Heidelberg University

Department of Physics and Astronomy

Stochastic Dynamics

Prof. Ulrich Schwarz
Winter term 2018/19
Last update: January 27, 2019

Contents

Chapter 1

Stochastic variables

In physics, there are basically three fields in which it is impossible to predict the exact outcome of a process: (1) for complex systems, where not all degrees of freedom can be controled, (2) for quantum systems, which are intrinsically non-deterministic, and (3) for deterministic chaos, when in principle the system evolves according to deterministic rules, but in such a way that even very small perturbations have large effects. Here we are concerned with the first case, that is we consider statistical systems where random processes result from fast time and small length scales which are not explicitly considered. The typical example is a large particle in contact with a bath of many small particles at finite temperature. We start with a discussion of the concept of probability and introduce stochastic variables. We then discuss important probabilities distributions and how to generate random variables from them. Finally we discuss the random walk on a line as a first step toward stochastic processes.

1.1 Axioms for probability

We start with an introduction into the basic concept of probability. Historically, problems involving probabilities have resulted in much confusion. A rigorous framework was presented by Kolmogorov in 1933, who introduced the axiomatic approach. Let us consider the set of all *possible events* Ω. We call A a particular *set of events* ($A \subset \Omega$) and $\omega \in \Omega$ a particular single event. The symbol $\emptyset$ denotes the *set of no events*.
For instance, when throwing a dice, ω could be getting a certain number, A getting an even number and Ω simply getting a number. Events can also be of a continuous nature like, for example, a Brownian particle being in volume ΔV around a certain position.
We assign a probability for A to happen, $p(A)$, that satisfies the following three axioms:

1. $p(A) \geq 0 \ \forall A$
2. $p(\Omega) = 1$
3. If A_i $(i = 1, 2, 3...)$ is a countable collection of non-overlapping sets

$$A_i \cap A_j = \emptyset \text{ for all } i \neq j$$

then

$$p(\bigcup_i A_i) = \sum_i p(A_i)$$

Consequently, we have:

1. if $\bar{A}$ is the complement of A, i.e. the set of all events not contained in A, then

$$p(\bar{A}) = 1 - p(A)$$

2. $p(\emptyset) = 0$

Relation to experiments

Assuming we can repeat an experiment as many times as we want, the probability for a certain outcome is identified with the frequency with which we measure this outcome in these independent experimental realizations.
A simple example is to consider a throwing dice experiment. If our six-sided-dice is perfectly symmetric, the probability of getting one of the six numbers is 1/6. Hence, after throwing the dice many times (N) we should obtain each number with approximately the same number of times (M), with M/N being very close to 1/6. And the higher N is, the closer it will be.
In order to measure probabilities experimentally, we have to construct an apparatus which gives frequencies in agreement with the axioms. Note that this implies a certain circularity of the procedure: a measurement process qualifies as random if it satisfies the axioms for random processes. Similar circularities are contained in all physical theories (e.g. Newtonian mechanics: a system is accepted as an inertial reference system if it obeys the laws of inertia).
Kolmogorov assumes the existence of *a priori probabilities*. Examples of such a priori probabilities are, for instance, the previous example (based on symmetry arguments) or a system in equilibrium, where within the context of statistical mechanics one assigns equal probabilities to equal volumes of the phase space. Then, one deduces the results from this reasonable *a priori* probabilities assigned and from the structure of the probability space (which is known as σ-algebra in mathematics).

1.2 Probability distributions

A *stochastic or random variable* is the outcome of a process that occurs with a certain probability. The mapping between the possible events and the probabilities defines a probability distribution. If the experimental results can be labelled with a discrete index $i \in \{1, 2, ...\}$, we will call the probability distribution *discrete*, while we will call it *continuous* if the experimental results correspond to points $x \in \Re$ on the real axis.

Discrete distributions

Throwing six-sided-dice: $i \in \{1, ..., 6\}$. Honest dice $\implies p_i = 1/6\ \forall i$. The axioms are trivially satisfied.
Flipping coin N times: Let's assign the probabilities p for the head and $q = 1 - p$ for the tail. The probability to get $k \in \{0, ..., N\}$ heads is given by the *binomial distribution*

$$p_k = \binom{N}{k} p^k q^{N-k} \tag{1.1}$$

where $\binom{N}{k} = \frac{N!}{k!(N-k)!}$ is the usual binomial coefficient. The binomial coefficient gives you the number of possible ways to obtain k heads. The factor $p^k q^{N-k}$ follows from the third axiom by considering the probability for a fixed sequence to occur.
The probabilities are manifestly positive. We can easily check the normalization condition using the binomial formula:

$$\sum_{k=0}^{N} p_k = \sum_{k=0}^{N} \binom{N}{k} p^k q^{n-k} = (p+q)^N = 1.$$

As in the total set of events we have events with every value of k, the total number of outcomes is given by the sum

$$\sum_{k=0}^{N} \binom{N}{k} = (1+1)^N = 2^N$$

as it should be.

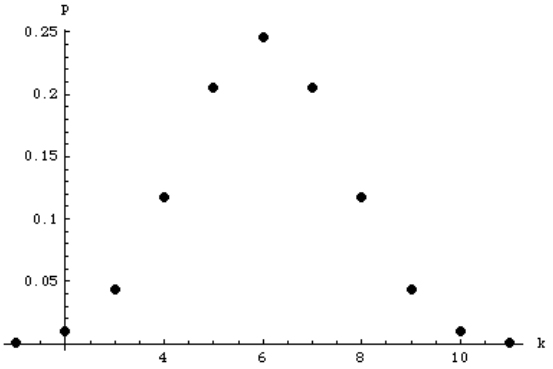


Figure 1.1: Binomial distribution for $N = 10$ and $q = 1/2$

Continuous distributions

The distribution function is given by a function $p(x)$ such that $p(x) \geq 0$ and $\int_{-\infty}^{\infty} dx p(x) = 1$ (axiom number 3 automatically satisfied by the linearity of the integral). This enables us to compute any other probability. For example, the probability of having $x \in [x_1, x_2]$ is

$$p(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} dx p(x)$$

and the probability of being around the point x_1 is

$$p(x_1 \leq x \leq x_1 + dx) = p(x_1)dx$$

We can easily recover the discrete case taking $p(x) = \sum_i p_i \delta(x - x_i)$ for the discrete values $\{x_i\}$ that x can take.
Example: particle diffusing along a line: with diffusion constant D and released at $x(t = 0) = \mu$. In this case we have the *Gaussian distribution*:

$$p(x) = \frac{1}{\sqrt{4\pi Dt}} \exp^{-(x-\mu)^2/4Dt} \tag{1.2}$$

We can see that it satisfies the initial condition ($\delta(x - \mu)$ as $t \to 0$) and that it is normalized (by using the Gauss integral, $\int_{-\infty}^{\infty} dx \exp^{-ax^2} = \left(\frac{\pi}{a}\right)^{1/2}$).

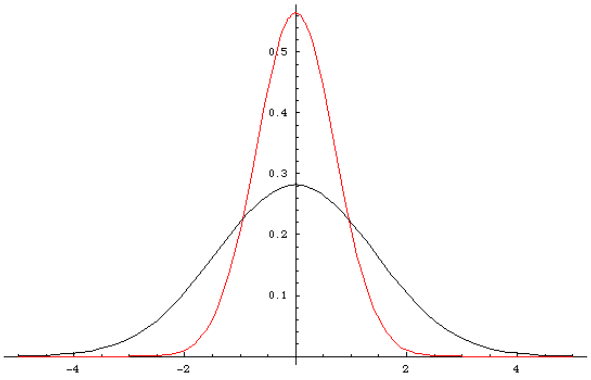


Figure 1.2: Gaussian distribution for a particle diffusing on a line. $Dt = 1/4$ for the red line and $1/2$ (later time) for the black one.

Given a probability distribution function, expectation values of an observable A can be calculated. For continuous probability densities we have

$$\langle A \rangle = \int_{-\infty}^{\infty} dx A(x) p(x) \tag{1.3}$$

while for the discrete case we have

$$\langle A \rangle = \sum_i A_i p_i \tag{1.4}$$

1.3 Characterising probability distributions

There are several quantities that are useful to characterize a given probability distribution and that we proceed to define (for continuous probability densities).

a) Moments

The moments are the quantities $\mu_n = \langle x^n \rangle$. They are often easily calculated and are useful to express other quantities of interest as we will see. The first moments are:

zero moment: $\mu_0 = 1$ (normalization)

first moment: $\mu_1 = \mu$ (mean or average)

second moment: $\mu_2 = \langle x^2 \rangle$

b) Variance

The variance or mean squared deviation (MSD) is defined as $\sigma^2 = \langle (x-\mu)^2 \rangle = \langle x^2 \rangle - 2\langle x \rangle \mu + \mu^2 = \langle x^2 \rangle - \mu^2$ and thus can be expressed in terms of the moments as $\sigma^2 = \mu_2 - \mu_1^2$. It measures how widely the distribution is spread around its average. The standard deviation or root mean square deviation (RMSD) is defined as $\sigma = \sqrt{\sigma^2}$.

c) Measures similar to mean

There are other quantities that do not give the mean value but, in some cases where getting the μ is difficult, they can be a very useful. Those are:

μ_{max}: most probable value of x

$\mu_{1/2}$: median defined by $\int_{-\infty}^{\mu_{1/2}} dx p(x) = \int_{\mu_{1/2}}^{\infty} dx p(x) = 1/2$

For a distribution with one dominating peak, μ, μ_{max} and $\mu_{1/2}$ take similar values.

d) Characteristic function

The characteristic function is defined as the Fourier transform of the probability density:

$$G(k) = \langle \exp^{ikx} \rangle = \int dx \exp^{ikx} p(x). \tag{1.5}$$

If all moments exist, we can expand the exponential:

$$G(k) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle \tag{1.6}$$

Therefore the characteristic function generates the moments:

$$\mu_n = \langle x^n \rangle = \frac{1}{i^n} \frac{d^n G(k)}{dk^n}\Big|_{k=0} \tag{1.7}$$

We say that $G(k)$ is the *generating function* of the moments.
Another further useful quantities that we introduce are the *cumulants* κ_n. We define them through their generating function $\ln G(K)$:

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

$$\kappa_n = \frac{1}{i^n} \frac{d^n (\ln G(k))}{dk^n}\Big|_{k=0} \tag{1.9}$$

The firsts cumulants are:

zero cumulant: $\kappa_0 = \ln G(k=0) = \ln 1 = 0$

first cumulant: $\kappa_1 = \frac{1}{i}\frac{G'}{G}\Big|_{k=0} = \mu$ (mean)

second cumulant: $\kappa_2 = \frac{1}{i^2}\frac{G''G - G'^2}{G^2}\Big|_{k=0} = \mu_2 - \mu_1^2 = \sigma^2$ (variance)

third cumulant: $\kappa_3 = \frac{1}{i^3}(G''' - 3G'G'' + 2G'^3)\Big|_{k=0} = \mu_3 - 3\mu_2\mu + 2\mu^3$

By continuing this scheme, one gets a one-to-one mapping between moments and cumulants.

1.4 Moments of the binomial distribution

As we have seen, the binomial distribution is given by

$$p_k = \binom{N}{k} p^k q^{N-k} \tag{1.10}$$

where $k \in \{0, ..., N\}$ and $p + q = 1$. Let's calculate the first moments of this distribution:

$$\langle k \rangle = \sum_{k=0}^{N} k \frac{N!}{k!(N-k)!} p^k q^{N-k} = Np \sum_{k=0}^{N} \frac{(N-1)!}{(k-1)!(N-k)!} p^{k-1} q^{N-k} \tag{1.11}$$

$$= Np \sum_{m=0}^{S} \frac{S!}{m!(S-m)!} p^m q^{S-m} = Np \tag{1.12}$$

where we did the change of variables $S = N - 1$ and $m = k - 1$.
There is a more elegant way to get this result that will be useful for future calculations:

$$\langle k \rangle = \sum_{k=0}^{N} \binom{N}{k} \big(p\frac{d}{dp}\big)(p^k q^{N-k}) = \big(p\frac{d}{dp}\big)(p+q)^N = Np(p+q)^{(N-1)} = Np \tag{1.13}$$

It is important to treat p and q as independent variables while we do the derivative. Otherwise, the first step would already be wrong. Only at the very end, when the derivative has been performed, we are allowed to substitute $q = p - 1$.
We use this trick to calculate the second moment:

$$\langle k^2 \rangle = \textstyle\sum_{k=0}^{N} \binom{N}{k} \big(p\frac{d}{dp}\big)^2 (p^k q^{N-k}) = \big(p\frac{d}{dp}\big)^2 (p+q)^N = Np + N(N-1)p^2$$
$$\implies \sigma^2 = \langle k^2 \rangle - \langle k \rangle^2 = Np - Np^2 = Npq$$

In summary, we found

$$\langle k \rangle = Np, \quad \sigma^2 = Npq \tag{1.14}$$

Another interesting quantity that we may be interested in is the relative width, σ/μ, of the distribution:

$$\frac{\sigma}{\langle k \rangle} = \frac{\sqrt{Npq}}{Np} = \sqrt{\frac{p}{q}} \frac{1}{N^{1/2}} \to 0 \text{ as } N \to \infty$$

We see that the relative width goes to zero for many realizations, that is, in the limit $N \to \infty$ (*self-averaging*).
The continuous limit ($N \to \infty$) of the binomial distribution can be taken in different ways. We will next see how we obtain the Gaussian and the Poisson distribution from those different limits.

1.5 Gaussian distribution

We first take the limit $N \to \infty$ of the binomial distribution with the probability p kept constant. One of the immediate observations is that the average $\langle k \rangle = Np \to \infty$ diverges. We are displacing the peak of the distribution to higher values of k. As we will see, the asymmetry of the original binomial distribution will vanish and we will obtain a perfectly symmetric distribution.
We will find the Gaussian distribution by expanding the binomial distribution around the value of k, k_m, that maximizes the probability (*method of steepest decent*). For convenience, we do not expand the p_k but its logarithm $\ln p_k$. Since the logarithm is a monotonic function, the maximum will not change.

$$\ln p_k = lnN! - \ln(N-k)! - \ln k! + k \ln p + (N-k) \ln q$$

Using the Stirling's formula $\ln N! = N \ln N - N + \frac{1}{2} \ln(2\pi N) + O(1/N)$ we find

$$\begin{aligned} 0 \stackrel{!}{=} \frac{d \ln p_k}{dk} &= \frac{d}{dk}[-\ln(N-k)! - \ln k!] + \ln p - \ln q = \\ &= -\frac{d}{dk}[k \ln k - k + (N-k)\ln(N-k) - (N-k)] + \ln(\tfrac{p}{q}) = \\ &= -[\ln k - \ln(N-k)] + \ln(\tfrac{p}{q}) = \ln\Big(\frac{(N-k)p}{kq}\Big) \implies k_m = Np = \langle k \rangle \end{aligned}$$

So the maximum occurs at the average. We now expand around this maximum to harmonic order (2nd):

$$\ln p_k \approx \ln p_k\Big|_{k=k_m} + \tfrac{1}{2} \frac{d^2}{dk^2} \ln p_k\Big|_{k=k_m} (k-k_m)^2$$

A linear term is absent around the extremum.
Using again the Stirling's formula (this time with the $\ln(2\pi N)$ term included), we find

$$\ln p_k\Big|_{k=k_m} \approx \Big(N\ln N - N - k\ln k + k - (N-k)\ln(N-k) + (N-k) + k\ln p + (N-k)\ln q + \tfrac{1}{2}\ln\frac{2\pi N}{2\pi k 2\pi(N-k)}\Big)\Big|_{k=k_m=Np} = N\ln N - Np\ln(Np) - Nq\ln(Nq) + Np\ln p + Nq\ln q + \tfrac{1}{2}\ln\frac{1}{2\pi Npq} = (N - Nq - Np)\ln N + \tfrac{1}{2}\ln\frac{1}{2\pi Npq} = \ln(2\pi Npq)^{-1/2}$$

and, using previous results,

$$\frac{d^2}{dk^2}\ln p_k\Big|_{k=k_m} = \frac{d}{dk}\ln\frac{(N-k)p}{kq}\Big|_{k=k_m} = -\frac{1}{Npq}$$

Therefore, we get

$$\ln p_k \approx \ln(2\pi Npq)^{-1/2} - \frac{1}{2Npq}(k-Np)^2 \tag{1.15}$$

Taking the exponential and using $\mu = pN$ and $\sigma^2 = Npq$, we finally obtain the *Gaussian distribution*

$$\boxed{p_k = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(k-\mu)^2/2\sigma^2}} \tag{1.16}$$

Note that this distribution is now perfectly symmetric around the mean and defined for all values of k. Moreover the distribution is normalized to 1. Comparison with equation 1.2 shows that for continuous diffusion on a line, the variance grows linear in time, $\sigma^2 = 2Dt$, which is a prominent feature of a random walk.

1.6 Poisson distribution

The second limit in the binomial distribution that we can take is $N \to \infty$ but this time $p \to 0$ ("rare events") in such a way that $\mu = Np = constant$ (in accordance to the literature we call that constant λ). Hence, unlike in the other limit, here we keep the position of the mean constant. The choice of $p = \frac{\lambda}{N}$ satisfies the requirement. Plugging it in the binomial distribution and taking the limit $N \to \infty$ gives us the Poisson distribution.

$$p_k = \frac{N!}{(N-k)!k!}p^k q^{N-k} = \frac{N(N-1)...(N-k+1)}{k!}\Big(\frac{\lambda}{N}\Big)^k\Big(1-\frac{\lambda}{N}\Big)^N\Big(\frac{1-\lambda}{N}\Big)^{-k}$$
$$= 1\Big(1-\frac{1}{N}\Big)...\Big(1-\frac{k-1}{N}\Big)\Big(\frac{\lambda^k}{k!}\Big)\Big(1-\frac{\lambda}{N}\Big)^N\Big(\frac{1-\lambda}{N}\Big)^{-k} \to \frac{\lambda^k}{k!}e^{-\lambda}$$

where we have used that in the limit $N \to \infty$

$$1\Big(1-\frac{1}{N}\Big)...\Big(1-\frac{k-1}{N}\Big) \to 1$$
$$\Big(\frac{1-\lambda}{N}\Big)^{-k} \to 1$$
$$\Big(1-\frac{\lambda}{N}\Big)^N \to e^{-\lambda}$$

Therefore,

$$\boxed{p_k = \frac{\lambda^k}{k!} e^{-\lambda}} \tag{1.17}$$

Normalization can be checked as follows

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} e^{\lambda} = 1$$

The first moment is:

$$\langle k \rangle = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

Higher moments can be derived recursively:

$$\lambda \frac{d}{d\lambda} \langle k^n \rangle = \sum_{k=0}^{\infty} \frac{k^n}{k!} e^{-\lambda} k \lambda^k - e^{-\lambda} k \lambda^{k+1} = \langle k^{n+1} \rangle - \lambda \langle k^n \rangle$$

$\implies \langle k^2 \rangle = \lambda + \lambda^2 \implies \sigma^2 = \lambda$ The variance and average are identical!
$\implies \langle k^3 \rangle = \lambda(\frac{d}{d\lambda} + 1)(\lambda + \lambda^2) = \lambda + 3\lambda^2 + \lambda^3$ etc
The Poisson distribution is completely determined by its first moment λ.

1.7 Central limit theorem

By analysing the properties of the Gaussian distribution we will find a very important result known as the *central limit theorem*.
First, we calculate its characteristic function and its cumulants. The characteristic function is given by:

$$G(k) = \int dx e^{ikx} p(x) = \int dx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}[x-(\mu+ik\sigma^2))^2 - (\mu+ik\sigma)+\mu^2]} = \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} e^{ik\mu - \frac{k^2\sigma^2}{2}}$$
$$\implies lnG(k) = ik\mu - \frac{k^2}{2}\sigma^2$$

We already see that we will only have two cumulants, since the generating function is of order two.
$\implies \kappa_1 = \frac{1}{i}\frac{d}{dk} \ln G(k)\Big|_{k=0} = \mu$; $\kappa_2 = \frac{1}{i^2}\frac{d^2}{dk^2} \ln G(k)\Big|_{k=0} = \sigma^2$; $\kappa_i = 0 \ \forall i \geq 2$
This is an important result. All probability distributions are completely characterized by their cumulants. Any distribution that has the first two cumulants non-vanishing and the other ones zero will be a Gaussian distribution. We will use this to prove the following:

Central limit theorem: *If a random variable is the sum of many independent random variables, then it has a Gaussian distribution.*

Proof:

Consider the random variable composed by sum of two independent random variables: $Y = X_1 + X_2$ Knowing the probability distributions p_{x_1} and p_{x_2} we can construct p_y as follows:

$$p_y(y) = \int dx_1 p_{x_1}(x_1) p_{x_2}(y - x_1) \implies G_y(k) = G_{x_1}(k) G_{x_2}(k)$$
$$\implies \ln G_y(k) = \ln G_{x_1}(k) + \ln G_{x_2}(k) \implies \kappa_n^y = \kappa_n^{x_1} + \kappa_n^{x_2}$$

Where, in the first line, we used the operation "convolution" and the fact that the Fourier transform of the convolution is the product of the Fourier transforms. This scheme can be iterated. If $Y = X_1 + X_2 + X_3 \implies \kappa_n^y = \kappa_n^{x_1} + \kappa_n^{x_2} + \kappa_n^{x_3}$
If we consider the random variable $Z = \frac{1}{N} \sum_{i=1}^{N} X_i$, then

$$\kappa_n^z = \frac{1}{N^n} \sum_{i=1}^{N} \kappa_n^{x_i}$$

where the $1/N^n$ comes from the fact that a cumulant like a moment involves a power of n for any scalar factor. Now consider that all the X_i random variables are the same process, i.e $X_i = X \ \forall i$, and $\langle X \rangle = \mu = 0$ and $\langle X^2 \rangle = \sigma^2$. Then, the variable Z is the average of N realizations of the process given by X and its cumulants are

$$\kappa_1^z = \langle Z \rangle = \langle X \rangle = 0$$

$$\kappa_2^z = \frac{1}{N^2} N \sigma^2 = \frac{1}{N} \sigma^2$$

$$\kappa_n^z \sim \frac{1}{N^{n-1}}$$

In the limit $N \to \infty$, only the first and second cumulants matter $\to$ Gaussian!

