

Non-equilibrium systems

Steffen Rulands

August 7, 2023

These are a compilation of lecture notes giving the foundations of non-equilibrium physics and stochastic processes. They are intended as reading material for the preparation of a biophysics practical involving the stochastic simulation of active Brownian motion.

Contents

1	Systems in an out of thermal equilibrium	5
1.1	Non-equilibrium systems	5
1.1.1	Probability and information	8
1.1.2	The fluctuation-dissipation theorem	18
2	Theory of stochastic processes	25
2.1	Stochastic processes	25
2.1.1	Langevin approach (stochastic differential equations)	27
2.1.2	Einstein approach	34

1 Systems in an out of thermal equilibrium

1.1 Non-equilibrium systems

Consider an isolated system. The second law of thermodynamics tells us that this system will evolve into a state characterized by a quantity termed entropy, S . The change in entropy vanishes if the process is completely reversible so that isolated systems always reach a state of thermodynamic equilibrium. Ludwig Boltzmann derived a microscopic description of the entropy by considering the probabilities of finding different microscopic configurations of systems composed of many particles. In the ensuing framework of statistical mechanics, the entropy is given by the number of microscopic configurations Ω that are compatible with a given macroscopic state of interest,

$$S = k_B \ln \Omega . \tag{1.1}$$

Immediately, many systems come to our mind that apparently break this law, including virtually any biological system. A striking example is the self-organization of cells during embryonic development, where an initially homogeneous cluster of cells gives rise to complex organs, such as the heart or the brain. But if isolated systems converge to thermodynamic equilibrium, what makes a system being out of thermal equilibrium? Several possibilities come to our mind:

1. The system may be closed but has not reached an equilibrium state yet. Examples here include the Belousov-Zhabotinsky (BZ) reaction and the universe we live in: there are spatial structures and dynamics present, indicating that it is clearly out of equilibrium, but is thought to ultimately reach a state of maximum entropy.
2. The system is not isolated, but *open*, i.e. it exchanges energy or particles with another system. However, this does not necessarily lead to an out-of-equilibrium state, as many open systems reach an equilibrium state that is described by the canonical or grand canonical ensemble of statistical mechanics. For the remainder of this section, we will study open systems that are permanently out of thermal equilibrium.

3. Finally, as a combination of the cases above, an open system might ultimately reach an equilibrium state but hasn't arrived there yet. This includes for example disordered systems, where many components interact in different ways. Such systems often exhibit *frustration*, where the energetic minima of different components are incompatible. Frustration can lead to a behavior where thermodynamic equilibrium is approached very slowly and might never be fully reached. Such systems serve, for example, as models for understanding deep learning.

As we know from statistical physics, there are many open systems that converge to equilibrium, for example when coupled to a heat or chemical bath. What then drives an open system out of equilibrium? To understand this, consider the dynamics of a single particle in a many-body system following Hamiltonian dynamics. How are the dynamics of this particle affected by the other particles?

1. The other particles will hinder the ballistic motion of the particle. From a coarse-grained perspective, this leads to friction;
2. Repeated collisions with other particles give rise to a fluctuating force on the particle.

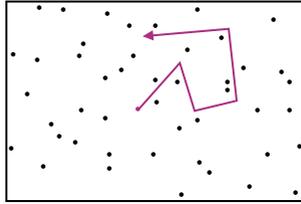


Figure 1.1: Collisions with other particles give rise to a random force. As a result of this fluctuating force, the particle will undergo a random trajectory.

The action of fluctuations and dissipation can be described in terms in a probabilistic manner, where its position is described by a random variable. Before we further study the properties of this random variable, let us remind ourselves of some basic concepts of probability theory.

1.1.1 Probability and information

We begin with the most elemental concept of probability theory: A *random variable* X represents possible outcomes S of a random phenomenon. S may be discrete (e.g. $S = \{x_1, x_2, \dots\}$) or continuous (e.g. $S = \mathbb{R}$). A specific realization of the random phenomenon is termed an *event*, $E \subseteq S$.

Example: Random variables

- A coin flip: $S = \{Head, Tail\}$;
- The number of students in a class room: $S = \mathbb{N}$.

Now we would like to quantify how likely different events are. To do this, we consider an ensemble of similarly prepared systems, meaning that we fix all degrees of freedom that we can control. The unspecified degrees of freedom, usually much larger in number, may still give rise to random outcomes¹. Consider an ensemble with a large number of realizations, N . Then we define the probability density function by computing a histogram:

$$p_X(x) = \frac{\left(\begin{array}{c} \text{number of systems with outcomes} \\ \text{between } x \text{ and } x + dx \end{array}\right)}{Ndx} \quad (1.2)$$

Often we drop the index X and simply write $p(x)$. For continuous random variables, the probability that a single realization takes a value x is, therefore, $p(x)dx$.

