of a stochastic random variable as some function X which sends the sample space S to the real line $\mathbb{R}$:

$$X : S \mapsto \mathbb{R}.$$

One must not confuse this with x, which are the possible values that can result.

Suppose that we have $p(x_i)$, which is the probability that one results in value x_i after some experiment. Then, we call p the *probability density function* or "pdf". The pdf has the following properties:

$$p(x_i) \geq 0, \qquad \forall i; \tag{1.3}$$

$$\sum_i p(x_i) = 1, \tag{1.4}$$

where the last expression is the statement of normality. We also make the definitions of the n^{th} *moment*

$$\mu_n \equiv \langle X^n \rangle \equiv \sum_i x_i^n p(x_i), \tag{1.5}$$

so that $\langle X \rangle$ is just the mean (usually denoted just μ), and the combination $\langle X^2 \rangle - \langle X \rangle^2$ is the variance σ^2; so that the standard deviation is σ.

Again, we must make the distinction between X the random variable, and x the actual value you can find.

Thus far, we have discussed discrete random variables; let us now consider continuous ones.

1.2.1 Continuous Random Variables

Now we have that $p(x)$ is some probability density function; where X is defined on some continuous interval $[a, b]$. Thus, the probability of an event within some range $[c, d]$ is

$$\int_c^d p(x)dx,$$

just the area under the plot of the pdf. We fairly obviously have normalisation via

$$\int_{\forall x} p(x)dx = 1,$$

where $\forall x$ denotes the entire range on which X is defined. Similarly, we have the moment

$$\langle X^n \rangle = \int_{\forall x} x^n p(x)dx.$$

It should be fairly obvious that a continuous distribution is just the limiting case of a discrete one. That is, if we let the number of events (possible choices) n tend to infinity, and the spacing between such events ϵ tend to zero (whilst keeping $n\epsilon$ finite); then all sums go over to integrals. Thus, the choice in using summations or integrals is just a notational convenience.

$$\sum_i p(x_i) \leftrightarrow \int p(x)dx.$$

We may also take the expectation value of some function of the random variable; so that

$$\langle f(X) \rangle = \int_{\forall x} f(x)p(x)dx.$$

1.2.2 Generating Function

We commonly take

$$f(x) = e^{ikx},$$

which is a function we denote the "characteristic function of X". We further denote

$$G(k) \equiv \left\langle e^{ikX} \right\rangle = \int_{\forall x} e^{ikx} p(x)dx. \tag{1.6}$$

This has the property that

$$|G(k)| \leq 1,$$

and that

$$G(0) = 1;$$

as one will notice that taking $k = 0$ just results in the normalisation condition. Also, taking the integration limits to $\pm\infty$ results an expression which is the fourier transform of the pdf

$$\int_{-\infty}^{\infty} e^{ikx} p(x) = \mathcal{F}\{p(x)\}.$$

Now, in (1.6), if we make the substitution that

$$e^y = \sum_{n=0}^{\infty} \frac{y^n}{n!},$$

then we see that

$$G(k) = \int_{\forall x} \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} x^n p(x) dx; \tag{1.7}$$

which, under a trivial reassignment, becomes

$$G(k) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \mu_n; \tag{1.8}$$

Where we have used (1.5) in identifying the moment-integral

$$\mu_n = \int x^n p(x) dx.$$

We see that, trivially, from the definition of μ_n, $\mu_0 = 1$ (as it is just normalisation). We have also suppresed the notation of $\forall x$ on the limit in the integral: it is implied unless otherwise specified.

One can see that from (1.8), if we differentiate the expression n-times, with respect to k, we will end up with μ_n (the factors of $n!$ cancel by differentiation). That is,

$$\langle X^n \rangle = \mu_n = \frac{1}{i^n} \left. \frac{d^n G}{dk^n} \right|_{k=0}. \tag{1.9}$$

Hence, if one posses the function $G(k)$, then one automatically has any moment one desires (via differentiation). Conversely, one can see that given the moments, if one computes the inverse fourier transform, one has the function $G(k)$. That is, we have a generating function for the moments. Thus, we call $G(k)$ the *generating function*.

Notice that the pdf and generating function relate via

$$p(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} G(k) e^{-ikx}. \tag{1.10}$$

1.2.3 Cumulants

We define

$$\ln G(k) = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \kappa_n, \tag{1.11}$$

where κ_n is the n^{th}-cumulant. Notice that the sum starts at unity (not zero, as before). This is because of the normalisation condition ($\kappa_0 = \mu_0 = 1$). Also in a similar way to before, we have

$$\kappa_n = \frac{1}{i^n} \frac{d^n}{dk^n} \ln G(k) \bigg|_{k=0}.$$

We can then compute the various cumulants

- Mean: $\kappa_1 = \mu_1$;
- Variance: $\kappa_2 = \mu_2 - \mu_1^2 = \sigma^2$;
- Skewness: $\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$;
- Kurtosis: $\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4$.

The Gaussian, or normal, pdf is given by

$$p(x) = \frac{1}{\sqrt{2\pi\kappa_2}} e^{-(x-\kappa_1)^2/2\kappa_2}, \tag{1.12}$$

and is the pdf for which $\kappa_n = 0$ for all $n \geq 3$.

The way I compute the κ_i is to use the function-of-a-function rule in the differential.

1.3 Multi-variable Probability Distribution Functions

Consider the case to begin with bivariate pdfs. That is, a system with two discrete stochastic random variables;

$$X(S) = \{x_1, x_2, \ldots\}; \qquad Y(S) = \{y_1, y_2, \ldots\}.$$

We shall denote the probability of event x_i and y_j occurring as $p(x_i, y_j)$. That is

$$p(x_i, y_j) \equiv P(x_i \cap y_j).$$

Thus, the normality condition is

$$\sum_{i,j} p(x_i, y_j) = 1.$$

We can define the *marginal pdf* as

$$p(x_i) = \sum_j p(x_i, y_j),$$

where we just sum over things we "aren't interested in". The version of Bayes' theorem is just

$$p(x_i|y_j) = \frac{p(x_i, y_j)}{p(y_j)};$$

That is, the probability of x_i occurring given that y_j has.

The continuous version of this is almost trivial:

$$\int_{\forall x} dx \int_{\forall y} dy p(x,y) = 1, \qquad p(x) = \int_c^d p(x,y)dy, \qquad p(y) = \int_a^b p(x,y)dx.$$

Bayes' theorem is just

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

where integrating this we can see the curious result

$$\begin{aligned} \int_{\forall x} p(x|y)dx &= \int \frac{p(x,y)}{p(y)}dx \\ &= \frac{1}{p(y)} \int p(x,y)dx \\ &= \frac{p(y)}{p(y)} \\ &= 1. \end{aligned}$$

We can also define some "conditional mean"

$$\langle X \rangle_{Y=y} \equiv \int_{\forall x} x p(x|y)dx.$$

Note the following integral

$$p(x|y) = \int_{\forall x'} \delta(x' - x) p(x'|y)dx' = \langle \delta(X - x) \rangle_{Y-y}.$$

Hence, we see that the expectation value of the delta-function is a conditional probability function.

The generalisation to arbitrary dimensions is fairly trivial. Given r stochastic variables, we have the pdf

$$p_r(x_1, x_2, \ldots, x_r).$$

Then, the probability that some subset $X_1, X_2, \ldots, X_s$ has certain values between $x_1 \to x_1 + dx_1, \ldots x_s \to x_s + dx_s$; regardless of the remaining $X_{s+1}, \ldots X_r$ (where we require $s < r$), is given by

$$p_s(x_1, x_2, \ldots, x_s) = \int dx_{s+1} dx_{s+1} \ldots dx_r p_r(x_1, x_2, \ldots, x_r). \tag{1.13}$$

This is the marginal probability density function for that subset. It is completely analogous to the bivariate case, where we integrate over the things that we dont care about, leaving us with the things that we do.

Also consider the conditional probability density function. Consider that $X_{s+1}, \ldots X_r$ have fixed values (i.e. they have values $x_{s+1}, \ldots, x_r$), then the probability that $x_1, \ldots x_s$ happen is just (by analogy to the bivariate case)

$$p_{s|r-s}(x_1, x_2, \ldots, x_s | x_{s+1}, \ldots, x_r) = \frac{p_r(x_1, x_2, \ldots, x_r)}{p_{r-s}(x_{s+1}, \ldots, x_r)}.$$

Again, by analogy, consider the expression

$$\int dx_1 dx_2 \ldots dx_s p_{s|r-s}(x_1, x_2, \ldots, x_s | x_{s+1}, \ldots, x_r) = \int dx_1 dx_2 \ldots dx_s \frac{p_r(x_1, x_2, \ldots, x_r)}{p_{r-s}(x_{s+1}, \ldots, x_r)} = 1;$$

Where we use (1.13) to integrate over the numerator (the denominator is a constant, as far as the integral is concerned). This also makes sense as we are integrating over the x's, whilst they are still unknown (all $x_i, i > s$ are fixed). This then leaves an integral over all unknowns, which should result in unity (normality).

1.4 Covariance

We define the covariance of two variables

$$Cov(X, Y) = \langle XY \rangle - \langle X \rangle \langle Y \rangle, \tag{1.14}$$

and it describes the relation of X with Y. We also write the covariance as

$$Cov(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle.$$

Similarly, we define the correlation function

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

1.5 The Central Limit Theorem

Suppose we are interested in estimating the mean of some stochastic random variable, X, and that we have made n independent observations $x_1, x_2, \ldots, x_n$. We estimate the mean via

$$\langle X \rangle = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

where this value will change as more values are taken. That is, the mean $\langle X \rangle$ is itself a stochastic variable. Thus, what is the probability density function of the mean?

The central limit theorem says that for a large enough sample n, the pdf is the normal (Gaussian) pdf. Let us prove it.

1.5.1 Proof

Let $X_1, X_2, \ldots, X_n$ be independent stochastic variables with the same mean $\mu = \mu_X$ and variance $\sigma^2 = \sigma_X^2$. Now, define some new stochastic quantity

$$Z \equiv \frac{X_1 + X_2 + \ldots + X_n}{n} - \mu,$$

and let us find its pdf.
Now, let

$$Z_i \equiv \frac{X_i - \mu}{n}; \quad Z = \sum_{i=1}^{n} Z_i, \tag{1.15}$$

and then the characteristic function for each Z_i is just

$$G_{Z_i}(k) = \int_{-\infty}^{\infty} e^{ikz_i} p(z_i) dz_i.$$

Now, the LHS of (1.15) leads us to trivially write that

$$G_{Z_i}(k) = \int_{-\infty}^{\infty} e^{ik(x_i - \mu)/n} p(x_i) dx_i. \tag{1.16}$$

Now, expanding the exponential term

$$e^w = \sum_{n=0}^{\infty} \frac{w^n}{n!} \quad \Rightarrow \quad e^{ik(x_i-\mu)/n} = 1 + \frac{ik(x_i - \mu)}{n} - \frac{k^2(x_i - \mu)^2}{2n^2} + \mathcal{O}\left(\tfrac{k^3}{n^3}\right).$$

Now, putting this back into (1.16), writing term by term

$$\begin{aligned} G_{Z_i}(k) = \quad & \int_{-\infty}^{\infty} p(x_i)dx_1 \\ & + \frac{ik}{n}\int_{-\infty}^{\infty}(x_i-\mu)p(x_i)dx_i \\ & - \frac{k^2}{2n^2}\int_{-\infty}^{\infty}(x_i-\mu)^2 p(x_i)dx_i \\ & + \mathcal{O}(\tfrac{k^3}{n^3}). \end{aligned} \tag{1.17}$$

Notice, if we have a very large sample (i.e. large n), then the higher-order terms are very small. Upon inspection, the second term vanishes

$$\begin{aligned} \int_{-\infty}^{\infty}(x_i-\mu)p(x_i)dx_i &= \int_{-\infty}^{\infty} x_i p(x_i)dx_i - \mu\int_{-\infty}^{\infty} p(x_i)dx_i \\ &= \mu - \mu = 0, \end{aligned}$$

due to normality, and that the moment μ is independent of n. Given this, the first and second terms of (1.17) are unity and zero, respectively. We also identify the integrand of the third term as $\langle (x_i-\mu)^2\rangle = \sigma_X^2$. Hence, we have that

$$G_{Z_i}(k) = 1 - \frac{k^2}{2n^2}\sigma_X^2 + \mathcal{O}(\tfrac{k^3}{n^3}),$$

which may be seen to be the expansion of the exponential

$$G_{Z_i}(k) = e^{-\frac{k^2}{2n^2}\sigma_X^2 + \mathcal{O}(\frac{k^3}{n^3})}.$$

Now, as the events are independent, the generating function of their sum is the product of the separate functions. That is

$$G_Z(k) = \prod_{i=1}^{n} G_{Z_i}(k),$$

which is just

$$G_Z(k) = e^{-\frac{k^2}{2n}\sigma_X^2 + \mathcal{O}(\frac{k^3}{n^2})}.$$

Therefore, in the limit $n \to \infty$, we ignore the terms $\mathcal{O}(n^{-2})$, giving the characteristic function

$$G_Z(k) = e^{-\frac{k^2}{2n}\sigma_X^2}.$$

And, as previously stated, the pdf is just the inverse Fourier transform of the characteristic function. Thus

$$p(x) = \int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{-\frac{k^2}{2n}\sigma_X^2},$$

which results in

$$p(x) = \sqrt{\frac{n}{2\pi\sigma_X^2}} e^{-\frac{nx^2}{2\sigma_X^2}};$$

nothing less than a Gaussian distribution.

We assumed in the proof that all measurements have the same mean and variance. If one removes the assumption, one is left with merely defining

$$\mu_i \equiv \langle X_i \rangle, \quad \sigma_i \equiv \left\langle (x_i - \mu_i)^2 \right\rangle, \quad \sigma_X^2 \equiv \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2.$$

1.5.2 Limitations of the Theorem

The central limit theorem holds if

- The moments are all independent of n and,
- The moments of $p(x_i)$ must be finite.

For example,

$$p(x_i) \equiv \frac{a_i}{\pi(a_i^2 + x_i^2)},$$

has an infinite second moment, and thus the central limit theorem is invalid in this case.

1.6 Time-dependent Systems

Most of the time we are interested in systems where $p(x,t)$; the probability that the stochastic variable X has value x at time t. That is, non-equilibrium systems. Thus, the formalism of the previous non-time-dependant section is easy to extend. The expectation value is just

$$\langle X(t) \rangle = \int_{\forall x} dx p(x,t),$$

the expectation value of a function of x

$$\langle f(X(t)) \rangle = \int_{\forall x} dx f(x) p(x,t),$$

and the n^{th} moment just

$$\langle X^n(t) \rangle = \int_{\forall x} dx x^n p(x,t).$$

Similarly, we can form joint probability distribution functions. Such as $p(x_1, t_1; x_2, t_2)$, which reads "the probability that X has value x_1 at time t_1, and value x_2 at time t_2". Also, we then see that

$$\langle X(t_1)X(t_2)\rangle = \int dx_1 dx_2 x_1 x_2 p(x_1, t_1; x_2, t_2). \tag{1.18}$$

Now, if the value of X at t_1 is independent of the value at t_2, then we can factorise the joint probability density

$$p(x_1, t_1; x_2, t_2) = p(x_1, t_1)p(x_2, t_2).$$

Thus, using this independent factorisation, we see that (1.18) may be written as

$$\begin{aligned}\langle X(t_1)X(t_2)\rangle &= \int dx_1 dx_2 x_1 x_2 p(x_1, t_1; x_2, t_2) \\ &= \int dx_1 x_1 p(x_1, t_1) \int dx_2 x_2 p(x_2, t_2) \\ &= \langle X(t_1)\rangle \langle X(t_2)\rangle .\end{aligned}$$

Notice that this gives the result that the covariance (or correlation, depending on who you speak to) of two independent quantities is zero.

This joint probability density fairly simply generalises to more than two values;

$$p(x_1, t_1; \ldots; x_i, t_i; \ldots; x_n, t_n)$$

is the probability that X has the value x_i at time t_i.
To get marginal probability densities, we can still just integrate out the things we aren't interested in, thus

$$p(x_1, t_1; \ldots; x_m, t_m) = \int dx_{m+1} \ldots dx_n p(x_1, t_1; \ldots; x_n, t_n), \quad m < n.$$

To get conditional probabilities,

$$p(x_1, t_1; \ldots; x_s, t_s | x_{s+1}, t_{s+1}; \ldots; x_n, t_n) = \frac{p(x_1, t_1; \ldots; x_n, t_n)}{p(x_{s+1}, t_{s+1}; \ldots; x_n, t_n)}.$$

If we just read out what this is saying: the probabilty of X having the values $x_1, t_1, \ldots x_s, t_s$, given that it already has the values $x_{s+1}, t_{s+1}, \ldots, x_n, t_n$, is the total probability divided by the probability that those things happened (those that we said have occurred).

We can also have conditional means,

$$\langle X(t_2)\rangle_{X(t_1)=x_1} = \int dx_2 x_2 p(x_2, t_2 | x_1, t_1);$$

which reads as the expectation value of X, at time t_2, given that it had value x_1 at time t_1. In a similar way to before, we see that

$$p(x_2, t_2 | x_1, t_1) = \langle \delta(X(t_2) - x_2)\rangle_{X(t_1)=x_1} .$$

As some more notation, we shall write cumulants with double brackets. That is,

$$\left\langle\left\langle X^2\right\rangle\right\rangle = \left\langle X^2\right\rangle - \langle X\rangle^2 , \quad \left\langle\left\langle X^3\right\rangle\right\rangle = \left\langle X^3\right\rangle - 3\langle X\rangle^2 \langle X\rangle + 2\langle X\rangle^3 .$$

1.6.1 Stationary Processes: Definition

We define a *stationary process* as one whose probability densities depend on time differences alone. Thus, that means that

$$p(x_n, t_n + \tau; \ldots; x_1, t_1 + \tau) = p(x_n, t_n; \ldots; x_1, t_1), \quad \forall n, \tau.$$

Thus, for example, taking $\tau = -t_1$, then

$$p(x_1, t_1 - t_1) = p(x_1, t_1) \quad \Rightarrow \quad p(x_1, t_1) = p(x_1, 0),$$

leaving the resultant probability time independent. As another example, consider

$$p(x_1, t_1; x_2, t_2) = p(x_1, 0; x_2, t_2 - t_1) = p(x_1, t_1 - t_2; x_2, 0).$$

Therefore notice that as

$$\langle X(t_1)X(t_2)\rangle = \int dx_1 dx_2 x_1 x_2 p(x_1, t_1; x_2, t_2)$$

is symmetric under $t_1 \leftrightarrow t_2$, it therefore only depends upon $|t_1 - t_2|$ when the associated process is stationary.

1.6.2 Gaussian Process: Definition

A process is Gaussian if all cumulants beyond second are zero. Then, such a process is fully specified by

$$\langle X(t_1)X(t_2)\rangle\,, \quad \langle X(t_1)\rangle\,.$$

2 Markov Processes

2.1 Introduction

A Markov process is one whose conditional probability density functions are only affected by the state of the system at a given time, and not by the states of the system at time prior to that time. That is,

$$p(x_{k+1}, t_{k+1}|x_k, t_k; \ldots; x_1, t_1)$$

depends on the state $X(t_k) = x_k$, but not the states $X(t_{k-1}) = x_{k-1}, \ldots, X(t_1) = x_1$, for all states k. Thus,

$$p(x_{k+1}, t_{k+1}|x_k, t_k; \ldots; x_1, t_1) = p(x_{k+1}, t_{k+1}|x_k, t_k). \tag{2.1}$$

Such processes may colloquially be thought of as those which "only remember current state of system, and not all previous states".

Let us see how to use this assumption. Consider the joint probability density function

$$\begin{aligned} &p(x_n, t_n; \ldots; x_1, t_1) \\ &= p(x_n, t_n|x_{n-1}t_{n-1}; \ldots; x_1, t_1)p(x_{n-1}, t_{n-1}; \ldots; x_1t_1), \end{aligned} \tag{2.2}$$

where we have used Bayes' theorem to write it in this way. Let us use (2.1), the "Markov assumption", to write the first expression on the RHS of (2.2) as

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

Let us use Bayes' theorem again, on the second expression on the RHS of (2.2)

$$\begin{aligned} &p(x_{n-1}, t_{n-1}; \ldots; x_1t_1) \\ &= p(x_{n-1}, t_{n-1}|x_{n-2}, t_{n-2}; \ldots; x_1, t_1)p(x_{n-2}, t_{n-2}; \ldots; y_1, t_1). \end{aligned}$$

Therefore, using these two expressions, we see that we can write (2.2) as

$$\begin{aligned} &p(x_n, t_n; \ldots; x_1, t_1) \\ &= p(x_n, t_n|x_{n-1}t_{n-1})p(x_{n-1}, t_{n-1}|x_{n-2}, t_{n-2}; \ldots; x_1, t_1)p(x_{n-2}, t_{n-2}; \ldots; y_1, t_1). \end{aligned}$$

Now, we see that we can repeat this whole process of using Bayes' theorem & the Markov assumption, on the two expressions on the far-RHS of the above. In doing so, we end up with

$$p(x_n, t_n; \ldots; x_1, t_1) = \prod_{i=1}^{n-1} p(x_{i+1}, t_{i+1}|x_i, t_i)p(x_1, t_1). \tag{2.3}$$

As another example, consider the conditional probability density function

$$p(x_{k+l}t_{k+l};\ldots;x_{k+1},t_{k+1}|x_k,t_k;\ldots;x_1,t_1)=\frac{p(x_{k+l},t_{k+l};\ldots;x_1,t_1)}{p(x_k,t_k;\ldots;x_1,t_1)},$$

after using Bayes' theorem.
Now, using our derived result, (2.3), we see that we can re-write the numerator & denominator as, using the implicit Markov assumption (in that we are allowed to do this at all)

$$\begin{aligned}p(x_{k+l},t_{k+l};\ldots;x_1,t_1) &= \prod_{i=1}^{k+l-1}p(x_{i+1}t_{i+1}|x_i,t_i)p(x_1,t_1),\\ p(x_k,t_k;\ldots;x_1,t_1) &= \prod_{i=1}^{k-1}p(x_{i+1}t_{i+1}|x_i,t_i)p(x_1,t_1).\end{aligned}$$

Therefore, using these, our original expression reads

$$p(x_{k+l}t_{k+l};\ldots;x_{k+1},t_{k+1}|x_k,t_k;\ldots;x_1,t_1)=\frac{\prod_{i=1}^{k+l-1}p(x_{i+1}t_{i+1}|x_i,t_i)p(x_1,t_1)}{\prod_{j=1}^{k-1}p(x_{j+1}t_{j+1}|x_j,t_j)p(x_1,t_1)},$$

some terms of which we see cancel down to leave

$$p(x_{k+l}t_{k+l};\ldots;x_{k+1},t_{k+1}|x_k,t_k;\ldots;x_1,t_1)=\prod_{i=k}^{k+l-1}p(x_{i+1}t_{i+1}|x_i,t_i). \tag{2.4}$$

Let us just write down the main results, again, together. These results only work when using the Markov assumption for the associated process.

$$p(x_n,t_n;\ldots;x_1,t_1)=\prod_{i=1}^{n-1}p(x_{i+1},t_{i+1}|x_i,t_i)p(x_1,t_1), \tag{2.5}$$

$$p(x_{k+l}t_{k+l};\ldots;x_{k+1},t_{k+1}|x_k,t_k;\ldots;x_1,t_1)=\prod_{i=k}^{k+l-1}p(x_{i+1}t_{i+1}|x_i,t_i). \tag{2.6}$$

We can apply these results to rather easy (but useful) examples.