1.8 Generating probability distributions

Gaussian distribution from the central limit theorem

A common random number generator has the uniform probability distribution

$$p(x) = 1 \text{ for } x \in [0, 1]$$

Remember that the probability of of getting the value x (plus a differential interval) is not $p(x)$ but $p(x)dx$. The mean and the variance are:

$$\mu = 1/2$$

$$\sigma^2 = \int_0^1 x^2 dx - \mu^2 = 1/12$$

From this random number generator we can construct the following random variable x':

$$x' = (x - 1/2)\sqrt{12}\sigma$$

which has mean $\mu = 0$ and variance σ. It is equally distributed in the interval $[-\frac{1}{2}\sqrt{12}\sigma, \frac{1}{2}\sqrt{12}\sigma]$.
Then, the random variable $Z = \frac{1}{\sqrt{N}}(X'_1 + ... + X'_N)$ is Gaussian distributed with mean $\mu = 0$ and variance σ^2:

$$\kappa_1^z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \kappa_1^{x'_i} = 0$$

$$\kappa_2^z = \left(\frac{1}{\sqrt{N}}\right)^2 \sum_{i=1}^{N} \kappa_2^{x'_i} = \kappa_2^x = \sigma^2$$

$$\kappa_n^z = O(\frac{1}{\sqrt{N}}) \ \forall \ n \geq 3$$

In practice, $N = 12$ is a sufficiently good approximation.
This way of generating probability distributions by doing a coordinate transformation is a very general and useful technique.

Using coordinate transformations

Consider a coordinate transformation $x \to y(x)$ and its inverse $y \to x(y)$. Then their probability distributions are related as follows:

$$p_y(y) = p_x(x(y))\left|\frac{dx}{dy}\right| \tag{1.18}$$

where $\left|\frac{dx}{dy}\right|$ is the Jacobian of the transformation[1]. We can see it as follows

$$1 = \int dx p_x(x) = \int dy \left|\tfrac{dx}{dy}\right| p_x(x(y)) = \int dy p_y(y)$$

Let $p_y(y)$ be given and assume $p_x(x)$ to be uniformly distributed in $[0, 1]$. Then, these three steps have to be followed in order to get the appropriate coordinate transformation:

(i) $\frac{dx}{dy} = p_y(y)$ ODE to be solved[2].

(ii) $x = \int_0^y dy' p_y(y')$ Integration

(iii) $y = y(x)$ Inversion

Example 1: Exponential distribution

We want to find a coordinate transformation that generates a exponential distribution out of an equally distributed one.

$$p_y(y) = we^{-wy} \ ; \ y \in [0, \infty] \implies x = \int_0^y dy' we^{-wy'} = 1 - e^{-wy}$$
$$\implies y = -\tfrac{1}{w}\ln(1 - x)$$

Equivalently, since $(1 - x)$ is also a equally distributed number between 0 and 1, one can use

$$y = -\tfrac{1}{w}\ln(x)$$

[1] which for scalar transformation it's the modulus of the derivative of x respect to y.
[2] note that $dx/dy \geq 0$ as $x(y)$ is a distribution function

Example 2: Lorentz distribution

$$p_y(y) = \frac{1}{\pi}\frac{1}{1+y^2} \implies x = \frac{1}{\pi}\arctan y \implies y = tan(\pi x)$$

Example 3: Box-Müller procedure for Gaussian distribution

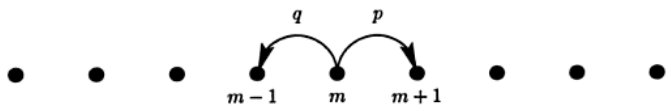
The method given above to generate a Gaussian probability distribution using the central limit theorem results to be slow to numerically compute. The Box-Müller procedure is a much faster way and one of the best procedures to create such a distribution.
Let x_1, x_2 be two random numbers uniformly distributed in $[0, 1]$. Consider the change of variables

$$\begin{array}{l} y_1 = \sqrt{-2\ln x_1}\cos(2\pi x_2) \\ y_2 = \sqrt{-2\ln x_1}\sin(2\pi x_2) \end{array} \iff \begin{array}{r} x_1 = \exp\big(-\frac{1}{2}(y_1^2+y_2^2)\big) \\ x_2 = \frac{1}{2\pi}\arctan\big(\frac{y_2}{y_1}\big) \end{array}$$

$$\implies detJ = \Big|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}\Big| = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y_2^2}$$

So y_1 and y_2 are Gaussian distributed!

1.9 Random walk on a line



The walker starts at lattice position $m = 0$ at $t = 0$. After a time Δt the walker jumps a distance Δx to the right with probability p and to the left with probability $q = 1 - p$. What is the probability p_m to be at the position m after N jumps? If the walker jumps r times to the right and l times to the left, we have $r - l = m$ and thus $r = m + l$. Moreover, $N = r + l = m + 2l$. Therefore, we have

$$l = \frac{N-m}{2}, \quad r = \frac{N+m}{2}$$

The desired probability is then given by

$$p_m = \frac{N!}{r!l!}p^r q^l = \frac{N!}{(N-r)!r!}p^r q^{N-r} \tag{1.19}$$

From the properties of the binomial distribution, we get immediately:

$$\langle r\rangle = Np, \quad \sigma_r^2 = Npq \tag{1.20}$$

$$\langle m\rangle = 2\langle r\rangle - N = 2N(p - 1/2) \tag{1.21}$$

$$\langle m^2\rangle = 4\langle r^2\rangle - 4\langle r\rangle N + N^2 = 4(Npq + N^2p^2) - 4N^2p + N^2 \tag{1.22}$$

$$= 4Npq + 4N^2(p - 1/2)^2 \tag{1.23}$$

$$\sigma_m^2 = 4Npq \tag{1.24}$$

For a symmetric jump ($p = 1/2$) we have $\langle m \rangle = 0$ and $\langle m^2 \rangle = \sigma_m^2 = N$, that is the variance grows linear with the number of jumps.
Now we take the continuous limit, that is Δx and Δt are sent to cero and the number of steps N to infinity. Then space and time become real numbers: $x = m\Delta x$, $t = N\Delta t$.
The *drift velocity* is defined as

$$v = \frac{\langle m \rangle \Delta x}{N \Delta t} = 2(p - \tfrac{1}{2})\frac{\Delta x}{\Delta t} = -2(q - \tfrac{1}{2})\frac{\Delta x}{\Delta t}$$

For ballistic motion: $p = 1 \implies v = \dfrac{\Delta x}{\Delta t}$ as it should be.
For a symmetric random walk: $p = 1/2 \implies \langle x \rangle = 0$ and $\langle x^2 \rangle = \Delta x^2 N = 2\dfrac{\Delta x^2}{2\Delta t} t = 2Dt$ where $D = \dfrac{\Delta x^2}{2\Delta t}$ is the *diffusion constant*. This corresponds to the result $\sigma^2 = 2Dt$ which we got before for continuous diffusion on a line, compare equation 1.2.
An important question is how far a particle can get in a given time t. For ballistic motion, this is $r = vt$, while for diffusive motion this is given by the RMSD $r = \sqrt{\langle x^2 \rangle} = \sqrt{2Dt}$. Thus this distance grows linear and like a square root for ballistic and diffusive motion, respectively, compare the plot.

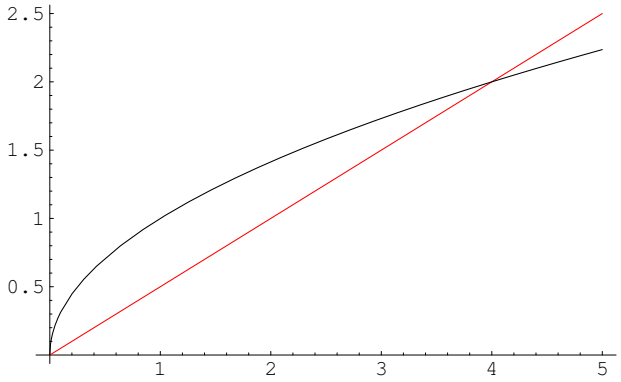


Figure 1.3: Behaviour of $r = \sqrt{\langle x^2 \rangle}$ as a function of time. In red the ballistic motion (linear) and in black the symmetric random walk (square root). Arbitrary units.

In general, the diffusion constant is given by

$$D = 2\frac{\Delta x^2}{\Delta t} pq \tag{1.25}$$

and the second moment by

$$\langle x^2 \rangle = 2Dt + (vt)^2 \tag{1.26}$$

Rate equation approach

The same problem can be approached through a rate equation (later *master equation*). From the requirement that the walker has to jump to the position m either from the left or the right, we can write

$$p(m, N+1) = p(m-1, N)p + p(m+1, N)q \tag{1.27}$$

We next replace p and q by v and D.

$$D = 2\frac{\Delta x^2}{\Delta t}pq = (2p-1)(1-p)\frac{\Delta x^2}{\Delta t} + (1-p)\frac{\Delta x^2}{\Delta t} = vq\Delta x + q\frac{\Delta x^2}{\Delta t}$$

Similarly for p:

$$D = 2\frac{\Delta x^2}{\Delta t}pq = (2q-1)(1-q)\frac{\Delta x^2}{\Delta t} + (1-q)\frac{\Delta x^2}{\Delta t} = -vp\Delta x + p\frac{\Delta x^2}{\Delta t}$$

Therefore we obtain

$$q = (D - vq\Delta x)\tfrac{\Delta t}{\Delta x^2}$$
$$p = (D + vq\Delta x)\tfrac{\Delta t}{\Delta x^2}$$

Plugging p and q in, we get

$$\frac{p(m, N+1) - p(m, N)}{\Delta t} = \frac{vp}{\Delta x}p(m-1, N) - \frac{vq}{\Delta x}p(m+1, N) \tag{1.28}$$
$$+ \frac{D}{\Delta x^2}\big(p(m+1, N) + p(m-1, N) - 2p(m, N)\big) + \Big(\frac{2D}{\Delta x^2} - \frac{1}{\Delta t}\Big)p(m, N) \tag{1.29}$$
$$= -vp\frac{p(m, N) - p(m-1, N)}{\Delta x} - vq\frac{p(m+1, N - p(m, N))}{\Delta x} \tag{1.30}$$
$$+ \frac{D}{\Delta x^2}\big(p(m+1, N) + p(m-1, N) - 2p(m, N)\big) \tag{1.31}$$
$$+ \Big(\frac{2D}{(\Delta x)^2} - \frac{1}{\Delta t} + \frac{vp}{\Delta x} - \frac{vq}{\Delta x}\Big)p(m, N) \tag{1.32}$$

The last term vanishes since

$$\Big(\frac{2D}{(\Delta x)^2} - \frac{1}{\Delta t} + \frac{vp}{\Delta x} - \frac{vq}{\Delta x}\Big) = \frac{4pq}{\Delta t} - \frac{1}{\Delta t} - \frac{2p(q-1/2)}{\Delta t} - \frac{2q(p-1/2)}{\Delta t} = 0$$

In the continuum limit one obtains the *diffusion equation with drift* (*Fick's equation* and later *Fokker-Planck equation*)

$$\boxed{\frac{\partial p(x,t)}{\partial t} = -v\frac{\partial p(x,t)}{\partial x} + D\frac{\partial^2 p(x,t)}{\partial x^2}} \tag{1.33}$$

For the initial condition $p(x, 0) = \delta(x)$, this equation is solved by the Gaussian distribution

$$p(x,t) = \frac{1}{\sqrt{4\pi Dt}}\exp\Big(-\frac{(x-vt)^2}{4Dt}\Big) \tag{1.34}$$

as obtained before in equation 1.2 without drift ($v = 0$). We conclude that the Gaussian with drift is the continuum limit for the jump problem on the line and that both give the linear increase of variance with time which is typical for random walks.

1.10 Summary

- Axioms:

  1. $p(A) \geq 0 \ \forall A$
  2. $p(\Omega) = 1$
  3. If A_i $(i = 1, 2, 3...)$ is a countable collection of non-overlapping sets

     $$A_i \cap A_j = \emptyset \text{ for all } i \neq j$$

     then

     $$p(\bigcup_i A_i) = \sum_i p(A_i)$$

- Binomial distribution:

$$p_k = \binom{N}{k} p^k q^{N-k} \tag{1.35}$$

  where $\binom{N}{k} = \frac{N!}{k!(N-k)!}$.

- Gaussian distribution:

$$p_k = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(k-\mu)^2/2\sigma^2} \tag{1.36}$$

- Poisson distribution:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda} \tag{1.37}$$

- Moments:

$$\mu_n = \langle x^n \rangle \tag{1.38}$$

- Variance:

$$\sigma^2 = \mu_2 - \mu_1^2 \tag{1.39}$$

- Characteristic function:

$$G(k) = \langle \exp^{ikx} \rangle = \int dx \exp^{ikx} p(x) \tag{1.40}$$

- Cumulants:

$$\kappa_n = \frac{1}{i^n} \frac{d^n (\ln G(k))}{dk^n} \Big|_{k=0} \tag{1.41}$$

- Central limit theorem: *If a random variable is the sum of many independent random variables, then it is Gaussian distributed.*

- Coordinate transformations:

  $x \to y(x)$ invertible. Probability distributions related as:

$$p_y(y) = p_x(x(y)) \Big| \frac{dx}{dy} \Big| \tag{1.42}$$

- Random walk on a line:

  Probability of being at the (discrete) position m after N jumps

$$p_m = \frac{N!}{r!l!} p^r q^l = \frac{N!}{(N-r)!r!} p^r q^{N-r} \tag{1.43}$$

  where

$$l = \frac{N-m}{2}, \quad r = \frac{N+m}{2}.$$

  Drift velocity:

$$v = \frac{\langle m \rangle \Delta x}{N \Delta t} = 2(p - \frac{1}{2}) \frac{\Delta x}{\Delta t} \tag{1.44}$$

  Diffusion constant:

$$D = 2 \frac{\Delta x^2}{\Delta t} pq. \tag{1.45}$$

  Second moment:

$$\langle x^2 \rangle = 2Dt + (vt)^2 \tag{1.46}$$

  where $t = N\Delta t$ and $x = m\Delta x$.

- Drift-diffusion equation:

$$\frac{\partial p(x,t)}{\partial t} = -v \frac{\partial p(x,t)}{\partial x} + D \frac{\partial^2 p(x,t)}{\partial x^2} \tag{1.47}$$

Chapter 2

Stochastic processes

In the last chapter we dealt with stochastic (or random) variables X, which had a realization x and a probability distribution $p(x)$. Here, we will go one step further and study *stochastic processes*, that are sequences of random variables $X_{t_1}, X_{t_2}, ..., X_{t_i}, ...$ Each time point t_i (that can be discrete or, dropping the sub-index, continuous) can have a different distribution function $p_{t_i}(x) = p(x, t_i)$. Even though we write x and t as the argument of p, they are completely different variables. x is a stochastic variable, while t is the time variable that plays the role of a parameter. The different realizations x of the random variable X as a function of time are called *trajectories*. The overriding aim of this chapter is to derive equations of motions for stochastic processes, like Newton's 2nd law in classical mechanics or Schrödinger's equation in quantum mechanics. Interestingly, there exists an integral equation (the Chapman-Kolmogorov equation) which is more general and from which we can derive the desired differential equations with the help of a few clear physics ideas. The Chapman-Kolmogorov equation is based on the central concept of a Markov process and we will start from here.

2.1 Markov processes and Chapman-Kolmogorov equation

Different physics will lead to different relations between the X_{t_i}. It is natural to introduce the *joint probability* of getting the trajectory $\{x_1, t_1; ...; x_n, t_n\}$ that we call $p_n(x_1, t_1; ...; x_n, t_n)$ (where the sub-index n is convenient to know that we are talking of a n-point (or n-time) probability distribution). Alternatively, we may write the same in a shorter way as $p_n(1; ...; n)$, where $i = (x_i, t_i)$.

Example: $p_2(1; 2)dx_1dx_2$ is the probability to be at position x_1 at time t_1 and at x_2 at time t_2.

Like before, from the joint probabilities we can define marginal and conditional probabilities. The simplest case of time correlations is the Markov[1] process, which

[1] Andrei Markov, 1856-1922.

is defined through the following requirement for a conditional probability:

$$p_n(n|(n-1);...;1) = p_2(n|(n-1)) \tag{2.1}$$

It says that the transition $(n-1) \to n$ is independent of the previous history. One says that the future is determined by the present (and not by the past) or that the Markov process has no memory.
The two-point conditional probability $p_2(n|(n-1))$ is called *transition probability* or *propagator*.
A Markov process is called *stationary* or *homogeneous* if $p_2(x,t;x',t')$ depends on time only through $(t-t')$.
A central property of Markov processes is that all n-point joint probabilities can be reduced to products of one and two-point probability distributions, e.g.

$$p_3(3;2;1) = p_3(3|2;1)p_2(2;1) = p_2(3|2)p_2(2|1)p(1). \tag{2.2}$$

Thus p_1 and p_2 are sufficient to determine a Markov process.
Using the definition of marginal probability and Markov process we can write

$$p_2(3;1) = p_2(3|1)p_1(1) = \int dx_2 p_3(3;2;1) = \int dx_2 p_2(3|2)p_2(2|1)p_1(1) \tag{2.3}$$

From here, it follows the *Chapman-Kolmogorov equation* (CKE):

$$\boxed{p_2(3|1) = \int dx_2 p_2(3|2)p_2(2|1)} \tag{2.4}$$

which is an integral law of motion. The time evolution from t_1 to t_2 and subsequently from t_2 to t_3 is identical to the evolution from t_1 to t_3. Later in this course we will show how the master, Fokker-Planck and Langevin equations follow as differential forms of the CKE.
Integrating the CKE times $p_1(1)$ over x_1 we obtain the second form of that equation, namely

$$p_1(3) = \int dx_2 p_2(3|2)p_1(2) \tag{2.5}$$

This equation shows that p_1 and p_2 are related to each other and one needs both to define a Markov process. This relation between p_1 and p_2 has to be satisfied for a stochastic process to be Markovian.
In the stationary limit, where time doesn't play any role, we have

$$p_{stat}(x) = \int dx_2 p_2(x,t_3|x_2,t_2)p_{stat}(x_2) \tag{2.6}$$

which can be interpreted as a kind of eigenvalue problem.
On a microscopic scale, physical process have memory. The Markov process property for the macroscopic scale is an idealization that makes it possible to define the entire process in terms of a few quantities. A process is non-Markovian when some essential element is missing. "The art of the physicist is to find those variables that are needed to make the description Markovian" (van Kampen).

2.2 Examples of Markov processes

Here the most fundamental stochastic processes are introduced. As they are Markov processes, they can be defined by the one and two-point correlation function.

Wiener process

This is an homogeneous process and, therefore, the two-point correlation function depends on time via $\tau \equiv t - t'$.

$$p_2(x,t|x',t') = \frac{1}{\sqrt{2\pi\tau}} e^{-(x-x')^2/2\tau} \tag{2.7}$$

$$p_1(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} \tag{2.8}$$

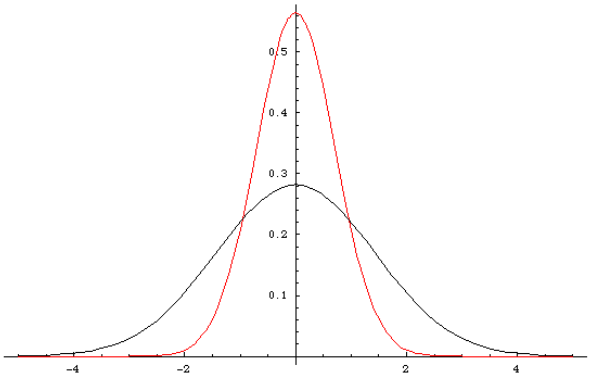


Figure 2.1: p_1 for different times ($t_{black} > t_{red}$).