In the following, we summarise some important properties of probability distributions:

¹In a quantum system, even fixing all degrees of freedom can give random outcomes

- As we are certain to observe any outcome of a random phenomenon the probability distribution is normalized, $\int p(x)dx = 1$.
- The *joint probability* that two random variables X and Y take values between x and $x + dx$, and y and $y + dy$, respectively, is $p(x, y)$.
- The *marginal probability* is defined as $p(x) = \int p(x, y)dy$.
- Two random variables, X and Y , are *independent* if the occurrence of one does not affect the other. In this case, the joint probability density factorizes $p(x, y) = p(x)p(y)$.
- The *conditional probability* that X takes a value x given that Y takes a value y is $p(x | y)$. This implies that $p(x | y)p(y) = p(x, y)$.

By noting that $p(x, y) = p(y, x)$ and substituting the formula relating conditional probabilities to joint probabilities we obtain *Bayes' theorem*,

$$p(x | y) = \frac{p(y | x)p(x)}{p(y)}. \quad (1.3)$$

Bayes theorem allows us to calculate a conditional probability $p(x | y)$, termed posterior, in terms of a conditional probability $p(y | x)$ and some marginal probabilities which might be easier to obtain.

Example: Do you have Covid?

As an example of Bayes' theorem, we ask what is the probability that someone has Covid given a positive (+) antigen test result. We use Bayes theorem to express this conditional probability in terms of the probability that a test result is positive given that someone has Covid. This probability is called the *true positive rate* and it is typically provided by manufacturers of diagnostic tests. For antigen tests, this is about 70%. We have

$$p(\text{Covid} | +) = \frac{p(+ | \text{Covid})p(\text{Covid})}{p(+)} . \quad (1.4)$$

How can we find the probability in the denominator? We write

$$p(+) = p(+ | \text{Covid})p(\text{Covid}) + p(+ | \text{not Covid})p(\text{not Covid}) . \quad (1.5)$$

The conditional probability $p(+ | \text{not Covid})$ is called *false positive rate*. It is reported by the manufacturers and typically is roughly 10%. If 1% of the population is infected with Covid, then the computation above shows that a positive test indicates an actual Covid infection with a probability of 6%. If you thought the result would have been a larger percentage, you presumably did not consider the relatively low base rate of 1%. This is known as

the base rate fallacy.

We can summarize important features of probability distributions and empirical data samples by calculating summary statistics. Some of the most useful summary statistics are:

1. The *mean* or expectation value of a random variable is $\langle x \rangle = \int xp(x)dx$;
2. The *variance* is a measure of the degree of variability of a random variable and it is defined as $\text{var}(x) = \langle (x - \langle x \rangle)^2 \rangle$.
3. The *covariance* measures the joint variability of two random variables, $\text{cov}(x, y) = \langle xy \rangle - \langle x \rangle \langle y \rangle$. The value of the covariance is positive if the change in the value of one of the random variables is associated with a change in the other random variable in the same direction. The value is negative if the associated change is in the opposite direction.

1.1.1.1 Entropy

The (Shannon) *entropy* quantifies our lack of knowledge about the outcome of a random variable. It is defined as

$$H(X) = - \int dx p(x) \log_2 p(x). \quad (1.6)$$

By definition, the entropy is formally identical to the expectation value of a quantity $\log_2 [1/p(x)]$,

$$H(X) \equiv \left\langle \log_2 \left(\frac{1}{p(x)} \right) \right\rangle. \quad (1.7)$$

The quantity $1/p(x)$ is small if the probability associated with the outcome x is high. It is large if the probability associated with the outcome x is low. It, therefore, represents the degree of surprise we have when learning about the outcome x . In other words, it is the amount of information that we learn by being informed about the outcome x . The Shannon entropy there is the average information content or surprise that we can expect over all possible outcomes of the experiment.

Example: Subject of study

Consider a random experiment, where we pick a random student from a lecture on statistical physics lecture. We ask the student what is their field of study. The response “physics” has low information content, because we expect this answer with a high probability. The corresponding entropy is low. If we now ask the same question in the canteen, the answer will be less predictable. Each answer has have higher information content and such that the entropy is higher.

Example: Coin toss

Consider a coin that shows a head with a probability q . In that case the entropy is $H = q \log_2(1 - q) + (1 - q) \log_2 q$. This means that the entropy is highest for $q = 1/2$ when individual outcomes are least predictable. The entropy vanishes if $q = 1$ where the coin toss becomes completely predictable.

