Stochastic Models of Complex Systems

CO905 - MSc in Complexity Science

Stefan Grosskinsky

Warwick, 2013

These notes and other information about the course are available on
http://www.warwick.ac.uk/~masgav/teaching/co905.html

Contents

References

[GS] G. Grimmett, D. Stirzaker: Probability and Random Processes (3rd edition), Oxford 2001
[Ga] C.W. Gardiner: Handbook of Stochastic Methods (3rd edition), Springer 2004
[G] G. Grimmett: Probability on Graphs, CUP 2010
http://www.statslab.cam.ac.uk/~grg/books/pgs.html
[MFH] V. Méndez, S. Fedotov, W. Horsthemke: Reaction-Transport Systems, Springer 2010

[MP] P. Mörters, Y. Peres: Brownian Motion, CUP 2010
http://people.bath.ac.uk/maspm/

[K] T.G. Kurtz, Lectures on Stochastic Analysis
http://www.math.wisc.edu/~kurtz/m735.htm

Introduction

In this module we will cover the basics to study models of complex systems with stochastic time evolution. There are two different origins of stochasticity:

- Classical mechanics: stochasticity due to lack of information
  In principle all components involved in the system are believed to follow a deterministic system of equations of motion. But in practice all microscopic details are not accessible and the unknown influences on the dynamics are approximated as effective random noise with a certain postulated distribution. The actual origin of the noise may be related to chaotic motion, i.e. deterministic time evolution with random initial data such as a dice or pendulum, or neglected interactions in a large system such as gases or fluids leading to a stochastic time evolution.

- Quantum mechanics: inherent stochasticity
  Even simple systems can only be described stochastically and the full microscopic details are inherently inaccessible (uncertainty principle). Mathematically, the state of such a system is therefore given by a complex probability density function (wave function), rather than a single element in the set of all possible configurations.

Examples. (gas, branching process)

In this course we only cover classical stochastic systems. After a general introduction to stochastic processes we will study some examples of particle systems with thermal interactions. The first and most classical example of this phenomenon is *Brownian motion* (see Gardiner, Section 1.2). In 1827 Robert Brown observed the irregular motion of small pollen grains suspended in water. A first satisfactory theoretical description of this phenomenon was given by Einstein in 1905. A mathematically idealized version of this is called the Wiener process and can be described by the theory of stochastic calculus which was developed in the 1950s by Itô. Due to the continuous state space of the system this theory is rather involved, and will be discussed towards the end of the module. Simpler to analyse are models with a discrete state space such as birth-death processes, which appear for example in predator-prey models in biology (see Gardiner, Section 1.3). In the first part of the course we concentrate on *Markov chains* (following [GS] Chapter 6), which are certain stochastic processes with discrete state space. We conclude the introductory section by two general definitions.

Definition 0.1 A *stochastic process* $X = (X_t : t \in \mathbb{T})$ is a family of random variables $X_t : \Omega \to S$ with *state space* S and time index set $\mathbb{T} \subseteq \mathbb{R}$.

A stochastic process $X : \mathbb{T} \times \Omega \to S$ is a function of two variables, time t and $\omega \in \Omega$. For fixed ω, the function $t \mapsto X_t(\omega)$ is called a *sample path*. The probability space Ω is arbitrary, but has to be big enough to encode all possible time evolutions. A canonical choice is the set of all possible sample paths $\Omega = \{f : \mathbb{T} \to S\}$, or often one requires some regularity of the functions f, such as continuity.

Definition 0.2 A stochastic process is a *Markov process* if for all $t_1 < t_2 < \ldots < t_n \in \mathbb{T}$, $n \in \mathbb{N}$, for all $s_1, \ldots, s_{n-1} \in S$ and all (measurable) $A \subseteq S$,

$$\mathbb{P}\big(X_{t_n} \in A_n \big| X_{t_1} = s_1, \ldots, X_{t_{n-1}} = s_{n-1}\big) = \mathbb{P}\big(X_{t_n} \in A_n \big| X_{t_{n-1}} = s_{n-1}\big) \,. \tag{0.1}$$

A Markov process is called *homogeneous* if for all (measurable) $A, B \subseteq S$ and $t > t' \in \mathbb{T}$

$$\mathbb{P}\big(X_t \in A \,\big|\, X_{t'} \in B\big) = \mathbb{P}\big(X_{t-t'} \in A \,\big|\, X_0 \in B\big) \,. \tag{0.2}$$

A homogeneous Markov process is called a *Markov chain*, if S is discrete.

In this course we will only deal with homogeneous Markov processes. We will concentrate on the choices $\mathbb{T} = \mathbb{N}$, $\mathbb{Z}$ for discrete time and $\mathbb{T} = [0, \infty)$, $\mathbb{R}$ for continuous time processes. Typical choices for state spaces are $S = \mathbb{Z}$ (e.g. random walk, birth-death processes), $\mathbb{N}$ (e.g. counting processes), $\mathbb{R}^d$ (e.g. Brownian motion).

Examples. (RW, Brownian motion, deck of cards)

1 Markov chains

1.1 General properties

Definition 1.1 For a Markov chain we define the *transition probabilities*

$$p_{ij}(t) := \mathbb{P}(X_t = j | X_0 = i) \in [0,1] \quad \text{for all } i,j \in S\,, \tag{1.1}$$

and the *transition 'matrices'* (which might be infinite)

$$P(t) := \big(p_{ij}(t) \,:\, i,j \in S\big) \in [0,1]^{|S|\times|S|}\,. \tag{1.2}$$

A homogeneous Markov chain starting at time $t = 0$ is uniquely determined by an initial distribution $\boldsymbol{\pi}(0)$ with $\pi_i(0) = \mathbb{P}(X_0 = i)$, $i \in S$ and the transition probabilities, because every joint probability can be written as

$$\begin{aligned}\mathbb{P}\big(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\big) &= \\ &= \sum_{i_0\in S, i_1\in A_1, .., i_n \in A_n} \pi_{i_0}(0) p_{i_0 i_1}(t_1) p_{i_1 i_2}(t_2 - t_1) \cdots p_{i_{n-1} i_n}(t_n - t_{n-1})\end{aligned} \tag{1.3}$$

for all $0 < t_1 < \ldots < t_n \in \mathbb{T}$ and $A_1, \ldots, A_n \subseteq S$. In particular, the distribution at time t is

$$\pi_j(t) = \mathbb{P}(X_t = j) = \sum_{i\in S} \pi_i(0)\, p_{ij}(t)\,, \quad \text{so} \quad \boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, P(t)\,. \tag{1.4}$$