The Wiener process can be interpreted as the diffusion of heat or the Brownian motion of a particle. A stationary probability p_{stat} does not exist.

Ornstein-Uhlenbeck process

$$p_2(x,t|x',t') = \frac{1}{\sqrt{2\pi(1-e^{-2\tau})}} \exp\Big(-\frac{(x-x'e^{-\tau})^2}{2(1-e^{-2\tau})}\Big) \tag{2.9}$$

$$p_1(x,t) = \frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp\Big(-\frac{(x-x_0e^{-t})^2}{2(1-e^{-2t})}\Big) \tag{2.10}$$

where x_0 is the initial condition. The stationary probability has the form:

$$p_{stat} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{2.11}$$

and the time-dependent mean and variance of this process are

$$\mu = x' e^{-\tau} \tag{2.12}$$

$$\sigma^2 = 1 - e^{-2\tau} \tag{2.13}$$

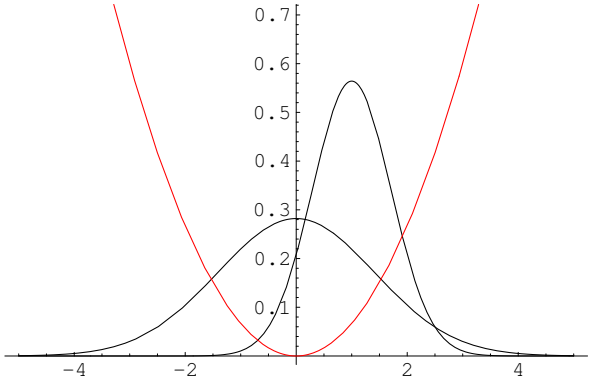


Figure 2.2: In black p_1 as a function of x for two different times. In red, an harmonic potential.

This process can be interpreted as the diffusion of a particle in a harmonic potential.

Poisson process

Here we have a discrete variable $n_{t_i} = 0, 1, ...$ which increases in time with a constant rate for the waiting time. Examples of such a process are the firing of a neuron, arrival of customers, impact of a rain drop, mutation in a population, etc.

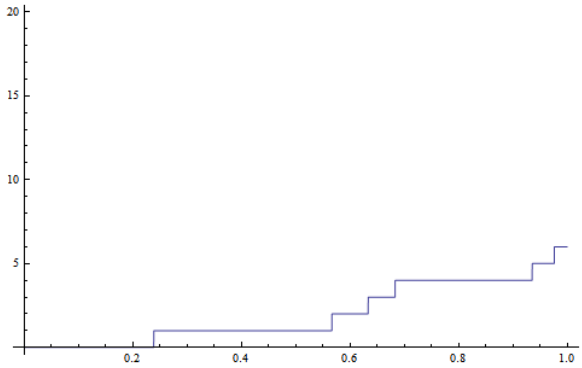


Figure 2.3: Sample path $N(t)$ of the Poisson process.

With $n \equiv n_2 - n_1 > 0$ and $\tau \equiv t_2 - t_1 > 0$, the process is defined by

$$p_2(n_2, t_2 | n_1, t_1) = \frac{(\alpha\tau)^n}{n!} e^{-\alpha\tau} \tag{2.14}$$

$$p_1(n_1, t_1) = \frac{(\alpha t_1)^{n_1}}{n_1!} e^{-\alpha t_1} \tag{2.15}$$

p_1 gives you the probability that exactly n_1 events have occurred in $[0, t_1]$.
If we compare it with the Poisson distribution

$$p_n = \frac{\lambda^n}{n!} e^{-\lambda} \tag{2.16}$$

we can identify $\lambda = \alpha t$. The mean grows linearly in time and α is the growth rate.
For small τ, the jump probability is

$$p_2(n+1, t+\tau | n, t) = \alpha\tau + O(\tau^2) \tag{2.17}$$

Therefore, α gives you the jump rate.
The Poisson process can also be interpreted as a "one-sided" random walk, in which the walker steps to the right only with probability per unit of time equal to α.

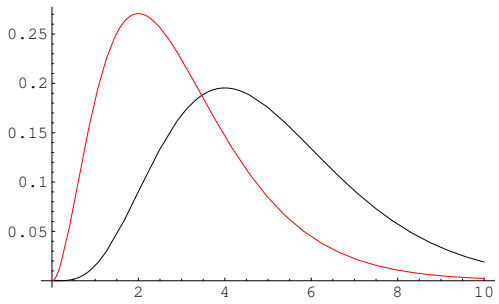


Figure 2.4: Poisson distribution as a function of time for $n = 2$ (red) and $n = 4$ (black). The maximum occurs at $t = n/\alpha = 2$ and $t = 4$ respectively.

2.3 Fokker-Planck equation

We can proceed from the Chapman-Kolmogorov equation to derive the central equation for continuous processes with diffusion, namely the Fokker-Planck equation (FPE). In order to obtain a differential equation from the CKE, we need to add statements on local properties. Here these local requirements are motivated by the short-time behaviour derived earlier for the random walk on a line. They are:

$$\int dx_2 (x_2 - x_1) p_2(x_2, t+\Delta t | x_1, t) = A(x_1, t)\Delta t + O(\Delta t^2) \tag{2.18}$$

$$\int dx_2 (x_2 - x_1)^2 p_2(x_2, t+\Delta t | x_1, t) = 2D(x_1, t)\Delta t + O(\Delta t^2) \tag{2.19}$$

Higher moments of the transition probabilities are assumed to vanish in linear order. Thus, for small times we deal with a normal distribution with mean A and variance $2D$. The coefficient A is the "drift coefficient" and D the "diffusion coefficient" (although they are actually functions)[2] The factor 2 in the second equation is a convention (some books like Gardiner or Honerkamp do not use it, but most modern treatments do).

[2]for higher dimension they become the "drift vector" and the "diffusion matrix".

We consider $x \in I \subset \Re$ and a test function $R(x)$ that vanishes together with its first derivative at the boundaries of the interval I.

$$\int_I dy R(y) \frac{\partial p_2(y,t|x',t')}{\partial t} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int dy R(y) \big[p_2(y,t+\Delta t|x',t') - p_2(y,t|x',t') \big]$$
$$\stackrel{CKE}{=} \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int dy R(y) \big[\int dz p_2(y,t+\Delta t|z,t) p_2(z,t|x',t') - p_2(y,t|x',t') \big]$$

Now, since y and z are close (because Δt is small), we expand R in $y-z$ up to second order according to the requirements stated above:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \Big\{ \int dy dz \big(R(z) + (y-z) \frac{\partial R(z)}{\partial z} + \frac{1}{2}(y-z)^2 \frac{\partial^2 R(z)}{\partial z^2} + ... \big) p_2(y,t+\Delta t|z,t) p_2(z,t|x',t')$$
$$- \int dy R(y) p_2(y,t|x',t') \Big\} = \int dz p_2(z,t|x',t') \Big\{ A(z,t) \frac{\partial R(z)}{dz} + D(z,t) \frac{\partial^2 R(z)}{\partial z^2} \Big\}$$

where for the last step we used that the first and last terms cancel and the requirement on the moments. Note that the linear assumption in Δt for the moments exactly cancels with the Δt from the time derivative.
In order to get rid of the derivatives of R, we use partial integration:

$$= \int dz R(z) \Big[- \frac{\partial}{\partial z} \big(A(z,t) p_2(z,t|x',t') \big) + \frac{\partial^2}{\partial z^2} \big(D(z,t) p_2(z,t|x',t') \big) \Big]$$

Comparing this with the first expression, the test function drops out and we obtain the *Fokker-Planck equation* (FPE), also known as *Kolmogorov equation* or *Smoluchowski equation*:

$$\boxed{\frac{\partial p_2(x,t|x',t')}{\partial t} = \big(- \frac{\partial}{\partial x} A(x,t) + \frac{\partial^2}{\partial x^2} D(x,t) \big) p_2(x,t|x',t')} \tag{2.20}$$

Since it is a partial differential equation (PDE), one needs initial and boundary conditions to solve it.

Comments on FPE

1. The derivatives act not only on A and D, but also on p_2. Therefore there might be more than two terms on the right hand side. If A and D are constants, there are only two terms and the FPE can be written as $\dot{p_2} = (-A\partial_x + D\partial_x^2) p_2$.

2. We can also write an equation for p_1. Multiplying the FPE on the right for $p_1(x',t')$ and integrating over x' we get:

$$\dot{p}_1(x,t) = (-\partial_x A + \partial_x^2 D) p_1(x,t) \tag{2.21}$$

   Thus the FPEs for p_1 and p_2 are formally the same.

3. The FPE can be written as a continuity equation:

$$\dot{p}_1 + \partial_x J = 0 \tag{2.22}$$
$$J \equiv (A - \partial_x D) p_1 \tag{2.23}$$

   with J being the *probability current*. Thus the FPE describes the flow of probability. Important boundary conditions are:

(i) Reflecting: $J|_{boundary} = 0$.
$A = 0$ and $D = constant \implies \partial_x p_1 = 0$

(ii) Absorbing: $p_1|_{boundary} = 0$

4. If viewed as a function of (x', t'), $p_2(x, t|x', t')$ satisfies the *adjoint or backward FPE*:

$$\partial_{t'} p_2(x, t|x', t') = \big(- A(x', t')\partial_{x'} - D(x', t')\partial^2_{x'}\big) p_2(x, t|x', t') \tag{2.24}$$

Note that there is no sign difference anymore and that derivatives do not act on A and D but only on p_2. That makes it a simpler equation than the FPE.

5. The derivation used here uses similar concepts (test function, integral, partial integration) as the weak formulation of the finite element method (FEM) for PDEs. Like the Schrödinger equation or the equations of continuum mechanics (Navier-Stokes equation, diffusion equation, Navier-Cauchy equation, etc) the FPE can be solved numerically by FEM-software (e.g. ANSYS, ABAQUS, Comsol Multiphysics, FEniCS, deal.II, DUNE, HighFlow3, etc).

Generalization to higher dimensions

Let $\vec{x}$ be a d-dimensional vector of random variables $\vec{x} = (x_1, ..., x_d)$. The requirements on the CKE for small times now read

$$\int d\vec{x}_2 (\vec{x}_2 - \vec{x}_1)_i p(\vec{x}_2, t + \Delta t|\vec{x}_1, t) = A_i(\vec{x}_1, t)\Delta t + O(\Delta t^2) \tag{2.25}$$

$$\int d\vec{x}_2 (\vec{x}_2 - \vec{x}_1)_i (\vec{x}_2 - \vec{x}_1)_j p(\vec{x}_2, t + \Delta t|\vec{x}_1, t) = 2D_{ij}(\vec{x}_1, t)\Delta t + O(\Delta t^2) \tag{2.26}$$

which define the drift vector A_i and the diffusion matrix D_{ij}. Proceeding analogously to the derivation of the one-dimensional FPE, one gets the multivariate FPE

$$\dot{p}_1(\vec{x}, t) = \Big(- \sum_{i=1}^{d} \partial_{x_i} A_i(\vec{x}, t) + \sum_{i,j=1}^{d} \partial_{x_i}\partial_{x_j} D_{ij}(\vec{x}, t)\Big) p_1(\vec{x}, t) \tag{2.27}$$

which can again be written as a continuity equation $\dot{p}_1 + \vec{\nabla}\vec{J} = 0$ with the probability flux $\vec{J}$ defined as

$$J_i = A_i - \sum_{J=1}^{d} \partial_{x_j} D_{ij} \tag{2.28}$$

The continuity equation tells us that the change in probability in a volume V is determined by the flux across the boundary in analogy to fluid dynamics. We can see it by integrating the continuity equation for $p_1(x, y, z, t)$ over a volume element V

$$0 = \int_V dV \frac{dp_1}{dt} + \int_V dV \vec{\nabla}\vec{J} = \frac{\partial}{\partial t} \int_V dV p_1 + \int_{A=\partial V} dA\vec{n} \cdot \vec{J} \tag{2.29}$$

where we have used Gauss' theorem to convert the volume into a surface integral.

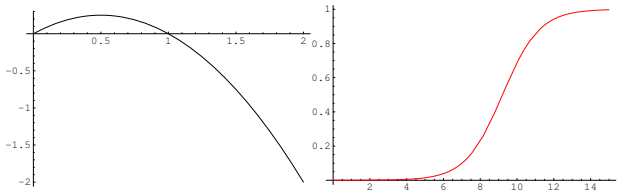
2.4 Examples for the FPE

Deterministic system

Consider a deterministic system and its trajectory $x(t)$ given by solving a differential equation of the form $\dot{x} = f(x)$. Here are two examples for such deterministic systems.

Example: Logistic growth $\dot{x} = x(1 - x/k)$

Logistic growth is the most popular model for population growth. A population without limitations in resources would grow exponentially. However, in a more realistic example, there is an upper limit to the number of individuals that the environment can support. The logistic growth shows an initially exponential behaviour but, then, a saturation begins and the growth slows down until it stops.



(a) $\dot{x}(x)$ for $k = 1$ (b) $x(t)$ for $k = 1$

Figure 2.5: Logistic growth

In order to get the deterministic limit for the FPE we need to take

$$p_1(y,t) = \delta(y - x(t)) \tag{2.30}$$

$$p_2(z,t|y,t') = \delta(z - x(t)) \text{ with } x(t') = y \tag{2.31}$$

Thus, the drift and diffusion constants are

$$A(x,t')\Delta t = \int dz(z-x)p_2(z,t'+\Delta t|x,t') = \int dz(z-x(t'))\delta(z-x(t'+\Delta t)) =$$

$$= x(t'+\Delta t) - x(t') = \dot{x}(t')\Delta t \implies \boxed{A(x,t') = \dot{x}(t') = f(x)}$$

$$D(x,t')\Delta t = \int dz(z-x)^2 p_2(z,t'+\Delta t|x,t') = \big(x(t'+\Delta t) - x(t')\big)^2 =$$

$$(\dot{x}\Delta t)^2 = O(\Delta t^2) \implies \boxed{D = 0}$$

Therefore, drift $A(x,t)$ is the deterministic part of the FPE (which we can also identify with the velocity) and diffusion $D(x,t)$ the noisy part. Therefore the FPE can be considered to be a drift-diffusion equation.

Wiener process

As previously stated, the two-point probability for the Wiener process is given by

$$p_2(x,t+\Delta t|x',t) = \frac{1}{\sqrt{2\pi\Delta t}} e^{-(x-x')^2/2\Delta t} \tag{2.32}$$

which is the solution of the FPE for $A = 0$ and $D = 1/2$ (diffusion equation)

$$\dot{p}_2 = \frac{1}{2}\partial_x^2 p_2 \tag{2.33}$$

Note that p_2 is not polynomial in Δt. If both diffusion and drift are present, we get the picture of a particle moving with velocity $v = A$ on which a Gaussian fluctuation with variance $2D$ is superimposed.

$$\implies x(t + \Delta t) = x(t) + A(x(t), t)\Delta t + (2D\Delta t)^{1/2}$$

Conclusions:

(i) Trajectories are continuous, $x(t + \Delta t) \overset{\Delta t \to 0}{\to} x(t)$.

(ii) Trajectories are nowhere differentiable (due to the $\Delta t^{1/2}$)

Diffusion in a gravitational field

Consider particles in a container with a viscous fluid with friction coefficient $\xi = 1$. We again consider the overdamped limit. The probability to find a particle at a height x is given by

$$\dot{p}_1 = \partial_x(Gp_1) + D\partial_x^2 p_1 \tag{2.34}$$

where G is the gravitational constant. In the stationary case,

$$\frac{\partial_x p_s}{p_s} = \partial_x \ln p_s = -\frac{G}{D} \implies p_s \sim e^{-Gx/D} = e^{-Gx/k_B T} \tag{2.35}$$

where in the last step we used the Einstein relation. This last expression is known as the *barometric formula*.

Ornstein-Uhlenbeck process

This is the first example with both diffusion and drift. It is the result of adding a linear term to the Wiener process. The FPE for this process is

$$\dot{p}_1 = \partial_x(\xi k x p_1) + D\partial_x^2 p_1 \tag{2.36}$$

The drift part can be physically interpreted as follows. Consider a spring submerged in a viscous fluid. The equation of motion is

$$F = m\ddot{x} = -kx - \frac{1}{\xi}v \tag{2.37}$$

where k is the spring constant and ξ the friction coefficient. In the over-damped limit (in which one roughly sends $m \to 0$)

$$v = -\xi k x = A \tag{2.38}$$

The stationary solution ($\dot{p}_s \overset{!}{=} 0$) is given by

$$\partial_x \underbrace{(\xi k x p_s + D\partial_x p_s)}_{=J} = 0 \tag{2.39}$$

Due to boundary conditions $J = const = 0$. Hence,

$$\frac{\partial_x p_s}{p_s} = -\frac{\xi k x}{D} = \partial_x \ln p_s \implies \boxed{p_s = \left(\frac{\xi k}{2\pi D}\right)^{1/2} e^{-\xi k x^2/2D}} \tag{2.40}$$

Note that this is the same result that we obtain in statistical physics when we consider a particle in an harmonic potential in a temperature bath, i.e the Boltzmann distribution $p_B(x) \sim e^{-V(x)/k_B T} = e^{-\frac{1}{2}kx^2/k_B T}$. Comparing the two we obtain the *Einstein relation*

$$\boxed{D = \frac{k_B T}{\xi}} \tag{2.41}$$

This is an example of the fluctuation-dissipation theorem. It connects two seemingly very different quantities, the diffusion constant D (fluctuation) with the friction coefficient ξ (dissipation).

Harmonic oscillator with damping

The equation of motion reads $m\ddot{x} = -kx - \gamma\dot{x}$. Using the velocity $v = \dot{x}$ we can write it as a vectorial differential equation $\dot{\vec{x}} = \vec{f}(\vec{x})$:

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{\gamma}{m} \end{pmatrix} \begin{pmatrix} x \\ v \end{pmatrix} \tag{2.42}$$

We obviously deal with a multivariate FPE. From the deterministic equation, we can read off the drift vector. We rewrite it as a matrix multiplication to see the linear character:

$$\vec{A} = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{\gamma}{m} \end{pmatrix} \vec{x} \tag{2.43}$$

What will be our diffusion matrix ? Here we note that the noise appears on the level of the second derivative, so we write

$$D = \begin{pmatrix} 0 & 0 \\ 0 & D_v \end{pmatrix} \tag{2.44}$$

In contrast, for an overdamped system we would write a diffusion constant D_x on the level of the first derivative.
Our multivariate FPE (also known as Kramers equation) now reads

$$\dot{p} = \partial_x(-vp) + \partial_v(\frac{k}{m}xp + \frac{\gamma}{m}vp + D_v\partial_v p) \tag{2.45}$$

The stationary solution is given by

$$p_s \sim e^{-\frac{\gamma}{2mD_v}v^2} e^{-\frac{\gamma k}{2m^2 D_v}x^2} \tag{2.46}$$

This gives the Boltzmann distribution if we have

$$\boxed{D_v = \frac{k_B T \gamma}{m^2}} \tag{2.47}$$

Indeed this diffusion constant has the correct physical dimension for $< v^2 >= 2D_v t$. As we will see later, it also corresponds to $D_x = k_B T/\gamma$ and $< x^2 >= 2D_x t$.

2.5 Master equation

The CKE does not only describe diffusive processes, but also jump processes. We now consider such jump process with the simplest possible law, namely that the transition (=jump) probability increases linearly with time. We write

$$p_2(x, t + \Delta t|x'', t) = \big(1 - a(x, t)\Delta t\big)\delta(x - x'') + W(x, x'', t)\Delta t + O(\Delta t^2) \quad (2.48)$$

where $\big(1 - a(x, t)\Delta t\big)$ is the probability to stay at the same point and $W(x, x'', t)$ is the probability to jump from x'' to x at a given time t (the starting point is written on the right and the end point at the left, so you should read this from right to left). These two probabilities can be related by considering the normalization condition:

$$\int dx p_2(x, t + \Delta t|x'', t) = 1 \implies \boxed{a(x'', t) = \int dx W(x, x'', t)} \quad (2.49)$$

Therefore, the larger the probability to jump, the smaller the probability to stay, as it should be.
Now, we combine it with the CKE:

$$p_2(x, t + \Delta t|x', t') = \int dx'' p_2(x, t + \Delta|x'', t)p_2(x'', t|x', t') =$$
$$= \big(1 - a(x, t)\Delta t\big)p_2(x, t|x', t') + \Delta t \int dx'' W(x, x'', t)p_2(x'', t|x', t') + O(\Delta t^2)$$

Writing $a(x, t)$ in terms of W we obtain the *master equation* (ME):

$$\boxed{\frac{\partial}{\partial t}p_2(x, t|x', t') = \int dx'' W(x, x'', t)p_2(x'', t|x', t') - \int dx'' W(x'', x, t)p_2(x, t|x', t')} \quad (2.50)$$

The first term is the gain term (all jumps to the position x) and the second term the loss term (all jumps away from position x). The ME is a balance equation, while the FPE is a continuity equation.
Since we are dealing with a jump process, it is more natural to write the ME in a discrete version. Let $p_n(t)$ be the probability to be at the site n at a given time and $W_{nn'}(t)$ the jump rate from the site n' to n. The discrete ME then reads

$$\boxed{\dot{p}_n(t) = \sum_{n'} \big(W_{nn'}p_{n'}(t) - W_{n'n}p_n(t)\big)} \quad (2.51)$$

Note the asymmetry of the sum and that there is no sum over the repeated index n on the second term.
The discrete version can be visualized as a weighted bidirectional graph. Every possible discrete state n is represented by a node of the graph while $W_{nn'}$ represents the directed edge connecting the node n' to the node n (with a certain weight). Some properties of the physical system simply follow from the topology of the graph, while others depend on the exact numbers of the weights. At any rate, a detailed treatment of master equations profits from graph theory.