An associated concept is that of the *conditional entropy*, which is the average uncertainty about a random variable X given an observation of the random variable Y ,

$$H(X | Y) = - \int dy p(y) \int dx p(x | y) \log_2 p(x | y). \quad (1.8)$$

This is essentially an average with respect to $p(y)$ of the Shannon entropy associated with the conditional probability $p(x|y)$.

1.1.1.2 The Kullback-Leibler divergence

The Kullback-Leibler divergence is a measure of the distance between two probability distributions $p(x)$ and $q(x)$ defined over

the same space S . It is defined as

$$D(p||q) = \int_{-\infty}^{\infty} dx p(x) \log \left(\frac{p(x)}{q(x)} \right) \quad (1.9)$$

$$\equiv \langle \log p(x) - \log q(x) \rangle_p. \quad (1.10)$$

By definition, we see that this is a weighted average of the logarithmic difference between two distributions, hence the interpretation as a distance measure. Usually, p is an empirical distribution obtained from a histogram and q is a theoretical model.

Example: Variational autoencoders

In machine learning, the Kullback-Leibler divergence is used extensively, for example in variational autoencoders. These architectures typically consist of two parts, an encoder, and a decoder. The encoder has the function of mapping the input features to the parameters of some distribution, for example, a Gaussian. A point is then sampled from this distribution and the decoder tries to reconstruct the input as precisely as possible from this sample. To enforce that the encoder learns a specific distribution, in this example a Gaussian, the Kullback-Leibler divergence is used in the loss function to compare the encoded distribution to the preferred one. When these are very different, the Kullback-Leibler term in the loss function will penalize the network, such that it

eventually learns the correct distribution.

1.1.1.3 Mutual information

Mutual information measures the amount of information that can be obtained about a random variable X from knowledge about another random variable Y ,

$$I(X; Y) \equiv \int dx dy p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right). \quad (1.11)$$

The mutual information can also be expressed in terms of the entropy and the conditional entropy,

$$I(X; Y) = H(Y) - H(Y | X). \quad (1.12)$$

The first term on the right-hand side can be interpreted as the amount of uncertainty about Y . The second term described what X does not tell us about Y . The difference between the two, the mutual information, therefore measures the amount of uncertainty about Y that is removed by knowing X .

1.1.1.4 Important probability distributions

Simple random experiments often produce widely used probability distributions. We now discuss a few of these distributions which appear in a wide range of applications. The most widely used probability distribution is the *normal distribution*, which is in the context of physics also often called the Gaussian distribution. The prevalence of the normal distribution stems from the fact that it is the limiting distribution of random experiments, where the random variable is the result of a summation of other random variables. The *central limit theorem* states that under quite general conditions the sum of N independent and identically distributed random variables follows a normal distribution,

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (1.13)$$

with mean μ and standard deviation σ , scaling as $1/\sqrt{N}$.

Consider a random experiment, where the random variable is given by the counting of statistically independent events. The number of events, k , in a given time interval then follows a *Poisson distribution* with probability density function

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (1.14)$$

where the parameter λ gives the average number of events in this time interval.

If we now consider the same experiment, but we measure the time x that passed between two successive events, then, if the events are statistically independent, this time follows an *exponential distribution*,

$$p(x) = \lambda e^{-\lambda x}. \quad (1.15)$$

Since these times must be positive we have the additional requirement that $p(x) = 0$ for $x < 0$.

1.1.2 The fluctuation-dissipation theorem

We are now in a position to describe the random variable representing the position of the particle. This random variable follows a stochastic generalization of Newton's equation, known as the

Langevin equation,

$$m\dot{v} = \underbrace{-m\gamma v}_{\substack{\text{Dissipative friction,} \\ \text{friction coefficient } \gamma}} + \underbrace{F}_{\substack{\text{Macroscopic force} \\ \text{acting on particle}}} + \underbrace{\xi(t)}_{\substack{\text{Randomly fluctu-} \\ \text{ating force}}}.$$

(1.16)

The Langevin equation contains a deterministic part, describing the macroscopic force exerted by friction, a part that is described by a random variable representing the fluctuating force stemming from random collisions with other particles. Due to the central limit theorem, the random variable $\xi(t)$ follows a normal (Gaussian) distribution with the following properties:

$$\langle \xi(t) \rangle = 0 \tag{1.17}$$

$$\langle \xi(t)\xi(t') \rangle = A\delta(t - t'), \tag{1.18}$$

where the amplitude A captures the strength or variance of the fluctuating force and the δ -function ensures that the random fluctuations are uncorrelated in time. To understand the connection between dissipation and fluctuations, we consider the dynamics in the absence of an external force, i.e. $F = 0$. The Langevin

equation in Fourier space then reads

$$-i\omega\tilde{v} = -m\gamma\tilde{v} + \tilde{\xi}, \quad (1.19)$$

with $\tilde{v} = \int dt \exp\{i\omega t\}v(t)$, and this has the formal solution

$$\tilde{v}(\omega) = \frac{1}{-im\omega + m\gamma}\tilde{\xi}(\omega) \equiv R(\omega)\tilde{\xi}(\omega). \quad (1.20)$$