Consider that

$$p(x_2,t_2)=\int dx_1 p(x_2,t_2;x_1,t_1),$$

but, using (2.5), we can rewrite the integrand as

$$p(x_2,t_2;x_1,t_1)=\prod_{i=1}^{1}p(x_{i+1},t_{i+1}|x_1,t_1)p(x_1,t_1),$$

thus leaving us with

$$p(x_2, t_2) = \int dx_1 p(x_2, t_2|x_1, t_1)p(x_1, t_1), \tag{2.7}$$

which is a useful (if obvious) result.

Next, consider how to prove that

$$p(x_3, t_3|x_1, t_1) = \int dx_2 p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1).$$

Now, using (2.5), we see that we have

$$p(x_3, t_3; x_2, t_2; x_1, t_1) = p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1)p(x_1, t_1),$$

integrating both sides with respect to x_2,

$$\int dx_2 p(x_3, t_3; x_2, t_2; x_1, t_1) = \int dx_2 p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1)p(x_1, t_1),$$

$$\Rightarrow \quad p(x_3, t_3; x_1, t_1) = \left(\int dx_2 p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1) \right) p(x_1, t_1).$$

Using Bayes' theorem on the LHS,

$$p(x_3, t_3; x_1, t_1) = p(x_3, t_3|x_1, t_1)p(x_1, t_1),$$

and therefore we see that

$$p(x_3, t_3|x_1, t_1) = \int dx_2 p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1) \tag{2.8}$$

which is thus proven (we cancelled off the common $p(x_1, t_1)$ factor).

The derived result (2.8) is called the *Chapman-Kolmogorov* (CK) equation. The CK equation tells us that in order to find the probability of transitions from $x_1 \to x_3$, we integrate over all possible ways of getting there, via x_2. That we integrate over x_2 is that we add up all possible ways of getting from x_1 at time t_1 to state x_3 at time t_3, via any possible intermediate x_2 (the integral sweeps over all possible intermediate states).

2.2 Markov Chains

Here we consider state variables to be discrete, denoted n. The Markov equations (2.7) and (2.8) are (moving to summations for discrete variables)

$$p(n_2, t_2) = \sum_{n_1} p(n_2, t_2|n_1, t_1)p(n_1, t_1), \tag{2.9}$$

$$p(n_3, t_3|n_1, t_1) = \sum_{n_2} p(n_3, t_3|n_2, t_2)p(n_2, t_2|n_1, t_1), \tag{2.10}$$

where we have the ordering $t_1 < t_2 < t_3$. The time parameter t is also discrete for Markov chains. We call the final equation, (2.10), the *Chapman-Kolmogorov equation*. Thus, under a trivial redefinition of the symbols, this reads

$$p(n, t+2|m, t) = \sum_{n'} p(n, t+2|n', t+1)p(n', t+1|m, t).$$

If we "say" what this equation is: the probability that the system is in state n, at time $t+2$, given that it was in state m at time t, is the sum over all possible intermediate states n' (where the system was in some state n' at time $t+1$) multiplied by the probability that the system started in state m at time t.
That is, if the system is going from some state m to n; starting at time t, ending at time $t+2$, then we sum over all possible intermediate states n' (which occur at time $t+1$), and then find the probability that the system moved from that intermediate state (i.e. n') to the state we are interested in (i.e. n).

By way of making this lot easier to denote, let us define

$$Q_{mn}(t) \equiv P(m, t+1|n, t), \tag{2.11}$$

as the probability that the system transitions at time t state n to state m, at time $t+1$. Thus, $Q_{mn}(t)$ describes a *transition probability matrix*, generally being a function of time.

If we use this notation in (2.9), as well as $P_n(t)$ denoting the same as $p(n, t)$ (i.e. the probability that the system is in state n at time t), then we have

$$P_n(t+1) = \sum_m Q_{nm}(t)P_m(t). \tag{2.12}$$

Therefore, using this, and given the transition matrix & state of the system at time t, one can find the probability that the system is in state n at time $t+1$.

Let us consider some examples, and how to formulate the transition matrix.

2.2.1 Random One-Dimensional Walk

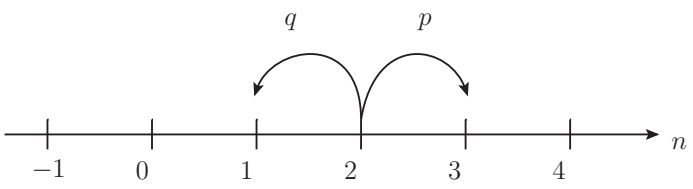

Figure 2.1: The 1D random walker. Starting at some position n_0, there is an associated probability of moving one step to the left or right.

Consider a line, which has been discretely divided up into bits, each of unit length. Each bit of the line has an associated number n. Let $n = 0, \pm 1, \pm 2, \ldots, \pm\infty$.

Now, suppose that a "walker" starts at a given point n_0, say, and moves a single place to either the left or right. The probability of moving left is q and to the right p. Then, we have that $p + q = 1$. Here, we model the probabilities as being independent of time.

So then, what is the probability of the walker being at site n at time t?

The way we have formulated this problem is in a "nearest neighbour" way, but there are many ways in which one could. That is, we have stipulated that if we start at n', then we can only move to $n = n' \pm 1$ in a single step.

So, the transition probability is

$$Q_{nn'} = \begin{cases} p, & n = n' + 1, \\ q, & n = n' - 1, \\ 0 & \text{else.} \end{cases}$$

Then, we see that if we write Q as a matrix, this simply reads

$$Q = \begin{pmatrix} 0 & q & 0 & 0 & \ldots \\ p & 0 & q & 0 & \\ 0 & p & 0 & q & \\ 0 & 0 & p & 0 & \\ \vdots & & & & \ddots \end{pmatrix},$$

where the dimension of Q will simply be the allowed values of n. We could also setup the problem so that the walker could stay put. That is, with probability r, the walker ends up at $n = n'$; in this case, we have

$$Q_{nn'} = \begin{cases} p, & n = n' + 1, \\ q, & n = n' - 1, \\ r, & n = n', \\ 0 & \text{else.} \end{cases} \quad \Rightarrow \quad Q = \begin{pmatrix} r & q & 0 & 0 & \ldots \\ p & r & q & 0 & \\ 0 & p & r & q & \\ 0 & 0 & p & r & \\ \vdots & & & & \ddots \end{pmatrix}.$$

Given this extra probability, we just need to make sure that $p + q + r = 1$.

Just to hammer the point home; $Q_{nn'}$ is the probability that the system transitions from state n' to state n, between times t and $t + 1$.

In this way, we can predict the state of the system at any time, given an initial state. That is, we can compute $P_n(t)$, the probability that the system is in the state n at time t, given the system was in state n at time $t = 0$, for all n.

Typically, suppose the system is in a pure state n_0 at time $t = 0$, then we have the initial condition that

$$P_n(0) = \delta_{nn_0},$$

where we notice that

$$\sum_n P_n(0) = \sum_n \delta_{nn_0} = 1,$$

confirming normalisation.

2.2.2 Urn Models

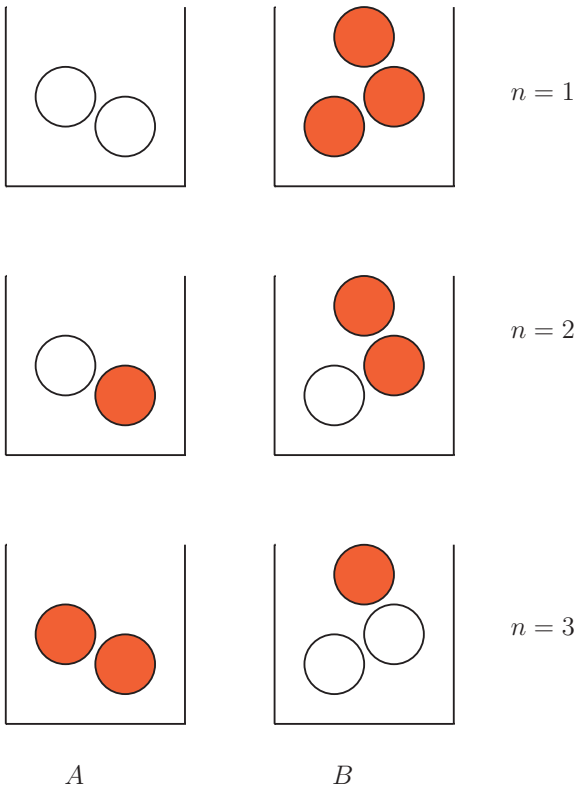

Figure 2.2: The possible configurations in the urn model. All pots on the left are A, and those on the right B. Notice that there are two types of "ball": red and white.

Suppose we have 2 pots, A and B, with 3 red balls and 2 white balls. The balls are to be distributed so that A always contains 2 balls, and B always 3 balls. Let us write down the states of the system, given these simple rules. We denote a white ball by W, and red by R.

- A contains 2W; B has 3R: $\{(W, W), (R, R, R)\}$. Denote this as $n = 1$.
- A contains 1W; 1R; B has 1W, 2R: $\{(W, R), (W, R, R)\}$. Denote this as $n = 2$.
- A contains 2R; B has 2W, 1R: $\{(R, R), (R, W, W)\}$. Denote this as $n = 3$.

So, what are the transition probabilities? We shall be picking up one ball from each of A and B, and replacing the ball into the other pot.

Suppose we start in the state $n = 1$. There is only one other state we can possibly go to, by our dynamics rule. Thus, the probability of going from state $n = 1$ to $n = 1$ is 0, from $1 \to 2$ is 1, from $1 \to 3$ is 0. Thus, we easily see that

$$Q_{11} = 0, \quad Q_{21} = 1, \quad Q_{31} = 0.$$

Suppose we start in state $n = 2$. We can go $2 \to 1$ by picking a W from pot B (with probability 1/3) and the R from pot A (with probability 1/2). Then, the probability is just $1/3 \times 1/2 = 1/6$. To make the transition $2 \to 2$, we require picking up a W from each (with probability 1/2 from A and 1/3 from B) or picking an R from each (1/2 from A, and 2/3 from B). Thus, the total probability of transition is $1/2 \times 1/3 + 1/2 \times 2/3 = 1/2$. Thus, $Q_{22} = 1/2$. To make the transition $2 \to 3$, we require picking the W from A, and R from B; thus probability of $Q_{32} = 1/2 \times 2/3 = 1/3$. Hence

$$Q_{12} = 1/6, \quad Q_{22} = 1/2, \quad Q_{32} = 1/3.$$

And finally, by similar logic, we can compute that

$$Q_{13} = 0, \quad Q_{23} = 2/3, \quad Q_{33} = 1/3.$$

And therefore, the transition matrix is

$$Q = \begin{pmatrix} 0 & 1/6 & 0 \\ 1 & 1/2 & 2/3 \\ 0 & 1/3 & 1/3 \end{pmatrix}.$$

Notice that columns add to unity (which they should).

So, consider that the system starts off in a pure state. Say it is in state $n = 3$. Then,

$$P(0) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Hence, the probability of where the system is, at time $t = 1$ is given by the multiplication

$$P(1) = QP(0) = \begin{pmatrix} 0 & 1/6 & 0 \\ 1 & 1/2 & 2/3 \\ 0 & 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2/3 \\ 1/3 \end{pmatrix}.$$

Further consider what the distribution of probabilities is, of the state of the system, after another time step

$$P(2) = QP(1) = Q^2P(0) = \begin{pmatrix} 1/9 \\ 20/36 \\ 1/3 \end{pmatrix};$$

That is, if the system starts off in state $n = 3$, then, after 2 "goes", one will find it in state $n = 1$ with probability 1/9, state $n = 2$ with probability 20/36, and state $n = 3$ with probability 1/3.

2.3 Stochastic Matrices

One may ask: is there a state that the system settles down in? That is, does

$$P(t) = Q^t P(0)$$

have one element which completely overwhelms all the others? Let us consider this. Infact, this statement only works if the transition matrix is independent of time. To continue, we must have a little aside on right- and left-eigenvectors of matrices.

2.3.1 Right & Left Eigenvectors

Let us proceed to calculate the right and left eigenvectors of some matrix. We will do this by example. Consider the matrix

$$M = \begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix},$$

which clearly is not a stochastic matrix (its columns don't sum to unity, and it has negative entries).

So, we compute its eigenvalues in the usual way; subtract λ from the diagonals, and set the determinant to zero;

$$\begin{vmatrix} -1-\lambda & 2 \\ -3 & 4-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \lambda^2 - 3\lambda + 2 = 0.$$

Thus, the characteristic equation (the one on the right, above) has solutions

$$\lambda^{(1)} = 1, \quad \lambda^{(2)} = 2.$$

We find what we will now call the *right eigenvectors* the way we "normally" find eigenvectors. That is, by solving

$$\begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}.$$

We shall denote the right eigenvector as $\psi^{(i)}$, corresponding to eigenvalue $\lambda^{(i)}$. So, we can fairly easily find that

$$\lambda^{(1)} = 1 \quad \Rightarrow \quad \psi^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and

$$\lambda^{(2)} = 2 \quad \Rightarrow \quad \psi^{(2)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Now, to find the *left eigenvectors* $\chi^{(i)}$, which corresponds to eigenvector $\lambda^{(i)}$, we solve

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix} = \lambda^{(i)} \begin{pmatrix} x & y \end{pmatrix},$$

for each $\lambda^{(i)}$. So, for $\lambda^{(1)} = 1$, we see that

$$-x - 3y = x \quad \Rightarrow \quad (3 \quad -2),$$

and that for $\lambda^{(2)} = 2$, we see that

$$x = -y \quad \Rightarrow \quad (1 \quad -1).$$

Now, one will notice that these eigenvectors are rows, rather than columns. We therefore denote the things above, $(\chi^{(i)})^T$. That is,

$$(\chi^{(1)})^T = (3 \quad -2) \quad \Rightarrow \quad \chi^{(1)} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$(\chi^{(2)})^T = (1 \quad -1) \quad \Rightarrow \quad \chi^{(2)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Now, by way of convenient notation, we denote right eigenvectors as kets $|\psi^{(i)}\rangle$, and left eigenvectors as bras $\langle\chi^{(i)}|$. Thus, notice the orthogonality of the two sets of eigenvectors:

$$\langle\chi^{(1)}|\psi^{(2)}\rangle = 0, \quad \langle\chi^{(2)}|\psi^{(1)}\rangle = 0.$$

2.3.2 Properties of Stochastic Matrices

As we saw with our Urn-model example, the columns of a stochastic matrix sum to unity. This is obviously the case, as something must happen. Mathematically, using the definition of Q, we see that this corresponds to

$$\sum_n Q_{nm} = \sum_n P(n, t+1|m, t) = 1,$$

so that the probability that the state transitions from state m at time t, to any other state n at time $t+1$, is unity.

Also, all entries of a stochastic matrix are non-negative. This is fairly obvious, as probabilities are non-negative;

$$Q_{nm} \geq 0.$$

Following this, if Q_1 & Q_2 are two stochastic matrices, then so is their product $Q_1 Q_2$. Furthermore, if Q is a stochastic matrix, then so is any power of that matrix, Q^t.

In our brief example, we saw that an eigenvalue $\lambda = 1$ appeared. This is a general property of stochastic matrices. All stochastic matrices have one eigenvalue which is unity. Corresponding to that eigenvalue, the left eigenvector is "unit". That is,

$$\lambda^{(1)} = 1, \quad (\chi^{(1)})^T = (1 \quad 1 \quad \ldots \quad 1).$$

This is called the *stationary state*. The eigenvalue equation for this is just

$$\sum_n \chi_n^{(1)} Q_{nm} = 1.\chi_m^{(1)}.$$

Note that our brief example was not a stochastic matrix, and therefore this left eigenvector did not appear.

If the system approaches a time-independent state, $P^{\rm st}$, say, then this is an equilibrium state. This means that no matter what the initial state of the system, it will tend towards some stationary state. That is,

$$P(t+1) = QP(t),$$

which is just

$$P^{\rm st} = QP^{\rm st},$$

which is merely an eigenvalue equation, corresponding to eigenvalue 1. That is,

$$QP^{\rm st} = 1.P^{\rm st}.$$

Therefore, for the eigenvalue $\lambda = 1$, we see that the right eigenvector is $P^{\rm st}$, and the left $\chi^{(1)}$, whereby their product is unity,

$$(1 \quad 1 \quad \dots \quad 1) \begin{pmatrix} P_1^{\rm st} \\ P_2^{\rm st} \\ \vdots \end{pmatrix} = \sum_n P_n^{\rm st} = 1.$$

Finally, all eigenvalues of a stochastic matrix have modulus ≤ 1,

$$|\lambda^{(i)}| \leq 1, \quad \forall i.$$

If the matrix is symmetric, $Q^T = Q$, then the left and right eigenvectors are the same.

2.3.3 Example: Urn Model

Let us compute the right & left eigenvectors for our previous transition matrix

$$Q = \begin{pmatrix} 0 & 1/6 & 0 \\ 1 & 1/2 & 2/3 \\ 0 & 1/3 & 1/3 \end{pmatrix},$$

for the Urn model. We compute eigenvalues via

$$\begin{vmatrix} -\lambda & 1/6 & 0 \\ 1 & 1/2-\lambda & 2/3 \\ 0 & 1/3 & 1/3-\lambda \end{vmatrix} = 0,$$

which results in the characteristic equation

$$\lambda^3 - \tfrac{5}{6}\lambda^2 - \tfrac{2}{9}\lambda + \tfrac{1}{18} = 0.$$

To solve this cubic, we first note that we know that one factor is $\lambda = 1$ (it being a stochastic matrix implies that there is one unit eigenvalue). Thus, we factorise;

$$(\lambda - 1)(\lambda^2 + a\lambda - \tfrac{1}{18}) = 0.$$

If we expand this out, and compare the powers of λ^2 with the original characteristic equation, we find that $a = \frac{1}{6}$. Therefore,

$$(\lambda - 1)(\lambda^2 + \tfrac{1}{6}\lambda - \tfrac{1}{18}) = 0,$$

which further factorises to

$$(\lambda - 1)(\lambda + \tfrac{1}{3})(\lambda - \tfrac{1}{6}) = 0.$$

Therefore, the 3 eigenvalues are

$$\lambda^{(1)} = 1, \quad \lambda^{(2)} = -\tfrac{1}{3}, \quad \lambda^{(3)} = \tfrac{1}{6}.$$

Now, corresponding to $\lambda^{(1)} = 1$, we know that the left eigenvector is just

$$(\chi^{(1)})^T = (1 \quad 1 \quad 1), \quad \lambda^{(1)} = 1.$$

So, to find the corresponding right eigenvector, $\psi^{(1)}$, we solve in the usual way;

$$\begin{pmatrix} 0 & 1/6 & 0 \\ 1 & 1/2 & 2/3 \\ 0 & 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1. \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

to give

$$\psi^{(1)} = \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix}.$$

This is the stationary state; but, to correctly normalise it so that $\langle \chi^{(1)} | \psi^{(1)} \rangle = 1$, we note that

$$(1 \quad 1 \quad 1) \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix} = 10; \quad \psi^{(1)} = P^{\text{st}} = \begin{pmatrix} 1/10 \\ 3/5 \\ 3/10 \end{pmatrix}.$$

And thus, we have found the stationary state. The other eigenvectors are fairly easily found to be

$$\lambda^{(2)} = -\tfrac{1}{3} \quad \Rightarrow \quad \psi^{(2)} = \begin{pmatrix} 1/6 \\ -1/3 \\ 1/6 \end{pmatrix}, \quad \chi^{(2)} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix};$$

$$\lambda^{(3)} = \tfrac{1}{6} \quad \Rightarrow \quad \psi^{(3)} = \begin{pmatrix} 4/15 \\ -4/15 \\ 8/15 \end{pmatrix}, \quad \chi^{(3)} = \begin{pmatrix} -3/2 \\ -1/4 \\ 1 \end{pmatrix}.$$

Remember that we get the $\chi^{(i)}$ in the form of a row vector, $(\chi^{(i)})^T$. To continue, we make the identification

$$\psi^{(i)} \longmapsto |\psi^i\rangle, \quad (\chi^{(i)})^T \longmapsto \langle\chi^{(i)}|.$$

And therefore, using this notation, we see orthonormality;

$$\langle\chi^{(1)}|\psi^{(3)}\rangle = 0.$$

Also note, as $\psi^{(2,3)}$ are orthogonal to $\chi^{(1)}$, their entries must sum to zero (which they do).

2.3.4 General Theory

We now prove various relations, in a very similar fashion to quantum theory. This theory does not rely on the matrix being stochastic; it is true for any matrix.

Orthogonality Let Q be an $m \times m$ matrix. Then, the eigenvalue equations (for both right & left eigenvectors) are

$$Q|\psi^{(i)}\rangle = \lambda^{(i)}|\psi^{(i)}\rangle, \quad \langle\chi^{(i)}|Q = \lambda^{(i)}\langle\chi^{(i)}|.$$

Then, forming the product with a bra-state on the first, and a ket-state on the second, we have that

$$\langle\chi^{(j)}|Q|\psi^{(i)}\rangle = \lambda^{(i)}\langle\chi^{(j)}|\psi^{(i)}\rangle, \quad \langle\chi^{(i)}|Q|\psi^{(j)}\rangle = \lambda^{(i)}\langle\chi^{(i)}|\psi^{(j)}\rangle.$$

Swapping the indices over, on the second expression, results in

$$\langle\chi^{(j)}|Q|\psi^{(i)}\rangle = \lambda^{(j)}\langle\chi^{(j)}|\psi^{(i)}\rangle.$$

Therefore, subtracting, we have

$$0 = \left(\lambda^{(i)} - \lambda^{(j)}\right)\langle\chi^{(j)}|\psi^{(i)}\rangle.$$

So, if $\lambda^{(i)} \neq \lambda^{(j)}$, then this simply reads

$$\langle\chi^{(j)}|\psi^{(i)}\rangle = 0, \quad \lambda^{(i)} \neq \lambda^{(j)}, \quad i \neq j.$$

If we have chosen normality, then we can say that

$$\langle\chi^{(j)}|\psi^{(i)}\rangle = \delta_{ij}, \tag{2.13}$$

which is the statement of orthonormality.

Completeness We can expand an arbitrary state (probability) into eigenstates of the matrix;

$$|P\rangle = \sum_{i=1}^{m} \alpha_i |\psi^{(i)}\rangle.$$

So, forming the product of this with a bra-state gives

$$\langle\chi^{(j)}|P\rangle = \sum_{i=1}^{m} \alpha_i \langle\chi^{(j)}|\psi^{(i)}\rangle,$$

which, by our orthonormality statement, is simply

$$\langle\chi^{(j)}|P\rangle = \sum_{i=1}^{m} \alpha_i \langle\chi^{(j)}|\psi^{(i)}\rangle = \sum_{i=1}^{m} \alpha_i \delta_{ij} = \alpha_j.$$

Therefore, we have the coefficients:

$$\alpha_i = \langle\chi^{(i)}|P\rangle. \tag{2.14}$$

Then, using this in our original expansion,

$$|P\rangle = \sum_{i=1}^{m} |\psi^{(i)}\rangle \alpha_i = \sum_{i=1}^{m} |\psi^{(i)}\rangle\langle\chi^{(i)}|P\rangle,$$

which leads us to state that

$$\sum_i |\psi^{(i)}\rangle\langle\chi^{(i)}| = 1. \tag{2.15}$$

2.3.5 Powers of a Matrix

If we multiply (2.15) by Q, from the LHS, we have

$$Q \sum_i |\psi^{(i)}\rangle\langle\chi^{(i)}| = 1.Q \quad \Rightarrow \quad Q = \sum_i \lambda^{(i)} |\psi^{(i)}\rangle\langle\chi^{(i)}|.$$

Now, consider that

$$Q^2|\psi^{(i)}\rangle = QQ|\psi^{(i)}\rangle = Q\lambda^{(i)}|\psi^{(i)}\rangle = \lambda^2|\psi^{(i)}\rangle,$$

then it is clear that

$$Q^t|\psi^{(i)}\rangle = \left(\lambda^{(i)}\right)^t |\psi^i\rangle.$$

Then, we easily see that

$$Q^t = \sum_i \left(\lambda^{(i)}\right)^t |\psi^{(i)}\rangle\langle\chi^{(i)}|.$$

Just to make notation a little easier, we shall express this as

$$Q^t = \sum_i \lambda_{(i)}^t |\psi^{(i)}\rangle\langle\chi^{(i)}|. \tag{2.16}$$

Therefore, we have a way in which we can compute the powers of a matrix; by forming the products of right with left eigenvectors.