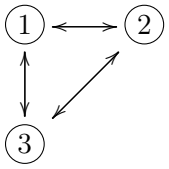
Stationary states

In many cases, our physical system arrives to an stationary state in which probabilities do not depend on time. The definition of a stationary state is $\dot{p}_n(t) = 0$ which can also be interpreted as an influx-outflux equilibrium:

$$\sum_{n'} W_{nn'} p_{n'} = \sum_{n'} W_{n'n} p_n \tag{2.52}$$

There are two types of equilibrium:

1. Local stationarity:

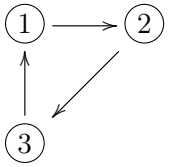


   Here each link is balanced by itself. Thus nodes are also balanced. This situation is called *detailed balance* and it is typical for thermodynamical equilibrium (follows from time-reversal invariance for closed physical systems).

   For systems in equilibrium in contact with a heat bath (Boltzmann distribution):

$$\frac{W_{n'n}}{W_{nn'}} = e^{-\Delta E/k_B T} \tag{2.53}$$

2. Global stationarity due to cycles:



   This case is only possible if loops exist in our system. It is also called a *non-equilibrium steady state* (NESS) and it is typical in driven systems. So we see that steady state does not mean equilibrium!

Matrix formulation

The ME is linear in p_i and can be solved by an eigenvalue analysis:

$$\dot{\vec{p}} = V\vec{p} \tag{2.54}$$

where

$$V_{ij} = \begin{cases} W_{ij} & i \neq j \\ -\sum_{l \neq i} W_{li} & i = j \end{cases} \tag{2.55}$$

The matrix V contains all the information of the graph (i.e. of our physical problem). Examples:

1. Two-state system: 

   The equations of motion are

$$\dot{p}_1 = W_{12}p_2 - W_{21}p_1 \tag{2.56}$$
$$\dot{p}_2 = W_{21}p_1 - W_{12}p_2 \tag{2.57}$$
$$V = \begin{pmatrix} -W_{21} & W_{12} \\ W_{21} & -W_{12} \end{pmatrix} \tag{2.58}$$

2. Three-state system:

$$\dot{p}_1 = W_{12}p_2 + W_{13}p_3 - W_{21}p_1 - W_{31}p_1 \tag{2.59}$$
$$V = \begin{pmatrix} -(W_{21}+W_{31}) & W_{12} & W_{13} \\ W_{21} & -(W_{12}+W_{32}) & W_{23} \\ W_{31} & W_{32} & -(W_{13}+W_{23}) \end{pmatrix} \tag{2.60}$$

   Notice that, according to the construction of V, all the columns of V sum to zero, the diagonal elements are non-positive and the off-diagonal ones are non-negative.

From (2.54) we see that

$$\vec{p}(dt) = (1+Vdt)\vec{p}(0) \implies \vec{p}(t) = \lim_{M\to\infty}(1+V\frac{t}{M})^M\vec{p}(0)$$
$$\implies \boxed{\vec{p}(t) = e^{Vt}\vec{p}(0)}$$

Therefore e^{Vt} is the operator that gives the time evolution[3].
A general solution can be expressed in terms of eigenvectors $\vec{p}_\lambda$ of V with eigenvalue λ as

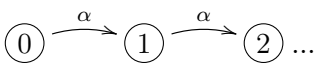
$$\vec{p}(t) = \sum_\lambda c_\lambda \vec{p}_\lambda(0)e^{\lambda t} \tag{2.61}$$

with $\vec{p}(0) = \sum c_\lambda \vec{p}_\lambda(0)$. $\lambda = 0$ describes the stationary state and $\lambda < 0$ describes different relaxation processes.

[3]Here there is a nice analogy with quantum mechanics. For symmetric (thus, hermitian) V, -V becomes the Hamiltonian of the quantum system. Therefore, the difference between the classical time evolution operator and the quantum one is only the i of the exponential. Directed edges (i.e. non-hermitian V which represent dissipation) can be treated quantum-mechanically with a quantum master equation ($\frac{d\rho}{dt} = -i[H,\rho] + \mathcal{L}$) in the context of quantum noise and decoherence.

2.6 Examples for the ME

(1) Poisson process



For the Poisson process we have

$$n = 0, 1, 2, ... \tag{2.62}$$

$$W_{nn'} = \alpha \delta_{n,n'+1} \tag{2.63}$$

The equations of motion are

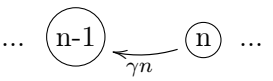
$$\dot{p}_n = \alpha p_{n-1} - \alpha p_n \quad \forall n \tag{2.64}$$

The solution of this equation is

$$p_n = \frac{(\alpha t)^n}{n!} e^{-\alpha t} \tag{2.65}$$

as can be easily checked. So we are dealing with a Poisson process with $< n >= \alpha t$.

(2) Radioactive decay



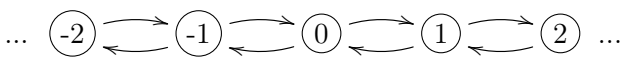
The number of nuclei decays in proportion to itself, thus we have

$$n = 0, 1, ..., n_0 \tag{2.66}$$

$$W_{nn'} = \gamma n \delta_{n,n'-1} \tag{2.67}$$

Thus the rate is linear in n. An exact solution exists (see below).

(3) Random walk on a line



For the symmetric random walk

$$n = -\infty, ..., \infty \tag{2.68}$$

$$W_{nn'} = \begin{cases} \alpha & \text{if } |n - n'| = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.69}$$

The equations are

$$\dot{p}_n = \alpha(p_{n-1} + p_{n+1} - 2p_n) \tag{2.70}$$

which is the discrete version of the diffusion equation

$$\dot{p}_n = \alpha \partial_n^2 p_n \tag{2.71}$$

An exact solution exists (see below). It is not the Gaussian, as expected in the continuum limit, but more complicated.

(4) Chemical reactions

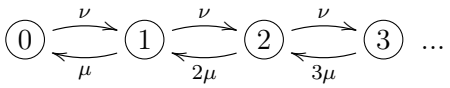
For small reaction volumes the number of molecules in the sample is crucial (not the concentrations). The ME applied to that field is often called *chemical master equation*. Here is a simple example:

$$\emptyset \xrightarrow{\nu} A$$
$$A \xrightarrow{\mu} \emptyset$$

This is a reaction in which we create a molecule A with a rate ν and it is destroyed with a rate μ. Now the number of the site represents the number of molecules A that we have. Therefore,

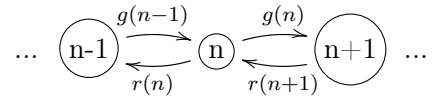
$$W_{nn'} = \begin{cases} \nu & \text{if } n = n' + 1 \\ \mu n' & \text{if } n = n' - 1 \end{cases} \tag{2.72}$$

with $n \geq 0$, where the n' is a combinatorial factor that takes into account that there are n' molecules A that can be destroyed.



One-step master equation

The three previous examples where examples of one-step master equation. There where only jumps from a site to a neighbouring site. It is convenient for later purposes to rename the jump rates and introduce $g(n) \equiv W_{n+1,n}$ (dissociation or death rate) and $r(n+1) \equiv W_{n,n+1}$ (association or birth rate).



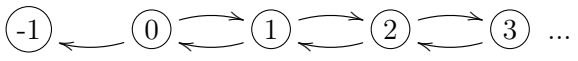
Typical applications are population dynamics, protein clusters, chemical reactions,... Like for the FPE, the boundaries are very important and key in determining the behaviour of the system. Often one has $n \geq 0$ (boundary at $n = 0$ like in the first and third examples).
Two basic boundary conditions are:

1. Reflecting boundary: $r(0) = g(-1) = 0$

   This is the case of the Poisson process for example, in which our system are the sites $[0, 1, 2...]$ and the sites $[-\infty, ..., -1]$ are completely disconnected.

2. Absorbing boundary: $g(-1) = 0$, $r(0) = c \neq 0$

   Here the site -1 acts as a sink. Some probability can go from the site 0 to the sink but it cannot go out.



   If we consider the probability $q = \sum_{n=0}^{\infty} p_n$ to be in the *allowed domain* defined as the sites $[0, \infty]$ (that is, we restrict ourselves to a sub-domain) we see that it always decreases and we have dissipative behaviour[4].

$$p_{-1} + q = 1 \implies \dot{q} = -\dot{p_{-1}} = -cp_0 < 0 \tag{2.73}$$

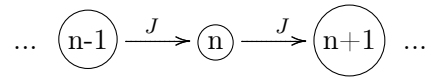
   The probability to be in the domain decreases until $p_0 = 0$.

Steady state for the one-step ME

$$\dot{p}_n = r(n+1)p_{n+1} + g(n-1)p_{n-1} - (r(n) + g(n))p_n \stackrel{!}{=} 0 \tag{2.74}$$

$$\implies J = g(n-1)p_{n-1} - r(n)p_n = g(n)p_n - r(n+1)p_{n+1} \tag{2.75}$$

thus the net current is the same everywhere.



Often one has $J = 0$ (e.g. detailed balance, reflecting boundary, etc). Then we can write the steady states p_n^* as

$$p_n^* = \frac{g(n-1)}{r(n)} p_{n-1}^* = \frac{g(n-1)g(n-2)...g(0)}{r(n)r(n-1)...r(1)} p_0^* \tag{2.76}$$

On the other hand, p_0^* follows from the normalization condition $\sum p_n^* = 1$.

Example 1: Queuing

We consider a queue, for example of customers in the supermarket or of tasks in a to-do-list.
Let $g(n) = \lambda$ be the arrival rate and $r(n) = \mu$ the rate of removal. We need $\lambda < \mu$ to get an steady state

$$p_n^* = \left(\frac{\lambda}{\mu}\right)^n p_0^* \tag{2.77}$$

[4]The probability in the whole system is of course conserved.

From the normalization we get the value of p_0^*

$$1 = \sum_{n=0}^{\infty} p_n^* = p_0^* \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = \frac{p_0^*}{1-\lambda/\mu} \implies p_0^* = 1 - \frac{\lambda}{\mu} \tag{2.78}$$

The probability to get served immediately is larger if few clients arrive (low λ) or if the service is faster (high μ). The average length of the queue is (with $r = \lambda/\mu$):

$$< n >= \sum_{n=0}^{\infty} n p_n^* = (1-r)(r\partial_r) \sum_{n=0}^{\infty} r^n = (1-r)(r\partial_r)\frac{1}{1-r} = \frac{r}{1-r} \tag{2.79}$$

Thus it grows linearly with small r and diverges at $r = 1$.

Example 2: Chemical reaction

Consider the following chemical reaction (molecules):

$$A + 2X \rightleftharpoons 3X$$

This reaction should, in principle, be described with a two-variable ME. However, we only focus and keep track of the X-molecules, so that we can use the one-step ME. This means that we treat A in the macroscopic limit and X in the microscopic limit.
Here the domain is $n \geq 2$, where n is the number of X-molecules, since we need at least two molecules in order for this reaction to take place.
The association and dissociation rates are

$$g(n) = k_1 C_A \frac{n}{\Omega}\frac{(n-1)}{\Omega} \tag{2.80}$$

$$r(n) = k_2 \frac{n}{\Omega}\frac{(n-1)}{\Omega}\frac{(n-2)}{\Omega} \tag{2.81}$$

where Ω is the volume of the system and C_A the concentration of A.
For $J = 0$ (detailed balance assumption), the steady probabilities read

$$p_n^* = \frac{g(n-1)}{r(n)} p_{n-1}^* = \frac{k_1 C_A (n-1)(n-2)\Omega^3}{k_2 n(n-1)(n-2)\Omega^2} p_{n-1}^* = \frac{K}{n} p_{n-1}^* = \frac{2K^{n-2}}{n!} p_2^* \tag{2.82}$$

where we have defined $K \equiv \dfrac{k_1 C_A \Omega}{k_2}$. Again, p_2^* is found through the normalization condition.

$$1 = \sum_{n=2}^{\infty} p_n^* = p_2^* \sum_{n=2}^{\infty} \frac{2K^{n-2}}{n!} = p_2^* \frac{2}{K^2}(e^K - 1 - K) \implies \tag{2.83}$$

$$p_n^* = \frac{K^n}{n!} e^- K \frac{1}{1-(1+K)e^{-K}} \overset{K \gg 1}{\longrightarrow} \frac{K^n}{n!} e^{-K} \tag{2.84}$$

We see that, for large K, it behaves like the Poisson distribution with a mean number of molecules

$$\langle n \rangle = K = \frac{k_1 C_A \Omega}{k_2} = C_x \Omega \tag{2.85}$$

where C_x is the concentration in the steady state ($C_x = \langle n \rangle / \Omega$). From here we can deduce the *law of mass action*:

$$\frac{C_x}{C_A} = \boxed{\frac{{C_x}^3}{C_A {C_x}^2} = \frac{k_1}{k_2}} = K_A \tag{2.86}$$

where K_A is the association constant. Thus in the limit of large systems we get the macroscopic law of mass action. However, the ME also gives us the stationary distribution for smaller systems which do not obey thermodynamics.

Equations for the moments

Also called "macroscopic equations" or "mean field approximation".
For the first moment

$$\begin{aligned}
&\frac{d}{dt}\langle n \rangle = \sum_{n=-\infty}^{\infty} n \dot{p}_n(t) = \sum_n n\big(r(n+1)p_{n+1} + g(n-1)p_{n-1} - (r(n)+g(n))p_n\big) = \\
&= \sum_n \big((n-1)r(n)p_n + (n+1)g(n)p_n - nr(n)p_n - ng(n)p_n\big) = \\
&= \sum_n \big(- r(n)p_n + g(n)p_n\big) = \langle g(n) \rangle - \langle r(n) \rangle
\end{aligned}$$

If $r(n)$ and $g(n)$ are linear, then the average can be drawn into the functions and we get an ODE for $\langle n \rangle$:

$$\boxed{\frac{d}{dt}\langle n \rangle = g(\langle n \rangle) - r(\langle n \rangle)} \tag{2.87}$$

Analogously, the higher moments can be calculated with the differential equation

$$\frac{d}{dt}\langle n^k \rangle = \langle \big((n-1)^k - n^k\big) r(n) \rangle + \langle \big((n+1)^k - n^k\big) g(n) \rangle \tag{2.88}$$

If r and g are linear in n, the RHS is a polynomial in n of order k. Therefore, it can be solved recursively using $\langle n^k \rangle$ and lower moments. If r and g are not linear, the system is not closed.
Consider the one-step ME with linear rates.

(i) First moment:

$$\frac{d}{dt}\langle n \rangle = g(\langle n \rangle) - r(\langle n \rangle) \tag{2.89}$$

(ii) Variance:

$$\frac{d}{dt}(\langle n^2 \rangle - \langle n \rangle^2) \equiv \frac{d}{dt}\langle \langle n^2 \rangle \rangle = \tag{2.90}$$

$$= \langle (1-2n) r(n) \rangle + \langle (2n+1) g(n) \rangle - 2\langle n \rangle (g(\langle n \rangle) - r(\langle n \rangle)) \tag{2.91}$$

For example:

1. Symmetric RW on a line: $r = g = 1$

$$\frac{d}{dt}\langle n\rangle = -1 + 1 = 0 \implies \langle n\rangle = 0 \tag{2.92}$$

$$\frac{d}{dt}\langle\langle n^2\rangle\rangle = 2 \implies \langle\langle n^2\rangle\rangle = 2t \tag{2.93}$$

2. Radioactive decay ("linear death process"):

   Suppose n_0 nuclei that can decay. Let $r(n) = \gamma n$ and $g(n) = 0$ be the death and birth rates. Then,

$$\frac{d}{dt}\langle n\rangle = -\gamma\langle n\rangle \implies \langle n\rangle = n_0 e^{-\gamma t} \tag{2.94}$$

   Recall that n takes discrete values and that a certain realization will not be continuous function of time. However, the average is. For the variance

$$\frac{d}{dt}\langle\langle n^2\rangle\rangle = -2\gamma\langle n^2\rangle + \gamma\langle n\rangle + 2\gamma\langle n\rangle^2 \implies \tag{2.95}$$

$$\langle\langle n^2\rangle\rangle = n_0(e^{-\gamma t} - e^{-2\gamma t}) \tag{2.96}$$

   The variance vanishes at small and large times and has a peak at intermediate times.

Exact solutions using generating functions

The generating function $G(z,t)$ is defined as

$$\boxed{G(z,t) \equiv \sum_{n=-\infty}^{\infty} z^n p_n(t)} \implies p_n(t) = \frac{1}{n!}\frac{\partial^n G(z,t)}{\partial z^n}\Big|_{z=0} \tag{2.97}$$

The discrete index n has been replaced by the continuous variable z. The normalization condition leads to $G(z=1,t) = 1$.
The moments can be expressed as

$$\langle n^k\rangle = (z\partial_z)^k G(z,t)\big|_{z=1} \tag{2.98}$$

We want to obtain a differential equation for $G(z,t)$. Multiplying the ME times z^n and summing over n we get the following PDE for $G(z,t)$:

$$\sum_{n=-\infty}^{\infty} z^n\Big(\dot{p}_n = r(n+1)p_{n+1} + g(n-1)p_{n-1} - (r(n)+g(n))p_n\Big) \implies \tag{2.99}$$

$$\frac{\partial G(z,t)}{\partial t} = \sum_{n=-\infty}^{\infty}\big(z^{n-1}r(n)p_n + z^{n+1}g(n)p_n - z^n(r(n)+g(n))p_n\big) \implies \tag{2.100}$$

$$\boxed{\frac{\partial G(z,t)}{\partial t} = \Big(\big(\frac{1}{z}-1\big)r(z\partial_z) + (z-1)g(z\partial_z)\Big)G(z,t)} \tag{2.101}$$

This equation can be solved for many important cases[5].

Example 1: Poisson process

Defined by $g(n) = \lambda$, $r(n) = 0$ and $n \geq 0$. Since $r(0) = 0$ we can consider the unbounded state space and the generating function

$$\partial_t G = \lambda(z-1)G \tag{2.102}$$

With the initial condition $p_n(t=0) = \delta_{nm}$, $m \geq 0$

$$G(z, t=0) = z^m \implies G(z,t) = e^{\lambda(z-1)t} z^m = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} z^{m+n} = \tag{2.103}$$

$$= e^{-\lambda t} \sum_{i=-m}^{\infty} \frac{(\lambda t)^{i-m}}{(i-m)!} z^i \implies \begin{cases} \boxed{p_n(t) = \dfrac{(\lambda t)^{n-m}}{(n-m)!} e^{-\lambda t}} & n \geq 0 \\ \boxed{p_n(t) = 0} & n < m \end{cases} \tag{2.104}$$

Again, with $G(z,t)$ we can calculate the moments. The first moment and variance are

$$\langle n \rangle = z\partial_z \big|_{z=1} = \lambda t + m \tag{2.105}$$

$$\langle n^2 \rangle = (z\partial_z)^2 G \big|_{z=1} = ... = (\lambda t + m)^2 + \lambda t \implies \langle\langle n^2 \rangle\rangle = \lambda t \tag{2.106}$$

Example 2: Radioactive decay

As seen above, the radioactive decay is defined through $r(n) = \gamma n$ and $g = 0$. Thus

$$\partial_t G = \gamma\big(\frac{1}{z} - 1\big) z \partial_z G \tag{2.107}$$

The initial condition $p_n(t=0) = \delta_{nn_0}$ (i.e. $G(z,0) = z^{n_0}$) gives

$$G(z,t) = \big(1 + (z-1)e^{-\gamma t}\big)^{n_0} = (1 - e^{-\gamma t})^{n_0} \Big(1 + \frac{z e^{-\gamma t}}{1 - e^{-\gamma t}}\Big)^{n_0} = \tag{2.108}$$

$$(1 - e^{-\gamma t})^{n_0} \sum_{n=0}^{n_0} \binom{n_0}{n} \Big(\frac{e^{-\gamma t}}{1 - e^{-\gamma t}}\Big)^n z^n = \sum_{n=0}^{n_0} \binom{n_0}{n} e^{-\gamma n t} (1 - e^{-\gamma t})^{n_0 - n} z^n \tag{2.109}$$

$$\implies \boxed{p_n(t) = \binom{n_0}{n} e^{-\gamma n t} (1 - e^{-\gamma t})^{n_0 - n}} \text{ for } 0 \leq n \leq n_0 \tag{2.110}$$

For example, for $n_0 = 2$ we have

[5] See book *Stochastic Models in Biology* by Narendra S. Goel and Nira Richter-Dyn, Springer 1974.

$$p_0 = (1 - e^{-\gamma t})^2 \tag{2.111}$$
$$p_1 = 2e^{-\gamma t}(1 - e^{-\gamma t}) \tag{2.112}$$
$$p_2 = e^{-2\gamma t} \tag{2.113}$$

and $p_0 + p_1 + p_2 = 1$.