$R(\omega)$ is defined as the *response function* of the system, and its imaginary part is in fact related to the fluctuations in the system. To see this, let us first define the spectral density as the Fourier transform of the correlation function,

$$S_{\xi}(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} \langle \xi(t)\xi(0) \rangle \quad (1.21)$$

$$= A, \quad (1.22)$$

where we used the Eq. (1.18) of the pairwise correlations to compute the integral. With this, we get

$$S_v(\omega) = \langle \tilde{v}(\omega)\tilde{v}(\omega') \rangle \quad (1.23)$$

$$= \frac{1}{|-i\omega m + \gamma m|^2} S_{\xi}(\omega) \quad (1.24)$$

$$= \frac{A}{m^2(\omega^2 + \gamma^2)}, \quad (1.25)$$

which can be transformed back to the time domain, yielding

$$\langle v(t)^2 \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega t} S_v(\omega) \quad (1.26)$$

$$= \frac{A}{2m^2\gamma^2}, \quad (1.27)$$

where in the last step we calculated the contour integral using the residual theorem.

In thermal equilibrium, the average kinetic energy per degree of freedom is connected to the temperature according to the *equipartition theorem*,

$$\frac{1}{2}m \langle v \rangle^2 = \frac{k_B T}{2}. \quad (1.28)$$

This leads to the *Einstein relation*

$$A = 2m\gamma k_B T, \quad (1.29)$$

which relates the strength of the fluctuating force to the temperature, T . The Einstein relation allows us to rewrite the spectral density in the form

$$S_v(\omega) = \frac{2m\gamma k_B T}{m^2(\omega^2 + \gamma^2)}. \quad (1.30)$$

This equation relates the generation of kinetic energy due to a

fluctuating force to its dissipation to heat due to friction. Fluctuations and dissipation are therefore linked, a manifestation of the *fluctuation dissipation theorem* of which the Einstein relation is a special case. In order to derive the general form of the fluctuation-dissipation theorem we note that

$$\text{Im } R(\omega) = \frac{m\omega}{m^2(\omega^2 + \gamma^2)} \quad (1.31)$$

$$= \frac{\omega}{m(\omega^2 + \gamma^2)}. \quad (1.32)$$

With this, we obtain the fluctuation-dissipation theorem in its general form,

$$S_v(\omega) = \frac{2k_B T}{\omega} \text{Im } R(\omega). \quad (1.33)$$

Therefore, in equilibrium, loss of energy by dissipation is strictly balanced by an intake of energy through fluctuations.

How can we then bring a system out of equilibrium? Let us couple the particle to an additional energy reservoir. To this end, we consider the example of self-propelled, or chemotactic, bacteria, such as *e.Coli*. Chemotactic bacteria consume chemical energy via nutrients. This energy is translated into a rotation of protein chains called flagella, and ultimately into a kinematic force $F = \varepsilon\kappa v$. The movement of the bacterium is then described

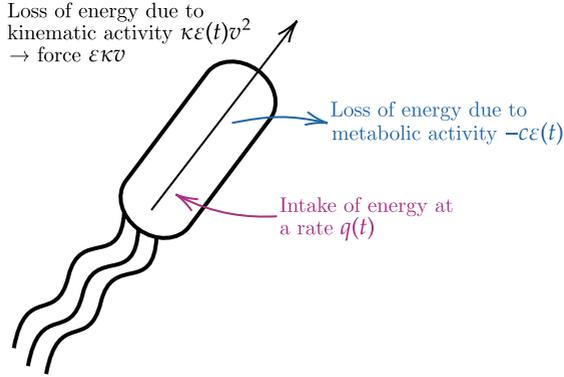


Figure 1.2: Many bacteria, such as e.Coli can move by converting chemical energy into rotational energy of their flagella. This gives rise to forward motion.

by a Langevin equation of the form

$$m\dot{v} = \underbrace{-m\gamma v + \epsilon\kappa v}_{\text{dissipation term}} + F_{\text{ext}} + \xi(t). \quad (1.34)$$

Chemotactic bacteria counteract dissipative friction by taking in chemical energy from the environment. They uncouple fluctuations and dissipation and therefore operate out of thermal equilibrium. As the total system is closed and obeys the second law of thermodynamics, this leads to production of entropy in the environment.