Example. (RW)

Proposition 1.1 *$P(0) = Id$ and the family $\big(P(t) \,:\, t \geq 0\big)$ satisfies the **Chapman-Kolmogorov equations**,*

$$P(t+t') = P(t)\, P(t') \quad \textit{for all } t, t', t+t' \in \mathbb{T}\,. \tag{1.5}$$

Proof. $p_{ij}(0) = \delta_{ij}$ by definition, and for all $t, t', t+t' \in \mathbb{T}$

$$\begin{aligned}p_{ij}(t+t') &= \mathbb{P}(X_{t+t'} = j | X_0 = i) = \\ &= \sum_{k\in S} \mathbb{P}(X_{t+t'} = j | X_0 = i, X_t = k)\, \mathbb{P}(X_t = k | X_0 = i) = \sum_{k\in S} p_{ik}(t)\, p_{kj}(t')\,,\end{aligned} \tag{1.6}$$

using the total probability sum rule, the Markov property and homogeneity. □

For discrete time with $\mathbb{T} = \mathbb{N}$ this leads to

$$\boxed{P(n+1) = P(1)\,P(n) = P(n)\,P(1) \quad \Rightarrow \quad P(n) = P^n\,,} \tag{1.7}$$

where we denote $P = P(1)$. Therefore a discrite time Markov chain is uniquely determined by the initial distribution $\boldsymbol{\pi}(0)$ and the transition matrix P, and in particular

$$\boxed{\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\,P^n\,.} \tag{1.8}$$

For a finite state chain with $|S| < \infty$ we then have

$$\boldsymbol{\pi}(n) = A_1 \mathbf{v}_1 \lambda_1^n + \ldots + A_{|S|}\mathbf{v}_{|S|}\lambda_{|S|}^n\,, \tag{1.9}$$

where the $\lambda_k \in \mathbb{C}$ are the eigenvalues with left eigenvectors $\mathbf{v}_k$ for P. The coefficients $A_k \in \mathbb{C}$ are determined by the initial condition $\boldsymbol{\pi}(0)$.

Example. (RW with various BCs)

For continuous time with $\mathbb{T} = [0, \infty)$ we require some regularity of the function $t \mapsto P(t)$ at $t = 0$. We only study processes where it is continuous and differentiable, i.e.

$$\lim_{t \searrow 0} P(t) = P(0) = Id \quad \text{and} \quad G := \lim_{t \searrow 0} \frac{P(t) - Id}{t} \quad \text{exists}\,, \tag{1.10}$$

so that $P(t) = Id + t\,G + o(t)$ for small t. Together with the Chapman-Kolmogorov equations this implies that

$$\frac{P(t+\Delta t) - P(t)}{\Delta t} = \frac{P(\Delta t) - Id}{\Delta t}\,P(t) = P(t)\frac{P(\Delta t) - Id}{\Delta t}\,, \tag{1.11}$$

and thus taking $\Delta t \searrow 0$, $P(t)$ is differentiable for all $t \geq 0$ and fulfills

$$\boxed{\frac{d}{dt}P(t) = G\,P(t) = P(t)\,G \quad \Rightarrow \quad P(t) = \exp(t\,G)\,.} \tag{1.12}$$

These are called *backward* and *forward equation*, respectively. Subject to the boundary conditions $P(0) = Id$, they often have a unique solution given by the matrix exponential

$$P(t) = \exp(t\,G) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\, G^k = Id + t\,G + \frac{t^2}{2}\, G^2 + \dots \ . \tag{1.13}$$

For example this is the case if $|S| < \infty$, and subject to certain technical conditions also for inifite state space S. Therefore a continuous-time Markov chain is uniquely determined by the initial distribution and the matrix G which is called the *generator* of the process.
The distribution at time t is then given by

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\ \exp(t\,G)\ . \tag{1.14}$$

Again, for a finite state chain we then have

$$\boldsymbol{\pi}(t) = A_1 \mathbf{v}_1 e^{\lambda_1 t} + \ldots + A_{|S|} \mathbf{v}_{|S|} e^{\lambda_{|S|} t}\ , \tag{1.15}$$

where the $\lambda_k \in \mathbb{C}$ are now the eigenvalues of G with left eigenvectors $\mathbf{v}_k$, and the coefficients $A_k \in \mathbb{C}$ are determined by the initial condition $\boldsymbol{\pi}(0)$.

What is the meaning of the entries of G and how do the sample paths of X look?

Entries of G. Assume that $X_t = i$. For small times Δt we have from (1.13)

$$p_{ij}(\Delta t) = g_{ij}\Delta t + o(\Delta t) \quad \text{for all } i \neq j \in S\ . \tag{1.16}$$

So during a time interval $(t, t+\Delta t)$ the chain jumps from state i to j with probability $g_{ij}\Delta t$, and $g_{ij} \geq 0$ can be interpreted as a jump rate. On the diagonal we have

$$p_{ii}(\Delta t) = 1 + g_{ii}\Delta t + o(\Delta t) \quad \text{for all } i \in S\ , \tag{1.17}$$

which gives the probability that nothing happens in the time interval $(t, t+\Delta t)$. By normalization we have

$$1 = \sum_{j\in S} p_{ij}(\Delta t) = 1 + \Delta t \sum_{j\in S} g_{ij} \quad \Rightarrow \quad \sum_{j\in S} g_{ij} = 0 \quad \text{for all } i \in S\ . \tag{1.18}$$

Therefore the diagonal entries of G are

$$g_{ii} = -\sum_{j\neq i} g_{ij} \leq 0 \quad \text{for all } I \in S\ , \tag{1.19}$$

and $|g_{ii}| \geq 0$ can be interpreted as the total rate to leave state i.

Sample paths. Assume that $X_0 = i$ and define the *holding time*

$$W_i := \inf\left\{t \geq 0\ :\ X_t \neq i\right\}\ , \tag{1.20}$$

i.e. the (random) time until a jump occurs. This is actually independent of t by homogeneity. If i is absorbing, $g_{ij} = 0$ for all $j \in S$ and $W_i = \infty$.