Example 3: Symmetric RW on a line

$$r = g = 1 \implies \partial_t G = \big(\frac{1}{z} + z - 2\big)G \implies G = e^{(z+\frac{1}{z}-2)t} \tag{2.114}$$

for the initial condition $G(z, t = 0) = 1$, which corresponds to $p_n(t = 0) = \delta_{n0}$. Normalization is satisfied since $G(z = 1, t) = 1$. G can be rewritten in terms of the modified Bessel functions I_n as

$$G(z, t) = e^{-2t} \sum_{n=-\infty}^{\infty} I_n(2t) z^n \tag{2.115}$$

which directly allows us to identify

$$\boxed{p_n(t) = e^{-2t} I_n(2t)} \tag{2.116}$$

This is the first time we are able to get $p(t)$ for the RW. For large times (stationary limit) we have that p_n goes to zero as

$$p_n(t) \to \frac{1}{\sqrt{4\pi t}} \tag{2.117}$$

Therefore, there is no stationary limit (corresponds to the Wiener process!).
The generating function contains all the information of our system and from it we can get all relevant quantities. The mean and variance are

$$\langle n \rangle = (z\partial_z) G(z, t)\big|_{z=1} = z e^{z+\frac{1}{z}-2}(1 - \frac{1}{z^2})t\big|_{z=1} = 0 \tag{2.118}$$
$$\langle n^2 \rangle = (z\partial_z)^2 G(z, t)\big|_{z=1} = z e^{z+\frac{1}{z}-2}(1 + \frac{1}{z^2})t\big|_{z=1} = 2t \tag{2.119}$$

which agree with the results obtained with the macroscopic equations.

Example 4: Linear birth-death process

For the radioactive decay there is only a decay rate. Now we add a birth rate:

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B$$

We define n to be the number of A molecules ($0 \leq n \leq N$). The birth and death rates are then $r = k_1 n$ and $g = k_2(N - n)$.

We want to use dimensionless time $k_1 t$. Dividing the one-step ME by k_1 we see that the new birth and death rates are $r = n$ and $g = \gamma(N - n)$, with $\gamma = k_2/k_1$. Then

$$\partial_t G = \Big(\big(\frac{1}{z} - 1\big)(z\partial_z) + (z-1)\gamma(N - z\partial_z) \Big) G \tag{2.120}$$

With the initial condition $p_n(t=0) = \delta_{nN}$ ($G(z, t=0) = z^N$) one gets

$$G(z,t) = \Big(\frac{(z-1)e^{-(1+\gamma)t} + (1+\gamma z)}{1+\gamma} \Big)^N \tag{2.121}$$

from which the following probability distributions are obtained

$$\boxed{p_n(t) = \binom{N}{n} \frac{(\gamma + e^{-(1+\gamma)t})^n (1 - e^{-(1+\gamma)t})^{N-n}}{(1+\gamma)^N}} \tag{2.122}$$

In the stationary limit

$$p_n^* = \binom{N}{n} \frac{\gamma^n}{(1+\gamma)^N} = \binom{N}{n} \underbrace{\Big(\frac{\gamma}{1+\gamma}\Big)^n}_{p} \underbrace{\Big(\frac{\gamma}{1+\gamma}\Big)^{N-n}}_{q} \tag{2.123}$$

which is the binomial distribution with probabilities p and q.
The moments are

$$\langle n \rangle = (z\partial_z) G\big|_{z=1} = \underbrace{\frac{\gamma}{1+\gamma} N}_{\text{stationary limit}} + \underbrace{\frac{N}{1+\gamma} e^{-(1+\gamma)t}}_{\text{exponential relaxation}} \tag{2.124}$$

$$\sigma^2 = \langle\langle n^2 \rangle\rangle = \frac{\langle n \rangle (1 - e^{-(1+\gamma)t})}{1+\gamma} \tag{2.125}$$

The variance vanishes for $t = 0$, $\gamma = 0$ or $\gamma \to \infty$. The larger the system is (N, $\langle n \rangle$), the smaller relative fluctuations $\frac{\sigma}{\langle n \rangle} \sim \langle n \rangle^{-1/2}$.

Differential CKE

In summary, we have shown that the CKE gives rise to the FPE (diffusion and drift) or the ME (jumps). However, both processes can also be combined into one equation, the *differential CKE*:

$$\boxed{\partial_t p(x,t) = (-\partial_x A + \partial_x^2 D) p(x,t) + \int dx' \big(W(x,x',t) p(x',t) - W(x',x,t) p(x,t) \big)} \tag{2.126}$$

2.7 Summary

- Joint probability $p_n(x_1, t_1; ...; x_n, t_n)$ of getting the trajectory $\{x_1, t_1; ...; x_n, t_n\}$.

  Marginal distribution:

$$p_x(x) \equiv \int p_2(x, y)dy \tag{2.127}$$

  Conditional distribution:

$$p_2(x|y) \equiv \frac{p_2(x, y)}{p_y(y)} \tag{2.128}$$

- Gaussian white noise: Gaussian-distributed random variable $\eta(t)$ defined through

$$\langle \eta(t) \rangle = 0 \tag{2.129}$$
$$\langle \eta(t)\eta(t') \rangle = \sigma^2 \delta(t - t') \tag{2.130}$$

- Markov process: no-memory process defined by

$$p_n(n|(n-1); ...; 1) = p_2(n|(n-1)) \tag{2.131}$$

  The *Wiener process* (W_t) is the most relevant Markov process

$$p_2(w, t|w', t') = \frac{1}{\sqrt{2\pi\tau}} e^{-(w-w')^2/2(t'-t)} \tag{2.132}$$
$$p_1(w, t) = \frac{1}{\sqrt{2\pi t}} e^{-w^2/2t} \tag{2.133}$$

- Chapman-Kolmogorov equation (CKE): Integral equation for Markov processes

$$p_2(3|1) = \int dx_2 p_2(3|2)p_2(2|1) \tag{2.134}$$

  From this equation and depending whether we have diffusion and drift or jumps one gets the Fokker-Planck equation or the master equation respectively.

- Fokker-Planck equation:

$$\frac{\partial p_2(x, t|x', t')}{\partial t} = \big(-\frac{\partial}{\partial x} A(x, t) + \frac{\partial^2}{\partial x^2} D(x, t)\big) p_2(x, t|x', t') \tag{2.135}$$

  where the the "drift coefficient" A and the "diffusion coefficient" D have been introduced through

$$\int dx_2(x_2 - x_1)p_2(x_2, t + \Delta t|x_1, t) = A(x_1, t)\Delta t + O(\Delta t^2) \tag{2.136}$$

$$\int dx_2(x_2 - x_1)^2 p_2(x_2, t + \Delta t|x_1, t) = 2D(x_1, t)\Delta t + O(\Delta t^2) \tag{2.137}$$

One can as well replace the two-point probability distribution in the FPE by $p(x, t)$.

The FPE can be written as a continuity equation:

$$\dot{p}_1 + \partial_x J = 0 \tag{2.138}$$

$$J \equiv (A - \partial_x D)p_1 \tag{2.139}$$

with J being the *probability current*

- Master equation:

$$\frac{\partial}{\partial t}p_2(x, t|x', t') = \int dx'' W(x, x'', t)p_2(x'', t|x', t') - \int dx'' W(x'', x, t)p_2(x, t|x', t') \tag{2.140}$$

where the probability to jump from x'' to x, $W(x'', x, t)$ is introduced via

$$p_2(x, t+\Delta t|x'', t) = \big(1-a(x, t)\Delta t\big)\delta(x-x'') + W(x, x'', t)\Delta t + O(\Delta t^2) \tag{2.141}$$

with $a(x'', t) = \int dx W(x, x'', t)$.

Since we are dealing with a jump process, it is more natural to write the ME in a discrete version:

$$\dot{p}_n(t) = \sum_{n'} \big(W_{nn'}p_{n'}(t) - W_{n'n}p_n(t)\big) \tag{2.142}$$

where $p_n(t)$ is the probability to be at the site n at a given time and $W_{nn'}(t)$ the jump rate from the site n' to n.

Chapter 3

Langevin equation and stochastic calculus

3.1 Brownian motion

In this chapter, instead of dealing with the probability distributions $p(x,t)$ of a random variable $X(t) = X_t$, we will start directly from the physical equation of motion and add a stochastic force to it. This was done by Paul Langevin (1872-1946) in 1906 for the first time for a Brownian particle:

$$\boxed{m\dot{v}(t) = -\xi v(t) + \underbrace{\sigma\eta(t)}_{random force}} \tag{3.1}$$

where $v = \dot{x}$ is velocity of the particle, the ξ is the friction coefficient, σ is the noise strength and $\eta(t)$ is Gaussian white noise motivated by the central limit theorem:

$$\langle \eta(t) \rangle = 0 \tag{3.2}$$
$$\langle \eta(t)\eta(t') \rangle = 2\delta(t-t') \tag{3.3}$$

Note the factor of two in the correlation function. There are different conventions in the literature.
Since η is a random variable, so is $v(t)$. One can write the following formal solution of Langevin's equation:

$$v(t) = v_0 e^{-\frac{\xi}{m}t} + e^{-\frac{\xi}{m}t} \int_0^t ds e^{\frac{\xi}{m}s} \frac{\sigma}{m} \eta(s) \tag{3.4}$$

which can be verified by substitution. It is important to note that we only used the property of linearity of the integral and we have not made any assumption of differentiability of η. Moreover, we have not dicussed yet how an integral over a random variable has to be defined.

Mean and correlation function

By drawing averages into the integral, we can calculate moments. For the average we get:

$$\boxed{\langle v(t)\rangle = v_0 e^{-\frac{\xi}{m}t}} \tag{3.5}$$

thus the average decays exponentially like in the deterministic case.
For the two-point correlation we get:

$$\langle v(t)v(t')\rangle = v_0^2 e^{-\frac{\xi}{m}(t+t')} + \frac{\sigma^2}{m^2} e^{-\frac{\xi}{m}(t+t')} \int_0^t ds \int_0^{t'} ds' e^{\frac{\xi}{m}(s+s')} 2\delta(s-s') = \tag{3.6}$$

$$= v_0^2 e^{-\frac{\xi}{m}(t+t')} + \frac{\sigma^2}{m^2} e^{-\frac{\xi}{m}(t+t')} \int_0^t ds 2e^{\frac{\xi}{m}2t} = v_0^2 e^{-\frac{\xi}{m}(t+t')} + \frac{\sigma^2}{m^2} e^{-\frac{\xi}{m}(t+t')} \frac{m}{\xi}(e^{\frac{\xi}{m}2t} - 1) \tag{3.7}$$

$$= e^{-\frac{\xi}{m}(t+t')}\Big(v_0^2 - \frac{\sigma^2}{m\xi}\Big) + \frac{\sigma^2}{m\xi} e^{-\frac{\xi}{m}(t'-t)} \tag{3.8}$$

Here we assumed $t < t'$ for the time order (in the opposite case, we had to invert the two; the general case could be formulated by using Heaviside step functions). The coefficient m/ξ is the relaxation time. Under the assumption $t, t' \gg m/\xi$, which is true in most cases ($m/\xi \sim 10^{-8}s$ for Brownian particles), the first term vanishes and we get

$$\boxed{\langle v(t)v(t')\rangle = \frac{\sigma^2}{m\xi} e^{-\frac{\xi}{m}(t'-t)}} \tag{3.9}$$

For large time differences, $t' \gg t$, it goes to zero, whereas for equal times it has a finite value

$$\langle v(t)^2\rangle = \frac{\sigma^2}{m\xi} \ . \tag{3.10}$$

Thus the random kicks from the environment keep the particle in motion.
In thermodynamics, the equipartition theorem tells us that in one dimension and in equilibrium

$$\frac{1}{2}m\langle v^2\rangle = \frac{1}{2}k_B T \implies \boxed{k_B T = \frac{\sigma^2}{\xi}} \tag{3.11}$$

which agrees with the Einstein relation previously found if we identify $\sigma^2 = D\xi^2$.

Mean squared displacement

$$\langle x(t)^2\rangle = \langle \int_0^t v(s)ds \int_0^t v(s')ds'\rangle \tag{3.12}$$

$$= \frac{\sigma^2}{m\xi} \int_0^t ds \Big(\int_0^s ds' e^{-\frac{\xi}{m}(s-s')} + \int_s^t ds' e^{-\frac{\xi}{m}(s'-s)} \Big) \tag{3.13}$$

$$= \frac{\sigma^2}{\xi^2} \int_0^t ds (2 - e^{-s\xi/m} - e^{-(t-s)\xi/m}) \tag{3.14}$$

$$= \frac{\sigma^2}{\xi^2} 2t + \frac{2\sigma^2 m}{\xi^3}(e^{-t\xi/m} - 1) \tag{3.15}$$

We can study two limiting cases.

(i) Short-time behaviour ($t \ll m/\xi$):

The first order terms cancel and one gets ballistic behaviour

$$\langle x(t)^2 \rangle \to \frac{\sigma^2}{m\xi} t^2 \tag{3.16}$$

(ii) Long-time behaviour ($t \gg m/\xi$):

Now one gets diffusive behaviour

$$\langle x(t)^2 \rangle \to \frac{2\sigma^2}{\xi^2} t \tag{3.17}$$

Since the diffusion constant is defined by $\langle x(t)^2 \rangle = 2Dt$, we get the Einstein relation

$$\boxed{D = \frac{\sigma^2}{\xi^2} = \frac{k_B T}{\xi}} \tag{3.18}$$

This relation tells us that the mobility of the particle and its diffusion constant are directly related through thermal energy (fluctuation-dissipation theorem).

Langevin's original derivation

Langevin's original derivation was even shorter.

$$m\ddot{x} = -\xi \dot{x} + F \tag{3.19}$$

where F is a fluctuating force satisfying $\langle F \rangle = 0$. Multiplying the above equation by x, he got

$$\frac{m}{2}\frac{d^2}{dt^2}x^2 - m\dot{x}^2 = -\xi \frac{1}{2}\frac{d}{dt}x^2 + Fx \tag{3.20}$$

Taking the expectation value and assuming no correlation between F and x gets rid of the random force. The velocity term can be dealt with by the equipartition theorem. We then have

$$\frac{m}{2}\frac{d^2}{dt^2}\langle x^2 \rangle + \frac{\xi}{2}\frac{d}{dt}\langle x^2 \rangle = k_B T \tag{3.21}$$

Again we can distinguish between two regimes. In the overdamped limit $t \gg m/\xi$ we can also get rid of the term containing the mass. We then have

$$\langle x^2 \rangle = 2\frac{k_B T}{\xi} t \implies \boxed{D = \frac{k_B T}{\xi}} \tag{3.22}$$

Comments

Since $\eta(t)$ is Gaussian white noise, only $\langle\eta(t)\rangle$ and $\langle\eta(t)\eta(t')\rangle$ are relevant. All higher cumulants of $\eta(t)$ and therefore of $v(t)$ vanish.
The velocity $v(t)$ is a random variable with mean and variance

$$\langle v(t)\rangle = v_0 e^{-t\xi/m} \tag{3.23}$$

$$\langle v(t)v(t')\rangle = \frac{\sigma^2}{m\xi} e^{-(t'-t)\xi/m} \tag{3.24}$$

with $t' > t$. This actually defines the Ornstein-Uhlenbeck process with steady state given by

$$p_s(v) = \left(\frac{\xi m}{2\pi\sigma^2}\right)^{1/2} e^{-v^2/(2\sigma^2/\xi m)} \tag{3.25}$$

In the over-damped limit, $m\ddot{x} \ll \xi\dot{x}$ or $t \gg \dfrac{m}{\xi}$, the Langevin equation reduces to

$$\boxed{\dot{x}(t) = \frac{\sigma}{\xi}\eta(t)} \implies x(t) = x_0 + \frac{\sigma}{\xi}\int_0^t \underbrace{\eta(t')dt'}_{\equiv dW_{t'}} \tag{3.26}$$

$W_t = \int_0^t dW_{t'}$ is the Wiener process with

$$\langle W_t\rangle = 0 \tag{3.27}$$

$$\langle W_t W_s\rangle = \int_0^t dt' \int_0^s ds' \langle\eta(t')\eta(s')\rangle = \int_0^t dt' \int_0^s dt 2\delta(t'-s') = \tag{3.28}$$

$$\theta(t-s)2\int_0^s dt' + \theta(s-t)2\int_0^t dt' = \theta(t-s)2s + \theta(s-t)2t = 2\min(t,s) \tag{3.29}$$

$$\implies \langle W_t^2\rangle = 2dt \implies dW_t \sim \sqrt{dt} \tag{3.30}$$

which is non-polynomial. The Wiener process is not differentiable.

3.2 Stochastic differential equations (SDEs)

Langevin's equation for the Brownian particle is an example of a *stochastic differential equation* (SDE). Due to the random force, each solution of such an equation is a different random trajectory. In physics, SDEs are also called *Langevin equations* (LEs).
Note that any one-dimensional n-order differential equation can be written as a n-dimensional system of first order differential equations. For example, Newton's equation with inertia can be written as two overdamped equations. For simplicity, here we restrict ourselves to the one-dimensional case. Then the general form of a SDE is:

$$\boxed{\dot{X}_t = \underbrace{a(X_t,t)}_{deterministic} + \underbrace{b(X_t,t)\eta(t)}_{stochastic}} \tag{3.31}$$

Without the noise (stochastic) term this would be an ODE. The noise amplitude $b(X_t, t)$ and the white noise $\eta(t)$ make it a SDE. It can also be written in the form

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t \tag{3.32}$$

We have to differ between two fundamentally different cases:

(i) *Additive noise*: $b = const = \sigma$

(ii) *Multiplicative noise*: $b = b(X_t)$

The t-dependence in b only appears when there are external fields in the system. We will not take them into account here.
How does a function $f(X_t)$ change with X_t in the case of additive noise ? Keeping in mind that $dX_t \sim \int \eta_t dt \sim \sqrt{dt}$, we write

$$\begin{aligned} df(X_t) &= f(X_t + dX_t) - f(X_t) && (3.33)\\ &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2 + O(dt^{3/2}) && (3.34)\\ &= f'(X_t)(adt + \sigma dW_t) + \frac{1}{2}f''(X_t)\sigma^2 dW_t^2 + O(dt^{3/2}) && (3.35)\\ &= \big(af'(X_t) + f''(X_t)\sigma^2\big)dt + f'(X_t)\sigma dW_t + O(dt^{3/2}) && (3.36) \end{aligned}$$

which is known as *Ito's formula*[1]. One has to take into account that the last equality is only valid in the mean-squared sense since we used $dW_t^2 = 2dt$. If we take $f(X_t) = X_t$, then $dX_t = adt + \sigma dW_t$, so we recover the equation from above.
For multiplicative noise, this (naively) suggests that

$$dX_t = a(X_t, t)dt + b(X_t)dW_t \ . \tag{3.37}$$

For small time steps, this suggests that we solve the equation step by step by treating $b(X_t)$ constant at every step and evaluate it at X_t. We now will see that this is not necessarily correct because stochastic integrals are more complicated than we are used to from classical analysis.

Stochastic integrals

Can an integral like $\int_0^t b(X_{t'})dW_{t'}$ be evaluated like a classical Riemann integral? We compare the result of the integral $\int_0^t W_{t'}dW_{t'}$ solved using Riemann integration with the expectation value of a discrete approximation. In analogy to how the Riemann integral is derived, one does a partition of time. We divide the interval of integration $[0, t]$ into n equidistant steps such that $0 = t_0 < t_1 < ... < t_n = t$. Then we define the following partial sum:

$$S_n \equiv \sum_{i=1}^{n} W_{\tau_i}(W_{t_i} - W_{t_{i-1}}) \tag{3.38}$$

[1] Itô Kiyoshi (1915-2008), Japanese mathematician.