Remark:

In fact, the rate of energy dissipation, J , which is a measure of how far away the system is from equilibrium, is equal to the magnitude of the deviation from the fluctuation-dissipation theorem,

$$J = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left[\tilde{C}(\omega) - 2k_B T \operatorname{Im} R(\omega) \right]. \quad (1.35)$$

Example: Fluctuation-dissipation relations

- Bacteria convert ATP into ADP in order to drive the rotation of flagella for self-propulsion. Energy is dissipated in this process, leading to an increase in temperature in the environment.
- In turbulence, velocity fluctuations of the flow are induced at large spatial scales. The kinetic energy of these fluctuations then cascades from large to small spatial scales. At a smaller spatial scale, these fluctuations dissipate by viscosity into heat.

2 Theory of stochastic processes

2.1 Stochastic processes

A stochastic process is a collection of random variables indexed by a “time” t , $X(t)$. An important class of stochastic processes are *Markov processes*, where the probability of finding the system in a state X at time t only depends on the previous state, but not on any other state further in the “past”,

$$P(x(t_i) | x(t_{i-1}), x(t_{i-2}), \dots) = P(x(t_i) | x(t_{i-1})). \quad (2.1)$$

In order to describe a stochastic process the best we can achieve is finding the probability that X takes a value x at time t , which we denote by $P(x, t)$. There are two strategies of how we can achieve this:

1. We may consider specific realisations of $X(t)$. In this case, the time evolution of $X(t)$ is given by equations of Langevin

type,

$$\dot{x} = f(x, t) + \xi(t), \quad (2.2)$$

where the first term, $f(x, t)$ describes the deterministic component of the time evolution and the second term, $\xi(t)$, is a stochastic process describing noise. We can then average over many realisations of ξ to obtain histograms that approximate $p(x, t)$. This approach was followed by Langevin.

2. We may also directly consider the probability $P(x, t)$ and derive differential equations describing its time evolution,

$$\frac{d}{dt}p(x, t) = F(p, x, t). \quad (2.3)$$

This approach was followed by Einstein in his work on Brownian motion.

We will study the connections between both approaches in more detail at a later point. A conceptual difference is that the Langevin approach assumes that fluctuations in the steady state are entirely determined by a “bath” represented by the noise term, $\xi(t)$. The Master equation approach on the other hand provides a general framework for general Markov processes.

2.1.1 Langevin approach (stochastic differential equations)

Langevin's equation for a Brownian particle is a simple example of a *stochastic differential equation* (SDE). As any system involving higher order derivatives with respect to time can be written in terms of a set of first order differential equations, the general form of stochastic differential equation is

$$\dot{x} = a(x, t) + b(x, t)\xi(t). \quad (2.4)$$

In this equation, $a(x, t)$ captures the deterministic part of the system, and $b(x, t)\xi(t)$ the stochastic part. Two classes of stochastic differential equations can be distinguished based on the explicit form of $b(x, t)$:

- A stochastic differential equation is termed to have *additive noise* if the amplitude of the noise term is independent of the value of $x(t)$, $b(x, t) = b \equiv \text{const}$;
- A stochastic differential equation is term to have *multiplicative noise* if the amplitude of the noise term is a function of $x(t)$ $b(x, t) = b(x)$.

We can also write stochastic differential equations in differential form,

$$dx = a(x, t)dt + b(x, t)dW(t), \quad (2.5)$$

where we have defined the noise increment $dW(t) \equiv \xi(t)dt$, which can be integrated to obtain $W(t) = \int_0^t \xi(t)dt$. $W(t)$ is called a *Wiener process* or, in the physics literature, *Brownian motion*. The Wiener process is a Markov process, such that the increments $dW(t)$ are independent of the current state $W(t)$.

The differential form is formally solved by

$$x(t) = x(0) + \int_{t_0}^t a(x, s)ds + \int_{t_0}^t b(x, s)dW_s. \quad (2.6)$$

This begs the question of how the integral with respect to dW_s should be interpreted? To see this, let us follow the approach taken for the definition of the Riemann integral and define a partition Π_n subdividing the time domain into n intervals. With this, we can define the *stochastic integral* as

$$\int_{t_0}^t f(s)dW_s = \lim_{|\Pi_n| \rightarrow 0} \sum_{j=1}^n f(t_j^*) (W_{t_{j+1}} - W_{t_j}). \quad (2.7)$$

The integral is then defined as a weighted sum, where the weights themselves are random variables. Therefore, the integral produces

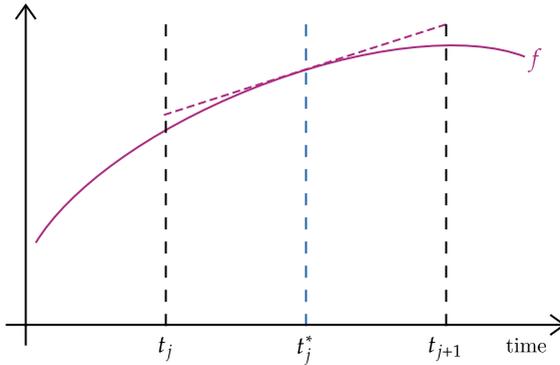


Figure 2.1: Discretization of a stochastic integral

a random variable. In contrast to Riemann integrals, for stochastic integrals the positions in a given interval at which we choose to evaluate the function f impact the value of the integral.