Proposition 1.2 *The random variable W_i is exponentially distributed with parameter $|g_{ii}|$ and if $|g_{ii}| > 0$, the probability that the chain jumps to $j \neq i$ after time W_i is $g_{ij}/|g_{ii}|$.*

Proof. W_i has 'loss of memory' property, i.e. for all $s, u > 0$

$$\mathbb{P}(W_i > s+u | W_i > s) = \mathbb{P}(W_i > s+u | X_s = i) = \mathbb{P}(W_i > u) \,, \tag{1.21}$$

where we have used the Markov property and homogeneity. Therefore

$$\mathbb{P}(W_i > s+u) = \mathbb{P}(W_i > u)\, \mathbb{P}(W_i > s) \,. \tag{1.22}$$

Analogous to the Chapman-Kolmogorov equations (1.5) this can be used to derive a differential equation for $\bar{F}(s) = \mathbb{P}(W_i > s)$ which has an exponential solution

$$\bar{F}(s) = \mathbb{P}(W_i > s) = e^{\lambda s} \quad \text{where} \quad \lambda = \bar{F}'(0) \,. \tag{1.23}$$

Together with (1.17) we get

$$\bar{F}'(0) = \lim_{\Delta t \searrow 0} \frac{\mathbb{P}(W_i > \Delta t) - 1}{\Delta t} = \lim_{\Delta t \searrow 0} \frac{p_{ii}(\Delta t) + o(\Delta t) - 1}{\Delta t} = g_{ii} \leq 0 \,, \tag{1.24}$$

and therefore $\mathbb{P}(W_i > s) = e^{-|g_{ii}|s}$ and $W_i \sim Exp(|g_{ii}|)$.
Now the probability that the chain jumps to j, conditioned on the event that it actually jumps somewhere in the time interval $(t, t+\Delta t]$, is given by

$$\begin{aligned} \mathbb{P}(X_{t+\Delta t} = j | X_t = i, W_i < \Delta t) &= \frac{\mathbb{P}(X_{t+\Delta t} = j | X_t = i)}{\mathbb{P}(W_i < \Delta t | X_t = i)} \simeq \\ &\simeq \frac{p_{ij}(\Delta t)}{1 - p_{ii}(\Delta t)} \to \frac{g_{ij}}{-g_{ii}} \quad \text{as } \Delta t \searrow 0 \,. \end{aligned} \tag{1.25}$$

So conditioned on jumping at time t we get the required result. □

Picture of sample path.

The chain jumps at the *jump time* $J_n = \sum_{k=0}^{n-1} W_{Y_k}$ to state $Y_n = X_{J_n}$.

$Y = (Y_n : n \in \mathbb{N})$ is called the *jump chain*, and it is a discrete time Markov chain with transition Matrix P^Y given by

$$p_{ij}^Y = \begin{cases} 0 & , \, i = j \\ g_{ij}/|g_{ii}| & , \, i \neq j \end{cases} \quad \text{if } g_{ii} > 0 \,, \quad \text{and} \quad p_{ij}^Y = \delta_{ij} \quad \text{if } g_{ii} = 0 \,. \tag{1.26}$$

So a continuous-time Markov chain can also be characterized by its jump chain Y and a sequence of independent exponentially distributed holding times $(W_{Y_n} : n \in \mathbb{N})$.

Examples. (Poisson, CTRW)

Using the forward equation (1.12) we can also get an evolution equation for the distribution,

$$\frac{d}{dt}\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\,\frac{d}{dt}P(t) = \boldsymbol{\pi}(0)\,P(t)\,G = \boldsymbol{\pi}(t)\,G \,. \tag{1.27}$$

This is called the *Master equation* and using (1.19) the coordinate form is given by

$$\boxed{\frac{d}{dt}\pi_i(t) = \sum_{j \neq i} \Big(\pi_j(t)\, g_{ji} - \pi_i(t)\, g_{ij} \Big) \,.} \tag{1.28}$$

For the *Poisson process* there exists also another characterization.

Proposition 1.3 *$X = (X_t : t \geq 0)$ is a Poisson process with rate λ if and only if it has stationary, independent increments, i.e.*

$$X_{t+t'} - X_{t'} \text{ is distributed like } X_t - X_0 \text{ and independent of } (X_s : s \leq t') \,, \tag{1.29}$$

and for each t, X_t has Poisson distribution with parameter λt, i.e. $\mathbb{P}(X_t = k) = \frac{(\lambda t)^k}{k!}\, e^{-\lambda t}$.

Proof. (differentiate $\pi_n(t)$)

1.2 Stationary distributions and reversibility

Definition 1.2 A probability distribution $\boldsymbol{\pi}^*$ is called *stationary* if $\boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^*$ for all $t \geq 0$.

This will play an important role in the long-time behaviour of Markov chains, since 'often' $\boldsymbol{\pi}(t) \to \boldsymbol{\pi}^*$. How and when this is true will be seen later.

Proposition 1.4 *π^* is stationary if and only if*

$$\begin{aligned} \pi^* P &= \pi^* \qquad \textit{for a discrete-time chain with transition matrix } P\,, \\ \pi^* G &= \mathbf{0} \qquad \textit{for a continuous-time chain with generator } G\,. \end{aligned} \tag{1.30}$$

Proof. Assume finite state space S. For discrete time this follows directly from $P(n) = P^n$. For continuous time we have

$$\begin{aligned} \boldsymbol{\pi}^* G = (0,\ldots,0) \quad &\Rightarrow \quad \boldsymbol{\pi}^* G^k = (0,\ldots,0) \quad \text{for all } k \geq 1 \\ &\Rightarrow \quad \sum_{k=1}^{\infty} \frac{t^k}{k!} \boldsymbol{\pi}^* G^k = (0,\ldots,0) \quad \text{for all } t \geq 0 \\ &\Rightarrow \quad \boldsymbol{\pi}^* \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k = \boldsymbol{\pi}^* \quad \text{for all } t \geq 0 \text{ since } G^0 = Id \\ &\Rightarrow \quad \boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^* \quad \text{for all } t \geq 0\,. \end{aligned} \tag{1.31}$$

The backward direction follows directly from differentiating $\boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^*$ at $t = 0$. □

Theorem 1.5 (Existence)
A Markov chain with finite state space S has at least one stationary distribution.

Proof. Based on linear algebra (an extended version of the Perron-Frobenius Theorem):
Since P and G have row sum 1 and 0, respectively, we have $P\mathbf{1} = \mathbf{1}$ and $G\mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ is the column vector with all entries 1. So 1 and 0 are eigenvalues of P and G, respectively, and thus there exist also corresponding left eigenvectors. These can be shown to have non-negative entries and can be normalized to be a stationary distribution.

Definition 1.3 State $i \in S$ *communicates with* state $j \in S$ if $p_{ij}(t) > 0$ for some $t \in \mathbb{T}$, and we write $i \to j$. States i and j are *connected* if $i \to j$ and $j \to i$, and we write $i \leftrightarrow j$.
The Markov chain is called *irreducible* if $i \leftrightarrow j$ for all $i, j \in S$.
A state i is called *absorbing*, if $i \not\to j$ for all $j \neq i$.