Example: Urn Model Let us return to the Urn models matrix & eigenvectors, to compute an arbitrary power of the matrix.

So, for $\lambda^{(1)} = 1$, we see that

$$|\psi^{(1)}\rangle\langle\chi^{(1)}| = \begin{pmatrix} 1/10 \\ 3/5 \\ 3/10 \end{pmatrix} (1 \quad 1 \quad 1) = \begin{pmatrix} 1/10 & 1/10 & 1/10 \\ 3/5 & 3/5 & 3/5 \\ 3/10 & 3/10 & 3/10 \end{pmatrix} \equiv Q_1.$$

For $\lambda^{(2)} = -1/3$, we have that

$$|\psi^{(2)}\rangle\langle\chi^{(2)}| = \begin{pmatrix} 1/6 \\ -1/3 \\ 1/6 \end{pmatrix} (3 \quad -1 \quad 1) = \begin{pmatrix} 1/2 & -1/6 & 1/6 \\ -1 & 1/3 & -1/3 \\ 1/2 & -1/6 & 1/6 \end{pmatrix} \equiv Q_2.$$

And finally, for $\lambda^{(3)} = 1/6$, we see that

$$|\psi^{3}\rangle\langle\chi^{3}| = \begin{pmatrix} 2/5 & 1/15 & -4/15 \\ 2/5 & 1/15 & -4/15 \\ -4/5 & -2/15 & 8/15 \end{pmatrix} \equiv Q_3.$$

So, an arbitrary power of Q may be found from

$$Q^t = Q_1 + (-\tfrac{1}{3})^t Q_2 + (\tfrac{1}{6})^t Q_3.$$

Now, we can notice a few things from this. First, note that the smallest eigenvalue will have little effect on the late-time behavior of the system. That is, for high t, the last term will be negligible. The next largest eigenvalue gives the dominant large t behavior. Second, notice that if any of the eigenvalues had been > 1, then the system would have diverged. Finally, the very-large t behavior is completely determined by Q_1.

Thus, we have a way of computing

$$P(t) = Q^t P(0).$$

2.4 Examples of Markov Chains

Here we shall consider some more transition matrices, and how boundary conditions effect their structure.

The Gamblers' Ruin Suppose a gambler starts out with $£n_0$, and makes a series of $£1$ bets against the house. The probability of winning each bet is p, and of loosing is $q = 1 - p$. If the gamblers capitol ever reaches zero, he is ruined & stops playing; he remains at zero.

This is obviously a random walk; but with an *absorbing boundary* at $n = 0$. As it stands, the transition matrix is of infinite dimension: there is no upper-stopping point. A variant

could be that once the gambler reaches a given amount, $£N$, he stops again. Thus, there are now two absorbing boundaries.

The transition matrix, naively (i.e. incorrectly) is

$$Q = \begin{pmatrix} 0 & q & 0 & 0 & & \dots & \\ p & 0 & q & 0 & & & \\ 0 & p & 0 & q & 0 & & \\ \vdots & & & \ddots & & & \\ & & & & 0 & q & 0 \\ & & & & p & 0 & q \\ & & \dots & & 0 & p & 0 \end{pmatrix}.$$

Now, one will notice that the first & last column do not add to unity. Also, consider that the absorbing boundary is at positions $n = 1, N$. Then, as the matrix stands, the element $Q_{21} = p$. That is, there is a probability of the walker moving from state $1 \to 2$. This is in contradiction with our absorbing boundary. Also, the matrix currently has $Q_{N-1,N} = q$; again, this is in contradiction with our absorbing boundary. The elements mentioned should be set to zero; and the element $Q_{11} = Q_{NN} = 1$. Thus,

$$Q = \begin{pmatrix} 1 & q & 0 & 0 & & \dots & \\ 0 & 0 & q & 0 & & & \\ 0 & p & 0 & q & 0 & & \\ \vdots & & & \ddots & & & \\ & & & & 0 & q & 0 \\ & & & & p & 0 & 0 \\ & & \dots & & 0 & p & 1 \end{pmatrix}.$$

is the correct transition matrix, for the two absorbing boundaries case.

Random Walk with Reflecting Boundaries Suppose that now, at a boundary, the walker "bounces off" or stays put. That is, if the walker is at the boundary, then on the next "go", there is a probability that he either moves off it, or stays on it; there is no probability that he moves through the barrier. Then, given that p is the probability of moving left, and q right, we see that we must have $Q_{11} = q$: it must move right at the boundary. Similarly, at the other boundary, $Q_{NN} = p$: it must move left. Therefore, this transition matrix looks like

$$Q = \begin{pmatrix} q & q & 0 & 0 & & \dots & \\ p & 0 & q & 0 & & & \\ 0 & p & 0 & q & 0 & & \\ \vdots & & & \ddots & & & \\ & & & & 0 & q & 0 \\ & & & & p & 0 & q \\ & & \dots & & 0 & p & p \end{pmatrix}.$$

Birth & Death Processes Suppose that at $t = 0$, there are n_0 bacteria in a colony. At each time step there is a probability μ that one dies, and λ that one lives.

This is clearly another example of a random walk.

2.4.1 The Ehrenfest Urn

Consider two containers, A & B, which contain molecules of the same gas. There are a total of N molecules in A & B.

The dynamics of the system is such that at each time step, a molecule is chosen at random from a container, and put in the other.

So, suppose that there are n' in A at time t. Then, at time $t + 1$, urn A will have one less molecule with a probability n'/N. It will obviously have increased in number with probability $1 - n'/N$. So, the transition matrix looks like

$$Q_{nn'} = \begin{cases} \frac{n'}{N} & n = n' - 1 \\ 1 - \frac{n'}{N} & n = n' + 1 \\ 0 & \text{else} \end{cases} .$$

That is, we have the transition probability for urn A to go from having n' molecules, to having n, where $n = n' \pm 1$. We shall look at the problem from the point of view of urn A. Obviously, B will be directly linked to A.

Notice that now, the transition probability depends on the current state of the system. That is, if there are more molecules in A, then there is a higher probability of choosing a molecule from A than from B.

We can deduce the elements of the transition matrix, for the case $N = 3$. That is, a system where there are 3 molecules in total.

- Q_{00}: probability of A going from 0 molecules to 0 molecules is $Q_{00} = 0$. This is obvious as a molecule must be placed in the container.
- Q_{01}: probability of A having 1 then 0 is $Q_{01} = 1/3$. There are only 1 in 3 molecules in A to begin with, so the probability that the single molecule in A is chosen is just $1/3$.
- Q_{21}: probability of A going from 1 to 2 molecules. This is the chance of picking one of the two molecules in B, and moving it over. Obviously, this is $2/3$.

We continue untill we have filled the whole transition matrix.

$$Q = \begin{pmatrix} 0 & 1/3 & 0 & 0 \\ 1 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1 \\ 0 & 0 & 1/3 & 0 \end{pmatrix} .$$

Now, we state (without proof), the eigenvalues of such a matrix, for general N:

$$\lambda^{(i)} = 1 - \frac{2i}{N}, \quad i = 0, 1, \ldots, N.$$

The right eigenvector, corresponding to the eigenvalue $\lambda = 1$ (i.e. the stationary state) has elements

$$P_n^{\text{st}} = \frac{N!}{(N-n)!n!}\frac{1}{2^N}.$$

Notice that this is a binomial distribution; which tends towards a normal distribution for $N \to \infty$. The variance of a normal distribution goes as $1/\sqrt{N}$. Therefore, for a system with many molecules, the system tends towards a state where the number of molecules is the same in each urn. This will be equilibrium, but with fluctuations of the order $1/\sqrt{N}$.

2.4.2 The Wright-Fisher Model

Suppose we have a population of individuals, at time t, who mate randomly, to produce a new generation at the next time step $t + 1$.

We shall focus on one particular gene, which may be in one of two states (which are called alleles). The two alleles are A & B.

The dynamics of the system are such that we select a gene at random, we copy it, place the copy in the new generation, and return the original to the parent population. We continue until there are the same number of genes in the new generation as in the old (i.e. gene number conservation), at time $t + 1$.

The idea is, that if the gene pool at time t is saturated with loads of one allele, and not very many of a second allele, then that second allele will be unlikely to be "picked" for copy into the next generation; thus, eventually, wiping out that allele. It is basically random genetic drift.

Suppose that at time t we have n' A alleles, and $N - n'$ B alleles (i.e. N in total). Then, the probability that in N trials, we get n A alleles, is given by the binomial distribution

$$\binom{N}{n} p^n q^{N-n},$$

where p is the probability of picking A and q of B,

$$p = \frac{n'}{N}, \quad q = 1 - p = 1 - \frac{n'}{N}.$$

Then, it is clear that the transition matrix for allele A (whereby we can trivially obtain that for B), is

$$Q_{nn'} = \binom{N}{n} p^n q^{N-n} = \binom{N}{n} \left(\frac{n'}{N}\right)^n \left(1 - \frac{n'}{N}\right)^{N-n}. \tag{2.17}$$

Now, we can start to analyse this expression a little.

Consider the expectation value at time $t+1$,

$$\langle n(t+1)\rangle = \sum_{n=0}^{N} nP_n(t+1).$$

Now, as the process is a Markov one, the probability of its current state can be written

$$\begin{aligned}\langle n(t+1)\rangle &= \sum_{n=0}^{N} nP_n(t+1) \\ &= \sum_{n=0}^{N} n \sum_{n'=0}^{N} Q_{nn'}P_{n'}(t) \\ &= \sum_{n'} \left(\sum_{n} nQ_{nn'} \right) P_{n'}(t).\end{aligned}$$

Now, we can write the bracketed quantity using (2.17);

$$\sum_{n} nQ_{nn'} = \sum_{n} n \binom{N}{n} p^n (1-p)^{N-n},$$

but this is just the expectation value of the binomial distribution, which is Np.

$$\sum_{n} nQ_{nn'} = Np = N\frac{n'}{N} = n'.$$

Therefore, the expectation value takes on a rather simple form,

$$\langle n(t+1)\rangle = \sum_{n'} n'P_{n'}(t) = \sum_{n} nP_n(t) = \langle n(t)\rangle .$$

That is,

$$\langle n(t+1)\rangle = \langle n(t)\rangle .$$

So, if we start off with n_0 A alleles, then

$$\langle n(t)\rangle = n_0.$$

This suggests that the system will not change over time, but one must understand that this is an "ensemble average", this is a value which is an average over many runs ("realisations") of the system.

We can also ask what the system looks like after a long time. Intuitively, it seems clear that the system will be entirely one allele or the other. Thus,

$$P^{\text{st}} = \begin{pmatrix} 1-\Pi \\ 0 \\ \vdots \\ 0 \\ \Pi \end{pmatrix},$$

where Π is the probability that A is fixed. That is,

$$\sum_{n'=0}^{N} Q_{nn'} P_{n'}^{\text{st}} = P_n^{\text{st}}.$$

We can intuitively calculate Π (this stationary state part has all "only been intuitive"), by considering the expectation value, and what happens to it as $t \to \infty$; the probability will go to the stationary state

$$\langle n(t) \rangle \to \sum_{n=0}^{N} n P_n^{\text{st}} = 0.(1 - \Pi) + N.\Pi.$$

But, we also have that $\langle n(t) \rangle = n_0$; therefore,

$$\Pi = \frac{n_0}{N}.$$

Now, the Wright-Fisher model has two independent right eigenvectors, both corresponding to $\lambda = 1$;

$$\begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix};$$

that is, all entries zero except the top & bottom ones. The left eigenvectors are

$$(1 \quad 1 \quad \dots \quad 1), \quad (0 \quad 1 \quad 2 \quad \dots \quad N).$$

So, for repeated eigenvalues, a linear combination of the eigenstates is also an eigenstate; but the combination must be chosen so that orthogonality is preserved.

2.5 The Master Equation

This is basically the continuous time version of Markov chains.

2.5.1 Derivation

The Chapman-Kolmogorov (CK) equation is

$$P(n, t + \Delta t | n_0, t_0) = \sum_{n'} P(n, t + \Delta t | n', t) p(n', t | n_0, t_0).$$

Now, we assume that

$$P(n, t + \Delta t | n', t) = \begin{cases} 1 - \kappa_n(t)\Delta t + \mathcal{O}(\Delta t)^2 & n = n', \\ w_{nn'}\Delta t + \mathcal{O}(\Delta t)^2 & n \neq n'. \end{cases} \tag{2.18}$$

Where $w_{nn'}$ will become clear that it is a transition rate. Now, by normalisation,

$$\begin{aligned} 1 &= \sum_n P(n, t+\Delta t|n', t) \\ &= 1 - \kappa_{n'}(t)\Delta t + \mathcal{O}(\Delta t)^2 + \sum_{n\neq n'} w_{nn'}\Delta t + \mathcal{O}(\Delta t)^2 \\ &= 1 - \kappa_{n'}(x) + \sum_{n\neq n'} w_{nn'}\Delta t \\ \Rightarrow \kappa_{n'}(t) &= \sum_{n\neq n'} w_{nn'}. \end{aligned}$$

Alternatively, switching the indices,

$$\kappa_n(t) = \sum_{n'\neq n} w_{n'n}. \tag{2.19}$$

Therefore, using (2.18), we see that the CK equation reads

$$\begin{aligned} P(n, t+\Delta t|n_0, t_0) = \quad & (1 - \kappa_n(t)\Delta t + \ldots)P(n, t|n_0, t_0) \\ & + \sum_{n'\neq n} w_{nn'}(t)P(n', t|n_0, t_0)\Delta t + \mathcal{O}(\Delta t)^2, \end{aligned}$$

which is easily rearranged to

$$\begin{aligned} \frac{P(n, t+\Delta t|n_0, t_0) - P(n, t|n_0, t_0)}{\Delta t} = \quad & -\kappa_n(t)P(n, t|n_0, t_0) \\ & + \sum_{n'\neq n} w_{nn'}(t)P(n', t|n_0, t_0) + \mathcal{O}(\Delta t). \end{aligned}$$

Now, in the limit that $\Delta t \to 0$, the LHS becomes a differential, and the RHS easily becomes

$$\frac{dP(n,t)}{dt} = -\kappa_n(t)p(n,t) + \sum_{n'\neq n} w_{nn'}(t)p(n', t),$$

using (2.19), the middle term is rewritten,

$$\frac{dP(n,t)}{dt} = -\sum_{n'\neq n} w_{n'n}P(n,t) + \sum_{n'\neq n} w_{nn'}(t)P(n', t).$$

Therefore, we have our *master equation*,

$$\frac{dP(n,t)}{dt} = \sum_{n'\neq n} w_{nn'}(t)P(n', t) - \sum_{n'\neq n} w_{n'n}P(n,t). \tag{2.20}$$

The interpretation of this is pretty simple. The first term is just the probability of transitioning from $n' \to n$, and the second just the other way around.

So, we say that the rate of change of probability of being in a state n is equal to the probability of making a transition into n, minus the probability of transitioning out of n.

This is all because of the Markov property that $w_{n'n}$ does not depend on previous states.

2.5.2 Relation to Markov Chains

Here we shall provide another derivation; but one that is much less rigorous than the previous one.

So, let us start from

$$P_n(t+1) = \sum_{n'} Q_{nn'} P_{n'}(t),$$

where the columns of the transition matrix add to unity;

$$\sum_{n} Q_{nn'} = 1.$$

So, if we write

$$P_n(t+1) - P_n(t) = \sum_{n'} Q_{nn'} P_{n'}(t) - \sum_{n'} Q_{n'n}(t) P_n(t),$$

and further that

$$P_n(t+1) - P_n(t) = \sum_{n' \neq n} Q_{nn'} P_{n'}(t) - \sum_{n' \neq n} Q_{n'n}(t) P_n(t),$$

whereby we notice that the $n = n'$ terms cancel. Now, let us take the time step to be Δt, rather than unity, and divide through by the time step;

$$\frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \sum_{n' \neq n} \frac{Q_{nn'}}{\Delta t} P_{n'}(t) - \sum_{n' \neq n} \frac{Q_{n'n}}{\Delta t} P_n(t).$$

Then, it is clear that taking the time step to zero, reduces the LHS to a differential of the probability, with respect to time. Furthermore, let us assume that on average, one event happens in the time step. To achieve this, let us assume

$$Q_{nn'} = w_{nn'} \Delta t + \mathcal{O}(\Delta t)^2, \quad n \neq n',$$

then, we have arrived at the master equation again;

$$\frac{dP_n}{dt} = \sum_{n' \neq n} w_{nn'}(t) P_{n'}(t) - \sum_{n' \neq n} w_{n'n}(t) P_n(t),$$

2.6 One Step Processes

These are a special class of processes, and are easier to analyse than the general case.

These systems have dynamical laws which only allow movement to the nearest states, rather than all states as in the general case. So, we have that $w_{nn'}$ and $w_{n'n}$ are zero, unless $n' = n \pm 1$. Hence, under this, we see that the master equation becomes

$$\frac{dP_n(t)}{dt} = w_{n,n+1} P_{n+1}(t) + w_{n,n-1} P_{n-1}(t) - w_{n+1,n} P_n(t) - w_{n-1,n} P_n(t).$$

Now, by way of convenient notation, we write

$$w_{n+1,n} \equiv g_n, \quad w_{n-1,n} \equiv r_n,$$

then the master equation reads

$$\frac{dP_n(t)}{dt} = r_{n+1}P_{n+1}(t) + g_{n-1}P_{n-1}(t) - (g_n + r_n)P_n(t). \tag{2.21}$$

Now, let us consider some examples.

2.6.1 The Decay Process

Suppose we have a sample of radioactive material, with n_0 at $t = 0$. How many are there at time t, given some decay rate γ? Then,

$$w_{nn'} = \begin{cases} \gamma n' & n = n' - 1, \\ 0 & \text{else.} \end{cases}$$

Therefore, we see that $w_{n+1,n} = r_n = \gamma n$ and $w_{n-1,n} = g_n = 0$. Thus, the master equation (2.21) reads

$$\frac{dP_n}{dt} = \gamma(n+1)P_{n+1}(t) - \gamma n P_n(t).$$

We shall state that the boundary condition is

$$P_n(0) = \delta_{n,n_0}.$$

Now, the mean, over many ensembles, is

$$\langle n(t) \rangle = \sum_n n P_n(t),$$

differentiating,

$$\frac{d}{dt}\langle n(t) \rangle = \sum_n n \frac{dP_n}{dt}.$$

We see that the RHS is just our master equation. Thus,

$$\frac{d}{dt}\langle n(t) \rangle = \gamma \sum_{n=0}^{\infty} n(n+1)P_{n+1}(t) - \gamma \sum_{n=0}^{\infty} n^2 P_n(t).$$

Now, we shall change variables, so that $m = n + 1$ in the first summation;

$$\frac{d}{dt}\langle n(t) \rangle = \gamma \sum_{m=1}^{\infty} m(m-1)P_m(t) - \gamma \sum_{n=0}^{\infty} n^2 P_n(t).$$

Then, shift the start of the summation back to zero, and rename back to n,

$$\frac{d}{dt}\langle n(t)\rangle = \gamma \sum_{n=0}^{\infty} n(n-1)P_n(t) - \gamma \sum_{n=0}^{\infty} n^2 P_n(t).$$

So, collecting terms, we see that

$$\frac{d}{dt}\langle n(t)\rangle = -\gamma \sum_{n=0}^{\infty} nP_n(t),$$

which is of course just

$$\frac{d}{dt}\langle n(t)\rangle = -\gamma \langle n(t)\rangle .$$

Now, let $x(t) \equiv \langle n(t)\rangle$; then, we have the familiar result

$$\frac{dx(t)}{dt} = -\gamma x(t).$$

That is, exponential decay. So, we see that this familiar result is only for an ensemble average: take a lot of experiments, and average over them.

Random Walk Here, we have $r_n = \mu$ and $g_n = \lambda$. That is, a fixed probability to move left or right, by one step. If the random walk is symmetric, then $\mu = \lambda$.

Birth & Death Processes Here, we shall take $g_n = bn$ (a birth rate), and $r_n = dn$ (a death rate). Again, this is a model choice. In reality, birth rate may decrease with n, as resources become scarce. For example, if

$$g_n = b(1 - \tfrac{n}{N})n,$$

then, once $n = N$, no more birth occurs, and the system stops. That is, N represents the maximum number of individuals that the system can have.

2.6.2 Linear One-step Processes & Their Solutions

A linear one-step process is one for whom

$$r_n = an + b, \quad g_n = cn + d.$$

These can be solved be introducing the generating function,

$$F(z,t) = \sum_n z^n P_n(t).$$

Note that before, we introduced the characteristic function (although we bought it in by considering integrals, we can just use a sum instead),

$$G(k,t) = \sum_n e^{ikn} P_n(t).$$

Then, we identify z with e^{ik}. So, differentiating the generating function,

$$\frac{\partial F}{\partial z} = \sum_n n z^{n-1} P_n(t),$$

and, setting $z = 1$, then we see that

$$\left.\frac{\partial F}{\partial z}\right|_{z=1} = \sum_n n P_n(t),$$

that is, the mean.

$$\left.\frac{\partial F}{\partial z}\right|_{z=1} = \langle n(t) \rangle .$$

To get this result for the characteristic function, G, we had to set $k = 0$. Let us differentiate the generating function a second time,

$$\frac{\partial^2 F}{\partial z^2} = \sum_n n(n-1) z^{n-2} P_n(t),$$

setting $z = 1$ again results in

$$\left.\frac{\partial^2 F}{\partial z^2}\right|_{z=1} = \sum_n n^2 P_n(t) - \sum_n n P_n(t) = \left\langle n^2(t) \right\rangle - \langle n(t) \rangle .$$

Note that these are not-quite cumulants. Let us solve a problem.