Remark: (Insert title)

To see this, consider taking the left end point, $t_j^* = t_j$, and the right end point $t_j^* = t_{j+1}$. For $f(t) \equiv W_t$ we find in the first

case

$$\left\langle \sum_{j=1}^n W_{t_j} (W_{t_{j+1}} - W_{t_j}) \right\rangle = \sum_{j=1}^n \langle W_{t_j} (W_{t_{j+1}} - W_{t_j}) \rangle \quad (2.8)$$

$$= \sum_{j=1}^n \langle W_{t_j} \rangle \langle W_{t_{j+1}} - W_{t_j} \rangle \quad (2.9)$$

$$= 0. \quad (2.10)$$

However, in the second case we have

$$\sum_{j=1}^n W_{t_{j+1}} (W_{t_{j+1}} - W_{t_j}) = \sum_{j=1}^n \langle (W_{t_{j+1}} - W_{t_j})^2 \rangle \quad (2.11)$$

$$= \sum_{j=1}^n (t_{j+1} - t_j) \quad (2.12)$$

$$= t - t_0, \quad (2.13)$$

where we used that we can subtract the expression from the first case, since it is equal to zero.

There are two common choices for the discretization of stochastic integrals:

1. The *Itô integral*: $t_j^* = t_j$, i.e. the left end point, and
2. the *Stratonovich integral*: $t_j^* = \frac{1}{2}(t_{j+1} - t_j)$, i.e. the mid

point).

Throughout this lecture we will use Itô's interpretation of the stochastic integral.

In addition to the integral, also the chain rule cannot be straightforwardly carried over from conventional calculus. For a given stochastic process, $x(t)$, what is the stochastic differential equation describing $f(x(t))$? To find out, we Taylor expand $f(x(t))$ and obtain

$$df(x(t)) \stackrel{\text{Taylor}}{=} \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dx + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} dx^2 + \dots \quad (2.14)$$

$$= \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dx + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (a^2 dt^2 + 2abdtdW + b^2 dW^2) + \dots \quad (2.15)$$

$$= \left(\frac{\partial f}{\partial t} + a \frac{\partial f}{\partial x} + \frac{b^2}{2} \frac{\partial^2 f}{\partial x^2} \right) dt + b \frac{\partial f}{\partial x} dW + \dots \quad (2.16)$$

In the second line we substituted the expression for dx , and in the last equality we used that $dx(t) \propto \sqrt{dt}$.

Remark:

The scaling $dx(t) \propto \sqrt{dt}$ means that the differential increments scale with the square root of the time increments. This follows from the fact that the typical distance covered by a Brownian motion, the standard deviation of its displacements, increases

with the square root of time.

Example: Geometric Brownian motion

Under healthy conditions, loss of cells in a tissue must be balanced by proliferation of other cells. Mutations on the DNA can, however, break this balance, such that mutated cells have a proliferative advantage over healthy cells. Deterministically, the concentration x of cancer cells in the tissue follows

$$\dot{x} = ax. \quad (2.17)$$

The rate of proliferation, a , is itself subject to many influences, such as metabolic states or biochemical signals from the environment,

$$a = \mu + \sigma\xi(t). \quad (2.18)$$

We can therefore write the time evolution of the concentration of tumor cells as an SDE, which in differential form reads

$$dx = \mu x dt + \sigma x dW(t). \quad (2.19)$$

After dividing by x we find

$$\frac{dx}{x} = \mu dt + \sigma dW(t), \quad (2.20)$$

to which we apply Itô's formula for $f(x) = \log x$,

$$dy = d(\log x) = \left[\mu - \frac{\sigma^2}{2} \right] dt + \sigma dW \quad (2.21)$$

$$= \frac{dx}{x} - \frac{\sigma^2}{2} dt, \quad (2.22)$$

where for the second equality we substituted the expression for dx/x . Therefore,

$$\frac{dx}{x} = d(\log x) + \frac{1}{2}\sigma^2 dt. \quad (2.23)$$

Integrating and using that $\int_0^t dx/x dx = -\mu t - \sigma W(t)$, we get

$$\log \frac{x}{x_0} = \int_0^t \frac{dx}{x} - \frac{1}{2}\sigma^2 t = \mu t + \sigma W(t) - \frac{1}{2}\sigma^2 t, \quad (2.24)$$

and it finally follows that

$$x(t) = x_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}. \quad (2.25)$$