Remark. The state space of a Markov chain can be decomposed into *communicating classes* S_k which are disjoint subsets of connected states such that $S = \cup_k S_k$. The chain is irreducible if there is only one such class $S_1 = S$.

Theorem 1.6 (Uniqueness)
An irreducible Markov chain has at most one stationary distribution.

Proof. Again based on linear algebra and Perron-Frobenius (see e.g. [GS], Section 6.4). The irreducible structure of the transition matrix P (discrete time) and the generator G (continuous time) can be shown to imply that the eigenvector to eigenvalue 1 and 0, respectively, are unique up to normalization.

Therefore, a finite state, irreducible Markov chain has a unique stationary distribution.

For inifinite state space S the chain can 'disappear at infinity' and there is no stationary distribution (see handout).

Examples. (RW with absorbing BC, on Z)

Definition 1.4 A Markov chain $(X_t : t \in \mathbb{T})$ is called *reverisble* w.r.t. $\boldsymbol{\pi}$ (or $\boldsymbol{\pi}$ is called *reverisble* for $(X_t : t \in \mathbb{T})$) if

$$\pi_i p_{ij}(t) = \pi_j p_{ji}(t) \quad \text{for all } t \in \mathbb{T} \quad \text{and} \quad i \neq j \in S\,. \tag{1.32}$$

Proposition 1.7 *If $\boldsymbol{\pi}$ is reversible then it is also stationary.*

Proof. For all $j \in S$ we have $\quad \big(\boldsymbol{\pi} P(t)\big)_j = \sum_{i \in S} \pi_i p_{ij}(t) = \sum_{i \in S} \pi_j p_{ji}(t) = \pi_j\,.$ □

Proposition 1.8 *$\boldsymbol{\pi}$ is reversible for the Markov chain $(X_t : t \in \mathbb{T})$ if and only if it fulfilles the **detailed balance** conditions*

$$\begin{aligned} &\textit{(discrete time)} && \pi_i p_{ij} = \pi_j p_{ji} \\ &\textit{(continuous time)} && \pi_i g_{ij} = \pi_j g j i \quad \textit{for all } i, j \in S\,. \end{aligned} \tag{1.33}$$

Proof. analogous to the proof of Prop. 1.4.

If $\boldsymbol{\pi}$ is reversible, the right-hand side of the master equation (1.28) vanishes term-wise.

Proposition 1.9 *Let $\big(X_t : t \in \begin{smallmatrix}\mathbb{Z}\\ \mathbb{R}\end{smallmatrix}\big)$ be a finite state irreducible Markov chain with transition matrix P^X (discrete time) or generator G^X (continuous time). Assume further that the chain is stationary (i.e. $X_t \sim \boldsymbol{\pi}^*$ for all t) which makes it possible to define it also for negative times. Then the **reversed chain** $Y = \big(Y_t : t \in \begin{smallmatrix}\mathbb{Z}\\ \mathbb{R}\end{smallmatrix}\big)$ with $Y_t = X_{-t}$ is a stationary Markov chain with*

$$\begin{aligned} &\textit{transition matrix} && p^Y_{ij} = \frac{\pi^*_j}{\pi^*_i}\, p^X_{ji} && \textit{(discrete time)} \\ &\textit{generator} && g^Y_{ij} = \frac{\pi^*_j}{\pi^*_i}\, g^X_{ji} && \textit{(continuous time)} \quad \textit{for all } i,j \in S\,. \end{aligned} \tag{1.34}$$

Proof. Using stationarity and the Markov property of X we get for discrete time

$$\begin{aligned} &\mathbb{P}\big(Y_{n+1} = i_{n+1} \,\big|\, Y_n = i_n, \dots, Y_0 = i_0\big) = \\ &\quad = \frac{\mathbb{P}(Y_k = i_k, 0 \le k \le n+1)}{\mathbb{P}(Y_k = i_k, 0 \le k \le n)} = \frac{\mathbb{P}(X_{N-k} = i_k, 0 \le k \le n+1)}{\mathbb{P}(X_{N-k} = i_k, 0 \le k \le n)} = \\ &\quad = \frac{\pi^*_{i_{n+1}} p_{i_{n+1} i_n} \cdots p_{i_1 i_0}}{\pi^*_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 i_0}} = \frac{\pi^*_{i_{n+1}} p_{i_{n+1} i_n}}{\pi^*_{i_n}} \end{aligned} \tag{1.35}$$

as required. Continuous time works analogously. □

Note that in general a time-reversed Markov chain is not necessarily a Markov chain, this only holds for stationary chains. Obviously, π^* is then also the stationary distribution for the reversed chain Y.

Prop. 1.9 together with (1.33) implies that a reversible Markov chain and its time-reversal are indistinguishable, i.e. $\big(Y_t : t \in \begin{smallmatrix}\mathbb{Z}\\ \mathbb{R}\end{smallmatrix}\big) \sim \big(X_t : t \in \begin{smallmatrix}\mathbb{Z}\\ \mathbb{R}\end{smallmatrix}\big)$ since they have the same transition probabilities

$$p^Y_{ij} = \frac{\pi^*_j}{\pi^*_i}\, p^X_{ji} = \frac{\pi^*_i}{\pi^*_i}\, p^X_{ij} = p^X_{ij}\,. \tag{1.36}$$

An analogous relation holds for rates in continuous time. The detailed balance relations (1.33) can be a useful tool to find stationary distributions. For certain Markov chains 'without loops', i.e. with a 'tree-like' structure of allowed transitions, every stationary distribution is also reversible.

Examples. (RW with PBC)

1.3 Ergodicity

Definition 1.5 Consider a Markov chain $(X_t : t \in \mathbb{T})$. Suppose that $X_0 = i \in S$ and define the *return time* (or *recurrence time*) T_i by

$$\begin{aligned} T_i &:= \min\{n \geq 1 : X_n = i\} \qquad \text{(discrete time)} \\ T_i &:= \inf\{t \geq J_1 : X_t = i\} \qquad \text{(continuous time)} . \end{aligned} \tag{1.37}$$

The *mean recurrence time* of state $i \in S$ is then $\mu_i = \mathbb{E}(T_i)$.