Example: Symmetric Random Walk Here, we suppose that the probability of going to the left or right is the same; $g_n = r_n = \alpha$ (also independent of n). So, the master equation for one step processes (2.21) reads

$$\frac{dP_n(t)}{dt} = \alpha P_{n+1}(t) + \alpha P_{n-1} - 2\alpha P_n(t).$$

We can get rid of α, by defining

$$\tau \equiv \alpha t,$$

so that

$$\frac{d}{dt} = \frac{d\tau}{dt}\frac{d}{d\tau} = \alpha \frac{d}{d\tau}.$$

Therefore, the master equation reads

$$\frac{dP_n(t)}{d\tau} = P_{n+1}(t) + P_{n-1} - 2P_n(t).$$

We shall also impose the initial condition $P_n(0) = \delta_{n,0}$; that is, to start at the origin. So then, if we multiply the master equation by z^n, and sum,

$$\sum_{n=-\infty}^{\infty} z^n \frac{dP_n(t)}{d\tau} = \sum_{n=-\infty}^{\infty} z^n P_{n+1}(t) + \sum_{n=-\infty}^{\infty} z^n P_{n-1} - 2\sum_{n=-\infty}^{\infty} z^n P_n(t).$$

Let us then redefine the summation variable on the first and second terms, to $m = n+1$ and $m = n-1$,

$$\sum_{n=-\infty}^{\infty} z^n \frac{dP_n(t)}{d\tau} = \sum_{m=-\infty}^{\infty} z^{m-1} P_m(t) + \sum_{m=-\infty}^{\infty} z^{m+1} P_m - 2\sum_{n=-\infty}^{\infty} z^n P_n(t).$$

Now, if we note that in the generating function, only the probability is a function of time t, differentiating it will result in the LHS of the above. So,

$$\frac{\partial F}{\partial \tau} = \sum_{m=-\infty}^{\infty} z^{m-1} P_m(t) + \sum_{m=-\infty}^{\infty} z^{m+1} P_m - 2\sum_{n=-\infty}^{\infty} z^n P_n(t).$$

In a similar vein, we see that we can write the RHS as a similar expression;

$$\frac{\partial F}{\partial \tau} = (z^{-1} + z - 2)F.$$

Therefore, we see that

$$\frac{1}{F}\frac{\partial F}{\partial \tau} = z^{-1} + z - 2.$$

This is simply integrated to give

$$\ln F = (z^{-1} + z - 2)\tau + \Omega(z),$$

where we note we have an arbitrary function of integration, rather than constant, as we integrated a partial differential. This gives

$$F(z, \tau) = \phi(z) e^{(z^{-1}+z-2)\tau},$$

where $\phi(z)$ is an arbitrary function. Putting t back in,

$$F(z, t) = \phi(z) e^{(z^{-1}+z-2)\alpha t}.$$

We can figure out the function $\phi(z)$, by considering that $P_n(0) = \delta_{n,0}$. So,

$$F(z, 0) = \sum_n z^n \delta_{n,0} = z^0.1 = 1$$

Therefore,

$$F(z,0) = \phi(z)e^0 = \phi(z),$$

hence,

$$\phi(z) = 1.$$

Therefore, the generating function is

$$F(z,t) = e^{(z^{-1}+z-2)\alpha t}.$$

Now, to find the probability function, $P_n(t)$, we either Taylor expand the exponential, or use contour integration. We shall not go through how to do it; but the answer we get is

$$P_n(t) = e^{-2\alpha t} I_{|n|}(2\alpha t),$$

where I_n is a modified Bessel function. Now, the mean is just

$$\langle n(t)\rangle = \left.\frac{\partial F}{\partial t}\right|_{z=1} = \left.(-z^{-2}+1)\alpha e^{(z^{-1}+z-2)\alpha t}\right|_{z=1} = 0.$$

Thus, the mean is zero. This can be intuitively seen; if we consider that the probability of going right is the same as going left, and that the system starts at zero, then we go left "as much" as right.

Similarly,

$$\langle n(n-1)(t)\rangle = \left\langle n^2(t)\right\rangle = \left.\frac{\partial^2 F}{\partial t^2}\right|_{z=1} = 2\alpha t.$$

Notice that

$$\sigma^2 = \left\langle n^2\right\rangle - \langle n\rangle^2 = \left\langle n^2\right\rangle,$$

and therefore that the root-mean-square distance goes as $t^{1/2}$.

Example: Simple Birth-Death Processes Here, let

$$r_n = dn, \quad g_n = bn;$$

then, the one-step master equation reads

$$\frac{dP_n}{dt} = d(n+1)P_{n+1}(t) + b(n-1)P_{n-1}(t) - n(b+d)P_n(t), \quad n > 0$$

Now, we must be careful at the boundary. Consider $n = 0$, then P_{-1} makes "no sense". So, we impose

$$\frac{dP_0}{dt} = dP_1(t).$$

So then, to solve our master equation, we do the usual thing. Multiply by $\sum_{n=0}^{\infty} z^n$,

$$\sum_{n=0}^{\infty} z^n \frac{dP_n}{dt} = d\sum_{n=0}^{\infty} z^n(n+1)P_{n+1}(t) + b\sum_{n=1}^{\infty} z^n(n-1)P_{n-1}(t) - (b+d)\sum_{n=1}^{\infty} nz^n P_n(t),$$

then, we switch summation indices;

$$\sum_{n=0}^{\infty} z^n \frac{dP_n}{dt} = d\sum_{n=0}^{\infty} z^{n-1} n P_n(t) + b\sum_{n=0}^{\infty} z^{n+1} n P_n(t) - (b+d)\sum_{n=0}^{\infty} n z^n P_n(t).$$

Now, notice that

$$\begin{aligned}
\sum_n n z^{n-1} P_n &= \frac{\partial}{\partial z}\sum_n z^n P_n, \\
\sum_n n z^{n+1} P_n &= z^2 \sum_n n z^{n-1} P_n \\
&= z^2 \frac{\partial}{\partial z}\sum_n z^n P_n, \\
\sum_n n z^n P_n &= z\sum_n n z^{n-1} P_n \\
&= z\frac{\partial}{\partial z}\sum_n z^n P_n.
\end{aligned}$$

Finally, we know that

$$F = \sum_n z^n P_n,$$

as thats how we defined the generating function. Therefore, the master equation reads

$$\frac{\partial F}{\partial t} = d\frac{\partial F}{\partial z} + bz^2\frac{\partial F}{\partial z} - (b+d)z\frac{\partial F}{\partial z},$$

which is easily rearranged into

$$\frac{\partial F}{\partial t} = (d - bz)(1-z)\frac{\partial F}{\partial z}.$$

The solution to this, we verify, rather than prove. Let us change variables,

$$u \equiv \frac{1-z}{d-bz}e^{(b-d)t}, \quad v \equiv z.$$

Then,

$$\begin{aligned}
\frac{\partial}{\partial z} &= \frac{\partial u}{\partial z}\frac{\partial}{\partial u} + \frac{\partial v}{\partial z}\frac{\partial}{\partial v} \\
&= \frac{\partial u}{\partial z}\frac{\partial}{\partial u} + \frac{\partial}{\partial v}, \\
\frac{\partial}{\partial t} &= \frac{\partial u}{\partial t}\frac{\partial}{\partial u} + \frac{\partial v}{\partial t}\frac{\partial}{\partial v} \\
&= \frac{\partial u}{\partial t}\frac{\partial}{\partial u}.
\end{aligned}$$

Thus, the partial differential equation that is our master equation, can be seen to just become

$$\frac{\partial F}{\partial v} = 0 \quad \Rightarrow \quad F(u,v) = \Phi(u),$$

where $\Phi(u)$ is an arbitrary function of u only. Then,

$$F(z,t) = \Phi\left(\frac{1-z}{d-bz}e^{(b-d)t}\right).$$

We now use the initial condition that at $t = 0$, $P_n(0) = \delta_{n,n_0}$. Then,

$$F(z,0) = \sum_{n=0}^{\infty} z^n \delta_{n,n_0} = z^{n_0}.$$

Therefore,

$$\Phi\left(\frac{1-z}{d-bz}\right) = z^{n_0}.$$

Now, if we let

$$x \equiv \frac{1-z}{d-bz} \quad \Rightarrow \quad z = \frac{1-dx}{1-bx},$$

Hence,

$$\Phi(x) = \left(\frac{1-dx}{1-bx}\right)^{n_0}$$

Thus, substituting z back in, we arrive at

$$F(z,t) = \left(\frac{d(z-1)e^{(b-d)t} - (bz-d)}{b(z-1)e^{(b-d)t} - (bz-d)}\right)^{n_0}.$$

If one were so inclined, then one would find $P_n(t)$ from this; however, we shall not!

We can deduce a couple of things from this expression for $F(z,t)$.
First, consider that

$$F(z,t) = \sum_{n=0}^{\infty} z^n P_n(t),$$

then, notice that setting $z = 0$ has the effect of picking out a single component from the sum,

$$F(0,t) = P_0(t).$$

That is, setting $z = 0$ in the generating function gives the probability of attaining the zero-state. Thus, for our generating function,

$$F(0,t) = P_0(t) = \left(\frac{d - de^{(b-d)t}}{d - be^{(b-d)t}}\right)^{n_0},$$

which is obviously the extinction probability.

Now, suppose that $d > b$ (i.e. death rate greater than birth rate). Then, as $t \to \infty$, $e^{(b-d)}t \to 0$. Hence, $P_0(t) \to 1$. This is an intuitively expected result.

Suppose that $b > d$. Then,

$$\lim_{t\to\infty} P_0(t) = \left(\frac{d}{b}\right)^{n_0}.$$

2.6.3 The Macroscopic Equation

Recall that for the decay process we found that

$$\frac{d}{dt}\langle n(t)\rangle = -\gamma \langle n(t)\rangle,$$

which is entirely deterministic. If we let $x(t) \equiv \langle n(t)\rangle$, then the solution to this is the well known

$$x(t) = x_0 e^{-\gamma t}.$$

These are non-stochastic deterministic equations; after taking an ensemble average. They will not tell you how each particle moves in a given "run" of the experiment, but they will tell you how that particle will move "on average". We can compute an analogous equation for the general one-step process.

Recall the master equation for one-step processes, (2.21). Then, let us multiply it by n, and sum over all n. Thus,

$$\sum_n n\frac{dP_n}{dt} = \sum_n n r_{n+1} P_{n+1}(t) + \sum_n n g_{n-1} P_{n-1}(t) - \sum_n n(g_n + r_n)P_n(t).$$

Let us change summation indices on the first and second sums. For the first, let $m \equiv n + 1$, and in the second let $q \equiv n - 1$. Then,

$$\frac{d}{dt}\sum_n nP_n = \sum_m (m-1) r_m P_m(t) + \sum_q (q+1) g_q P_q(t) - \sum_n n(g_n + r_n)P_n(t).$$

It is clear that we may change the indices back to n's. Once doing that, we see that terms cancel, leaving

$$\frac{d}{dt}\sum_n nP_n = \sum_n g_n P_n(t) - \sum_n r_n P_n(t),$$

which is just the expectation values (ensemble average),

$$\frac{d}{dt}\langle n(t)\rangle = \langle g_n\rangle - \langle r_n\rangle. \tag{2.22}$$

To check that this is consistent with the previous case for the decay process, recall that $g_n = 0$, and $r_n = \gamma n$. We thus see that this general expression holds for our specifically derived case of the decay process. This equation is called the *macroscopic equation*

We have been a little sloppy in reference to boundaries. Suppose we have boundaries at $n = 0, N$. Then, if we define $r_0 \equiv 0$ (i.e. no transitions from 0 to -1), $g_{-1} \equiv 0$ (i.e. no transitions from -1 to 0), $r_{N+1} = g_N \equiv 0$ (i.e. no transitions from $N+1$ to N or N to $N+1$), then all results hold.

Lets consider the standard few examples again.

Example: Random Walk We have that $r_n = \mu, g_n = \lambda$. Therefore, the macroscopic equation is

$$\frac{d}{dt}\langle n(t)\rangle = \lambda - \mu.$$

Therefore, this is easy to solve; resulting in

$$\langle n(t)\rangle = (\lambda - \mu)t + n_0.$$

Example: Birth & Death Processes Here, we have that $r_n = dn, g_n = bn$. Hence, the macroscopic equation,

$$\frac{d}{dt}\langle n(t)\rangle = (b-d)\langle n(t)\rangle,$$

which has solution

$$\langle n(t)\rangle = n_0 e^{(b-d)t}.$$

Now, for a more complicated example.

2.6.4 Long Term Behaviour of Non-linear Processes

Consider

$$r_n = \alpha n^2, \quad g_n = \beta,$$

where α, β are constants. Then,

$$\frac{d}{dt}\langle n\rangle = \beta - \alpha\langle n^2\rangle;$$

where we suppress implied time-dependance. Now this is hard to solve, as we do not know the expectation value of n^2. So, we instead look at the long-term behavior of $P_n(t)$. The central limit theorem tells us that for a big enough system, the probability distribution will tend towards Gaussianity. Therefore, the initial $\delta-$style peak will eventually "widen" into

a Gaussian, probably moving position as it does so. And, as a simplification, we set all fluctuations (i.e. the variance) to zero. Hence,

$$\sigma^2 = \left\langle n^2 \right\rangle - \langle n \rangle^2 = 0 \quad \Rightarrow \quad \left\langle n^2 \right\rangle = \langle n \rangle^2 .$$

This carries on to the other moments as well;

$$\left\langle n^3 \right\rangle \to \langle n \rangle^3 \quad \left\langle n^4 \right\rangle \to \langle n \rangle^4 \ldots$$

Then, our macroscopic equation reads

$$\frac{d}{dt} \langle n \rangle = \beta - \alpha \langle n \rangle^2 .$$

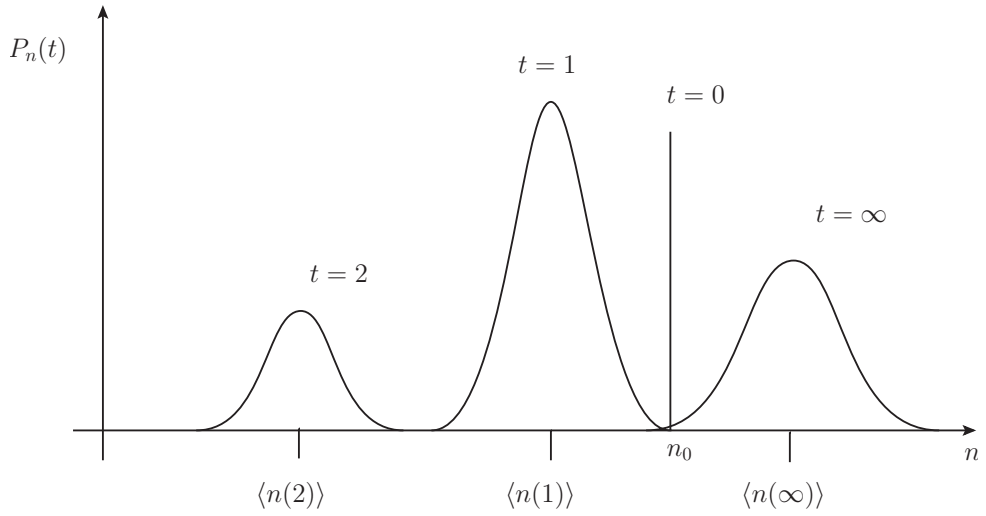

Figure 2.3: Evolution of the probability distribution, to a Gaussian. The distribution starts off as an infinite spike, then spreads out over time. The shape changes, as does the poisition.

2.6.5 Boundaries

Let us consider the boundary conditions of a one step-process. There are obviously a few types. The ranges of n are specified as being one of 3 types.

- All integers $-\infty < n < \infty$. Such a system needs no boundary conditions, as there are no boundaries.
- Semi-infinite $n = 0, 1, 2, \ldots$. An example of such a system is a bacteria colony obeying the simple birth-death rules.

- Finite $0 \leq n \leq N$. That is, some sort of boundary at $n = 0, N$. For example, when we wrote $r_n = dn, g_n = bn(1 - \frac{n}{N})$.

Then, we can have a few types of boundary. Reflecting where the walker is reflected from the boundary (whereby total probability is conserved); and absorbing, where the walker is removed from the system as soon as he reaches the boundary.

Absorbing boundaries do not conserve the total probability.

2.6.6 Stationary States of the Master Equation

Consider the difference between a stationary state & the equilibrium state. The equilibrium state is the state the system will "settle down to", if all external influences are removed. Now, consider the stationary state. Suppose a metal bar is heated at both ends, with each end held at different temperatures. Then, given enough time, the temperature of the bar will have settled down, so that the temperature at a given point is constant; and the variation of temperature along the bar is only a function of position, and not of time. Then, this state is the stationary state, and is obviously not the equilibrium state (as we are applying heat to the bar ends). We shall denote the stationary state P_n^{st}. Obviously, as the stationary state

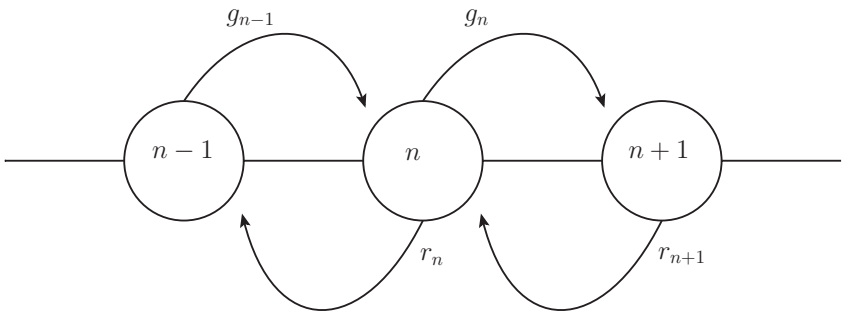

Figure 2.4: The movement from one state to the next, for a one-step process

is not a function of time (as we just discussed), then

$$\frac{dP_n^{\text{st}}}{dt} = 0.$$

Therefore, the master equation reads

$$r_{n+1}P_{n+1}^{\text{st}} + g_{n-1}P_{n-1}^{\text{st}} - (g_n + r_n)P_n^{\text{st}} = 0,$$

which trivially rearranges to

$$r_{n+1}P_{n+1}^{\text{st}} - g_nP_n^{\text{st}} = r_nP_n^{\text{st}} - g_{n-1}P_{n-1}^{\text{st}}, \quad \forall n.$$

Now, let us define the LHS as $-J_{n+1}$ and the RHS as $-J_n$ (where the signs are introduced by convention, which we will soon see). Then, we see that

$$J_{n+1} = J_n, \quad \forall n.$$

Therefore, we see that J_n is independent of n. So then, we have that

$$J_{n+1} = g_n P_n^{\text{st}} - r_{n+1} P_{n+1}^{\text{st}},$$

which has the interpretation of being the net flow of probability from state n to state $n+1$. Hence, we call call J the *probability current* (without the subscript as it is independent of n). Thus,

$$r_{n+1} P_{n+1}^{\text{st}} - g_n P_n^{\text{st}} = -J.$$

Now, a reflecting boundary at $n = 0$ is one for whom $J = 0$. This must be the case, as there is no probability flow from $n = 0$ to $n = -1$; thus, as the current is conserved, we must have $J = 0$ everywhere. Therefore, we see that

$$r_1 P_1^{\text{st}} = g_0 P_0^{\text{st}} \quad r_2 P_2^{\text{st}} = g_1 P_1^{\text{st}} \quad \dots \quad r_n P_n^{\text{st}} = g_{n-1} P_{n-1}^{\text{st}} \dots$$

Therefore, from the first expression, we see that

$$P_1^{\text{st}} = \frac{g_0}{r_1} P_0^{\text{st}},$$

and from the second that

$$P_2^{\text{st}} = \frac{g_1}{r_2} P_1^{\text{st}},$$

thus, combining that

$$P_2^{\text{st}} = \frac{g_1 g_0}{r_2 r_1} P_0^{\text{st}}.$$

Therefore, iterating up, we see that

$$P_n^{\text{st}} = \frac{g_{n-1} g_{n-2} \dots g_0}{r_n r_{n-1} \dots r_1} P_0^{\text{st}}. \tag{2.23}$$

Now then, we have that all probabilities must sum to unity,

$$1 = \sum_{n=0}^{N} P_n^{\text{st}},$$

Therefore, taking out the 0-state,

$$1 = P_0^{\text{st}} + \sum_{n=1}^{N} P_n^{\text{st}}.$$

So, inserting our expression for $P_n^{\rm st}$,

$$1 = P_n^{\rm st}\left(1 + \sum_{n>0} \frac{g_{n-1}g_{n-2}\cdots g_0}{r_n r_{n-1}\ldots r_1}\right),$$

which easily rearranges to

$$P_0^{\rm st} = \frac{1}{1+\sum_{n>0}\frac{g_{n-1}g_{n-2}\cdots g_0}{r_n r_{n-1}\ldots r_1}}. \tag{2.24}$$

Hence, between this and (2.23), we are able to compute all parts to the stationary state probability.

2.7 Solution to Master Equation Under Detailed Balance

Let us return to the general master equation,

$$\frac{dP_n(t)}{dt} = \sum_{n'\neq n} w_{nn'}P_{n'}(t) - \sum_{n'\neq n} w_{n'n}P_n(t),$$

then, the stationary state clearly is when

$$\sum_{n'\neq n} w_{nn'}P_{n'}^{\rm st} = \sum_{n'\neq n} w_{n'n}P_n^{\rm st}.$$

This is because the stationary state is the state which does not depend upon time.

Now, transition rates, $w_{nn'}$, satisfy *detailed balance*, where detailed balance is the statement that

$$w_{nn'}P_{n'}^{\rm st} = w_{n'n}P_n^{\rm st} \quad \forall n, n'.$$

If one says in words what this statement is: "at stationarity, the flow of probability into a state n from n' is equal to the flow of probability into n' from n".

For a one-step process, we have that $n' = n+1$, which clearly corresponds to

$$r_{n+1}P_{n+1}^{\rm st} = g_n P_n^{\rm st},$$

which is exactly what we found for the system with reflecting boundaries ($J = 0$).

Now, let us write the master equation in a more concise form. Let us introduce the transition matrix

$$W_{mn}(t) \equiv w_{mn}(t) - \delta_{mn}\sum_{n'\neq n} w_{n'n}(t). \tag{2.25}$$

Then, multiplying this by $P_n(t)$, and summing over n,

$$\begin{aligned}\sum_n W_{mn}(t)P_n(t) &= \sum_{n\neq m} P_n(t)w_{mn}(t) - \sum_n \sum_{n'\neq n} \delta_{mn}P_n(t)w_{n'n}(t) \\ &= \sum_{n\neq m} P_n(t)w_{mn}(t) - \sum_{n'\neq m} P_m(t)w_{n'm}(t) \\ &= \frac{dP_m}{dt}.\end{aligned}$$

Therefore, we see that we can write the master equation as

$$\frac{dP_m}{dt} = \sum_n W_{mn}(t)P_n(t), \tag{2.26}$$

which is a form very close to the previous Markov chains.

Let us consider the properties of the matrix W.

- As $w_{nm} \geq 0$, then, if $n \neq m$, the Kronecker-delta term does not contribute, leaving $W_{nm} \geq 0$. That is, all off-diagonal elements of W are greater-than or equal-to zero.
- Consider summing W_{nm} up a column. That is,

  $$\sum_n W_{nm} = \sum_n w_{nm} - \sum_n \delta_{nm} \sum_{n'\neq m} w_{n'm},$$

  which is,

  $$\sum_n W_{nm} = \sum_n w_{nm} - \sum_{n'} w_{n'm} = 0.$$

  Thus, we see that all columns of W add to zero.

To see what the last property "does" to an example W, consider a 3x3 case. Its form must be

$$W = \begin{pmatrix} -a-b & c & e \\ a & -c-d & f \\ b & d & -e-f \end{pmatrix}.$$

Notice that we can write the continuous Markov equation, (2.26), in ket-notation,

$$\frac{d}{dt}|P(t)\rangle = W(t)|P(t)\rangle.$$

Where $W(t)$ is a time dependent operator. Notice that this equation is of very similar form to the Schrodinger equation. In writing this ket-form, we use

$$P_n(t) = \langle n|P(t)\rangle, \quad W_{nn'} = \langle n|W|n'\rangle.$$

Now, in principle, we can solve the continuous Markov equation as we did the discrete equation. When we solved the discrete equation, we implicitly assumed that there existed a complete set of eigenvalues and eigenstates. This is not always guaranteed. However, if the matrix is symmetric, then this method will always work.