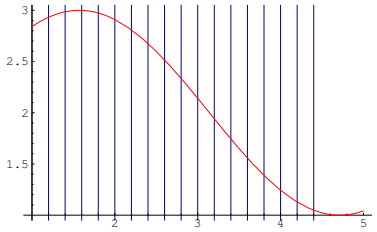


Figure 3.1: Discretization of time t_i.

where $\tau_i \in [t_{i-1}, t_i]$. We can parametrize τ_i as

$$\tau_i = \alpha t_i + (1-\alpha) t_{i-1} \tag{3.39}$$

with $0 \leq \alpha \leq 1$. Since S_n would take different values every time we evaluate it, one considers its expectation value

$$\langle S_n \rangle = \sum_{i=1}^{n} \langle W_{\tau_i}(W_{t_i} - W_{t_{i-1}}) \rangle = 2 \sum_{i=1}^{n} \big(\min(\tau_i, t_i) - \min(\tau_i, t_{i-1}) \big) \tag{3.40}$$

$$= 2 \sum_{i=1}^{n} (\tau_i - t_{i-1}) = 2\alpha \sum_{i=1}^{n} (t_i - t_{i-1}) = 2\alpha t = t + 2\big(\alpha - \frac{1}{2}\big) t \tag{3.41}$$

We see that the expectation value (average) of the partial sum depends on α even in the limit $n \to \infty$. This is a fundamental difference to Riemann integration, which gives a result independent of α:

$$\int_0^t W_{t'} dW_{t'} = \frac{1}{2} W_t^2 \overset{ms}{=} t \tag{3.42}$$

where *ms* stands for *in the mean squared sense*.
Compared with the result from the averaged partial sum, the Riemann result corresponds to taking $\alpha = \frac{1}{2}$. However, our first naive expectation corresponded to $\alpha = 0$. What is correct ? The answer is: both are correct, this is an additional degree of freedom we have with stochastic integrals. With other words, the definition is not complete without specifying a value for α[2].
In general, there is an infinite number of possible stochastic integrals. However, the choices $\alpha = 0$ and $\alpha = \frac{1}{2}$ seem to be special. We therefore differ between two different *interpretations*:

1. *Itô interpretation*: $\alpha = 0$

   Preferred by mathematicians. There is no anticipation of the future and the system jumps from one time step to the next. From the physics point of view, this is a good assumption e.g. for population modelling.

[2]The situation is similar in lattice gas models, where we also have to specify the update procedure (sequential, parallel, random) to define the complete model.

Within this interpretation, stochastic integrals are defined as[3]

$$\boxed{\int_0^t b(X_{t'},t')dW_{t'} \equiv \lim_{n\to\infty}\sum_{i=1}^{n} b(X_{t_{i-1}},t_{i-1})(W_{t_i}-W_{t_{i-1}})} \tag{3.43}$$

The SDE is written as

$$\dot{X}_t = a(X_t,t) + b(X_t,t)\eta_t \tag{3.44}$$

2. *Stratonovich interpretation*: $\alpha = \frac{1}{2}$

   Preferred by physicists. Agrees with normal calculus and corresponds to the case of coloured noise in the limit of vanishing correlation time[4] (Wong-Zakai theorem). Therefore it is a good assumption for continuous time processes like random walks.

   According to this interpretation, the stochastic integral is defined as

$$\boxed{\int_0^t b(X_{t'},t')dW_{t'} \equiv \lim_{n\to\infty}\sum_{i=1}^{n} b\Big(\frac{X_{t_i}-X_{t_{i-1}}}{2},t_{i-1}\Big)(W_{t_i}-W_{t_{i-1}})} \tag{3.45}$$

   The SDE is then written with the following notation:

$$\dot{X}_t = a(X_t,t) + b(X_t,t)\circ\eta_t \tag{3.46}$$

   where $\circ$ is a newly introduced symbol.

In computer programs , we cannot evaluate b at $\dfrac{X_{t+dt}-X_t}{2}$ since it involves future times. We need to do a Taylor expansion:

$$dX_t = adt + b(X_t+\alpha dX_t)dW_t = adt + b(X_t)dW_t + b'(X_t)\alpha dX_t dW_t + O(dt^{3/2}) \tag{3.47}$$

$$\stackrel{ms}{=} \underbrace{(a+2\alpha b'b)dt}_{drift} + \underbrace{bdW_t}_{diffusion} + O(dt^{3/2}) \tag{3.48}$$

Note that for $\alpha \neq 0$ we get an additional drift term.
One can deal with the Stratonovich interpretation in the *Itô version* (relevant for computing programs) by solving

$$\dot{X}_t = \underbrace{(a+b'b)}_{\equiv A} + b\eta_t \tag{3.49}$$

Here the general drift $A = a + b'b$ has been defined ($A = a$ for additive noise and $A = a + b'b$ for multiplicative noise). Our standard choice will be Stratonovich interpretation in the Itô version.

[3]Strictly speaking the limit should be a mean square limit ms-$\lim_{n\to\infty} S_n$. It is defined as the following limit in the mean. Let $S_n(t)$ be a sequence of random variables. One says ms-$\lim_{n\to\infty} S_n = S \iff \lim_{n\to\infty}\langle (S_n - S)^2\rangle = 0$ (convergence in mean square $\implies$ convergence in mean)

[4]A coloured noise $\xi(t)$ is defined by having $\langle \xi(t)\xi(t')\rangle = f(t-t')$, where f is a fast decaying function of $t-t'$ (not $\delta(t-t')$). So it has finite but short correlation time.

Equivalence to the FPE

We are interested in the relation between the FPE and SDE frameworks. Given a SDE for X_t, what equation governs the corresponding probability density function $p(x,t)$ associated with X_t? dX_t is a normally distributed random variable with mean Adt and second moment $2b^2$. Therefore the diffusion constant $D = b^2$ and the FPE reads

$$\boxed{\dot{p}_2 = (-\partial_x A + \partial_x^2 b^2) p_2} \tag{3.50}$$

where $p_2 = p_2(x,t|x_0,t_0)$. In the SI $A = a + b'b$:

$$\dot{p}_2 = (-\partial_x a - \partial_x b'b + \partial_x^2 b^2) p_2 \implies \boxed{\dot{p}_2 = (-\partial_x a + \partial_x b \partial_x b) p_2} \tag{3.51}$$

Note that the derivatives act to the complete right. Therefore, we see that with FPE as naively written down corresponds to the Itô interpretation, while in practise we usually are interested in the Stratonovich interpretation. Therefore it is essential to add the additional drift term.

3.3 Physical example for multiplicative noise

Consider a colloidal particle (e.g. polystyrene bead with a typical radius of $1\mu m$) in a container with water. One can use a laser trap to force the particle to move only in one dimension (vertical z-direction). Due to the damping by the water, this is an overdamped system (no inertia, low Reynolds-number hydrodynamics). Force and velocity are therefore connected by a linear relation, $v = MF$, with mobility M.
What are the forces acting in the z-direction?

(i) Gravity pulls the particle down

(ii) Electrostatic repulsion pushes it up

This leads to a potential with a minimum at a certain distance z_0 to the bottom of the container (where $z = 0$):

$$V(z) = \underbrace{A_0 e^{-kz}}_{electrostat.rep.} + \underbrace{B_0 z}_{gravitation} \tag{3.52}$$

What is then the mobility M of the particle?
There are two cases to distinguish:

1. Far from the wall:

   In that case we can solve the hydrodynamic equations in a infinitely extended space.

   $$\implies \boxed{M_0 = \frac{1}{6\pi\eta R}} \tag{3.53}$$

   where η is the viscosity and R the radius of the particle. This is the Stokes-mobility.

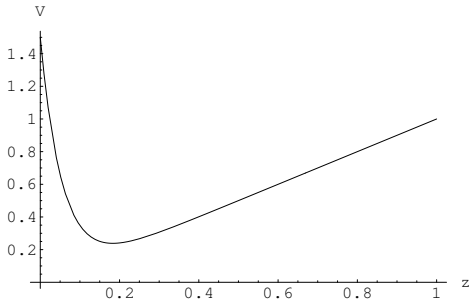


Figure 3.2: Shape of the potential created by the electrostatic repulsion and the gravitation (arbitrary units and parameters).

The mobility is related to the friction coefficient ξ_0 as $M_0 = \frac{1}{\xi_0}$. Using the Einstein relation we get the *Stokes-Einstein relation*:

$$\boxed{D_0 = \frac{k_B T}{\xi_0} = \frac{k_B T}{6\pi\eta R}} \tag{3.54}$$

2. Close to the wall:

   In that case the flow fields are very different from the previous case and one has to take the no-slip boundary condition at the wall into account. The hydrodynamics calculation gives

$$\boxed{D(z) = D_0\Big(1 + \frac{R}{(z-R)}\Big)^{-1}} \tag{3.55}$$

   Thus we have multiplicative noise since our diffusion coefficient is space-dependent.

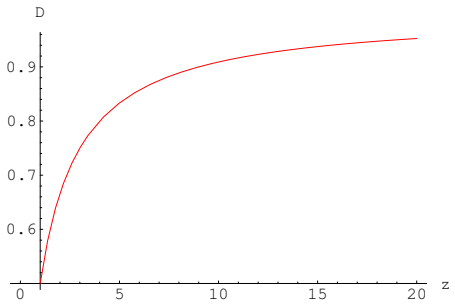


Figure 3.3: Behaviour of the diffusion coefficient as a function of z. It vanishes at $z = R$ and then increases with distance from the wall towards the free case far away from the wall.

Chapter 4

Black-Scholes theory for finance

One of the most important applications of the theory of stochastic processes is the field of economics and finances. We start with a short historical review [1]:

- Although the theory of random walks applied to physics is often credited to the 1905 paper by Einstein, Louis Bachelier in 1900 already developed such a theory for the stock market in his PhD-thesis *Theorie de la speculation* (officially supervised by Henri Poincare). However, his results very forgotten and rediscovered only by Ito in 1944 for mathematics and in the 1950s in economics, mainly through Paul Samuelson.

- MIT professor Paul Samuelson also used the analogy between physics and finance and introduced random walk concepts into economics in the 1940s. He was awarded the Nobel Prize in economics in 1970.

- In 1962 MIT professor Ed Thorp published his book *Beat the Dealer: A Winning Strategy for the Game of Twenty-One* and in 1969 he started the first hedge fund, Princeton/Newport partner based on statistical methods. This hedge fund was extremely successful and scored double-digit increases even if the stock market was going down. He even managed to get through the 1987 crisis without much damage, but only because Thorp himself managed the funds by hand on Black Monday. Hedge funds manage risks in their portfolios by distributing it over diverse assets through mathematical and highly secret algorithms. The biggest US hedge fund is Bridgewater Associates with more than 1.000 employees and more than 100 billion USD. Together the hedge fund industry in the US handles around 2 trillion USD.

[1] For a nice introduction into the subject, compare the popular book *The Quants: the maths geniuses who brought down Wall Street* by Scott Patterson, Random House Business Books 2011, which describes the history leading up to the Lehman crisis in 2007/08. There are now also several movies dealing with the crash from 2008, namely *Margin Call* from 2011 starring Kevin Spacey and Demi Moore and *The big short* from 2015 starring Christian Bale and Brad Pitt. The 2008 movie *21*, based on the book *Bringing down the house*, also stars Kevin Spacey and deals with the MIT team playing backjack at Las Vegas based on statistical methods, similar to the story of Ed Thorp. Here we closely follow chapter 5, Modeling the Financial Market, from the book by Wolfgang Paul and Joerg Baschnagel, *Stochastic Processes: From Physics to Finance*, 2nd edition 2013, who in turn recommend Wilmott, Deynne and Howison, *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press 1993.

- In 1970 Chicago professor Eugene Fama like Paul Samuelson before also looks at stocks as random walks and formulates the efficient market hypothesis: all information is immediately absorbed by the market and thus stocks cannot be predicted in a deterministic manner, because nobody has privileged information that can be turned into a profit. This means that the stock market should behave like a Markov process.

- In 1973 Fischer Black and Myron Scholes published their famous work on pricing stock options based on the concept of a geometrical random walk. Independently Robert Merton (son of Harvard historian of science Robert Merton) came to the same conclusions. Scholes and Merton earned the Nobel Prize in economics in 1997, after Black had died in 1995 from cancer (from 1984 until his death, he worked for Goldman Sachs managing quantitative trading strategies). Interestingly, Ed Thorp used an unpublished formula for his hedge fund that is very similar to the one by Black and Scholes.

- In the 1980s the cold war ended and more and more quants moved from the military into the financial industries. This initiated a development of computerized trading that was blamed as one of the reasons for the 1987 crash. Hedge funds also played a central role for the 2008 crash (AKA the Lehman crisis).

- Already in 1963, Benoit Mandelbrot had discovered that large price changes occur more often than predicted by Gaussian models (*heavy* or *fat tails*). As an alternative, he suggested a Levy distribution. This work too was not really noticed much, but the subject became very important again after the stock market crashes in 1987 and 2008. Today many financial models go beyond Gaussian random walks. For example, in 1995 Mantegra and Stanley analyzed stock market data and suggested a truncated Levy distribution with exponent $\alpha = 1.4$ working up to 6σ.

Here we give an introduction into Black-Scholes theory, which is the standard way to price options and the starting point for more advanced schemes. Consider an asset like cotton or a foreign currency that is traded on the stock market. At any time t, it has a spot price $S(t)$. In an efficient market, this should be a Markov process, because nobody can predict the future in a deterministic manner. Bachelier suggested a Wiener process for $S(t)$, but a more realistic model is the geometrical random walk, as realized by Samuelson, Black, Scholes, Merton and others:

$$dS = \mu S dt + \sigma S dW \tag{4.1}$$

Here μ is the drift or average growth rate, and σ the fluctuation amplitude, called volatility in finance. We note that because also the second term is linear in S, fluctuations grow with the stock and we deal with an Ito-process with multiplicative noise:

$$dS = a(S)dt + b(S)dW \tag{4.2}$$

In contrast to the simple random walk with $b = constant$, the random walk with $b = \sigma S$ is geometrical in nature. In other words, not the absolute change dS is

relevant, but the relative change dS/S, called the return. We thus change variables from S to $f(S) = \ln S$. Ito's formula predicts the change in a function of S:

$$df(S) = (\partial_t f + a\partial_S f + \frac{1}{2}b^2\partial_S^2 f)dt + b\partial_S f dW \tag{4.3}$$

With $f = \ln S$, $a = \mu S$, $b = \sigma S$ we get

$$d\ln S = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW \tag{4.4}$$

Thus not S, but $\ln S$ performs a Wiener process, with drift $\mu - \frac{1}{2}\sigma^2$ and variance σ^2. For S this means that it has a log-normal distribution. The log-normal distribution of $y = e^x$ is defined by

$$p_G(x)dx = p_{LN}(y)dy \tag{4.5}$$

With $y_0 = e^{<x>}$, it reads

$$p_{LN}(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y}\exp -\frac{(\ln(y/y_0))^2}{2\sigma^2} \tag{4.6}$$

Most importantly, y has only positive values. One can show that y_0 is the median and that the distribution peaks at a smaller value of y. In general, log-normal distributions arise if a processes is not the sum, but the product of many independent random processes, as it is typical for growth processes. Stock market prices can usually be fitted well with such a distribution.

How can we use the geometrical random walk to derive a rewarding strategy for the stock market? We now explain this for the European call option. Here the holder obtains an option from the writer to buy an asset (the underlying with spot price $S(t)$) for a prescribed price K (the strike price) at the expiry data T in the future (typically one year). An American option would allow the holder to exercise the option at any time prior to expiry. A related instrument are futures, in which both parties assume an obligation. Options and futures are examples of derivatives. Historically they were first used in a broad manner in the Dutch tulip market in the 17th century which crashed in 1637).

How should the writer set the price C of the call option? The classical solution is Black-Scholes theory. We consider $C = C(S,t)$ and first discuss the boundary conditions. If $S(T) < K$, the option will not be exercised, because the holder can buy at a cheaper price on the market. Thus the price C should be cero. If $S(T) > K$, the option will be exercised and a fair price C would be $S(T) - K$ (note that the procedure is fair and transparent for both parties, the benefit arises because one wants to sell and other other wants to buy). Thus $C(S,T) = max(S(T) - K, 0)$. If $S = 0$, the price will never change again and thus $C(0,t) = 0$. Finally if $S \to \infty$, we have $S > K$ with certainty and $C(\infty, t) = S$.

To address the time evolution of $C = C(S,t)$, we again use Ito's formula:

$$dC = (\partial_t C + \mu S\partial_S C + \frac{1}{2}(\sigma S)^2\partial_S^2 C)dt + \sigma S\partial_S C dW \tag{4.7}$$

With $dS = a(S)dt + b(S)dW$, this amounts to

$$dC = (\partial_t C + \frac{1}{2}(\sigma S)^2\partial_S^2 C)dt + \partial_S C dS \tag{4.8}$$

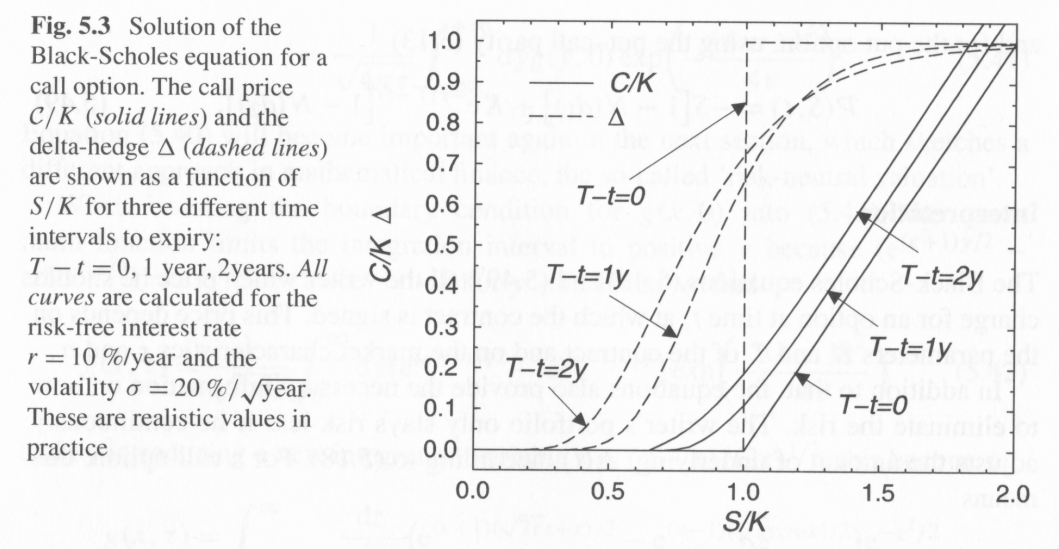


Fig. 5.3 Solution of the Black-Scholes equation for a call option. The call price C/K (*solid lines*) and the delta-hedge Δ (*dashed lines*) are shown as a function of S/K for three different time intervals to expiry: $T-t=0$, 1 year, 2years. *All curves* are calculated for the risk-free interest rate $r=10$ %/year and the volatility $\sigma=20\,\%/\sqrt{\text{year}}$. These are realistic values in practice

Figure 4.1: Plot of Black-Scholes equations from Paul and Baschnagel book.