For continuous-time processes the condition of being larger than the first jump time ensures that T_i is non-zero and gives the time of first return to i after an excursion. Note that $T_i = \infty$ can happen with positive probability, e.g. if the chain gets stuck in an absorbing state and does not return to i. If $T_i < \infty$ with probability 1 still $\mu_i = \infty$ is possible, i.e. the distribution of T_i does have a heavy tail with infinite expectation.

Theorem 1.10 *For an irreducible Markov chain with finite state space $\mu_i \in (0, \infty)$ for all $i \in S$, and the unique stationary distribution $\boldsymbol{\pi}^*$ is given by*

$$\begin{aligned} \pi_i^* &= \frac{1}{\mu_i} \qquad \textit{(discrete time)} \\ \pi_i^* &= \frac{\mathbb{E}(W_i)}{\mu_i} = \frac{1}{\mu_i |g_{ii}|} \qquad \textit{(continuous time)} , \end{aligned} \tag{1.38}$$

where $W_i \sim Exp(|g_{ii}|)$ is the holding time of state i.

Proof. see e.g. [GS] pp 229 - 230.

In particular, this implies that $\pi_i^* > 0$ for all $i \in S$ and is determined by the average fraction of the time the chain spends in state i. For discrete-time processes the holding time can be thought of being 1 (one time step).

Picture of sample path.

Definition 1.6 A discrete time Markov chain is called *aperiodic* if for all $i \in S$, $p_{ii}(n)$ is eventually positive, i.e.

$$\text{there exists } N_i \in \mathbb{N} \text{ such that } p_{ii}(n) > 0 \text{ for all } n \geq N_i . \tag{1.39}$$

Remark. Note that for irreducible Markov chains, aperiodicity implies that for all $i, j \in S$, $p_{ij}(n)$ is eventually positive.

Example. (RW)

Note that for continuous time there is no issue of periodicity, since

$$\text{if } i \to j \quad \text{then} \quad p_{ij}(t) > 0 \quad \text{for all } t > 0\,. \tag{1.40}$$

This is because $i \to j$ is equivalent to

$$g_{ii_1} g_{i_1 i_2} \cdots g_{i_{n-1} j} > 0 \quad \text{for some } i_1, \ldots i_{n-1} \in S, n \in \mathbb{N}\,, \tag{1.41}$$

which implies that $\quad p_{ij}(t) \geq p_{ii_1}(t/n) \cdots p_{i_{n-1}j}(t/n) > 0\,.$

Theorem 1.11 *An irreducible (aperiodic) Markov chain with finite state space is* ***ergodic****, i.e. it has a unique stationary distribution* $\boldsymbol{\pi}^*$ *and*

$$p_{ij}(t) = \mathbb{P}(X_t = j \,|\, X_0 = i) \to \pi^*_j \quad \textit{as } t \to \infty\,, \quad \textit{for all } i, j \in S\,. \tag{1.42}$$

Proof. For discrete time this follows from the **Perron-Frobenius Theorem**:
If P is the transition matrix of a finite state, aperiodic, irreducible Markov chain then
(i) $\lambda_1 = 1$ is a single eigenvalue of P
(ii) and the remaining (complex) eigenvalues $\lambda_2, \ldots, \lambda_{|S|}$ satisfy $|\lambda_j| < 1$.
Note that (i) includes uniqueness of the stationary distribution claimed in Theorem 1.6.
Suppose further that all the eigenvalues are distinct, then P can be diagonalized, i.e.

$$B\,P\,B^{-1} = \Lambda = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_{|S|} \end{pmatrix} \tag{1.43}$$

where the rows of B are the left and the columns of B^{-1} are the right eigenvectors of P, normalized such that $B\,B^{-1} = Id$. Thus

$$P^n = \left(B^{-1}\Lambda B\right)^n = B^{-1} \begin{pmatrix} \lambda_1^n & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_{|S|}^n \end{pmatrix} B \to B^{-1} \begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix} B \tag{1.44}$$

as $n \to \infty$, since $\lambda_1 = 1$ and $|\lambda_i| < 1$ for all $i > 1$. Since the first column of B^{-1} is $\mathbf{1}$ (right eigenvector to $\lambda_1 = 1$), the right-hand side is equal to the matrix

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix} B = \begin{pmatrix} \pi^*_1 \dots \pi^*_{|S|} \\ \vdots \\ \pi^*_1 \dots \pi^*_{|S|} \end{pmatrix} \quad \text{which implies the statement.} \tag{1.45}$$

The proof can be extended to more general cases and works similar for continuous time by using the corresponding jump chain. □

Remark. Theorem 1.11 implies that for every initial distribution $\boldsymbol{\pi}(0)$,

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, P(t) \to \boldsymbol{\pi}^* \quad \text{as } t \to \infty\,. \tag{1.46}$$

In general, the distribution at time t is the solution to a linear first order equation. Therefore it is given by a linear combination of left eigenvectors $\mathbf{v}_i$ of the transition matrix P (discrete time) or the generator G (continuous time) of the form (1.9) and (1.15). The largest eigenvalue is $\lambda_1 = 1$ for discrete and $\lambda_1 = 0$ for continuous time, with left eigenvector $\mathbf{v}_1 = \boldsymbol{\pi}^*$ and coefficient $A_1 = 1$.

Example. (RW on finite interval)

Theorem 1.12 (Ergodic Theorem)
Let $X = (X_t, t \in \mathbb{T})$ be an ergodic Markov chain with unique stationary distribution π^*. Then for every observable $f : S \to \mathbb{R}$

$$\frac{1}{t}\int_0^t f(X_s)\, ds \quad \textit{or} \quad \frac{1}{N}\sum_{n=0}^{N} f(X_n) \quad \longrightarrow \quad \mathbb{E}_{\pi^*}(f) \quad \textit{as } t, N \to \infty\,. \tag{1.47}$$

Proof. see e.g. [GS], chapter 9.5

So stationary expectations can be approximated by time averages over long periods. This is the basis for Markov chain Monte Carlo (MCMC) which is used to sample from the stationary distribution of a chain. In particular, using indicator functions $f(X_t) = \delta_{X_t,i}$ the right-hand side of (1.47) is equal to π^*_i. To improve the speed of convergence in practice, the chain is run for a while before starting to sample (equilibration).