In other words, the solution is the exponential of a Wiener process with drift μ

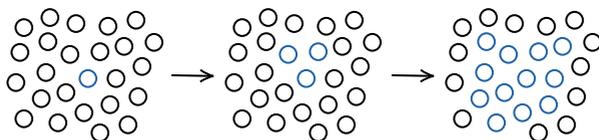


Figure 2.2: The evolution of the number of tumor cells (blue) in healthy cell tissue (black) can be described by a geometric Brownian motion stochastic process.

2.1.2 Einstein approach

We now take a complementary point of view and ask how the probability density function $p(x, t)$ changes over time. More specifically, we seek to derive differential equations for the probability $p(x, t | x_0, t_0)$ of ending up in state x at time t when starting at x_0 at time t_0 . The stochastic process $X(t)$ can take multiple paths from x_0 and x , so that $p(x, t | x_0, t_0)$ is obtained by summing up contributions of all possible paths. The mathematical equivalent of this statement is called the Chapman-Kolmogorov equation,

$$p(x, t | x_0, t_0) = \int dx' p(x, t | x', t') p(x', t' | x_0, t_0) \quad (2.26)$$

The goal is now to derive an equation for the time evolution of $p(x, t)$, i.e. an equation of the form $\frac{d}{dt}p(x, t) = \dots$. To do this, consider the case $t' \rightarrow t$. The gain of probability in state x is

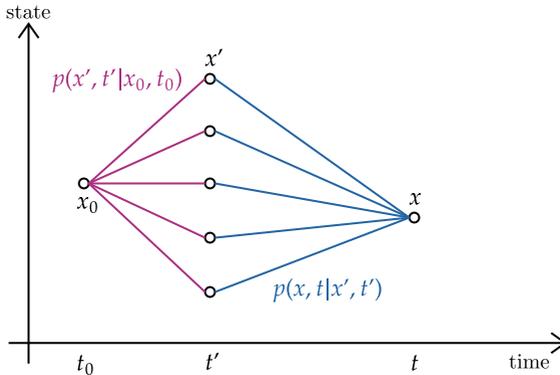


Figure 2.3: Multiple stochastic paths from x_0 to x are possible, such that the total probability is given by the sum over all these paths.

counteracted by a loss of probability to other states. We then obtain the so-called Master equation,

$$\frac{d}{dt}p(x, t) = \underbrace{\int dx' w(x|x')p(x', t)}_{\text{flux into state } x} + \underbrace{\int dx'' w(x''|x)p(x, t)}_{\text{flux out of state } x} \quad . \quad (2.27)$$

Example: Gene expression

Certain parts of the DNA called genes are read out by special molecules (polymerases) and transcribed into mRNA molecules. These mRNA molecules are then translated to proteins, which perform a biological function. What is the probability that n

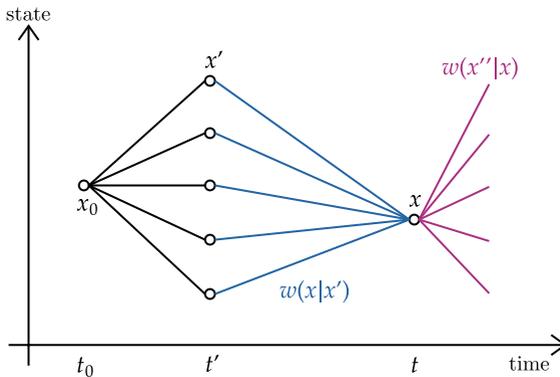


Figure 2.4: The gain of probability to be in state x at time t coming from state x' is counteracted by a loss of probability to other states.

such mRNA molecules are produced in a time interval t ? To answer this question we make two assumptions, namely that the production events are statistically independent and that initially there are 0 molecules. The Master equation for this process reads:

$$\frac{d}{dt}p(n, t) = \lambda [p(n - 1, t) - p(n, t)]. \quad (2.28)$$

We will solve this equation by introducing a *characteristic function*

$$G(s, t) = \langle e^{ins} \rangle = \sum_n p(n, t) e^{ins}. \quad (2.29)$$

After substituting this into the Master equation we obtain

$$\partial_t G(s, t) = \lambda [e^{is} - 1] G(s, t), \quad (2.30)$$

which is solved by

$$G(s, t) = e^{\lambda t [e^{is} - 1]}. \quad (2.31)$$

This characteristic function can be transformed back, yielding a Poisson distribution,

$$p(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (2.32)$$

Remark: Detailed balance

In a closed and isolated system, the dynamics will reach a steady state that is typically described by a Boltzmann distribution,

$$0 = \int dx' w(x|x') p(x', t) - \int dx'' w(x|x'') p(x, t). \quad (2.33)$$

Using time reversal symmetry of the microscopic dynamics it can be shown that in equilibrium this equation even holds for every

element of the sums independently,

$$\frac{w(x | x')}{w(x' | x)} = \frac{p_{\text{eq}}(x)}{p_{\text{eq}}(x')}. \quad (2.34)$$

Systems for which this holds are said to obey *detailed balance*.