We next introduce the hedging strategy of the writer. We assume that he owns a fraction $0 \leq \Delta(t) \leq 1$ of the underlying, which he adjusts dynamically to respond to the market. If S rises, Δ should increase, and vice versa. The purchase has to be paid from a cash amount $P(t)$, which also could be invested without risk with interest rate r. Typically $r < \mu$. The writer's wealth is the sum of the two and for a risk-free strategy, it should be equal to the option price:

$$W(t) = \Delta(t)S + P(t) = C(S,t) \tag{4.9}$$

If we assume that Δ is adjusted on a slow time scale, we have

$$dC = \Delta(t)dS + rP(t)dt \tag{4.10}$$

Comparing with the stochastic time evolution from above, we get

$$\Delta(t) = \partial_S C, rP(t) = \partial_t C + \frac{1}{2}(\sigma S)^2 \partial_S^2 C \tag{4.11}$$

Using $P(t) = C - \Delta S = C - (\partial_S C)S$ (a Legendre transform without information loss), we finally get

$$\partial_t C + \frac{1}{2}(\sigma S)^2 \partial_S^2 C + rS\partial_S C - rC = 0 \tag{4.12}$$

This deterministic PDE for C is the celebrated Black-Scholes equation. Surprisingly, it is independent of drift μ, which effectively has been replaced in the risk-free strategy by the interest rate r. Of course it strongly depends on volatility σ. Together with the boundary conditions from above, the solution is fully determined. In fact it can be obtained analytically using first non-dimensionalization, then transformation to a diffusion equation and finally its solution by the Green's function method. Here we only give the end result of this procedure:

$$C(S,t) = SN(d_1) - Ke^{-r(T-t)}N(d_2) \tag{4.13}$$

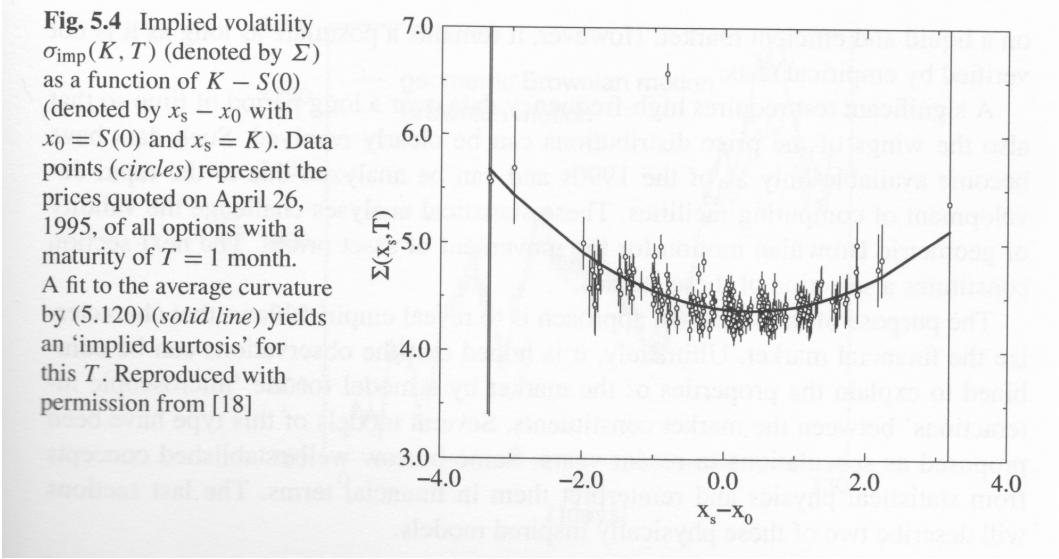


Fig. 5.4 Implied volatility $\sigma_{\text{imp}}(K, T)$ (denoted by Σ) as a function of $K - S(0)$ (denoted by $x_s - x_0$ with $x_0 = S(0)$ and $x_s = K$). Data points (*circles*) represent the prices quoted on April 26, 1995, of all options with a maturity of $T = 1$ month. A fit to the average curvature by (5.120) (*solid line*) yields an 'implied kurtosis' for this T. Reproduced with permission from [18]

Figure 4.2: Plot of volatility smile from Paul and Baschnagel book.

where the first term is the Delta-hedge and the second the cash amount. $N(x)$ is the cumulative distribution function of the normal distribution

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} dz e^{-z^2/2} = \frac{1}{2}(1 + erf(x/\sqrt{2})) \tag{4.14}$$

Finally we have defined

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \tag{4.15}$$

$$d_2 = d_1 - \sigma\sqrt{T-t} = \frac{\ln(S/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \tag{4.16}$$

$$\tag{4.17}$$

With these formulae, $C(S,t)$ is known and can be plotted for typical values of r, σ and T, compare Figure 4.1. One sees that even for $S(t) = K$, the writer already has to ask for a finite call price C and to build up a finite Delta-hedge Δ to be prepared for fluctuations.

Finally we comment on the main deficiencies of Black-Scholes theory. First volatility is not constant and increases with $mod(K - S(0))$ (the so-called volatility smile), compare Figure 4.2. Second price variations are not Gaussian, but have heavy tails, compare Figure 4.3. However, in times of normal business and if sampling rate is sufficiently small, the theory works quite well.

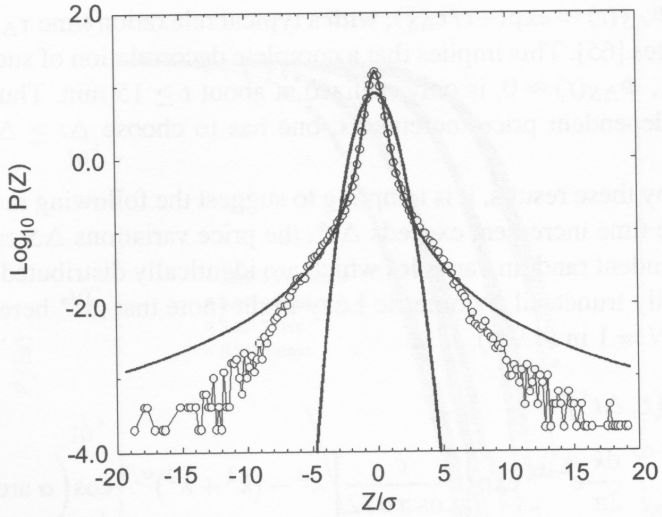


Fig. 5.6 Probability distribution $p_{\Delta t}(\ell)$ of the price variations ℓ $[= S(t + \Delta t) - S(t)]$ for the S&P500, determined from all records between 1984 and 1989 (about 1.5 million records). $p_{\Delta t}(\ell)$ (denoted by $P(Z)$) is plotted versus ℓ/σ for $\Delta t = 1$ min, where σ $(= 0.0508)$ is the standard deviation calculated from the data points. The Gaussian distribution (*thick solid line*) corresponding to this value of σ is compared to the best fit with a Lévy distribution (*thin solid line*). The Lévy distribution (with $\alpha = 1.4$ and $c = 3.75 \times 10^{-3}$, see (5.83)) gives a much better representation of the data for $\ell/\sigma \leq 6$. If $\ell/\sigma \geq 6$, the distribution of the index decays approximately exponentially. Reproduced with permission from [129]

Figure 4.3: Plot of stock market variations from Paul and Baschnagel book.

Chapter 5

Noise-induced transitions

The Langevin equation is a stochastic differential equation (SDE) which naively might be considered to be simply an ordinary differential equation (ODE) for some deterministic process extended by a noise term. This view might lead to the expectation that noise adds only a small perturbation to a deterministic reference system. In this chapter we will see that in fact (multiplicative) noise can qualititively change the behaviour of a system[1].
Recall the Fokker-Planck equation (FPE) written as a continuity equation:

$$\dot{p} + \partial_x J = 0 \tag{5.1}$$
$$J \equiv (A - \partial_x D)p \tag{5.2}$$

For stationary states ($\dot{p} = 0$, thus $\partial_x J = 0$) and reflecting boundary conditions (no flux over the boundary, thus flux vanishes everywhere) we have from $J = 0$

$$\partial_x p_s = \frac{A - \partial_x D}{D} p_s \implies \partial_x \ln p_s = \frac{A - \partial_x D}{D} \tag{5.3}$$
$$\implies \boxed{p_s = C \exp \int^x dx' \frac{A - \partial_{x'} D}{D}} \tag{5.4}$$

This suggest to define a *stochastic potential* $U(x)$ such that $p_s \propto e^{-U(x)/k_B T}$.
Now we distinguish between two cases: additive and multiplicative noise.

5.1 Additive noise

For additive noise and a potential (which always exists in one dimension) we have $\partial_x D = \partial_x \sigma^2 = 0$, $A = a = -\partial_x V$ where V is the real potential. Therefore

$$p_s = C e^{-V(x)/\sigma^2} = \frac{1}{Z} e^{-V(x)/k_B T} \tag{5.5}$$

where the last equality (Boltzman distribution) comes from thermodynamics. We recover the Einstein relation as before. Again we see that small noise amounts to

[1] For more examples for noise-induced transitions, e.g. the genetic model, check the book by W. Horsthemke and R. Lefever on exactly this subject, Springer 1984.

having low temperature. At low T the system minimizes V. On the other hand, at high T the system distributes over large regions in phase space. In this case, the stochastic and the real potential are identical.

Example: Double well (or Mexican hat) potential

The potential and the corresponding LE are:

$$V(x) = -\frac{\lambda}{2}x^2 + \frac{g}{4}x^4 \tag{5.6}$$

$$\dot{x} = \lambda x - gx^3 + \sigma\eta_t \tag{5.7}$$

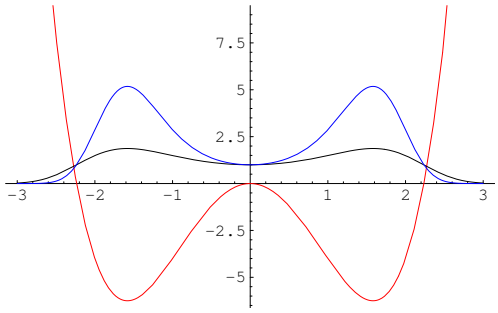


Figure 5.1: Double well potential (red) and stationary probability distributions (blue and black), $p_s = e^{-V/\sigma^2}$, for two different values of σ ($\sigma_{black} > \sigma_{red}$).

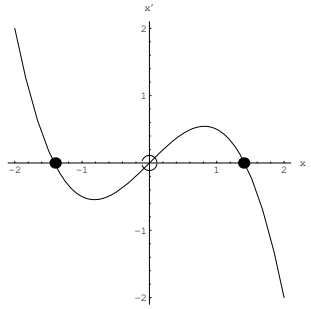


Figure 5.2: Deterministic dynamic equation for the double well potential, $\dot{x} = \lambda x - gx^3$. The three points where $\dot{x} = 0$ are the three fixed points: two stable ($x_s = \pm a$) and one unstable ($x_u = 0$).

5.2 Multiplicative noise

We consider the Stratonovich case. Here we have $A = a + b'b$ and $D = b^2$, therefore $A = a + \frac{1}{2}\partial_x D$. The stationary distribution function thus is

$$p_s = C \exp \int^x dx' \frac{A - \partial_{x'} D}{D} = C \exp \int^x dx' \frac{a - \frac{1}{2}\partial_{x'} D}{D} \tag{5.8}$$

Thus, the minimum (the stationary state a low temperature) is at $a - \frac{1}{2}\partial_x D = 0$ rather than at $a = -\partial_x V = 0$. For the Ito case, the sign would be different.
The correction term due to the multiplicative noise shifts the position of the stationary state away from the minimum of the potential. In general, noise can even destroy minima or create new minima. In the language of non-linear dynamics, it can change the nature of the fixed points. In particular, it can lead to different bifurcation behaviour (a bifurcation occurs when fixed points change their character or appear/disappear). Such qualitative changes are called *noise-induced transitions*.

5.3 Example: Verhulst model (or logistic growth)

This is the simplest and most important model for population growth. The deterministic equation is (population size $x \geq 0$)

$$\dot{x} = \underbrace{\lambda x}_{\text{growth}} - \underbrace{x^2}_{\text{limited resources}} \tag{5.9}$$

with a stable point at $x = \lambda$.

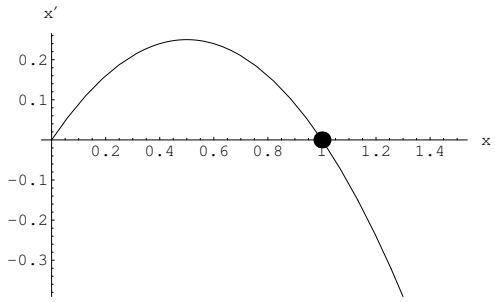


Figure 5.3: $\dot{x}$ as a function of x. In the plot the stable point is at $x_s = \lambda = 1$.

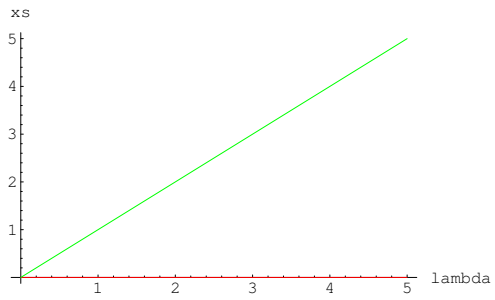


Figure 5.4: Bifurcation diagram for $x \geq 0$. In green the stable fixed points and in red the unstable ones.

The equation is solved by

$$x(t) = \frac{x_0 e^{\lambda t}}{1 + \frac{x_0}{\lambda}(e^{\lambda t} - 1)} \tag{5.10}$$

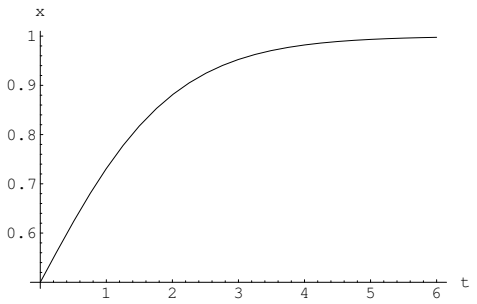


Figure 5.5: Deterministic solution $x(t)$. For an stochastic model of the same system one could expect just some perturbations around this trajectory. Is it always the case?

Now consider some fluctuations in the growth rate $\lambda \to \lambda + \sigma\eta_t$ (representing e.g. noise in the availability of food). The LE is then:

$$\dot{x} = \lambda x - x^2 + \underbrace{\sigma x \eta_x}_{\text{multiplicative noise}} \tag{5.11}$$

that is

$$a = \lambda x - x^2 \tag{5.12}$$

$$b = \sigma x \iff D = \sigma^2 x^2 \tag{5.13}$$

For the equivalent description using the FPE we have:

1. Itô interpretation:
$$\dot{p} = -\partial_x[(\lambda x - x^2)p] + \sigma^2 \partial_x^2 x^2 p \tag{5.14}$$

2. Stratonovich interpretation:
$$\dot{p} = -\partial_x[(\lambda x - x^2 + \sigma^2 x)p] + \sigma^2 \partial_x^2 x^2 p \tag{5.15}$$

Thus we deal with multiplicative noise.
We investigate the stationary state in the Stratonovich case. We now calculate the stochastic potential:

$$-\frac{U}{k_B T} = \int^x dx' \frac{a - \frac{1}{2}\partial_{x'} D}{D} = \int^x dx' \Big(\frac{\lambda x' - x'^2}{\sigma^2 x'^2} - \frac{1}{2}\partial_{x'} \ln(\sigma^2 x'^2) \Big) = \tag{5.16}$$

$$= \Big(\frac{\lambda}{\sigma^2} - 1 \Big) \ln x - \frac{x}{\sigma^2} + const \tag{5.17}$$

$$\implies \boxed{p_s = C x^{(\frac{\lambda}{\sigma^2} - 1)} e^{-x/\sigma^2}} \tag{5.18}$$

A calculation of the moments give:

$$\langle x \rangle_s = \int dx x p_s(x) = \lambda = x_s \tag{5.19}$$

$$\langle x^2 \rangle_s = \lambda^2 + \lambda\sigma \implies \langle (x - \langle x \rangle_s)^2 \rangle_s = \lambda\sigma^2 \tag{5.20}$$

Thus, nothing changes at the level of the first moment: $x_s = \lambda$ is a stable solution for $\lambda > 0$ as for the deterministic equation. The difference is in the higher moments and it makes $p_s(x)$ change its shape as a function of λ at $\lambda_c = \sigma^2$. The most probable value changes from $x_m = 0$ to $x_m = \lambda - \sigma^2$ at $\lambda = \lambda_c$.

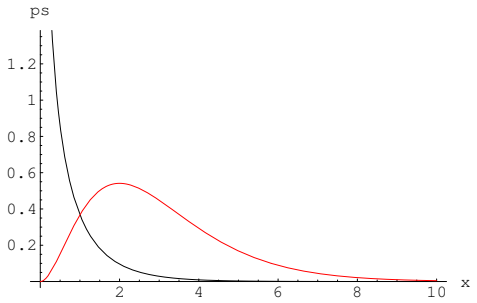


Figure 5.6: $p_s(x)$ for $\lambda < \lambda_c \equiv \sigma^2$ in black and for $\lambda > \lambda_c$ in red.

This also leads to a new bifurcation diagram with two transitions: at $\lambda = 0$, the population starts to grow in average, but only at $\lambda = \lambda_c$ there is a transition to a finite population size as most probable outcome. This second transition is *noise-induced*: it vanishes for $\sigma = 0$ (no noise). Between $\lambda = 0$ and $\lambda = \lambda_c$ the population is destabilised by noise: the variance $\langle (x - \langle x \rangle_s)^2 \rangle_s = \lambda\sigma^2 > \lambda^2 = \langle x \rangle_s^2$, that is the fluctuations are larger that the mean and the population is not stable.

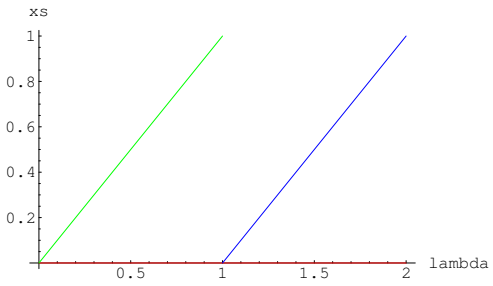


Figure 5.7: The most probable value (blue) is different from the average value (red) and is finite only above $\lambda = \lambda_c$. This defines a *noise-induced transition* at $\lambda = \lambda_c = \sigma^2$.

Chapter 6

First passage time problems

Often one is interested in the question when a stochastic process reaches a certain threshold:

- When does the Neckar reach the 3.5 m level?
- When does a stock reach the buy/sell limit?
- When does my laptop break down?
- When does a Brownian particle leave a domain?
- When does a biomolecular bond break?
- When does an ion channel open?

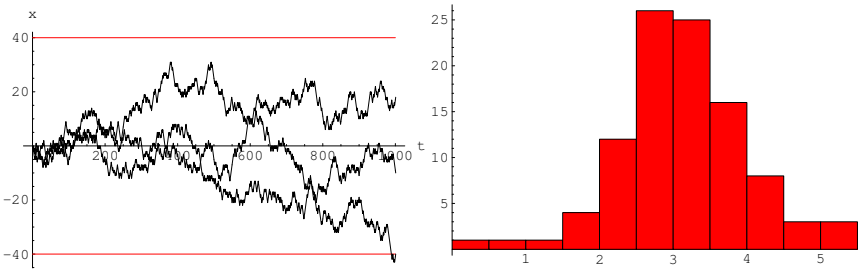
This time is called the *first passage time* (FPT). Like for the stochastic variable, there is a probability distribution (the *first passage time distribution*) from which one can calculate the *mean first passage time* (MFPT)[1].

6.1 FPT for Fokker-Planck equation

Random walk on a line

As a first simple example we consider a 1D particle performing a symmetric random walk with step size δ and jump time τ (e.g. transcription factor diffusion on DNA). We ask ourselves when the particle will reach the boundaries $x = a$ or $x = b$ as a function of a certain initial position. We ask for $T_1(x)$, the mean first passage time that it takes for the particle to reach the boundaries if it started at the position x.

[1] FPT-theory is discussed in all major books on stochastic processes, in particular in the one by van Kampen. For a recent review see Srividya Iyer-Biswas and Anton Zilman, First Passage processes in cellular biology, Advances in Chemical Physics, Volume 160, John Wiley and Sons 2016.



(a) Three random walks for $x \in (a, b)$ (b) Histogram with the frequencies of the first contact times

Figure 6.1: First passage time

Here a simple recursive relation holds:

$$T_1(x) = \tau + \frac{1}{2}\big[T_1(x+\delta) + T_1(x-\delta)\big] \tag{6.1}$$

$$\Longrightarrow \frac{1}{\delta^2}\big[T_1(x+\delta) + T_1(x-\delta) - 2T_1(x)\big] + \frac{2\tau}{\delta^2} = 0 \tag{6.2}$$

$$\Longrightarrow \boxed{T_1''(x) + \frac{1}{D} = 0} \tag{6.3}$$

where the last equation is obtained in the limit $\delta, \tau \to 0$ with $D = \delta^2/2\tau = const.$ This equation is an ODE for T_1 with a polynomial solution of second order in x. To solve it, we need boundary conditions, for example

(i) reflecting $\rightarrow T_1'(x_0) = 0$

(ii) absorbing $\rightarrow T_1(x_0) = 0$

For example, if we consider $a = 0$ and $b > 0$ being absorbing boundaries, then

$$T_1(x) = \frac{1}{2D}(bx - x^2) \tag{6.4}$$

For right or left side reflective, respectively, one gets:

$$T_1 = \frac{1}{2D}(2bx - x^2) \tag{6.5}$$

$$T_1 = \frac{1}{2D}(b^2 - x^2) \tag{6.6}$$

If the particle is released randomly, one still has to average over the starting position x. Being both sides absorbing:

$$T_1 = \frac{1}{b}\int_0^b T_1(x)dx = \frac{b^2}{12D} \tag{6.7}$$

The scaling with b^2/D is clear for dimensional reasons and reflects the RW-nature.

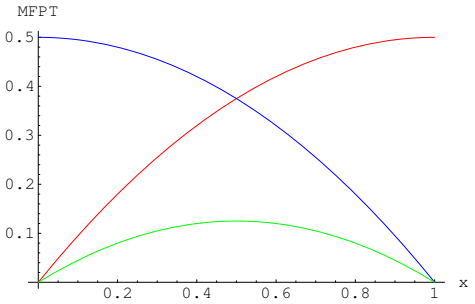


Figure 6.2: $T_1(x)$ for both boundaries absorbing (green), right side reflecting (red) and left side reflecting (blue). a and b have been set to 0 and 1, respectively.