Now, W, in general, is not symmetric. Then, finding some symmetric matrix, V say, that is symmetric (from W) will enable us to use previous methods. Infact, as we shall see, a process satisfying the detailed balance condition allows this.

Now, let us define a new matrix,

$$\begin{aligned} V_{mn} &\equiv \sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}} W_{mn} \\ &= \sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}} \left(w_{mn} - \delta_{mn} \sum_{n' \neq n} w_{n'n} \right). \end{aligned} \tag{2.27}$$

Now, to see that V_{mn} is symmetric, we shall use the detailed balance condition

$$w_{nm} P_m^{\rm st} = w_{mn} P_n^{\rm st},$$

which we rewrite in the form

$$w_{nm} \sqrt{\frac{P_m^{\rm st}}{P_n^{\rm st}}} = w_{mn} \sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}}.$$

Now then, notice that V_{mn} can be rewritten, using this

$$\begin{aligned} V_{mn} &= \sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}} w_{mn} - \delta_{mn} \sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}} \sum_{n' \neq n} w_{n'n} \\ &= \sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}} w_{mn} - \delta_{mn} \sum_{n' \neq n} w_{n'n} \\ &= \sqrt{\frac{P_m^{\rm st}}{P_n^{\rm st}}} w_{nm} - \delta_{mn} \sum_{n' \neq n} w_{n'n} \\ &= V_{nm}. \end{aligned}$$

The steps follow as the Kronecker-delta reduce the square-root term to unity. Therefore, we see that V_{mn} is symmetric, if the detailed balance assumption holds. Thus, $V = V^T$.

Hence, using (2.27), the master equation reads

$$\frac{dP_m}{dt} = \sum_n V_{mn} \sqrt{\frac{P_m^{\rm st}}{P_n^{\rm st}}} P_n,$$

which is easily rewritten as

$$\frac{d}{dt}\frac{P_m}{\sqrt{P_m^{\rm st}}} = \sum_n V_{mn}\frac{P_n}{\sqrt{P_n^{\rm st}}}.$$

If we then define

$$\tilde{P}_n \equiv \frac{P_n}{\sqrt{P_n^{\rm st}}},$$

the master equation reads

$$\frac{d\tilde{P}_m}{dt} = \sum_n V_{mn}\tilde{P}_n.$$

Now this is in a form which we can solve using previous methods. This is because the operator, V is symmetric. Note that this, in ket-notation, reads

$$\frac{d}{dt}|\tilde{P}(t)\rangle = V|\tilde{P}(t)\rangle.$$

The *formal solution* to this, is

$$|\tilde{P}(t)\rangle = e^{Vt}|\tilde{P}(0)\rangle,$$

assuming that V is time independent. We can check this easily,

$$\frac{d}{dt}|\tilde{P}(t)\rangle = Ve^{Vt}|\tilde{P}(0)\rangle = V|\tilde{P}(t)\rangle.$$

This formal solution is actually pretty useless. We can explicitly solve the equation using eigenvalues and eigenvectors.

Now, as V is symmetric, the left- and right-eigenvectors are identical. That is,

$$V|\phi^{(i)}\rangle = \mu^{(i)}|\phi^{(i)}\rangle, \quad \langle\phi^{(i)}|V = \langle\phi^{(i)}|\mu^{(i)}.$$

The "superscript with bracket" notation reads *the i^{th} right-eigenvector is $|\phi^{(i)}\rangle$, with eigenvalue $\mu^{(i)}$*. The eigenvectors are orthonormal, and satisfy the completeness relation,

$$\langle\phi^{(i)}|\phi^{(j)}\rangle = \delta_{ij}, \quad \sum_i |\phi^{(i)}\rangle\langle\phi^{(i)}| = I.$$

Also, since the matrix V is symmetric, all eigenvalues are real.

The stationary state is the state which is independent of time. Therefore, the time derivative of the stationary state is zero. Hence, the stationary state satisfies

$$\sum_n V_{mn}\tilde{P}_n^{\rm st} = 0.$$

We can write this as an eigenvalue equation, so that

$$V|\tilde{P}^{\rm st}\rangle = 0.|\tilde{P}^{\rm st}\rangle.$$

That is, the state $\tilde{P}_n^{\rm st}$ is an eigenstate of V, with eigenvalue 0. Notice the difference in notation between $|P\rangle$ and P_n. We generally use the former to refer to the entire state P, whereas the latter is referring to a single component of P.

Now, we can take our "formal solution",

$$|\tilde{P}(t)\rangle = e^{Vt}|\tilde{P}(0)\rangle,$$

insert a unity,

$$|\tilde{P}(t)\rangle = e^{Vt}\sum_i |\phi^{(i)}\rangle\langle\phi^{(i)}|\tilde{P}(0)\rangle.$$

Now, we note that as

$$e^{At}|\psi^{(i)}\rangle = e^{a^{(i)}}|\psi^{(i)}\rangle,$$

then

$$|\tilde{P}(t)\rangle = \sum_i e^{\mu^{(i)}t}|\phi^{(i)}\rangle\langle\phi^{(i)}|\tilde{P}(0)\rangle.$$

Now, by completeness, we can write that a single component of $\tilde{P}$ is

$$\begin{aligned}
\tilde{P}_n(t) &= \langle n|\tilde{P}(t)\rangle \\
&= \langle n|\sum_i e^{\mu^{(i)}t}|\phi^{(i)}\rangle\langle\phi^{(i)}|\tilde{P}(0)\rangle \\
&= \langle n|\sum_{i,m} e^{\mu^{(i)}t}|\phi^{(i)}\rangle\langle\phi^{(i)}|m\rangle\langle m|\tilde{P}(0)\rangle \\
&= \sum_{i,m} e^{\mu^{(i)}t}\langle n|\phi^{(i)}\rangle\langle\phi^{(i)}|m\rangle\langle m|\tilde{P}(0)\rangle \\
&= \sum_i e^{\mu^{(i)}t}\phi_n^{(i)}\sum_m \phi_m^{(i)}\tilde{P}_m(0).
\end{aligned}$$

We now see that the n^{th} component of the state, at some time t is found by summing over the n^{th} component of all eigenstates, multiplied by the sum over all components of that eigenvector multiplied by all components of the original state. Thus, we have derived that

$$\tilde{P}_n(t) = \sum_i e^{\mu^{(i)}t}\phi_n^{(i)}\sum_m \phi_m^{(i)}\tilde{P}_m(0). \tag{2.28}$$

Notice that the only time-dependent term is the exponential, which carries the eigenvalue $\mu^{(i)}$.

This is, again, very close to the discrete Markov equation solution. The difference between this continuous solution, and the discrete solution, is that $e^{\mu^{(i)}t}$ replaces $(\lambda^{(i)})^t$. That is, the correspondence between continuous and discrete is

$$\begin{array}{ccc}
\lambda^{(i)} & \longleftrightarrow & e^{\mu^{(i)}} \\
\text{discrete} & & \text{continuous.}
\end{array}$$

In the discrete case, we had that all eigenvalues lay between zero and unity, $0 \leq \lambda^{(i)} \leq 1$. In the continuous case, we thus have that all eigenvalues are less than zero $\mu^{(i)} < 0$. Infact, if one recalls that the case $-1 \leq \lambda^{(i)} \leq 0$ corresponded to a solution which flicked between states, then this transfers over, as it corresponds to complex $\mu^{(i)}$, which is an oscillatory term.

We can also investigate the correspondence between our eigenstates $\phi^{(i)}$ and the left- and right-eigenstates $\chi^{(i)}, \psi^{(i)}$.

Before we continue, it is worth stressing that the bracketed-superscript denotes the "number" of the vector (i.e. first, second, third, etc), and the subscript the component of that vector.

Now, the continuous eigenstates satisfy

$$\sum_n V_{mn}\phi_n^{(i)} = \mu^{(i)}\phi_m^{(i)},$$

or, in terms of W,

$$\begin{aligned} \sum_n \sqrt{\frac{P_n^{\text{st}}}{P_m^{\text{st}}}} W_{mn}\phi_n^{(i)} &= \mu^{(i)}\phi_m^{(i)} \\ \Rightarrow \sum_n \sqrt{P_n^{\text{st}}} W_{mn}\phi_n^{(i)} &= \sqrt{P_m^{\text{st}}}\mu^{(i)}\phi_m^{(i)}. \end{aligned} \tag{2.29}$$

Notice that if we define

$$\psi_n^{(i)} \equiv \sqrt{P_n^{\text{st}}}\phi_n^{(i)},$$

then (2.29) reads

$$\sum_n W_{mn}\psi_n^{(i)} = \psi_m^{(i)}\mu^{(i)}.$$

Now, if we use the symmetry of V, then we also have that

$$\begin{aligned} \sum_n V_{nm}\phi_n^{(i)} &= \mu^{(i)}\phi_m^{(i)} \\ \Rightarrow \sum_n \sqrt{\frac{P_m^{\text{st}}}{P_n^{\text{st}}}} W_{nm}\phi_n^{(i)} &= \mu^{(i)}\phi_m^{(i)} \\ \Rightarrow \sum_n W_{nm}\frac{\phi_n^{(i)}}{\sqrt{P_n^{\text{st}}}} &= \mu^{(i)}\frac{\phi_m^{(i)}}{\sqrt{P_m^{\text{st}}}}. \end{aligned} \tag{2.30}$$

Then, if we define

$$\chi_n^{(i)} \equiv \frac{\phi_n^{(i)}}{\sqrt{P_n^{\text{st}}}},$$

then (2.30) reads

$$\sum_n \chi_n^{(i)} W_{nm} = \mu^{(i)}\chi_m^{(i)}.$$

Therefore, we have an interpretation of the left- and right-eigenvectors of W, in terms of the eigenvectors of V. To summarise the different states of each matrix V, W:

- The symmetric V has eigenstates, the i^{th} vector has its n^{th} component $\phi_n^{(i)}$. The right- and left-eigenstates of V are identical, so we do not distinguish between them.

- The matrix W has different right- and left-eigenstates. The left-eigenstate is denoted $\chi_n^{(i)}$, and right as $\psi_n^{(i)}$. They relate back to the eigenstates $\phi_n^{(i)}$ of V via

$$\psi_n^{(i)} \equiv \sqrt{P_n^{\text{st}}}\phi_n^{(i)}, \quad \chi_n^{(i)} \equiv \frac{\phi_n^{(i)}}{\sqrt{P_n^{\text{st}}}}.$$

- All eigenstates of both V, W correspond to eigenvalues $\mu^{(i)}$.

As before, the long-term behaviour of the system is characterised by the eigenstate corresponding to the eigenvalue closest to 0. This is because all eigenvalues of the continuous master equation are negative, and a large-negative number will cause its contribution to die off exponentially quickly.

Also as before, the stationary state is such that

$$\chi_n^{(1)} = 1, \quad \psi_n^{(1)} = P_n^{\text{st}}, \quad \forall n.$$

Example: Random Walker Consider a random walker, with 4 lattice sites. The transition probabilities are (a given)

$$\begin{aligned} w_{21} = 1, \quad w_{32} = \tfrac{3}{4}, \quad w_{43} = \tfrac{3}{4}, \\ w_{34} = 1, \quad w_{23} = \tfrac{1}{4}, \quad w_{12} = \tfrac{1}{4}. \end{aligned}$$

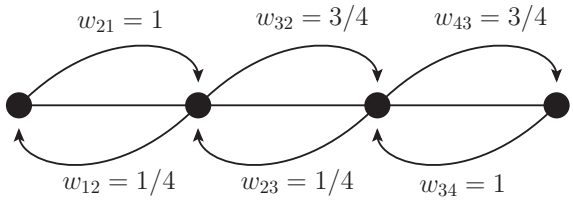

Figure 2.5: The 1D random walker, with 4 lattice sites. Also shown are the transition rates from each state.

Then, using (2.25),

$$W_{mn}(t) \equiv w_{mn}(t) - \delta_{mn} \sum_{n' \neq n} w_{n'n}(t),$$

we see that the matrix W is

$$W = \begin{pmatrix} -1 & 1/4 & 0 & 0 \\ 1 & -1 & 1/4 & 0 \\ 0 & 3/4 & -1 & 1 \\ 0 & 0 & 3/4 & -1 \end{pmatrix}.$$

Notice that the columns add to zero.

Let us find wether or not detailed balance holds. The stationary state is the state for whom

$$W \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

the solution of which is

$$\begin{pmatrix} a \\ 4a \\ 12a \\ 9a \end{pmatrix}.$$

Now, to find a, we consider the normalisation condition between the stationary state and the unit state. That is,

$$(1 \quad 1 \quad 1 \quad 1) \begin{pmatrix} a \\ 4a \\ 12a \\ 9a \end{pmatrix} = 1,$$

thus $26a = 1$, and hence $a = 1/26$. Therefore, the stationary state is

$$|P^{\rm st}\rangle = \begin{pmatrix} 1/26 \\ 2/13 \\ 6/13 \\ 9/26 \end{pmatrix}.$$

That is, after enough time, the system will settle down to lattice point 1 once in 26-goes, lattice point 2 twice in 13-goes, lattice point 3 six times in 13-goes, and point 4 in nine times from 26-goes. That is, we have the probability of finding the system in a given lattice site, after a long time.

To determine if detailed balance holds, we must test

$$w_{nm} P^{\rm st}_m = w_{mn} P^{\rm st}_n,$$

for all n. So,

$$\begin{aligned} w_{21} P^{\rm st}_1 = w_{12} P^{\rm st}_2 \quad &\Rightarrow \quad 1.\tfrac{1}{26} = \tfrac{1}{4}.\tfrac{2}{13} \\ w_{23} P^{\rm st}_3 = w_{32} P^{\rm st}_2 \quad &\Rightarrow \quad \tfrac{1}{4}.\tfrac{6}{13} = \tfrac{3}{4}.\tfrac{2}{13} \\ &\ldots \end{aligned}$$

Which we find all to be true. Therefore, detailed balance holds.

To find the eigenvalues of W, we must determine

$$\begin{vmatrix} -1-\mu & 1/4 & 0 & 0 \\ 1 & -1-\mu & 1/4 & 0 \\ 0 & 3/4 & -1-\mu & 1 \\ 0 & 0 & 3/4 & -1-\mu \end{vmatrix} = 0.$$

The characteristic equation is thus

$$(1+\mu)^4 - \tfrac{9}{16}(1+\mu)^2 + \tfrac{3}{16} = 0.$$

We can "read off" some eigenvalues. We can see that $\mu = 0, -2$ are easily attained from this. To get the other two, define $x \equiv (1+\mu)^2$. Then, the characteristic equation is

$$x^2 - \tfrac{19}{16}x + \tfrac{3}{16} = 0,$$

which factorises to

$$(x-1)(x-\tfrac{3}{16}) = 0.$$

Therefore, from this, we have three more values of μ:

$$\mu = 0, -1 \pm \frac{\sqrt{3}}{4}.$$

Therefore, all the eigenvalues are

$$\mu = 0, -1 + \frac{\sqrt{3}}{4}, -1 - \frac{\sqrt{3}}{4}, -2.$$

To get V from W, we must multiply each element of W by $\sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}}$. This results in

$$V = \begin{pmatrix} -1 & 1/2 & 0 & 0 \\ 1/2 & -1 & \sqrt{3}/4 & 0 \\ 0 & \sqrt{3}/4 & -1 & \sqrt{3}/2 \\ 0 & 0 & \sqrt{3}/2 & -1 \end{pmatrix}.$$

To find the eigenvectors, say $\phi^{(i)}$ (i.e. of V), we use the relation

$$\chi_n^{(i)} = \frac{\phi_n^{(i)}}{\sqrt{P_n^{\rm st}}}.$$

Now, we can easily read off the eigenvector $\phi^{(1)}$, as we know that $\chi_n^{(1)} = 1$, and we know what the $P_n^{\rm st}$ are (the components of the stationary state). Therefore,

$$|\phi^{(1)}\rangle = \begin{pmatrix} (1/26)^{1/2} \\ (2/13)^{1/2} \\ (6/13)^{1/2} \\ (9/26)^{1/2} \end{pmatrix}.$$

Suppose that the system starts in state 1 at $t = 0$. That is, $P_n(0) = \delta_{1n}$. Then, from the expression

$$\tilde{P}_n(t) = \sum_{i,m} e^{\mu^{(i)}t} \phi_n^{(i)} \phi_m^{(i)} \tilde{P}_m(0),$$

and

$$\tilde{P}_n = \frac{P_n}{\sqrt{P_n^{\rm st}}},$$

then

$$P_n(t) = \sum_{m,i} e^{\mu^{(i)}t} \phi_n^{(i)} \phi_m^{(i)} \sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}} P_m(0).$$

If we put the initial condition in, we see that only the case with $m = 1$ contributes,

$$P_n(t) = \sum_{i=1}^{4} e^{\mu^{(i)}t} \phi_n^{(i)} \phi_1^{(i)} \sqrt{\frac{P_n^{\rm st}}{P_1^{\rm st}}}.$$

We can simplify this. Note that

$$\phi_n^{(i)} = \chi_n^{(i)} \sqrt{P_n^{\rm st}}, \quad \chi_n^{(1)} = 1,$$

then

$$\phi_n^{(1)} = \sqrt{P_n^{\rm st}}, \quad \phi_1^{(1)} = \sqrt{P_1^{\rm st}}.$$

Hence,

$$\begin{aligned} P_n(t) &= e^{\mu^{(1)}t} \phi_n^{(1)} \phi_1^{(1)} \sqrt{\frac{P_n^{\rm st}}{P_1^{\rm st}}} + \sum_{i=2}^{4} e^{\mu^{(i)}t} \phi_n^{(i)} \phi_1^{(i)} \sqrt{\frac{P_n^{\rm st}}{P_1^{\rm st}}} \\ &= e^{\mu^{(1)}t} P_n^{\rm st} + \sum_{i=2}^{4} e^{\mu^{(i)}t} \phi_n^{(i)} \phi_1^{(i)} \sqrt{\frac{P_n^{\rm st}}{P_1^{\rm st}}}. \end{aligned}$$

Finally, if we have ordered our eigenvalues in order of "least negativity", then $e^{(0)} = 0$, and

$$P_n(t) = P_n^{\rm st} + \sum_{i=2}^{4} e^{\mu^{(i)}t} \phi_n^{(i)} \phi_1^{(i)} \sqrt{\frac{P_n^{\rm st}}{P_1^{\rm st}}}.$$

Then, if $\mu^{(2)}$ is less negative than $\mu^{(3)}, \mu^{(4)}$, then that term will be the main contributor at large time. Therefore, at large time,

$$\lim_{t\to\infty} P_n(t) = P_n^{\rm st} + e^{\mu^{(2)}t} \phi_n^{(2)} \phi_1^{(2)} \sqrt{\frac{P_n^{\rm st}}{P_1^{\rm st}}}.$$

Example: Periodic Lattice Let us consider a random walk on a periodic lattice. That is, we identify one end with the other. Then, the walker walks on a circular line. We define the transition rates

$$w_{21} = 3/4, \quad w_{32} = 3/4, \quad w_{43} = 3/4, \quad w_{14} = 3/4,$$
$$w_{12} = 1/4, \quad w_{23} = 1/4, \quad w_{34} = 1/4, \quad w_{41} = 1/4.$$

Therefore, from these, we can write the transition matrix,

$$W = \begin{pmatrix} -1 & 1/4 & 0 & 3/4 \\ 3/4 & -1 & 1/4 & 0 \\ 0 & 3/4 & -1 & 1/4 \\ 1/4 & 0 & 3/4 & -1 \end{pmatrix},$$

whereby we introduce the diagonal elements so that columns sum to zero.

Now, the stationary state is found by solving

$$W \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

which easily solves to give

$$a = b = c = d,$$

thus,

$$P^{\mathrm{st}} = \begin{pmatrix} a \\ a \\ a \\ a \end{pmatrix}.$$

We find the value of a by requiring normalisation,

$$(1 \quad 1 \quad 1 \quad 1) \begin{pmatrix} a \\ a \\ a \\ a \end{pmatrix} = 1.$$

Therefore, we see that $a = 1/4$. Therefore, the stationary state is

$$P^{\mathrm{st}} = \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}.$$

Now, we must check that the condition for detailed balance holds. That is, we must check that

$$w_{nn'} P^{\mathrm{st}}_{n'} = w_{n'n} P^{\mathrm{st}}_{n}$$

holds for all n. Now, as $P^{\rm st}_{n'} = P^{\rm st}_n = 1/4$, this reduces to the requirement that

$$w_{nn'} = w_{n'n},$$

which clearly is not true for all n. Therefore, we see that detailed balance does not hold. Therefore, we cannot construct a symmetric matrix from which to solve the master equation. That is, there is always some current of probability.

2.8 Summary

The theory of Markov processes came from the Chapman-Kolmogorov equation,

$$p(x_f, t_f|x_i, t_i) = \int dx_m p(x_f, t_f|x_m, t_m) p(x_m, t_m|x_i, t_i),$$

which can be thought of as being the probability of going from state i to state f, over any possible intermediate state m.

Discrete Time For cases where time is discrete, we wrote the CK equation as

$$P_n(t+1) = \sum_m Q_{nm}(t) P_m(t),$$

where the transition matrix has elements

$$Q_{mn}(t) = P(m, t+1|n, t),$$

describing the probaility of making a transition from state n to state m. The right eigenvectors of Q are denoted $|\psi\rangle$, and the left $\langle\chi|$. We showed that powers of the matrix Q can be computed via

$$Q^t = \sum_i \lambda_i^t |\psi^i\rangle\langle\chi^i|,$$

where λ_i is the eigenvalue of Q. All eigenvalues are less than or equal to one, in magnitude. The right-eigenvector corresponding to an eigenvalue of 1, is the stationary state. The elements of the left-eigenvector, with this eigenvalue, are all unity. Given the initial state of the system, $P(0)$, all subsequent states may be computed via

$$P(t) = Q^t P(0).$$

Continuous Time We wrote the CK equation, for systems who have continuous time, in the form of a master equation,

$$\frac{dP_n(t)}{dt} = \sum_{n' \neq n} w_{nn'}(t) P_{n'}(t) - \sum_{n' \neq n} w_{n'n}(t) P_n(t),$$

where $w_{nn'}$ is the rate of transitions from the state n' to the state n.

One-Step Processes If we have systems for whom states only transition to other states in their immediate vicinity, then we write the master equation as

$$\frac{dP_n(t)}{dt} = r_{n+1}P_{n+1}(t) + g_{n-1}P_{n-1}(t) - (g_n + r_n)P_n(t),$$

where

$$w_{n+1,n} \equiv g_n, \quad w_{n-1,n} \equiv r_n.$$

Linear one-step processes are solved by introducing the generating function,

$$F(z,t) = \sum_n z^n P_n(t),$$

and solving the subsequent differential equation. The rate of change of the expectation value of a stochastic variable was shown to be

$$\frac{d}{dt}\langle n(t)\rangle = \langle g_n\rangle - \langle r_n\rangle .$$

The stationary state is the state which is independent of time,

$$\frac{dP_n^{\rm st}}{dt} = 0.$$

Detailed Balance We can solve the master equation, if detailed balance holds,

$$w_{nn'}P_{n'}(t) = w_{n'n}P_n(t) \quad \forall n, n'.$$

To aid solving the master equation, we define the transition matrix

$$W_{mn}(t) = w_{mn}(t) - \delta_{mn}\sum_{n'\neq n} w_{n'n}(t),$$

the then master equation can be written as

$$\frac{dP_m(t)}{dt} = \sum_n W_{mn}(t)P_n(t).$$

We further write

$$V_{mn} = \sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}}W_{mn} = V_{nm},$$

which, if detailed balance holds, allows us to compute the solution to the master equation. The stationary state is the state which corresponds to a zero eigenvalue of V. If the eigenvectors of V are $\phi^{(i)}$, then the solution to the master equation is

$$P_n(t) = \sum_{i,m} e^{\mu^{(i)}t}\phi_n^{(i)}\phi_m^{(i)}P_m(0)\sqrt{\frac{P_n^{\rm st}}{P_m^{\rm st}}},$$

only if detailed balance holds.