Further remarks on periodicity
The *period* $d(i)$ of a state $i \in S$ is defined as

$$d(i) := gcd\big\{t \geq 1 \,:\, p_{ii}(t) > 0\big\}\,, \tag{1.48}$$

the greatest common divisor of the epochs at which return is possible.
For an irreducible Markov chain all states have the same period (for a proof see p.224 in [GS]). If the chain is aperiodic we have $d(i) = 1$ for all $i \in S$. Note, however, that the requirement $d = 1$ is weaker than Def. 1.6 and is often used as an alternative definition of aperiodicity.
Ergodicity follows again from a more general version of the **Perron-Frobenius Theorem**:
If P is the transition matrix of a finite state irreducible Markov chain with period d then
(i) the d complex roots of unity are eigenvalues of P,

$$\lambda_1 = w^0 = 1,\ \lambda_2 = w^1, \ldots, \lambda_d = w^{d-1} \quad \textit{where} \quad w = e^{2\pi i/d}\,, \tag{1.49}$$

(ii) and the remaining eigenvalues $\lambda_{d+1}, \ldots, \lambda_{|S|}$ satisfy $|\lambda_j| < 1$.

1.4 Countably infinite state spaces

For infinite state space S, the Markov chain can 'get lost at infinity', and therefore not have a stationary probability distribution.
Let $X_0 = i$ and T_i be the time of first return to state i as defined in Def. 1.5 for continuous and discrete time.

Definition 1.7 A state $i \in S$ is called

$$\begin{aligned}
\textit{transient}, \text{ if} \quad & \mathbb{P}(T_i = \infty) > 0\,, \\
\textit{null recurrent}, \text{ if} \quad & \mathbb{P}(T_i < \infty) = 1 \quad \text{and} \quad \mathbb{E}(T_i) = \infty\,, \\
\textit{positiv recurrent}, \text{ if} \quad & \mathbb{P}(T_i < \infty) = 1 \quad \text{and} \quad \mathbb{E}(T_i) < \infty\,.
\end{aligned} \tag{1.50}$$

Theorem 1.13 *Let X be an irreducible Markov chain. Then all states are either transient, null recurrent or positive recurrent. X has a unique stationary distribution if and only if it is positive recurrent. In this case X is also ergodic (subject to aperiodicity in case of discrete time).*

Proof. see Section 6.2 in [GS]

Examples. (BD process)

So the positive recurrent Markov chains behave like chains with finite state space concerning their stationary distributions and long time behaviour. Transient continuous time chains can get lost at infinity even in finite time. This phenomenon is called *explosion*. Define the *explosion time*

$$J_\infty := \lim_{n\to\infty} J_n = \sum_{i=1}^{\infty} W_{Y_i} \in (0,\infty] \,, \quad \text{where} \quad Y_i = X_{J_i} \quad \text{is the jump chain.} \tag{1.51}$$

This is a random variable that usually takes the value ∞, and we say that the chain is *non-explosive* if $\mathbb{P}(J_\infty = \infty) = 1$. For example this is the case if $|S| < \infty$ or $\sup_{i\in S} |g_{ii}| < \infty$.

Example. (birth chain)

Theorem 1.14 Polya's Theorem.
The simple random walk (in discrete or continuous time) on $S = \mathbb{Z}^d$ is null recurrent for $d = 1, 2$ and transient for $d \geq 3$.

Proof. The simplest proof follows an interesting connection between Markov chains and electrical network theory. Consider a continuous time SRW on $\mathbb{Z}^d$ with rate q across each bond (discrete time works analogously). The *resistence* of a bond is then given by $r = 1/q$, and *Kirchhoff's laws* apply, i.e. the total resistence r is given by

$$r = r_1 + r_2 \qquad \text{for resistences in series (sums of waiting times)} \tag{1.52}$$

$$1/r = 1/r_1 + 1/r_2 \qquad \text{for resistences in parallel (sums of rates) .} \tag{1.53}$$

Denote by $R(n)$ the total resistence between the origin and the set $\{x \in \mathbb{Z}^d : \|x\|_1 = n\}$, and by $R = \lim_{n\to\infty} R(N)$ the resistence to infinity. It can be shown that the SRW is transient if and only if $R < \infty$, and a rough estimates gives

$$R(n) \approx \left(n^{d-1}\, \tfrac{1}{n}\right)^{-1} = n^{2-d} \tag{1.54}$$

counting n^{d-1} parallel paths of length n to reach distance n from the origin. This implies $R = \infty$ for $d = 1$ and $R = 0$ for $d \geq 3$, for $d = 2$ a more careful analysis reveals logarithmic corrections that lead also to $R = \infty$.
For more details and infinite expected return times see e.g. [G]. □

2 Stochastic particle systems

2.1 Basic examples

Stochastic particle systems, or often also called interacting particle systems (IPS), are Markov chains where the state space has a particular structure. Let $\Lambda = \{1, \dots, L\}$ be a finite set of L points, which we call *lattice*. The state space is given by the set of all configurations

$$\eta = \big(\eta(x) : x \in \Lambda\big) \in S = \{0,1\}^L \quad \big(\text{often also written as } \{0,1\}^\Lambda\big) \,, \tag{2.1}$$

where $\eta(x) = 1$ is interpreted as the presence of a particle or an infection at site x. The dynamics of the IPS we consider is given by local continuous-time transitions of the following two kinds:

$$\begin{aligned} \eta &\to \eta^x \quad \text{with rate} \quad c(\eta, \eta^x) \,, \\ \eta &\to \eta^{xy} \quad \text{with rate} \quad c(\eta, \eta^{xy}) \end{aligned} \tag{2.2}$$

for all $x, y \in \Lambda$, where we use the shorthand notation

$$\eta^x(z) = \begin{cases} \eta(z) & , \, z \neq x \\ 1 - \eta(x) & , \, z = x \end{cases} \qquad \text{and} \qquad \eta^{xy}(z) = \begin{cases} \eta(z) & , \, z \neq x, y \\ \eta(y) & , \, z = x \\ \eta(x) & , \, z = y \end{cases} \,, \tag{2.3}$$

so that η^x corresponds to *creation/annihilation* of a particle at site x and η^{xy} to *motion* of a particle between x and y. In the following $q_{x,y} \geq 0$ will be transition rates of a continuous time random walk on Λ, and (Λ, Q) can be interpreted as a graph with adjacency matrix $Q = (q_{x,y} : x, y \in \Lambda)$. To avoid degeneracies we will always assume that

$$\textbf{the walk on } \Lambda \textbf{ with rates } q_{x,y} \textbf{ is irreducible} \,, \tag{2.4}$$

so that particles/infections can reach all parts of the lattice.