In most cases, the Master equation is not directly solvable. However, if the transition kernel $w(x | x')$ decays sufficiently fast in $|x - x'|$ and the distribution $p(x, t)$ is smooth on the scale of typical jump sizes, we can Taylor expand the Master equation in $x - x'$. Setting $\Delta x \equiv x - x'$ we can write

$$\partial_t p(x, t) = \int d\Delta x [w(x | x - \Delta x)p(x - \Delta x, t) - w(x + \Delta x | x)p(x, t)]. \quad (2.35)$$

Now note that since $\int d\Delta x [w(x + \Delta x | x) - w(x - \Delta x | x)] = 0$, we can Taylor expand $w(x | x - \Delta x)p(x - \Delta x, t)$ around x , obtaining the *Kramers-Moyal expansion* of the Master equation,

$$\partial_t p(x, t) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \partial_x^n \alpha_n(x) p(x, t). \quad (2.36)$$

The expansion coefficients $\alpha_n(x)$ are the moments of the transition

probabilities,

$$\alpha_n(x) = \int d\Delta x w(x + \Delta x | x) \Delta x^n. \quad (2.37)$$

If we truncate this expansion at second order we obtain the *Fokker-Planck equation*

$$\partial_t p(x, t) = -\partial_x \alpha_1(x) p(x, t) + \frac{1}{2} \partial_x^2 \alpha_2(x) p(x, t). \quad (2.38)$$

The two terms on the right hand side are called the drift ($\partial_t \langle x \rangle = \langle \alpha_1 \rangle$) and diffusion terms, respectively. Usually, the Fokker-Planck equation is a good approximation for the Master equation but it fails in describing the tails of probability distributions.

Example: Poisson process

Introducing a rescaled variable $x \equiv n/N$ with $p(x, t) dx = p(n, t) dn$, we obtain a Fokker-Planck approximation to the Master equation of the Poisson process,

$$\partial_t p(x, t) = -\frac{\lambda}{N} \partial_x p(x, t) + \frac{\lambda}{2N^2} \partial_x^2 p(x, t), \quad (2.39)$$

which is solved by

$$p(x, t) = \frac{N}{\sqrt{2\pi\lambda t}} e^{-\frac{1}{2\lambda t}(Nx - \lambda t)^2}. \quad (2.40)$$

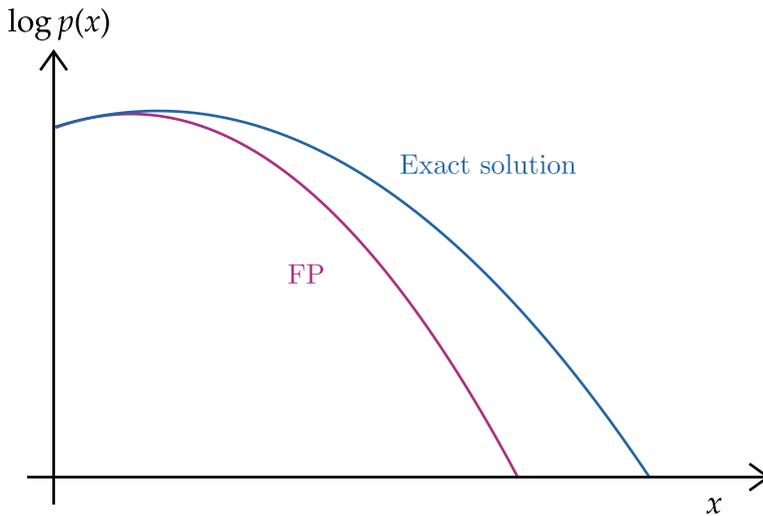


Figure 2.5: Comparison of the exact solution of the Poisson process (purple) and its approximation by the Fokker-Planck equation (blue).

Literature

1. C. Gardiner, *Stochastic Methods, A Handbook for the Natural and Social Sciences*. Springer Berlin, Heidelberg Cambridge University Press, Cambridge (2009)
2. M. Cates, J. Tailleur, *Motility-Induced Phase Separation*. Annual Review of Condensed Matter Physics 6 (1): 219–244 (2015).