We have so far discussed Markov processes for whom time is both continuous and discrete. However, in both cases, the state space has been discrete. That is, there has always been a set number of values that the system can be in. We shall now consider systems for whom the state space is continuous.

3 Drift & Diffusion

3.1 Introduction

Frequently, we do not want to work with a discrete state space such as

$$n = \ldots, -2, -1, 0, 1, 2, \ldots.$$

Now, we shall introduce this by example, but will go on to derive the main equation in a more rigorous manner.

In the last section, we introduced the Wright-Fisher model, which in essence, reduces to picking N genes to form the next generation of a population. The model worked via the set of rules

- Choose an individual at random,
- Copy it,
- Put copy into generation $t + 1$,
- Put original back into generation t,
- Repeat until generation $t + 1$ has as many individuals as generation t.

We then found that

$$Q_{nn'} = \binom{N}{n} p^n q^{N-n}, \quad p \equiv \frac{n'}{N}.$$

A model which is in some sense similar, but is a one-step process, was introduced by Moran in 1958.

3.1.1 The Moran Model

At a given time, two individuals are chosen. One is designated the parent, which is copied to create an offspring, the other is sacrificed in order to make way for the new offspring.

Now, suppose that there are n' individuals of type A at time t, and that there are $N - n'$ of type B. Then, the different ways in which we can pick out two individuals are

- Pick two A's. Then, the composition of the population is unchanged. This happens

  $$\left(\frac{n'}{N}\right)\left(\frac{n'-1}{N-1}\right)$$

  of the time.

- Pick two B's. Then, again, the composition does not change. This happens

$$\left(\frac{N-n'}{N}\right)\left(\frac{(N-1)-n'}{N-1}\right)$$

of the time.

- If A and B are picked, then this happens

$$\frac{n'}{N}\left(\frac{(N-1)-(n'-1)}{N-1}\right)+\left(\frac{N-n'}{N}\right)\left(\frac{n'}{N-1}\right).$$

This may be written as

$$2\frac{n'}{N}\frac{N-n'}{N-1},$$

and corresponds to either $n=n'+1$ or $n=n'-1$, where n is the number of A's.

Then, we can write the transition matrix as

$$Q_{nn'}=\begin{cases}\frac{n'}{N}\left(\frac{N-n'}{N-1}\right) & n=n'+1,\\ 1-\frac{2n'(n'-N)}{N(N-1)} & n=n',\\ \frac{n'}{N}\frac{N-n'}{N-1} & n=n'-1.\end{cases}$$

We can see that these all sum to unity. Now, under the master equation formulation,

$$w_{nn'}=\begin{cases}\frac{n'}{N}\left(\frac{n'-N}{N-1}\right) & n=n'+1,\\ \frac{n'}{N}\frac{n'-N}{N-1} & n=n'-1.\end{cases}$$

That is, a symmetric one-step process, with

$$g_n=r_n=\frac{n}{N}\left(\frac{N-n}{N-1}\right).$$

Hence, notice that the macroscopic equation looks like

$$\frac{d}{dt}\langle n\rangle=\langle g_n\rangle-\langle r_n\rangle=0,$$

thus, the ensemble average is a constant, $\langle n\rangle=n_0$, say.

We can now derive the diffusion approximation, for this specific model.

The master equation reads

$$\frac{dP(n,t)}{dt}=r_{n+1}P(n+1,t)+g_{n-1}P(n-1,t)-(r_n+g_n)P(n,t),$$

where we have used the notation $P(n,t) \equiv P_n(t)$. Therefore, using our g_n, r_n,

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

Now, we make the substitution $x \equiv n/N$. In making this substitution, we see that when we had $n = 0, 1, 2, 3, \ldots$, we now have $x = \frac{0}{N}, \frac{1}{N}, \frac{2}{N}, \ldots$. Then, as $N \to \infty$, it is clear that x becomes continuous. Before we substitute x into the above master equation, we shall make clear what it is we are doing to $P(n+1,t)$. So,

$$P(n+1,t) \longmapsto P(x+\tfrac{1}{N},t) = P(x,t) + \frac{1}{N}\frac{\partial P}{\partial x} + \frac{1}{2N^2}\frac{\partial^2 P}{\partial x^2} + \ldots,$$

that is, more or less a Taylor expansion. Note that we have change the definition of the argument of P slightly. Hence, using this, the RHS of the master equation reads

$$\frac{N}{N-1}\left[\left(x+\frac{1}{N}\right)\left(1-x-\frac{1}{N}\right)\left(P(x,t) + \frac{1}{N}\frac{\partial P}{\partial x} + \frac{1}{2N^2}\frac{\partial^2 P}{\partial x^2} + \ldots\right) + \left(x-\frac{1}{N}\right)\left(1-x+\frac{1}{N}\right)\left(P(x,t) - \frac{1}{N}\frac{\partial P}{\partial x} + \frac{1}{2N^2}\frac{\partial^2 P}{\partial x^2} + \ldots\right) - 2x(1-x)P(x,t)\right].$$

Upon expansion, one finds that

$$\frac{\partial P(x,t)}{\partial t} = \frac{N}{N-1}\left[\frac{1}{N^2}\frac{\partial^2}{\partial x^2}\left(x(1-x)P\right) + \mathcal{O}\left(\frac{1}{N^3}\right)\right].$$

Now, let us redefine time,

$$\tau \equiv \frac{2t}{N(N-1)}.$$

Hence, the master equation easily becomes

$$\frac{\partial P(x,\tau)}{\partial \tau} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[x(1-x)P(x,\tau)\right] + \mathcal{O}\left(\frac{1}{N}\right),$$

then, if we take $N \to \infty$, we can neglect the last term. Therefore, we have

$$\frac{\partial P(x,\tau)}{\partial \tau} = \frac{\partial^2}{\partial x^2}\left[D(x)P(x,\tau)\right], \quad D(x) \equiv \frac{1}{2}x(1-x).$$

That is, a diffusion equation, with diffusion "constant" $D(x)$. Therefore, we have made the systems state space, and time, continuous, on $0 \leq x \leq 1$.

We shall now discuss a general derivation of the diffusion equation.

3.2 The Fokker-Planck Equation

Let us start from the Chapman-Kolmogorov equation,

$$P(x, t + \Delta t|x_0, t_0) = \int dx' P(x, t + \Delta t|x', t)P(x', t|x_0, t_0).$$

In the following discussion, we shall ignore the initial conditions, as they will not play a role.

Now, let $x' = x - \Delta x$, then, the integrand of the CK equation is written

$$P(x, t + \Delta t|x - \Delta x, t)P(x - \Delta x, t),$$

or, alternatively, as

$$P(x - \Delta x + \Delta x, t + \Delta t|x - \Delta x, t)P(x - \Delta x, t).$$

Now, consider Taylor expanding a function of the form,

$$f(x - \Delta x; \Delta x, t, t + \Delta t) = \sum_{\ell=0}^{\infty} \frac{(-\Delta x)^\ell}{\ell!} \frac{\partial^\ell}{\partial x^\ell} f(x; \Delta x, t, t + \Delta t).$$

Also, notice that as $x' \sim \Delta x$ (note that x is constant), then

$$\int dx' \longmapsto \int d(\Delta x).$$

Now, we shall introduce *jump moments*, defined as

$$M_\ell(x, t, \Delta t) \equiv \int d(\Delta x)(\Delta x)^\ell P(x + \Delta x, t + \Delta t|x, t). \tag{3.1}$$

Therefore, using this, the CK equation reads

$$P(x, t + \Delta t) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!} \frac{\partial^\ell}{\partial x^\ell} \left[M_\ell(x, t, \Delta t)P(x, t) \right]. \tag{3.2}$$

Let us consider what the jump moments look like. Changing variables, $\Delta x = z - x$, we see that

$$M_\ell(x, t, \Delta t) = \int dz (z - x)^\ell P(z, t + \Delta t|x, t),$$

taking $\ell = 0$,

$$M_0(x, t, \Delta t) = \int dz P(z, t + \Delta t|x, t),$$

which is just the sum over all probabilities, which is unity. That is, $M_0 = 1$. An integrand with this property is just a delta-function. That is,

$$P(z, t|x, t) = \delta(z - x).$$

Now, we can expand the jump moments in powers of Δt, such that

$$M_\ell(x,t,\Delta t) = D^{(\ell)}(x,t)\Delta t + \mathcal{O}(\Delta t)^2, \quad \ell \geq 1. \tag{3.3}$$

Compare this with the transition matrix expansion in terms of the transition rates

$$Q_{nn'} = w_{nn'}\Delta t + \mathcal{O}(\Delta t)^2.$$

Therefore, using this expansion, the master equation reads

$$P(x,t+\Delta t) = P(x,t) + \sum_{\ell=1}^{\infty} \frac{(-1)^\ell}{\ell!}\frac{\partial^\ell}{\partial x^\ell}\left(D^{(\ell)}(x,t)P(x,t)\right)\Delta t + \mathcal{O}(\Delta t)^2$$

where we had to be careful of the $\ell = 0$ term. Taking the $P(x,t)$ term to the other side, and dividing by Δt results in

$$\frac{P(x,t+\Delta t) - P(x,t)}{\Delta t} = \sum_{\ell=1}^{\infty} \frac{(-1)^\ell}{\ell!}\frac{\partial^\ell}{\partial x^\ell}\left(D^{(\ell)}(x,t)P(x,t)\right) + \mathcal{O}(\Delta t).$$

Then, if we let $\Delta t \to 0$, the LHS becomes a partial differential, and we ignore the $\mathcal{O}(\Delta t)$-term on the RHS. That is,

$$\frac{\partial P}{\partial t} = \sum_{\ell=1}^{\infty} \frac{(-1)^\ell}{\ell!}\frac{\partial^\ell}{\partial x^\ell}\left(D^{(\ell)}(x,t)P(x,t)\right), \tag{3.4}$$

which is known as the *Kramers-Moyal* expansion. In deriving this, all we assumed was the Markov assumption, and that the Taylor expansion holds.

In many cases, calculation of the jump moments M_ℓ shows that they are negligible for $\ell > 2$. Therefore, we may truncate the Kramers-Moyal expansion after second order, to obtain the Fokker-Planck equation,

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\left(A(x,t)P\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(B(x,t)P\right), \tag{3.5}$$

where

$$A \equiv D^{(1)}, \quad B \equiv D^{(2)},$$

Although we have included the time dependance of A, B, they are frequently independent of time. This is the Fokker-Planck equation.

Examples of A and B come from simple diffusion and the Moran model

- Simple diffusion: has $A = 0$ and $B = 2D$, where D is a constant.
- Moran model: under $N \to \infty$, has $A = 0$ and $B = x(1-x)$.

3.2.1 Computing Jump Moments

So, we must be able to compute the jump moments, for a given process. Let us consider how to do so.

Let us use the previous notation for a conditional mean,

$$\langle x(t)\rangle_{x(t_0)=x_0} = \int dx x P(x,t|x_0 t_0),$$

where the notation reads "the mean value of the stochastic variable x at time t, given that it had value x_0 at time t_0".

So, let us go back to the definition of the jump moment, (3.1), making the substitution $z \equiv x + \Delta x$, so that

$$M_\ell(x,t,\Delta t) = \int dz (z-x)^\ell P(z,t+\Delta t|x,t).$$

Now, notice that we can take the binomial expansion

$$(z-x)^\ell = \sum_{\ell_1=0}^{\infty} \binom{\ell}{\ell_1} z^{\ell_1} (-x)^{\ell-\ell_1},$$

so that the jump moment is

$$M_\ell(x,t,\Delta t) = \int dz \sum_{\ell_1=0}^{\ell} \binom{\ell}{\ell_1} z^{\ell_1} (-x)^{\ell-\ell_1} P(z,t+\Delta t|x,t).$$

Now, we take take out of the integral, things that do not depend upon z, giving

$$M_\ell(x,t,\Delta t) = \sum_{\ell_1=0}^{\ell} \binom{\ell}{\ell_1} (-x)^{\ell-\ell_1} \int dz z^{\ell_1} P(z,t+\Delta t|x,t),$$

however, we notice that the resulting integral is just the conditional mean of z^{ℓ_1}, at time $t+\Delta t$, given that it had value x at time t. That is,

$$\int dz z^{\ell_1} P(z,t+\Delta t|x,t) = \left\langle z^{\ell_1}(t+\Delta t)\right\rangle_{z(t)=x},$$

we change variables so that we use x rather than z,

$$\int dx x^{\ell_1} P(x,t+\Delta t|x,t) = \left\langle x^{\ell_1}(t+\Delta t)\right\rangle_{x(t)=x}.$$

Therefore, the jump moment can be written as

$$M_\ell(x,t,\Delta t) = \sum_{\ell_1=0}^{\ell} \binom{\ell}{\ell_1} (-x)^{\ell-\ell_1} \left\langle x^{\ell_1}(t+\Delta t)\right\rangle_{x(t)=x}.$$

Now, notice that we can pull the $(-x)^{\ell-\ell_1}$ inside the expectation value (recall that it was just a constant with respect to the integral), so that

$$M_\ell(x,t,\Delta t) = \sum_{\ell_1=0}^{\ell} \binom{\ell}{\ell_1} \left\langle (-x)^{\ell-\ell_1} x^{\ell_1}(t+\Delta t)\right\rangle_{x(t)=x},$$

the RHS of which we notice is just the binomial expansion of

$$M_\ell(x,t,\Delta t) = \left\langle [x(t+\Delta t)-x]^\ell \right\rangle_{x(t)=x}.$$

Therefore, to calculate the jump moments, we need to compute the moments at time $t+\Delta t$, given that the system is fixed to be in state x at time t.

To see what this means, suppose that a stochastic process is described by a one-step master equation, with the transition probabilities r_n, g_n. Recall that for a one-step process, we expand the probability out as

$$P(n,t+\Delta t|n',t) = \begin{cases} 1-(r_{n'}+g_{n'})\Delta t + \mathcal{O}(\Delta t)^2 & n=n' \\ g_{n'}\Delta t + \mathcal{O}(\Delta t)^2 & n=n'+1 \\ r_{n'}\Delta t + \mathcal{O}(\Delta t)^2 & n=n'-1 \end{cases}$$

Then, the associated conditional mean is written

$$\begin{aligned} \langle n(t+\Delta t)\rangle_{n(t)=n'} &= \sum_n nP(n,t+\Delta t|n',t) \\ &= n'(1-(r_{n'}+g_{n'})\Delta t) + (n'+1)g_{n'}\Delta t \\ &\quad +(n'-1)r_{n'}\Delta t + \mathcal{O}(\Delta t)^2 \\ &= (n'+g_{n'}-r_{n'})\,\Delta t + \mathcal{O}(\Delta t)^2, \end{aligned}$$

hence, one can easily see that

$$\langle n(t+\Delta t)-n'\rangle_{n(t)=n'} = (g_{n'}-r_{n'})\,\Delta t + \mathcal{O}(\Delta t)^2.$$

Infact, its not too hard to see that

$$\left\langle [n(t+\Delta t)-n']^\ell \right\rangle_{n(t)=n'} = \left(g_{n'}+r_{n'}(-1)^\ell\right)\Delta t + \mathcal{O}(\Delta t)^2.$$

Hence, dropping the prime, we arrive at a very useful relation for one-step processes,

$$\begin{aligned} M_\ell(x,t,\Delta t) = \frac{1}{N^\ell}\left\langle [x(t+\Delta t)-x]^\ell \right\rangle_{n(t)=n} = \\ \left(g_n+r_n(-1)^\ell\right)\Delta t + \mathcal{O}(\Delta t)^2. \end{aligned} \tag{3.6}$$

We can then read off that the coefficient of Δt is just $D^{(\ell)}$, using (3.3),

$$D^{(\ell)}(x,t) = \left(g_n+r_n(-1)^\ell\right); \tag{3.7}$$

in most cases $D^{(\ell)}$ will not depend on t.

3.2.2 Simple Diffusion

Let us consider a simple, symmetric, random walk; with $r_n = g_n = \alpha$.

To make the diffusion approximation, let $x = nL$, where L is a step size, so that

$$\begin{aligned} M_\ell(x,t,\Delta t) &= \left\langle [x(t+\Delta t) - x]^\ell \right\rangle_{x(t)=x} \\ &= L^\ell \left\langle [n(t+\Delta t) - n]^\ell \right\rangle_{n(t)=n} \\ &= L^\ell \left(\alpha + (-1)^\ell \alpha\right) \Delta t + \mathcal{O}(\Delta t)^2, \end{aligned}$$

which allows us to read off that

$$D^{(\ell)} = L^\ell \left(\alpha + (-1)^\ell \alpha\right),$$

where we note that there is neither x nor t dependance. Hence, a few values of $D^{(\ell)}$ are

$$D^{(1)} = 0, \quad D^{(2)} = 2L^2\alpha, \quad D^{(3)} = 0, \quad D^{(4)} = 2L^4\alpha, \quad \ldots$$

Therefore, the Kramers-Moyal expansion is

$$\frac{\partial P}{\partial t} = \frac{2L^2\alpha}{2!}\frac{\partial^2 P}{\partial x^2} + \frac{2L^4\alpha}{4!}\frac{\partial P^4}{\partial x^4} + \ldots$$

Now, we want to let $L \to 0$, without killing off the entire equation. So, we rescale time,

$$\tau \equiv L^2 t \quad \Rightarrow \quad \frac{\partial}{\partial t} = \frac{\partial \tau}{\partial t}\frac{\partial}{\partial \tau} = L^2 \frac{\partial}{\partial \tau}.$$

Hence, the Kramers-Moyal expansion looks like

$$\frac{\partial P}{\partial \tau} = \frac{2\alpha}{2!}\frac{\partial^2 P}{\partial x^2} + \frac{2L^2\alpha}{4!}\frac{\partial P^4}{\partial x^4} + \ldots,$$

letting $L \to 0$ results in

$$\frac{\partial P}{\partial \tau} = \alpha \frac{\partial^2 P}{\partial x^2}.$$

This is a diffusion equation, with diffusion constant α. Therefore, we have derived that a simple symmetric one-step process can be modeled as a diffusion process.

3.2.3 The Moran Model

In the Moran model discussed above, we have that

$$g_n = r_n = \frac{n(N-n)}{N(N-1)}.$$

Now, we introduce

$$x \equiv \frac{n}{N},$$

so that the jump moments may be written as

$$\begin{aligned} M_\ell(x,t,\Delta t) &= \left\langle [x(t+\Delta t)-x]^\ell \right\rangle_{x(t)=x} \\ &= \frac{1}{N^\ell}\left\langle [n(t+\Delta t)-n]^\ell \right\rangle_{n(t)=n}. \end{aligned}$$

Furthermore, if we notice that in (3.7), the only non-zero $D^{(\ell)}$ are those with ℓ even (as $r_n = g_n$).

Now, notice that with $n = xN$,

$$r_n = g_n = \frac{xN(1-x)}{N-1},$$

so that

$$g_n - r_n = 0, \quad g_n + r_n = \frac{2xN(1-x)}{N-1}$$

And then that

$$D^{(\ell)} = \frac{1}{N^\ell}\left(g_n + (-1)^\ell r_n\right).$$

Hence, a few of the $D^{(\ell)}$ are

$$\begin{aligned} D^{(1)} = 0, \quad D^{(2)} &= \frac{2}{N(N-1)}x(1-x), \\ D^{(3)} = 0, \quad D^{(4)} &= \frac{2}{N^3(N-1)}x(x-1). \end{aligned}$$

Therefore, the Kramers-Moyal expansion is

$$\frac{\partial P}{\partial t} = \frac{2}{N(N-1)}\frac{1}{2!}\frac{\partial^2}{\partial x^2}\left(x(1-x)P\right) + \mathcal{O}\left(\frac{1}{N^3(N-1)}\right).$$

Now, we want to be able to let $N \to \infty$, without killing off the entire equation. So, to do so, we rescale time

$$\tau \equiv \frac{2t}{N(N-1)},$$

so that the Kramers-Moyal expansion reads

$$\frac{\partial P}{\partial \tau} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(x(x-1)P\right),$$

after letting $N \to \infty$. This is the same result we arrived at earlier.

3.2.4 The Moran Model: With Mutation

Now, in the original Moran model, we had n' A alleles and $N - n'$ B alleles. We then picked two at random, one of which was copied, the other destroyed. Both the copy and the parent were then returned to the big pot of alleles.

Let us now suppose that the copy process is not perfect. That is, we introduce a small mutation probability, so that the offspring can contain a different allele to the parent.

For example, if the parent is type A, then the offspring is of type A with probability $(1-\hat{u})$, but of type B with probability $\hat{u}$. That is, $\hat{u}$ is the mutation probability from A to B. Similarly, we let $\hat{v}$ be the mutation probability from B to A.

Then, what are the g_n, r_n in this case? Let us be systematic in setting this up.

When does the number of A alleles increase by one, in a single time step? There are two ways this can happen

1. Eliminate B and pick A as parent, with no mutation. Thus, the probability of this happening is
$$(1-\hat{u})\frac{n'(N-n')}{N(N-1)}.$$
2. Eliminate B and pick B as parent, with mutation. Then, the probability of this happening is
$$\hat{v}\frac{(N-n')(N-1-n')}{N(N-1)}.$$

Therefore, the total probability of increasing is the sum of these two cases,

$$g_{n'} = (1-\hat{u})\frac{n'(N-n')}{N(N-1)} + \hat{v}\frac{(N-n')(N-1-n')}{N(N-1)}.$$

When does the number of A alleles decrease by one, in a single time step? Again, there are two ways in which this can happen

1. Eliminate A and pick B as parent, with no mutation; with probability
$$(1-\hat{v})\frac{n'(N-n')}{N(N-1)}.$$
2. Eliminate A and pick A as parent, with mutation; with probability
$$\hat{u}\frac{n'(n'-1)}{N(N-1)}.$$

Therefore, the probability of decrease is the sum of these two probabilities

$$r_{n'} = (1-\hat{v})\frac{n'(N-n')}{N(N-1)} + \hat{u}\frac{n'(n'-1)}{N(N-1)}.$$

From hereon, everything starts to get a little messy.