Definition 2.1 Let $q_{x,y} \in \{0, \lambda\}$ for all $x, y \in \Lambda$ with *infection rate* $\lambda > 0$, and write $x \sim y$ if $q_{x,y} = \lambda$. The *contact process* (CP) is an IPS with rates

$$c(\eta, \eta^x) = \eta(x) + \lambda\big(1 - \eta(x)\big) \sum_{y \sim x} \eta(y) \quad \text{for all } x \in \Lambda \,. \tag{2.5}$$

Sites with $\eta(x) = 1$ are interpreted as being infected and recover independently at rate 1, and healthy sites with $\eta(x) = 0$ get infected by each of their neighbours independently at rate λ. A short way of writing this is

$$1 \xrightarrow{1} 0 \quad \text{and} \quad 0 \xrightarrow{\lambda \# \mathrm{IN}} 1 \,. \tag{2.6}$$

Properties.

- The CP has one absorbing state, $\eta = \underline{0} = (0, \dots, 0)$ (all healthy), and is not irreducible. For finite Λ, $\underline{0}$ can be reached from all $\eta \in S$ and therefore $\boldsymbol{\pi}^* = \delta_{\underline{0}}$ is the unique stationary distribution and

  $$\boldsymbol{\pi}(t) \xrightarrow{t \to \infty} \delta_{\underline{0}} \quad \text{for all } \lambda \geq 0 \text{ and all initial conditions } \boldsymbol{\pi}(0) \,. \tag{2.7}$$

  Therefore, the CP on finite lattices is ergodic and the infection eventually dies out.

- For infinite Λ (e.g. $\Lambda = \mathbb{Z}^d$), it can be shown that there exists another stationary distribution $\boldsymbol{\pi}^*$ with $\rho = \mathbb{E}_{\boldsymbol{\pi}^*}\big(\eta(x)\big) > 0$ where the infection persists, provided that the infection rate λ is greater than a critical value $\lambda_c \in (0, \infty)$. The loss of ergodicity in infinite system limits is also called *ergodicity breaking*. Depending on the lattice structure there can be further different cases, for details see e.g. [G].

- Let T be the time of extinction of the epidemic on a finite lattice of size L. It can be shown that, starting from a positive density of infected sites, as $L \to \infty$

$$\begin{aligned} \mathbb{E}(T) &\propto \log L \quad , \quad \text{for } \lambda < \lambda_c \, , \\ \mathbb{E}(T) &\propto L^{\alpha} \quad , \quad \text{for } \lambda = \lambda_c \quad \text{with some } \alpha > 0 \, , \\ \mathbb{E}(T) &\propto e^{CL} \quad , \quad \text{for } \lambda > \lambda_c \quad \text{with some } C > 0 \, . \end{aligned} \tag{2.8}$$

  Furthermore, for $\lambda > \lambda_c$ one has convergence of $T/\mathbb{E}(T) \to Exp(1)$ to an exponential random variable in the limit $L \to \infty$. This phenomenon is known as *metastability*, i.e. the process converges to a metastable distribution and is apparently in equilibrium before it reaches the absorbing state $\underline{0}$ by fluctuations on a much longer timescale. In general, the transitions between metastable states are Markovian in the limit of large systems, enabling an effective description on a highly reduced state space.

Graphical construction.

Definition 2.2 The *(linear) voter model* (VM) is an IPS with rates

$$c(\eta,\eta^x) = \sum_{y\sim x} q_{y,x}\Big(\eta(x)\big(1-\eta(y)\big) + \big(1-\eta(x)\big)\eta(y)\Big) \quad \text{for all } x\in\Lambda\,. \tag{2.9}$$

$\eta(x) = 0,1$ are interpreted as two opinions, and x adopts the opinion of y independently at rate $q_{x,y}$, in short

$$1_x 0_y \overset{q_{x,y}}{\longrightarrow} 0_x 0_y \quad \text{and} \quad 0_x 1_y \overset{q_{x,y}}{\longrightarrow} 1_x 1_y\,. \tag{2.10}$$

Properties.

- The linear VM is symmetric under relabelling opinions $0 \leftrightarrow 1$ and is *dual* to a system of coalescing random walkers, which can be seen from the graphical construction in reversed time (see below).
- The VM is not ergodic and has two absorbing states $\eta = \underline{0}, \underline{1}$ (all of the same opinion). It is therefore also not ergodic (even on finite lattices), and

  $$\boldsymbol{\pi}(t) \overset{t\to\infty}{\longrightarrow} \alpha\delta_{\underline{0}} + (1-\alpha)\delta_{\underline{1}} \quad \text{where} \quad \alpha\in[0,1] \quad \text{depends on the initial condition.} \tag{2.11}$$

  Eventually only one opinion will survive, and the r.h.s. denotes all stationary distributions of the finite process.
- For inifinite Λ there may be other stationary distributions $\boldsymbol{\pi}^*$ on $\{0,1\}^\Lambda$ with $\rho = \mathbb{E}_{\boldsymbol{\pi}^*}\big(\eta(x)\big) \in (0,1)$ where both opinions persist.
- There are many generalizations of the linear VM, including non-linear majority rules or models with more than two opinions such as the Axelrod model.

Graphical construction.

Definition 2.3 The *exclusion process* (EP) is an IPS with rates

$$c(\eta, \eta^{xy}) = q_{x,y}\eta(x)\big(1-\eta(y)\big) \quad \text{for all } x, y \in \Lambda\ . \tag{2.12}$$

So particles $(\eta(x) = 1)$ jump to empty sites $(\eta(y) = 0)$ independently with rate $q_{x,y}$, in short

$$1_x 0_y \xrightarrow{q_{x,y}} 0_x 1_y\ . \tag{2.13}$$

For $\Lambda \subseteq \mathbb{Z}^d$ and $q_{x,y} \equiv q > 0$ only if $x \sim y$ (spatially homogeneous nearest neighbour jumps) the EP is called *simple* (SEP). If in addition $q_{x,y} = q_{y,x}$ for all $x, y \in \Lambda$ it is called *symmetric* (SSEP), otherwise *asymmetric* (ASEP). If $d = 1$ and $q_{x,y} = q\delta_{x+1,y}$ it is called *totally asymmetric* (TASEP).

Properties.

- The EP conserves the number of particles and is therefore not irreducible on the state space $S = \{0,1\}^\Lambda$. On the subspaces

  $$S_N = \Big\{\eta : \sum_{x\in\Lambda} \eta(x) = N\Big\} \quad \text{for } N = 0, \dots, L \tag{2.14}$$

  the EP is irreducible and ergodic with a unique stationary distribution $\boldsymbol{\pi}^*_{L,N}$.
  For the SEP it can be shown that $\boldsymbol{\pi}^*_{L,N} = 1/\binom{L}{N}$ is uniform on all possible configurations.