MFPT for a 1D FPE

We now consider the same situation for a $1D$ FPE with arbitrary diffusion and drift. Recall the FPE and the adjoint FPE:

$$\frac{\partial p_2(x,t|x',t')}{\partial t} = \big(-\partial_x A(x,t) + \partial_x^2 D(x,t)\big) p_2(x,t|x',t') \tag{6.8}$$

$$\frac{\partial p_2(x,t|x',t')}{\partial t'} = \big(-A(x',t')\partial_{x'} - D(x',t')\partial_{x'}^2\big) p_2(x,t|x',t') \tag{6.9}$$

We define the probability that the particle is still in $[a,b]$ at time t when it started at x at time $t=0$:

$$G(t|x) \equiv \int_a^b dx' p(x',t|x,0) \tag{6.10}$$

If we consider an ensemble of particles, then G is the number of remaining particles. Boundary conditions on G:

$$p(x',0|x,0) = \delta(x'-x) \implies G(0|x) = 1 \text{ in } (a,b) \tag{6.11}$$

$$b \text{ absorbing} \implies G(t|b) = 0 \tag{6.12}$$

$$a \text{ reflecting} \implies \partial_x G(t|a) = 0 \tag{6.13}$$

Because b is absorbing, $G(t|x)$ decreases in time to $G(\infty|x) = 0$. We define the probability per time (rate) at which the boundary is reached as

$$f(t|x) = -\partial_t G(t|x) \tag{6.14}$$

Thus, the MFPT is

$$T_1(x) = \int_0^\infty dt t f(t|x) = -\int_0^\infty dt t \partial_t G(t|x) = \int_0^\infty dt G(t|x) \tag{6.15}$$

where partial integration has been used in the last step. The constant terms at the boundaries vanish because of the boundary conditions on G.

Similarly, for the higher moments we get

$$T_n(x) = \int_0^\infty dt t^n f(t|x) = -\int_0^\infty dt\ t^n \partial_t G(t|x) = n\int_0^\infty dt\ t^{n-1} G(t|x) \tag{6.16}$$

Like before, the constant terms vanish.
Now we want to derive a differential equation for G:

$$-\partial_t p(x',t|x,0) \overset{t-shift}{=} -\partial_t p(x',0|x,-t) \overset{AFPE}{=} \big[-A(x)\partial_x - D(x)\partial_x^2\big] p(x',0|x,-t) \tag{6.17}$$

$$\overset{t-shift}{=} \big[-A(x)\partial_x - D(x)\partial_x^2\big] p(x',t|x,0) \tag{6.18}$$

$$\overset{\int_a^b dx'}{\Rightarrow} \partial_t G = \big[A(x)\partial_x + D(x)\partial_x^2\big] G(t|x) \tag{6.19}$$

$$\overset{\int_0^\infty dt}{\Rightarrow} \underbrace{G(\infty|x)}_{=0} - \underbrace{G(0|x)}_{=1} = \big[A(x)\partial_x + D(x)\partial_x^2\big] T_1(x) \tag{6.20}$$

$$\Rightarrow \boxed{\big[A(x)\partial_x + D(x)\partial_x^2\big] T_1(x) = -1} \tag{6.21}$$

This is the *MFTP-equation* for the FPE. It is not a PDE like the FPE, but a non-homogeneous ODE. Therefore solving for the MFPT is easier than solving for the complete FPE.
For the higher moments, we combine

$$T_n(x) = n\int_0^\infty dt\ t^{n-1} G(t|x) \tag{6.22}$$

and

$$\partial_t G = \big[A(x)\partial_x + D(x)\partial_x^2\big] G(t|x) \tag{6.23}$$

to get

$$\big[A(x)\partial_x + D(x)\partial_x^2\big] T_n(x) = n\int_0^\infty dt\ t^{n-1}\partial_t G(t|x) = -nT_{n-1}(x) \tag{6.24}$$

a formula which for $n = 1$ gives the result for the MFPT from above with $T_0(x) = 1$. Thus we have a hierarchy of equations that can be solved recursively: once we have solved for moment n, we can go up to moment $n+1$.
We can check that the MFPT equation recovers the example discussed above by plugging in $A = 0$ and $D = const$ (this corresponds to the Wiener process):

$$\implies T_1''(x) + \frac{1}{D} = 0 \tag{6.25}$$

as obtained above.

The general solution for the MFPT-equation is:

$$T_1(x) = \int_x^b dz \frac{1}{\phi(z)} \int_a^z dy \frac{\phi(y)}{D(y)} \tag{6.26}$$

$$\phi(z) \equiv \exp\Big(\int^z dx \frac{A(x)}{D(x)} \Big) \tag{6.27}$$

with a being reflecting and b absorbing. We can check right away that indeed $T_1(b) = 0$ and $T'(a) = 0$. Note the definition of ϕ which we have seen before with the stochastic potential, namely in the form $\phi(x) = e^{-U(x)/k_B T}$. We check the general formula by explicit calculation:

$$(A\partial_x + D\partial_x^2)T = \underbrace{A\frac{-1}{\phi(x)} \int_a^x dy \frac{\phi(y)}{D(y)} + D\frac{\phi'(x)}{\phi^2(x)} \int_a^x dy \frac{\phi(y)}{D(y)}}_{=0} - \underbrace{D\frac{1}{\phi(x)}\frac{\phi(x)}{D(x)}}_{=1} = -1 \tag{6.28}$$

using $\phi' = \phi \dfrac{A}{D}$.

Example 1: Wiener process

For the symmetric RW $A = 0$ and $D = ctnt.$ The left boundary is taken to be at $a = 0$:

$$\implies \phi = 1 \ , \ T_1 = \frac{1}{D} \int_x^b dz \int_a^z dy = \frac{1}{2D}(b^2 - x^2) \tag{6.29}$$

$$x = 0 \implies T_1 = \frac{b^2}{2D} \implies b^2 = 2DT_1 \tag{6.30}$$

in accordance to previous results.
p, G and f can be calculated analytically as infinite sums over exp-functions. Therefore the complete solution is not easy, but getting the MFPT is relatively easy. However, the distribution for $f(t|x)$ (see Figure 6.3) is very broad, with standard deviation $O(b^2/D)$ like the mean. Therefore it is always good to check also for higher moments.

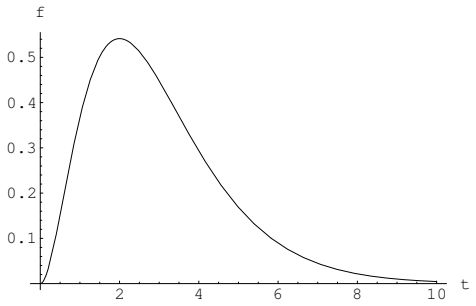


Figure 6.3: $f(t|x)$ as a function of t.

Example 2: Ornstein-Uhlenbeck process

It has $A = -\lambda x$ and $D = const.$ We take $a = x = 0$. Then

$$\phi(z) = e^{-\lambda z^2/(2D)} \tag{6.31}$$

$$T_1 = \frac{1}{D}\int_0^b dz e^{\lambda z^2/(2D)} \int_0^z dy e^{-\lambda y^2/(2D)} \tag{6.32}$$

$$\approx \frac{\sqrt{\pi}}{2}\frac{e^{\eta^2}}{\eta}\left(1 + \frac{1}{2\eta^2} + ...\right) \tag{6.33}$$

for $\eta^2 = \dfrac{\lambda b^2}{2D} \gg 1$. T_1 strongly increases with both λ or b.

Example 3: Escape over a barrier (Kramers problem)

In 1D, we always have a potential with $A = -\partial_x V$. We consider additive noise, $D = \sigma^2 = const$:

$$\dot{x} = -\frac{\partial V}{\partial x} + \sigma\eta_t \tag{6.34}$$

For the Kramers problem, we consider a double well potential, with a metastable state at the left and the equilibrium state at the right (see Figure 6.4). We want to know the MFPT for a particle in such a potential to go to the equilibrium position when it is originally at the metastable position. For this, it has to escape over the transition state barrier.

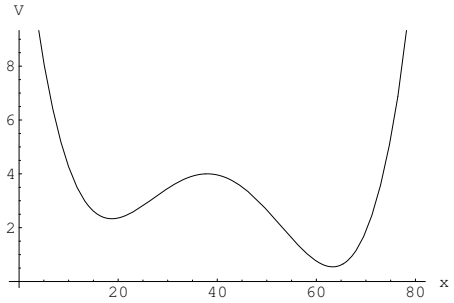


Figure 6.4: Double well potential. The minimum at the left is at x_0 and we take it as the starting position. The transition state is at x_1 and the absorbing state somewhere to the right, at b (exact position does not matter here). The reflecting position is at a at the left (again exact position does not matter).

Let the minimum of the left well be the starting point x_0. The *transition state* for this system is the point x_1 with the extremum. The boundaries are $a = -\infty$ on the left of the starting point, which will be reflective, and b on the right of x_1, which is

absorbing. Therefore we have

$$\phi(z) = \exp\Big(\int^z dx \frac{-V'(x)}{D}\Big) = e^{-V(z)/D} \tag{6.35}$$

$$T_1 = \frac{1}{D}\int_{x_0}^{b} dz e^{V(z)/D}\int_{-\infty}^{z} dy e^{-V(y)/D} \approx \frac{1}{D}\Big(\int_{x_0}^{b} dz e^{V(z)/D}\Big)\Big(\int_{-\infty}^{x_1} dy e^{-V(y)/D}\Big) \tag{6.36}$$

$$\approx \frac{1}{D} e^{V(x_1)/D}\Big(\frac{2\pi D}{V''(x_1)}\Big)^{1/2} e^{-V(x_0)/D}\Big(\frac{2\pi D}{V''(x_0)}\Big)^{1/2} \tag{6.37}$$

where the upper bound for the second integral is placed at x_1, because the largest contribution to the outer integral is generated at x_1 and the inner integral is flat around x_1. Here we have used the saddle point approximation or method of steepest descend twice:

$$\int_{-\infty}^{\infty} dy e^{-V(y)} \approx e^{-V(y_0)}\int dy e^{-\frac{1}{2}V''(y_0)(y-y_0)^2} = e^{-V(y_0)}\Big(\frac{2\pi}{V''(y_0)}\Big)^{1/2} \tag{6.38}$$

One finally obtains:

$$\boxed{T_1 = \frac{2\pi}{\sqrt{V''(x_0)V''(x_1)}} e^{(V(x_1)-V(x_0))/D} \sim e^{\Delta V/D} \sim e^{\Delta V/k_B T}} \tag{6.39}$$

where the last step follows with the Einstein relation. The escape time grows exponentially with the barrier height. In chemistry, this treatment is also known as *Eyring theory*. The exponential relation between escape time and inverse temperature is known as *Arrhenius law*. This result is very famous and many applications exist in physics, chemistry and biology for thermally assisted escape over a transition state barrier (including the fields of catalysis in chemistry and of enzymes in biochemistry). A related subject is extreme value statistics, which is important for e.g. insurance companies.

6.2 FPT for one-step master equation

We consider the one-step master equation from section 2.6 and ask when the process reaches $n = N$ after starting at $n = m$ at $t = 0$. Thus N is an absorbing boundary. Let $p_{n,m}(t)$ be the general solution of the master equation with the absorbing boundary. Then $p_{N,m}(t) = 0$.
Similar to the situation with the FPE, we define the following quantities:

1. Rate for reaching position N at time t: $f_{N,m}(t) = g(N-1)p_{N-1,m}(t)$.
2. Probability to reach N at all: $\pi_{N,m}(t) = \int_0^\infty dt f_{N,m}(t)$.
3. Mean first passage time to reach N: $\tau_{N,m}(t) = \int_0^\infty dt t f_{N,m}(t)$.

We next denote the solution of the unconstrained process (no absorbing boundary) as $q_{n,m}(t)$. For example, for the symmetric RW we have $q_{n,m}(t) = e^{-2t} I_{n-m}(2t)$. We then decompose the path from m to n into those which go directly, and those which first go to N:

$$q_{n,m}(t) = p_{n,m}(t) + \int_0^t dt' f_{N,m}(t') q_{n,N}(t-t') \ . \tag{6.40}$$

For $n = N$, this *renewal equation* becomes

$$q_{N,m}(t) = \int_0^t dt' f_{N,m}(t') q_{N,N}(t-t') \tag{6.41}$$

because $p_{N,m}(t) = 0$ from the boundary condition. We use the Laplace transform $q(s) = \int_0^\infty dt e^{-st} q(t)$ to get

$$q_{N,m}(s) = f_{N,m}(s) q_{N,N}(s) \ . \tag{6.42}$$

Therefore

$$f_{N,m}(s) = \frac{q_{N,m}(s)}{q_{N,N}(s)} = \int_0^\infty dt e^{-st} f_{N,m}(t) \tag{6.43}$$

and thus

$$\pi_{N,m} = f_{N,m}(s=0), \ \tau_{N,m} = -f'_{N,m}(s=0) \ . \tag{6.44}$$

As an example, we consider the symmetric RW, for which we find:

$$q_{n,m}(s) = \int_0^\infty dt e^{-(s+2)t} I_{n-m}(2t) = \frac{(1 + \frac{s}{2} + \frac{1}{2}\sqrt{s(s+4)})^{m-n}}{\sqrt{s(s+4)}} \tag{6.45}$$

Therefore

$$f_{N,m}(s) = (1 + \frac{s}{2} + \frac{1}{2}\sqrt{s(s+4)})^{m-N} \ . \tag{6.46}$$

We set $m = 0$ and find

$$\pi_{N,0} = f_{N,m}(s=0) = 1, \ \tau_{N,0} = -f'_{N,m}(s=0) = \infty \ . \tag{6.47}$$

We conclude that the particle reaches position N with certainty, but in average after infinite time.

This result is part of the famous 1921 theorem by George Polya which states that random walks are recurrent in one and two dimensions, but not in three or higher dimensions. The recurrence probability in three dimensions is around 34 percent and an analytical but very complicated formula exists for this number. He also showed that the expectation value for the recurrence time is infinite in both one and two dimensions. Today there are several proofs available for the Polya theorem. The essential difference between two and three dimensions is that the harmonic series $\sum_{i=1}^\infty 1/i$ diverges, while $\sum_{i=1}^\infty 1/i^2$ is finite (value $\pi^2/6$) (compare e.g. the book by Häggström). These sums appear if one counts how many paths there are for the probability flow from the origin to infinity (following the Kirchhoff laws for current flow in an electrical network).

The mean first passage time to reach the absorbing boundary at the right position $n = N = R$ will become finite if a reflecting boundary exists at the left position

$n = L$. Again we consider a partice that starts at $n = m$. At each position m, the particle jumps left with probability $g_m/(g_m + r_m)$ and right with probability $r_m/(g_m + r_m)$. For the probability to reach the right boundary R when starting at m, we therefore have a recursive relation:

$$\pi_{R,m} = \frac{g_m}{g_m + r_m}\pi_{R,m+1} + \frac{r_m}{g_m + r_m}\pi_{R,m-1} \ . \tag{6.48}$$

We rearrange to

$$g_m(\pi_{R,m+1} - \pi_{R,m}) + r_m(\pi_{R,m-1} - \pi_{R,m}) = 0 \ . \tag{6.49}$$

With a reflecting boundary at $n = L = 0$, we have $r_0 = 0$. Therefore $\pi_{R,1} = \pi_{R,0}$. We now increase m by one at a time and recursively find that all $\pi_{R,m} = \pi_{R,R} = 1$. There with the reflecting boundary at the left, the particle now reaches the absorbing right boundary with certainty.
We now make a similar argument for the mean first passage time, but this time we also have to consider the fact that the particle can not move (this costs time, but not probability):

$$\pi_{R,m} = \Delta t + g_m \Delta t \tau_{R,m+1} + r_m \Delta t \tau_{R,m-1} + (1 - (g_m + r_m)\Delta t)\tau_{R,m} \tag{6.50}$$

We rearrange to

$$g_m \underbrace{(\tau_{R,m+1} - \tau_{R,m})}_{=:\Delta_m} + r_m \underbrace{(\tau_{R,m-1} - \tau_{R,m})}_{=-\Delta_{m-1}} = -1 \ . \tag{6.51}$$

Thus we again have a recursive relation:

$$\Delta_m = -\frac{1}{g_m} + \frac{r_m}{g_m}\Delta_{m-1} \ . \tag{6.52}$$

From $r_0 = 0$ we have $\Delta_L = -1/g_L$. The next step gives us

$$\Delta_{L+1} = -\frac{1}{g_{L+1}} - \frac{r_{L+1}}{g_{L+1}}\frac{1}{g_L} \ . \tag{6.53}$$

By recursion we get

$$\Delta_m = -\sum_{\mu=L}^{m} \frac{r_m \dots r_{\mu+1}}{g_m \dots g_{\mu+1}}\frac{1}{g_\mu} \ . \tag{6.54}$$

Using $\tau_{R,R} = 0$ we arrive at our final result:

$$\tau_{R,m} = -\sum_{\nu=m}^{R-1} \Delta_\nu = \sum_{\nu=m}^{R-1}\sum_{\mu=L}^{m} \frac{r_m \dots r_{\mu+1}}{g_m \dots g_{\mu+1}}\frac{1}{g_\mu} \ . \tag{6.55}$$

Thus we have found an exact formula for the MFPT for the one-step master equation with reflecting boundary at L. The same result can also be derived using Laplace-transforms.

For the symmetric random walk, we have $r = g = 1$ and therefore

$$\tau_{R,m} = \sum_{\nu=m}^{R-1} (\nu + 1 - L) = (1 - L)(R - m) + \frac{(R-1)R}{2} - \frac{(m-1)m}{2} =\approx \frac{R^2}{2} \quad (6.56)$$

for $R \gg m$. This is indeed the expected scaling for a random walk.
Another important example is the linear birth-death process with $0 \leq n \leq N$, $r_n = n$ and $g_n = \gamma(N - n)$. Here one would place the reflecting boundary at the right (population cannot become larger than N) and the absorbing boundary at the left (population cannot recover when reaching 0). A similar formula as above can be derived for the MFPT, both by recursion or by Laplace transform. One finds [2]

$$T = \sum_{i=1}^{N} \frac{1}{r_i} + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\prod_{k=j-i}^{j-1} g_k}{\prod_{k=j-i}^{j} r_k} \; . \quad (6.57)$$

The first part is the one for $\gamma = 0$, also known as the harmonic number, which scales as $\ln N$ for large N. The second part increases lifetime by the birth or rebinding rate γ. Overall T scales exponentially with N and as γ^{N-1} with the rebinding rate. To increase population lifetime, we should make it large and/or increase birth rate. For $N = 2$, we get

$$T = \frac{1}{2}(3 + \gamma) \; . \quad (6.58)$$

This result can be verified by direct calculation.

[2] T. Erdmann and U.S. Schwarz, Stochastic dynamics of adhesion clusters under shared constant force and with rebinding, Journal of Chemical Physics 121: 8997, 2004.

Chapter 7

Path integral formulation

Stochastic dynamics can be formulated also in the language of path integrals, similar to quantum field theory. This formulation immediately opens the door to such powerful methods as Feynman integrals and renormalization group theory. Traditionally it has been used to calculate correlation and response functions. More recently it has been used a lot in stochastic thermodynamics, for example to prove the integral fluctuation theorem[1].
We start with the Chapman-Kolmogorov equation and apply it repeatedly to small time steps $\Delta t = (t'' - t')/N$. Thus we have

$$p_2(x'', t''|x', t') = \int dx_1 \dots dx_{N-1} \prod_{\alpha=0}^{N-1} p_2(x_{\alpha+1}, t_{\alpha+1}|x_\alpha, t_\alpha) \ . \tag{7.1}$$

This is effectively an integration over all paths from x' to x'', that is a path integral (also known as functional integral). For a stochastic process with drift $A(x,t)$ and noise amplitude $\sqrt{D(t)}$ (additive noise), one can convert this to

$$p_2(x'', t''|x', t') = \int Dx e^{-S[x(t)]} \tag{7.2}$$

with a suitably defined path integral Dx and the action

$$S[x(t)] = \int_{t'}^{t''} dt \frac{(\dot{x} - A)^2}{2D} \tag{7.3}$$

where the integrand can be called the Lagrange function. One sees that the deterministic path defined by $\dot{x} = A$ dominates the outcome, because it has $S = 0$, and that fluctuations are suppressed exponentially, but become more important with larger noise.
We note that in quantum field theory, one would write a similar path integral for the transition probability:

$$p_2 = \int D\Phi e^{\frac{i}{\hbar} S[x(t)]} \tag{7.4}$$

[1]This field is reviewed in Udo Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, Reports on Progress in Physics 75: 126001, 2012.

where the path integral is now over all possible wavefunctions Φ. For the free theory, the Lagrange function would be

$$L = \frac{1}{2}(\partial_\mu \Phi)(\partial^\mu \Phi) - \frac{1}{2}m^2\Phi^2 \ . \tag{7.5}$$

Thus QFT corresponds to SD with a complex time (Wick rotation).