Now, we use the re-scaling $n \equiv xN$, so that

$$\begin{aligned} g_n &= (1-\hat{u})\frac{N}{N-1}x(1-x) + \hat{v}\frac{N}{N-1}(1-x)\left(1-x-\frac{1}{N}\right), \\ r_n &= (1-\hat{v})\frac{N}{N-1}x(1-x) + \hat{u}\frac{N}{N-1}x\left(x-\frac{1}{N}\right). \end{aligned}$$

Hence, their sum and difference are found to be

$$\begin{aligned} g_n - r_n &= \frac{N}{N-1}\left[-\hat{u}x\left(1-\frac{1}{N}\right) + \hat{v}(1-x)\left(1-\frac{1}{N}\right)\right], \\ g_n + r_n &= \frac{N}{N-1}\left[2x(1-x) + \hat{u}x\left(2x-1-\frac{1}{N}\right) - \hat{v}x\left(2-2x-\frac{1}{N}\right)\right]. \end{aligned}$$

So, if we write the Kramers-Moyal expansion (up to second order), we see that

$$\begin{aligned} \frac{\partial P}{\partial t} = \quad & \frac{-1}{N}\frac{N}{N-1}\frac{\partial}{\partial x}\left\{\left[-\hat{u}x\left(1-\frac{1}{N}\right) + \hat{v}(1-x)\left(1-\frac{1}{N}\right)\right]P\right\} \\ & + \frac{1}{2!N^2}\frac{N}{N-1}\frac{\partial^2}{\partial x^2}\left\{\left[2x(1-x) + \hat{u}x\left(2x-1-\frac{1}{N}\right)\right.\right. \\ & \left.\left. -\hat{v}x\left(2-2x-\frac{1}{N}\right)\right]P\right\}. \end{aligned}$$

Now, if we rescale time, such that

$$\tau \equiv \frac{2t}{N(N-1)},$$

then the Kramers-Moyal expansion reads

$$\begin{aligned} \frac{\partial P}{\partial \tau} = \quad & -\frac{1}{2}N\frac{\partial}{\partial x}\left\{\left[-\hat{u}x\left(1-\frac{1}{N}\right) + \hat{v}(1-x)\left(1-\frac{1}{N}\right)\right]P\right\} \\ & + \frac{1}{4}\frac{\partial^2}{\partial x^2}\left\{\left[2x(1-x) + \hat{u}x\left(2x-1-\frac{1}{N}\right) - \hat{v}x\left(2-2x-\frac{1}{N}\right)\right]P\right\}. \end{aligned}$$

Now, let us consider the first term,

$$-\frac{1}{2}N\frac{\partial}{\partial x}\left\{\left[-\hat{u}x\left(1-\frac{1}{N}\right) + \hat{v}(1-x)\left(1-\frac{1}{N}\right)\right]P\right\},$$

then, if we do our usual thing of letting $N \to \infty$, then this expression diverges. So, let us rescale the $\hat{u}, \hat{v}$ such that

$$u \equiv \frac{\hat{u}N}{2}, \quad v \equiv \frac{\hat{v}N}{2},$$

then, this first term reads

$$-\frac{\partial}{\partial x}\left\{\left[-ux\left(1-\frac{1}{N}\right)+v(1-x)\left(1-\frac{1}{N}\right)\right]P\right\}.$$

Now, we require that u, v are finite as $N \to \infty$. This is equivalent to requiring that $\hat{u}, \hat{v}$ are very small. For example, if $N \sim 10^5$ and $\hat{u}, \hat{v} \sim 10^{-6}$, then this holds. Only under this condition does the diffusion approximation hold.

So, letting $N \to \infty$, this first term becomes

$$-\frac{\partial}{\partial x}\left\{\left[-ux+v(1-x)\right]P\right\}.$$

Furthermore, notice that the second term of the expansion has terms of order $\hat{u}, \hat{v}$, which obviously go to zero under this approximation. Therefore, the Kramer-Moyal expansion reads

$$\frac{\partial P}{\partial \tau} = -\frac{\partial}{\partial x}\left\{\left[-ux+v(1-x)\right]P\right\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(x(x-1)P\right),$$

which is a Fokker-Planck equation with

$$A(x) = -ux + v(1-x), \quad B(x) = x(1-x).$$

Notice that without mutation, $A(x)$ is non-existent. Also recall that this is only valid if the diffusion approximation is valid.

3.3 Properties of the Fokker-Planck Equation

We use the acronym FPE to denote the Fokker-Planck equation.

We have seen that in one-dimension, the FPE takes on the form

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\left[A(x,t)P\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[B(x,t)P\right], \tag{3.8}$$

where $B(x,t)$ are defined

$$B(x,t)\Delta t + \mathcal{O}(\Delta t)^2 = \left\langle \left[x(t+\Delta t) - x\right]^2 \right\rangle_{x(t)=t} \geq 0.$$

Hence, the first thing we see is that $B(x,t) \geq 0$.

Also, for most processes, A and B will not have an explicit time dependance.

3.3.1 Probability Current

Now, we can rewrite the FPE in the form

$$\frac{\partial P}{\partial t} + \frac{\partial J}{\partial x} = 0, \tag{3.9}$$

where we have defined

$$J(x,t) \equiv A(x,t)P - \frac{1}{2}\frac{\partial}{\partial x}\left[B(x,t)P\right]. \tag{3.10}$$

Now, it is clear that (3.9) is a continuity equation, where J is the probability current.

If we compare this continuity equation with the master equation for a one-step process,

$$\frac{dP_n}{dt} = J_n - J_{n+1},$$

where $J_n \equiv g_{n-1}P_{n-1} - r_nP_n$, then we see an analogy with the above continuity equation. That is, there is a direct correspondence with the continuous differential of J, with this discrete difference. That is, we can write this master equation using some discrete differential operator,

$$\frac{dP_n}{dt} + \Delta_n J = 0, \quad \Delta_n J \equiv J_{n+1} - J_n.$$

3.3.2 Boundary Conditions

Suppose the system is defined on some interval, so that $x \in (a,b)$. Then, one can conceive of some types of boundaries; how the system behaves at these boundaries.

Reflecting Boundaries Here, there is no net flow of probability across the boundaries $x = a$ and $x = b$. Hence, the current at the boundary is zero,

$$J(a,t) = J(b,t) = 0.$$

A consequence of this is that the normalisation of $P(x,t)$ is preserved over time. Let us integrate the continuity equation,

$$\int_a^b \frac{\partial P}{\partial t}dx + \int_a^b \frac{\partial J}{\partial x}dx = 0,$$

which easily gives

$$\frac{d}{dt}\int_a^b P(x,t)dx + J(b,t) - J(a,t) = 0.$$

Now, by having reflecting boundaries, the final two terms are zero, leaving the statement that

$$\frac{d}{dt}\int_a^b P(x,t)dx = 0,$$

which is the statement that the normalisation of the probability is constant (i.e. independent of time). Hence, if the probability is initially normalised,

$$\int_a^b P(x,0)dx = 1,$$

then it retains that normalisation throughout its evolution,

$$\int_a^b P(x,t)dx = 1, \quad \forall t.$$

Absorbing Boundary Conditions Here, we say that the probability of being at the boundary is zero,

$$P(a,t) = P(b,t) = 0.$$

However, if the boundaries are at $x = \pm\infty$, then we stipulate that

$$\lim_{x\to\pm\infty} P(x,t) = 0,$$

that is, the probability must decay for large x. This decay means that P is still normalisable. This also immediately implies that

$$\lim_{x\to\pm\infty} \frac{\partial P}{\partial x} = 0, \quad \lim_{x\to\pm\infty} J(x,t) = 0.$$

Infact, this is only true if A or B do not diverge as $x \to \pm\infty$.

3.3.3 The Adjoint Operator

We may write the FPE as

$$\frac{\partial P}{\partial t} = LP,$$

where

$$L \equiv -\frac{\partial}{\partial x}A(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}B(x,t). \tag{3.11}$$

So, now we have an equation which looks very much like Schrodingers equation, of quantum mechanics. The main difference is that there are no i's or $\hbar$'s. Infact, one can view this as the Schrodinger equation, but in complex time.

Now, the operator L is not necessarily Hermitian (it is real, but may not be symmetric). Its Hermitian conjugate is

$$L^\dagger = A(x,t)\frac{\partial}{\partial x} + \frac{1}{2}B(x,t)\frac{\partial^2}{\partial x^2}.$$

3.3.4 The Stationary State

From now on, we shall assume that A and B are independent of time t. That is, they are $A(x), B(x)$ only.

The stationary state $P^{\rm st}$ is that state which is independent of time. Therefore, the FPE reduces to

$$-\frac{d}{dx}\left[A(x)P^{\rm st}(x)\right]+\frac{1}{2}\frac{d^2}{dx^2}\left[B(x)P^{\rm st}(x)\right]=0,$$

where we have deliberately moved from partial to normal differentials. Now, notice that the continuity equation then becomes,

$$\frac{\partial P^{\rm st}}{\partial t}+\frac{\partial J}{\partial x}=0 \quad \Rightarrow \quad \frac{\partial J}{\partial x}=0,$$

that is, the current J is constant. Now, if we look at the definition of J, (3.10), then we see that

$$J=A(x)P^{\rm st}(x)-\frac{1}{2}\frac{\partial}{\partial x}\left[B(x)P^{\rm st}(x)\right]=\text{const.}$$

Now, in the case of *reflecting boundaries*, we have that $J=0$. Hence, the above simply reads

$$A(x)P^{\rm st}(x)=\frac{1}{2}\frac{\partial}{\partial x}\left[B(x)P^{\rm st}(x)\right].$$

This is just

$$2A(x)P^{\rm st}(x)=B'(x)P^{\rm st}(x)+B(x)\frac{dP^{\rm st}(x)}{dx},$$

which rearranges easily into

$$2\frac{A(x)}{B(x)}-\frac{B'(x)}{B(x)}=\frac{1}{P^{\rm st}(x)}\frac{dP^{\rm st}(x)}{dx}.$$

We can then integrate this,

$$\int dx\,\left(2\frac{A(x)}{B(x)}-\frac{B'(x)}{B(x)}\right)=\int dx\,\frac{1}{P^{\rm st}(x)}\frac{dP^{\rm st}(x)}{dx},$$

which gives

$$\begin{aligned}\int dx\,\left(2\frac{A(x)}{B(x)}-\frac{B'(x)}{B(x)}\right) &= \int\frac{dP^{\rm st}(x)}{P^{\rm st}(x)}\\ &= \ln P^{\rm st}(x);\end{aligned}$$

which is

$$\begin{aligned}\ln P^{\rm st}(x) &= \int 2\frac{A(x)}{B(x)}dx-\int\frac{dB(x)}{dx}\frac{dx}{B(x)}\\ &= \int 2\frac{A(x)}{B(x)}dx-\int\frac{dB(x)}{B(x)}.\end{aligned}$$

Hence,

$$\ln P^{\rm st}(x) = \int 2\frac{A(x)}{B(x)}dx - \ln B(x).$$

Therefore, this gives (also putting primes on dummy variables),

$$P^{\rm st}(x) = \frac{\mathcal{N}}{B(x)} e^{2\int_a^x \frac{A(x')dx'}{B(x')}}, \tag{3.12}$$

where $\mathcal{N}$ is some normalisation.

Brownian motion Consider a pollen grain in the overdamped limit, whereby inertial forces are small compared to viscous forces. The pollen grain moves in a potential $V(x)$ and is described by an FPE, with

$$A(x) = -\frac{dV}{dx}, \quad B(x) = 2\alpha k_B T;$$

where T is the temperature of the fluid the grains are immersed in. So, the stationary state (3.12) is given by

$$\begin{aligned} P^{\rm st}(x) &= \frac{\mathcal{N}}{2\alpha k_B T} e^{2\int_a^x \frac{-\frac{dV}{dx'}dx'}{2\alpha k_B T}} \\ &= Ce^{-\frac{V(x)\alpha^{-1}}{k_B T}}. \end{aligned}$$

Notice how similar this final expression is to the Boltzmann distribution.

3.3.5 Transformation to Schrodinger-like Equation

Based on our experience with the master equation, let us define

$$\tilde{P}(x,t) \equiv \frac{P(x,t)}{\sqrt{P^{\rm st}(x)}}. \tag{3.13}$$

Now, for simplicity, let us assume that B is independent of x. Then, we find that

$$-B\frac{\partial \tilde{P}}{\partial t} = -\frac{B^2}{2}\frac{\partial^2 \tilde{P}}{\partial x^2} + U(x)\tilde{P},$$

where

$$U(x) \equiv \frac{1}{2}\left[A(x)\right]^2 + \frac{B}{2}\frac{dA}{dx}. \tag{3.14}$$

This is equivalent to

$$\frac{\partial \tilde{P}}{\partial t} = H\tilde{P}, \tag{3.15}$$

where

$$H \equiv \frac{B}{2}\frac{\partial^2}{\partial x^2} - B^{-1}U(x). \tag{3.16}$$

This is equivalent to the Hamiltonian in the Schrodinger equation of quantum mechanics. Notice that this H is Hermitian, $H^\dagger = H$.

Hence, a one-dimensional stochastic process, under certain conditions (such as A and B having no explicit time dependance, and B being independent of x), is equivalent to quantum mechanics in imaginary time. The correspondence is

$$\hbar \longleftrightarrow B,$$

with unit mass. Infact, recall that fluctuations in quantum mechanics are of the order $\hbar$; and we had an example (Brownian motion) in which $B \sim k_B T$. Hence, we see that there are temperature fluctuations of order $k_B T$.

One can show that

$$H = \left(P^{\rm st}\right)^{-1/2} L \left(P^{\rm st}\right)^{1/2},$$

which bears strong resemblance to

$$V_{mn} = \left(P_m^{\rm st}\right)^{-1/2} W_{mn} \left(P_n^{\rm st}\right)^{1/2},$$

of the master equation formulation. We had that W was not symmetric (hence, non-Hermitian), and that V was Hermitian. Similarly, here, we have that L is non-Hermitian, but H is Hermitian.

Let us reconsider the overdamped Brownian particle, with

$$A(x) = -V'(x), \quad B = 2D.$$

Hence, the FPE looks like

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left[V'(x)P\right] + D\frac{\partial^2 P}{\partial x^2}, \tag{3.17}$$

and the "potential" $U(x)$ in the "Schrodinger equation",

$$U(x) = \frac{1}{2}\left[V'(x)\right]^2 - DV''(x).$$

3.3.6 Transformation to Adjoint Equation

Suppose we define

$$Q(x,t) \equiv \frac{P(x,t)}{P^{\rm st}(x)}, \tag{3.18}$$

then, we find that $Q(x,t)$ satisfies the adjoint equation

$$\frac{\partial Q}{\partial t} = L^\dagger Q. \tag{3.19}$$

3.3.7 Time Independent FPE

Suppose we can write

$$P(x,t) = P^{(\mu)}(x)e^{-\mu t},$$

then, as

$$\frac{\partial P}{\partial t} = -\mu P^{(\mu)}(x)e^{-\mu t} = -\mu P(x,t),$$

that is,

$$\frac{\partial P}{\partial t} = -\mu P(x,t), \tag{3.20}$$

we therefore see that

$$LP^{(\mu)} = -\mu P^{(\mu)};$$

where we have used L above (3.11). Therefore, we have put the FPE into an eigenvalue problem (compare with the time independent Schrodinger equation $\hat{H}\psi = E\psi$). Now, due to the linearity of the FPE, we can develop any solution to it into the general solution, as a sum over eigenfunctions,

$$P(x,t) = \sum_{\mu} c_\mu P^{(\mu)}(x)e^{-\mu t}, \tag{3.21}$$

where $P^{(\mu)}(x)$ are the eigenfunctions, μ the eigenvalues, and c_μ constants. Since the operator L is non-Hermitian, we need to define right- and left-eigenfunctions,

$$\begin{aligned}
\hat{L}|P^{(\mu)}\rangle &= -\mu|P^{(\mu)}\rangle, & (3.22)\\
\langle Q^{(\mu)}|\hat{L} &= -\mu\langle Q^{(\mu)}|, & (3.23)\\
\hat{L}^\dagger|Q^{(\mu)}\rangle &= -\mu|Q^{(\mu)}\rangle, & (3.24)\\
\langle P^{(\mu)}|\hat{L}^\dagger &= -\mu\langle P^{(\mu)}|. & (3.25)
\end{aligned}$$

where P and Q are the right- and left-eigenfunctions respectively.

We can, as usual, prove orthogonality and completeness of the eigenstates.

Orthogonality Let us take (3.22) and multiply through by $\langle Q^{(\mu')}|$, so that

$$\langle Q^{(\mu')}|\hat{L}|P^{(\mu)}\rangle = -\mu\langle Q^{(\mu')}|P^{(\mu)}\rangle.$$

Similarly, let us multiply (3.23) by $|P^{(\mu)}\rangle$, and prime unprimed quantities, such that

$$\langle Q^{(\mu')}|\hat{L}|P^{(\mu)}\rangle = -\mu'\langle Q^{(\mu')}|P^{(\mu)}\rangle.$$

First, notice that the LHS of both expressions are identical. Let us then subtract these two expressions, leaving

$$(\mu - \mu')\langle Q^{(\mu')}|P^{(\mu)}\rangle = 0.$$

Hence, if $\mu \neq \mu'$, we must have that

$$\langle Q^{(\mu')}|P^{(\mu)}\rangle = 0, \quad \mu \neq \mu'.$$

Now, we cannot say anything about those states with $\mu = \mu'$. If we normalise the states, by the Gram-Schmidt orthogonalisation procedure (say), then we can say that

$$\langle Q^{(\mu')}|P^{(\mu)}\rangle = \delta_{\mu\mu'}. \tag{3.26}$$

Thus the statement of orthogonality has been found. Now, we can go a little further. Consider that

$$I = \int dx|x\rangle\langle x|,$$

where I is just the identity operator. The integral must be understood to go over the entire range upon which the system is defined. Then, inserting the identity operator between the bra- and ket-state of the orthogonality statement,

$$\int dx\langle Q^{(\mu')}|x\rangle\langle x|P^{(\mu)}\rangle = \delta_{\mu\mu'}.$$

Now, we use the notation of projecting an operator into a particular coordinate representation,

$$\langle Q^{(\mu')}|x\rangle = Q^{(\mu')}(x).$$

Technically, this should be the complex conjugate, but, the functions are real. So, we hence have

$$\int dx\, Q^{(\mu')}(x)P^{(\mu)}(x) = \delta_{\mu\mu'}. \tag{3.27}$$

This "orthogonality relation" is ok, but it is between two "different" functions. Let us get it in terms of the same functions. Now, recall the definition of Q from (3.18). So, using this, the above easily becomes

$$\int dx\, \frac{P^{(\mu')}(x)P^{(\mu)}(x)}{P^{\rm st}(x)} = \delta_{\mu\mu'}. \tag{3.28}$$

Equivalently, we have that

$$\int dx\, P^{\rm st}(x)Q^{(\mu')}(x)Q^{(\mu)}(x) = \delta_{\mu\mu'}. \tag{3.29}$$

Now, both of these expressions are "true" orthonormality expressions. That is, the eigenfunctions $P^{(\mu)}$ and $Q^{(\mu)}$ are orthonormal, up to a *weight function*. That is, the weight function for the $P^{(\mu)}$ is $1/P^{\rm st}$; the weight function for the $Q^{(\mu)}$ is $P^{\rm st}$. Infact, if we recall the definition of $\tilde{P}$ from (3.13), we can see that (3.28) can be written

$$\int dx\, \tilde{P}^{(\mu')}(x)\tilde{P}^{(\mu)}(x) = \delta_{\mu\mu'}.$$

Hence, we see that the $\tilde{P}^{(\mu)}$ are orthonormal, without the need for a weight function.

Completeness If we recall the general solution (3.21), and if we multiply through by $\int dx\, Q^{(\mu')}(x)$, then we have

$$\int dx\, Q^{(\mu')}(x)P(x,t) = \sum_{\mu} c_\mu e^{-\mu t} \int dx\, Q^{(\mu')}(x)P^{(\mu)}(x),$$

the RHS integral of which is just the orthonormality relation (3.27), so that

$$\int dx\, Q^{(\mu')}(x)P(x,t) = \sum_{\mu} c_\mu e^{-\mu t}\delta_{\mu\mu'}.$$

Hence,

$$\int dx\, Q^{(\mu')}(x)P(x,t) = c_{\mu'} e^{-\mu' t}. \tag{3.30}$$

So, if we insert this back into the general solution (being careful in priming integration variables), then

$$P(x,t) = \sum_{\mu} \int dx'\, Q^{(\mu)}(x')P(x',t)P^{(\mu)}(x),$$

which we rearrange slightly into

$$P(x,t) = \int dx'\, P(x',t) \left\{ \sum_{\mu} Q^{(\mu)}(x')P^{(\mu)}(x) \right\}.$$

Now, as this must hold for all $P(x,t)$, then we must have that the bracketed quantity is a delta-function,

$$\sum_{\mu} Q^{(\mu)}(x')P^{(\mu)}(x) = \delta(x - x'); \tag{3.31}$$

whereby

$$\begin{aligned} P(x,t) &= \int dx'\, P(x't)\delta(x - x') \\ &= P(x,t). \end{aligned}$$

Hence, we have the completeness relation.

Imposing the Initial Condition Suppose that $P(x,t|x_0,t_0)$ satisfies the initial condition

$$P(x,t_0|x_0,t_0) = \delta(x - x_0).$$

Further more, we also have that

$$P(x,t|x_0,t_0) = \sum_{\mu} c_\mu e^{-\mu t} P^{(\mu)}(x).$$

Now, from (3.30), we therefore have that

$$\begin{aligned} c_\mu e^{-\mu t_0} &= \int dx\, Q^{(\mu)}(x) P(x, t_0 | x_0, t_0) \\ &= \int dx\, Q^{(\mu)}(x) \delta(x - x_0) \\ &= Q^{(\mu)}(x_0). \end{aligned}$$

Therefore, this determines c_μ,

$$c_\mu = Q^{(\mu)}(x_0) e^{\mu t_0}.$$

Hence, putting this back into the general solution,

$$P(x, t | x_0, t_0) = \sum_\mu Q^{(\mu)}(x_0) P^{(\mu)}(x) e^{-\mu(t - t_0)}.$$

Hence, we now have a general solution, with the initial conditions "built in".

Stationary State The stationary state is the state for whom the probability distribution is independent of time. That is, from (3.20), we see that this corresponds to $\mu = 0$. Hence, the general solution reduces to

$$P(x) = P^{\text{st}}(x) = Q^{(0)}(x_0) P^{(0)}(x).$$

From which we see that

$$Q^{(0)}(x_0) \frac{P^{(0)}(x)}{P^{\text{st}}(x)} = 1,$$

however, the fractioned-quantity is just the definition of $Q^{(0)}(x)$, so that this reads

$$Q^{(0)}(x_0) Q^{(0)}(x) = 1.$$

From which we say that

$$Q^{(0)}(x) = 1.$$

3.3.8 Example: Brownian Motion in a Potential

Consider the FPE

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x} [A(x, t) P] + \frac{B}{2} \frac{\partial^2 P}{\partial x^2},$$

where B is a constant. Suppose we have that

$$A(x) \equiv -\gamma x + \frac{D}{x}, \quad B \equiv 2D,$$

where γ, D are constants. This is just like Brownian motion in the potential

$$V'(x) = -A(x) = \gamma x - \frac{D}{x}.$$

That is, integrating,

$$V(x) = \frac{1}{2}\gamma x^2 - D \ln x.$$

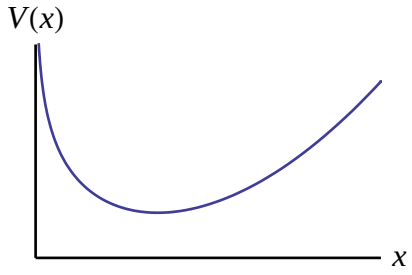

Figure 3.1: The potential that the Brownian particle "feels".

The minimum of the potential is easily seen to be at

$$x = \sqrt{\frac{D}{\gamma}}.$$

Let us say that the system is defined on $x \in (0, \infty)$, so that $J = 0$. So, the stationary state, from (3.17) is just

$$\frac{d}{dx}\left[V'(x)P^{\rm st}\right] + D\frac{d^2P^{\rm st}}{dx^2} = 0,$$

or, integrating,

$$V'(x)P^{\rm st} + D\frac{dP^{\rm st}}{dx} = 0.$$

Solving this to

$$P^{\rm st}(x) = \mathcal{N}e^{-V(x)/D},$$

using our potential,

$$\begin{aligned} P^{\rm st}(x) &= \mathcal{N}e^{-\gamma^2x^2/2D}e^{\ln x} \\ &= \mathcal{N}xe^{-\gamma^2x^2/2D}. \end{aligned}$$

Where $\mathcal{N}$ is a normalisation constant. Now, due to normalisation, the integral of the stationary state over all space is unity. Hence,

$$\int_0^\infty dx\,\mathcal{N}xe^{-\gamma^2x^2/2D} = 1.$$

To do the integral, make the substitution

$$z \equiv \frac{\gamma x^2}{2D} \quad \Rightarrow \quad dz = \frac{\gamma x}{D} dx.$$

Hence, using this, the integral allows us to find that

$$\mathcal{N} = \frac{\gamma}{D}.$$

Therefore, the stationary state is

$$P^{\rm st}(x) = \frac{\gamma x}{D} e^{-\gamma^2 x^2/2D}.$$

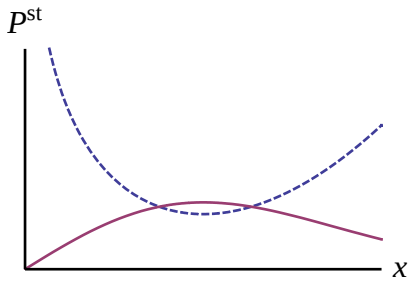

Figure 3.2: The stationary probability distribution of the Brownian particle. Also shown is the potential (dashed line) for comparison. Notice that where the probability is peaked, the potential is minimum; which should be obvious: most particles will be in the minimum of the potential.

The adjoint equation, with our expressions for $A(x)$ and B, looks like

$$\frac{\partial Q}{\partial t} = \left(-\gamma x + \frac{D}{x} \right) \frac{\partial Q}{\partial x} + D \frac{\partial^2 Q}{\partial x^2}.$$

If we then write that

$$Q(x,t) = Q^{(\mu)}(x) e^{-\mu t},$$

then the adjoint equation can be written

$$D \frac{d^2 Q^{(\mu)}}{dx^2} + \left(-\gamma x + \frac{D}{x} \right) \frac{dQ^{(\mu)}}{dx} + \mu Q^{(\mu)} = 0.$$

We can make this look nicer, by making the same change of variables to z. Using this, the above differential equation looks like

$$z \frac{d^2 Q^{(\mu)}}{dz^2} + (1-z) \frac{dQ^{(\mu)}}{dz} + \left(\frac{\mu}{2\gamma} \right) Q^{(\mu)} = 0.$$

This equation is of the form of *Laguerres' equation*. This equation also crops up in the radial part of the Hydrogen atom.

The relevant solutions (i.e. those which satisfy the boundary conditions) only exist if

$$\frac{\mu}{2\gamma} = n, \quad n = 0, 1, 2, \ldots.$$

The solutions to Laguerres' equation are called *Laguerre polynomials*, denoted $L_n(z)$. The first few are

$$L_0(z) = 1, \quad L_1(z) = 1 - z, \quad L_2(z) = 1 - 2z + \frac{1}{2}z^2, \quad \ldots$$

We have the correspondence that

$$Q^{(n)}(x) = L_n\left(\gamma x^2/2D\right), \quad n = 0, 1, 2, \ldots.$$

Notice that these states are discrete.

4 Stochastic Differential Equations

We shall introduce stochastic differential equations by example. We consider the classical example of the Brownian motion of a particle.

4.1 Brownian Motion Described by the Langevin Equation

Brownian motions classic example is that of the motion of a pollen grain in water. The pollen grain appears to undergo random motion, because it is kicked by the water molecules. This is the stochastic motion of the pollen grain.

So, we could say that at time $t = 0$, the velocity of the pollen grain is $\mathbf{v}(t = 0) = \mathbf{v}_0$. What do we expect the average values of $\mathbf{v}(t)$ and $\mathbf{v}^2(t)$ to be, at large times?

The average of $\mathbf{v}(t)$ is zero; any motion in one direction will probably be cancelled out by motion in the opposite, over an ensemble average. We can make a bit of a guess at the value of $\langle \mathbf{v}^2 \rangle$. Recall that

$$\frac{1}{2} m \langle v \rangle_{eq}^2 = \frac{3}{2} k_B T,$$

where we use the equipartition theorem for gases. Hence, we have

$$\langle v \rangle_{eq}^2 = \frac{3 k_B T}{m}.$$

So, the particle moves deterministically, but with random collisions with molecules of the fluid. Let us write the equation of motion, for a single Brownian particle:

$$m\ddot{\mathbf{r}} = -\alpha \dot{\mathbf{r}} - \nabla V + \mathbf{F}(t).$$

The α-term is a friction term, ∇V the potential and $\mathbf{F}$ some random force, due to collisions with other molecules. Let us use a simple model, whereby the particle does not feel a potential, so that the equation of motion is just

$$m\ddot{\mathbf{r}} = -\alpha \dot{\mathbf{r}} + \mathbf{F}(t),$$

or, using the notation that $\mathbf{v} \equiv \dot{\mathbf{r}}$, then

$$m\dot{\mathbf{v}} = -\alpha \mathbf{v} + \mathbf{F}(t). \tag{4.1}$$

This is known as a *Langevin equation*, and must be solved, subject to $\mathbf{v}(0) = \mathbf{v}_0$. This is a stochastic differential equation, due to the presence of the "random term". In component form, this reads

$$m\dot{v}_i = -\alpha v_i + F_i(t).$$

Now, to complete the description of the system, the force $F_i(t)$ has be specified stochastically. That is, we must define the moments,

$$\langle F_i(t)\rangle\,, \quad \langle F_i(t)F_j(t')\rangle\,, \quad \ldots. \quad i,j \in \{1,2,3\}\,.$$

We shall assume that

$$\langle F_i(t)\rangle = 0, \quad \langle F_i(t)F_j(t')\rangle = 2D\delta_{ij}\delta(t-t'), \tag{4.2}$$

where D is a constant, and all higher order moments are zero. The delta-functions mean that there is no correlation in space or time, for the collision force. That is, the force F_i becomes uncorrelated over times of a few τ, where

$$\tau \sim \frac{\text{mean molecular separation}}{\text{mean molecular velocity}} \sim 10^{-13}\text{seconds}.$$

That is,

$$\langle F_i(t)F_j(t+s)\rangle \approx 0, \quad s >> \tau.$$

We can write this a little better, as

$$\langle F_i(t)F_j(t+\Delta t)\rangle = k(\Delta t),$$

where the function k is effectively a spike, centred on t, width of order τ.

Such a specification makes the collision force $F_i(t)$ Gaussian; we call such a force *white noise*, or just the *noise term*. That is, if one took the Fourier transform of the force, one would find contributions from all modes.

So, to summarise, we have the stochastic differential equation with initial condition

$$m\dot{v}_i = -\alpha v_i + F_i(t), \quad v_i(0) = v_{i,0}.$$

We also impose the Gaussian white noise condition on the collision force parameter,

$$\langle F_i(t)\rangle = 0, \quad \langle F_i(t)F_j(t')\rangle = 2D\delta_{ij}\delta(t-t'),$$

where this gives a complete description of the Brownian motion of a particle.

4.2 The Solution to the Langevin Equation Describing Brownian Motion

Let us restrict our attention to one dimension, then the Langevin equation becomes

$$\dot{v} = -\gamma v + \eta(t),$$

where we rescale the damping and noise term as

$$\gamma \equiv \frac{\alpha}{m}, \quad \eta(t) \equiv \frac{F(t)}{m}.$$

We subject the noise term to

$$\langle \eta(t) \rangle = 0, \quad \langle \eta(t)\eta(t') \rangle = \frac{2D}{m^2}\delta(t-t').$$

Now, we have to solve $\dot{v} + \gamma v = \eta(t)$ subject to the initial condition $v(0) = v_0$. So, if we multiply the Langevin equation by the integrating factor

$$e^{\int \gamma dt} = e^{\gamma t},$$

the Langevin equation becomes

$$\frac{d}{dt}\left[v(t)e^{\gamma t}\right] = e^{\gamma t}\eta(t).$$

We can solve this by multiplying and integrating, to

$$\int_0^t dt' \frac{d}{dt'}\left[v(t')e^{\gamma t'}\right] = \int_0^t dt' e^{\gamma t'}\eta(t'),$$

which gives

$$v(t)e^{\gamma t} - v(0) = \int_0^t dt' e^{\gamma t'}\eta(t').$$

Using the initial condition that $v(0) = v_0$, and rearranging, this is just

$$v(t) = v_0 e^{-\gamma t} + e^{-\gamma t}\int_0^t dt' e^{\gamma t'}\eta(t').$$

Let us find the ensemble average $\langle v(t) \rangle$. The expectation value of the first term on the RHS is just itself (it is a fixed number). Now, the ensemble average and integral commute, so that

$$\langle v(t) \rangle = v_0 e^{-\gamma t} + e^{-\gamma t}\int_0^t dt' e^{\gamma t'} \left\langle \eta(t') \right\rangle,$$

but, $\langle \eta(t') \rangle = 0$, hence,

$$\langle v(t) \rangle = v_0 e^{-\gamma t}.$$

Let us now compute $\langle v^2(t) \rangle$. So,

$$v^2(t) = v_0^2 e^{-2\gamma t} + 2v_0 e^{-2\gamma t}\int_0^t dt' e^{\gamma t'}\eta(t') + e^{-2\gamma t}\int_0^t dt' e^{\gamma t'}\eta(t')\int_0^t dt'' e^{\gamma t''}\eta(t''),$$

and hence the expectation value

$$\begin{aligned}\left\langle v^2(t)\right\rangle &= v_0^2 e^{-2\gamma t} + 2v_0 e^{-2\gamma t}\int_0^t dt' e^{-\gamma t'}\eta(t') + \\ &\quad e^{-2\gamma t'}\left\langle \int_0^t dt' e^{\gamma t'}\eta(t')\int_0^t dt'' e^{\gamma t''}\eta(t'')\right\rangle,\end{aligned}$$

which we write as

$$\begin{aligned}\left\langle v^2(t)\right\rangle &= v_0^2 e^{-2\gamma t} + 2v_0 e^{-2\gamma t}\int_0^t dt' e^{-\gamma t'}\left\langle\eta(t')\right\rangle + \\ &\quad e^{-2\gamma t'}\int_0^t dt'\int_0^t dt'' e^{\gamma(t'+t'')}\left\langle\eta(t')\eta(t'')\right\rangle.\end{aligned}$$

We can then use our conditions for $\langle\eta(t)\rangle, \langle\eta(t)\eta(t')\rangle$, so that

$$\begin{aligned}\left\langle v^2(t)\right\rangle &= v_0^2 e^{-2\gamma t} + e^{-2\gamma t'}\frac{2D}{m^2}\int_0^t dt'\int_0^t dt'' e^{\gamma(t'+t'')}\delta(t'-t'') \\ &= v_0^2 e^{-2\gamma t} + e^{-2\gamma t'}\frac{2D}{m^2}\int_0^t dt' e^{2\gamma t'} \\ &= v_0^2 e^{-2\gamma t} + e^{-2\gamma t}\frac{2D}{m^2}\frac{1}{2\gamma}\left[e^{2\gamma t'}\right]_0^t \\ &= v_0^2 e^{-2\gamma t} + \frac{D}{\gamma m^2}e^{-2\gamma t}\left(e^{2\gamma t}-1\right).\end{aligned}$$

Hence,

$$\left\langle v^2(t)\right\rangle = v_0^2 e^{-2\gamma t} + \frac{D}{\gamma m^2}e^{-2\gamma t}\left(e^{2\gamma t}-1\right).$$

By the definition of our constants, we can write

$$\frac{D}{m^2\gamma} = \frac{Dm}{m^2\alpha} = \frac{D}{m\alpha},$$

hence,

$$\begin{aligned}\left\langle v^2(t)\right\rangle &= v_0^2 e^{-2\gamma t} + \frac{D}{m\alpha}e^{-2\gamma t}\left(e^{2\gamma t}-1\right) \\ &= v_0^2 e^{-2\gamma t} + \frac{D}{m\alpha}\left(1-e^{-2\gamma t}\right).\end{aligned}$$

So, suppose that $t\to\infty$, then this becomes

$$\lim_{t\to\infty}\left\langle v^2(t)\right\rangle = \frac{D}{m\alpha}.$$

Thus, we see that $\langle v(t)\rangle = 0$ but $\langle v^2(t)\rangle \neq 0$. This is because of the underlying equilibrium. The other terms show the dynamics as the system approaches equilibrium. Also, as $t \to \infty$, the Brownian particle thermally equilibrates with the fluid. Hence, let us notate

$$\lim_{t\to\infty} \left\langle v^2(t)\right\rangle = v_{eq}^2.$$

By the equipartition theorem, in 1D, we have that

$$\frac{1}{2}mv_{eq}^2 = \frac{1}{2}k_BT,$$

where T is the temperature of the fluid. Hence,

$$\frac{1}{2}m\frac{D}{m\alpha} = \frac{1}{2}k_BT \quad \Rightarrow \quad D = \alpha k_BT.$$

Therefore, we have a rather interesting relation. We are able to link the random fluctuations of the fluid, to the damping and temperature of the fluid. That is, we have D which is an "unknown microscopic" quantity, and α, T which are both "known macroscopic" quantities. This is an example of the *dissipation/fluctuation theorem.*

Using this dissipation/fluctuation relation, we have that

$$\left\langle v^2(t)\right\rangle = v_0^2e^{-2\gamma t} + \frac{k_BT}{m}\left(1 - e^{-2\gamma t}\right).$$

We can also compute the correlation function,

$$\langle v(t_1)v(t_2)\rangle = v_0^2e^{-\gamma(t_1+t_2)} + e^{-\gamma(t_1+t_2)}\int_0^{t_1} dt' \int_0^{t_2} dt'' e^{\gamma(t'+t'')}\left\langle \eta(t')\eta(t'')\right\rangle.$$

Now, we apply a "trick" to the integrals. First, we put in our correlation between the η's;

$$\int_0^{t_1} dt' \int_0^{t_2} dt'' e^{\gamma(t'+t'')}\left\langle \eta(t')\eta(t'')\right\rangle \longmapsto \int_0^{t_1} dt' \int_0^{t_2} dt'' e^{\gamma(t'+t'')}\delta(t'-t'').$$

Then, suppose that $t_2 > t_1$, then, we see that

$$\begin{aligned}\int_0^{t_1} dt' \int_0^{t_2} dt'' e^{\gamma(t'+t'')}\delta(t'-t'') &= \int_0^{t_1} dt' \int_0^{t_1} dt'' e^{\gamma(t'+t'')}\delta(t'-t'') \\ &\quad + \int_0^{t_1} dt' \int_{t_1}^{t_2} dt'' e^{\gamma(t'+t'')}\delta(t'-t'').\end{aligned}$$

Now, the second term is always zero, as t' is never equal to t''. Hence,

$$\int_0^{t_1} dt' \int_0^{t_2} dt'' e^{\gamma(t'+t'')}\delta(t'-t'') = \int_0^{t_1} dt' \int_0^{t_1} dt'' e^{\gamma(t'+t'')}\delta(t'-t'').$$

Therefore, using this, we see that

$$\langle v(t_1)v(t_2)\rangle = v_0^2 e^{-\gamma(t_1+t_2)} + \frac{D}{\alpha m}\left(e^{-\gamma|t_1-t_2|} - e^{-\gamma|t_1+t_2|}\right).$$

Notice that as $t_1, t_2 \to \infty$,

$$\lim_{t_1,t_2\to\infty} \langle v(t_1)v(t_2)\rangle = \frac{k_B T}{m} e^{-\gamma|t_1-t_2|}.$$

Hence, the correlation function in equilibrium.

As this is such a "useful trick", let us write the main steps again. We take the integral, and split up the integral we deem to have the largest upper limit

$$\begin{aligned}\int_0^{t_1} dt' \int_0^{t_2} dt'' f(t',t'')\delta(t'-t'') &= \int_0^{t_1} dt' \left[\int_0^{t_1} dt'' f(t',t'')\delta(t'-t'') + \right. \\ &\left. \int_{t_1}^{t_2} dt'' f(t',t'')\delta(t'-t'')\right].\end{aligned}$$

In the second integral, we see that $t' \in (0, t_1)$ and $t'' \in (t_1, t_2)$, where there is no overlap in the limits. So, an integral over such limits, with a delta-function as an integrand, results in zero. Therefore,

$$\begin{aligned}\int_0^{t_1} dt' \int_0^{t_2} dt'' f(t',t'')\delta(t'-t'') &= \int_0^{t_1} dt' \int_0^{t_1} dt'' f(t',t'')\delta(t'-t'') \\ &= \int_0^{t_1} dt' f(t',t').\end{aligned}$$

This will generally greatly simplify integrals.

4.3 Comments on the Langevin Equation

Here we shall present some generalisations, comments, and other examples of using the Langevin equation.

The Langevin equation, allowing for an external potential $V'(x)$ (in 1D) reads

$$m\ddot{x} = -\alpha\dot{x} - V'(x) + F(t).$$

Frequently, the $m\ddot{x}$ term is unimportant compared to $\alpha\dot{x}$. This is called the *overdamped limit*, so that the Langevin equation reads

$$\alpha\dot{x} = -V'(x) + F(t),$$

If there is no external potential, then this simply reads

$$\alpha\dot{x} = F(t).$$

If one finds the variance of $x^2(t)$, then one finds that

$$\left\langle\left\langle x^2(t)\right\rangle\right\rangle = 2D't, \quad D' \equiv \frac{k_B T}{\alpha},$$

which is the usual random walk result.

Let us consider if the stochastic process described by the Langevin equation, is a Markov process.

For example, consider the Langevin equation

$$m\dot{v} + \alpha v = F(t).$$

This is a Markov process if $F(t)$ is white noise, and is not Markovian if $F(t)$ is not white noise. Now, notice that the differential may be written

$$\dot{v}(t) = \frac{v(t+\tau) - v(t)}{\tau},$$

so that the Langevin equation reads

$$mv(t+\tau) - mv(t) = -\alpha\tau v(t) + \tau F(t), \quad \tau \to 0.$$

As this equation is of first order, and if $F(t)$ is uncorrelated (i.e. has no memory of past times), then this is Markovian. Similarly,

$$\dot{x} = -V'(x) + F(t)$$

is Markovian if $F(t)$ is white noise. Now,

$$m\ddot{x} + \alpha\dot{x} + V'(x) = F(t)$$

is not Markovian, because upon expansion of the differential, one finds

$$m\left[x(t+\tau) - 2x(t) + x(t-\tau)\right] = -\tau^2 V'(x) + \ldots,$$

which depends not only upon the current state of the system, but the state of the system in (one time-step) in the past. This second order property makes the Langevin equation non-Markovian. To get around this problem, we could define $v(t) \equiv \dot{x}(t)$, so that the 1D Langevin equation becomes 2D,

$$\dot{x} = v(t), \quad \dot{v} = -\gamma v - \frac{V'(x)}{m} + \frac{F(t)}{m},$$

that is, two 1D differential equations. If $F(t)$ is white noise, then this system is Markovian.

4.4 Equivalence to the Fokker-Planck Equation

Recall the Kramers-Moyal expansion (3.4),

$$\frac{\partial P}{\partial t} = \sum_{\ell=1}^{\infty} \frac{(-1)^\ell}{\ell!} \frac{\partial^\ell}{\partial x^\ell} \left(D^{(\ell)}(x,t) \right),$$

where the $D^{(\ell)}$ were defined as

$$\left\langle [x(t+\Delta t) - x(t)]^\ell \right\rangle_{x(t)=x} = M_\ell(x,t,\Delta t) = D^{(\ell)}(x,t)\Delta t + \mathcal{O}(\Delta t)^2.$$

Now, we can use the Langevin equation to compute the jump moments. As an example, consider the Langevin equation

$$\dot{v} = -\gamma v + \eta(t),$$

where

$$\langle \eta(t) \rangle = 0, \quad \langle \eta(t')\eta(t'') \rangle = \frac{2D}{m^2}\delta(t'-t'').$$

So, let us integrate the Langevin equation, from t to $t+\Delta t$; giving

$$\int_t^{t+\Delta t} \dot{v}(t')dt' = -\gamma \int_t^{t+\Delta t} v(t')dt' + \int_t^{t+\Delta t} \eta(t')dt',$$

which easily gives

$$v(t+\Delta t) - v(t) = -\gamma v(t)\Delta t + \mathcal{O}(\Delta t)^2 + f(t), \quad f(t) \equiv \int_t^{t+\Delta t} \eta(t')dt'.$$

Now, notice that

$$\langle f(t) \rangle = 0,$$

and that

$$\begin{aligned} \left\langle f^2(t) \right\rangle &= \int_t^{t+\Delta t} dt' \int_t^{t+\Delta t} dt'' \, \langle \eta(t')\eta(t'') \rangle \\ &= \frac{2D}{m^2} \int_t^{t+\Delta t} dt' \\ &= \frac{2D}{m^2}\Delta t. \end{aligned}$$

Furthermore, it is easy to convince ones-self that

$$\left\langle f^\ell(t) \right\rangle = \mathcal{O}(\Delta t)^\ell, \quad \ell \geq 3.$$

Therefore, using these, one can easily see that the moments are

$$\begin{aligned} \langle [v(t+\Delta t) - v(t)] \rangle_{v(t)=v} &= -\gamma v \Delta t + \mathcal{O}(\Delta t)^2, \\ \left\langle [v(t+\Delta t) - v(t)]^2 \right\rangle_{v(t)=v} &= \frac{2D}{m^2}\Delta t + \mathcal{O}(\Delta t)^2, \\ \left\langle [v(t+\Delta t) - v(t)]^\ell \right\rangle_{v(t)=v} &= \mathcal{O}(\Delta t)^\ell, \quad \ell \geq 3. \end{aligned}$$

Now, the $D^{(\ell)}$ is the coefficient of Δt. Hence, we read off that

$$\begin{aligned} D^{(1)} &= -\gamma v, \\ D^{(2)} &= 2\tilde{D}, \quad \tilde{D} \equiv \frac{D}{m^2}, \\ D^{(\ell)} &= 0, \quad \ell \geq 3. \end{aligned}$$

Therefore, we see that the Kramers-Moyal expansion naturally collapses into the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial v}\left(-\gamma v P\right) + \frac{1}{2!}\frac{\partial^2}{\partial v^2}\left(2\tilde{D}P\right),$$

or, cleaning up slightly,

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial v}\left(\gamma v P\right) + \frac{\partial^2}{\partial v^2}\left(\tilde{D}P\right).$$

The solution to this, which may be verified, is

$$P(v,t|v_0,0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(v-v_0 e^{-\gamma t}\right)^2}{2\sigma^2}}, \quad \sigma^2(t) \equiv \frac{D}{\alpha m}\left(1 - e^{-2\gamma t}\right),$$

which is a Gaussian. That the solution is a Gaussian is a product of the master equation being linear in v.

Now, as another example, consider the overdamped Brownian particle, in potential $V(x)$,

$$\dot{x} = -V'(x) + F(t).$$

Now, an identical argument as in the previous section, finds that

$$D^{(1)} = -V'(x), \quad D^{(2)} = 2D, \quad D^{(\ell)} = 0, \quad \ell \geq 3.$$

Which gives the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left(V'(x)\right) + D\frac{\partial^2 P}{\partial x^2}.$$

This is the FPE claimed to be describing Brownian motion, from the previous section.