- Of particular importance for systems with a conserved quantity is the corresponding stationary current

  $$j_{x,y} := \mathbb{E}_{\boldsymbol{\pi}^*_{L,N}}\big(c(\eta,\eta^{xy})\big) = q_{x,y}\,\mathbb{E}_{\boldsymbol{\pi}^*_{L,N}}\big(\eta(x)(1-\eta(y))\big)\ . \tag{2.15}$$

  For the SEP we have

  $$j_{x,y} = q\,\frac{N}{L}\Big(1 - \frac{N-1}{L-1}\Big) \to qz\,\rho(1-\rho) \tag{2.16}$$

  in the *thermodynamic limit* $L, N \to \infty$, such that $N/L \to \rho \in [0,1]$.
  In general, in this limit stationary probabilities and correlation functions factorize, i.e.

  $$\boldsymbol{\pi}^*_{L,N}\Big(\eta(x_1)\cdots\eta(x_n) = 1\Big) = \frac{N}{L}\cdots\frac{N-n}{L-n} \to \rho^n \quad \text{for all fixed } n \in \mathbb{N}\ . \tag{2.17}$$

  So $\boldsymbol{\pi}^*_{L,N}$ converges locally to the *product measure* ν_ρ on $\{0,1\}^{\mathbb{Z}^d}$ under which the $\eta(x)$ are iid $Be(\rho)$ random variables.

- The ASEP is one of the most studied so-called *driven diffusive systems* in non-equilibrium statistical mechanics and has various connections to many other models, such as surface growth, directed percolation or random polymer models.

Graphical construction.

General properties of the time evolution of IPS.
Let us focus on IPS with flip dynamics such as CP and VM, transport systems like the EP can be treated analogously. Due to the local dynamics, the master equation of an IPS with state space $S = \{0,1\}^\Lambda$ can be written as

$$\frac{d}{dt}\pi_\eta(t) = \sum_{x\in\Lambda} \Big(\pi_{\eta^x}(t)\, c(\eta^x,\eta) - \pi_\eta(t)\, c(\eta,\eta^x) \Big) \quad \text{for all } \eta \in S\,. \tag{2.18}$$

This is the coordinate form of the general vector equation $\frac{d}{dt}\boldsymbol{\pi}(t) = \boldsymbol{\pi}(t)G$, and the r.h.s. provides a compact notation for the generator G with intuitive gain and loss terms. Due to the fact that only one site can flip at a time, most of the entries of G are 0.

To get an equation for the time evolution of an observable $f : S \to \mathbb{R}$, we note that

$$\mathbb{E}_{\boldsymbol{\pi}(t)}(f) = \sum_{\eta\in S} \pi_\eta(t)\, f(\eta) = \boldsymbol{\pi}(t) \circ f \tag{2.19}$$

can be written as a scalar product. Using the master equation we get

$$\frac{d}{dt}\mathbb{E}_{\boldsymbol{\pi}(t)}(f) = \big(\boldsymbol{\pi}(t)G\big) \circ f = \boldsymbol{\pi}(t) \circ (G\, f) = \mathbb{E}_{\boldsymbol{\pi}(t)}(G\, f)\,. \tag{2.20}$$

When acting on the observable f (column vector) rather than the distribution $\boldsymbol{\pi}(t)$ (row vector) one often also writes $\mathcal{L}$ for the generator of the process as an operator on observables. The action on f leads to another function $\mathcal{L}f$ (column vector) and takes a particularly simple form:

$$(\mathcal{L}f)(\eta) = \sum_{x\in\Lambda} c(\eta,\eta^x)\Big(f(\eta^x) - f(\eta)\Big)\,, \tag{2.21}$$

which can be interpreted as a discrete derivative of the function f under the dynamics of the process. There is a fully developed mathematical theory of generators and corresponding semigroups for stochastic particle systems, see [G] and references therein for details.

Example. (opinions in the linear VM)

For a general continuous-time Markov chain on a state space S with jump rates g_{ij} as introduced in Section 1, the generator is given by

$$\mathcal{L}f(i) = \sum_{j \in S} g_{ij}\big(f(j) - f(i)\big)\,, \quad i \in S\,, \tag{2.22}$$

for all functions $f : S \to \mathbb{R}$. Like the master equation, this also has an intuitive interpretation as a 'discrete derivative' describing the change of f under all possible jumps out of state i weighted by the jump rates.

Further (non-examinable) remarks:
For infinite lattices such as $\Lambda = \mathbb{Z}^d$ the state space $S = \{0,1\}^\Lambda$ is **not** countable and therefore the process $(\eta_t : t \geq 0)$ is not a Markov chain and the master equation becomes meaningless. Still, the dynamics can be defined via the generator $\mathcal{L}$ and equation (2.20) for a suitably large set of observables $f : S \to \mathbb{R}$ (usually continuous functions). The generator (2.21) is a well defined convergent sum for *local* functions f that depend on η only through a finite set of coordinates in Λ, such as $f(\eta) = \eta(x)$. In the $\|.\|_\infty$-norm, any continuous function can be written as a limit of such functions.
Very important in this context is the fact that S is a *compact* set, since it is the countable product of compact sets $\{0,1\}$ (Tychonoff's theorem). This also implies that the set of probability measures $\mathcal{M}_1(S)$ on S is compact, and therefore $\boldsymbol{\pi}(t)$ has a subsequential limit as $t \to \infty$. It is easy to see that any such limit has to be a stationary distribution, which implies existence of the latter for general IPS with compact state space. Such processes are also called *Feller processes*.

2.2 The Ising model and Markov chain Monte Carlo

The **Ising model** is a very basic model for magnetic behaviour from Statistical Mechanics. The state space is given by $S = \{-1, 1\}^\Lambda$ and configurations are denoted by $\sigma = (\sigma_x : x \in \Lambda)$, consisting of *spin variables* $\sigma_x = \pm 1$ interacting on a lattice (or graph) Λ. The interaction is defined by an *energy function* (or *Hamiltonian*) $H : S \to \mathbb{R}$, and the equilibrium distribution of the spins is given by

$$\pi(\sigma) = \frac{1}{Z} e^{-\beta H(\sigma)} \quad \text{where} \quad Z = \sum_{\sigma \in S} e^{-\beta H(\sigma)} \tag{2.23}$$

is the normalizing constant called *partition function*. $\beta \geq 0$ is interpreted of the inverse *temperature* in the system, and for

$$\begin{aligned} \beta \to \infty \quad & \text{very low temperature} \quad \Rightarrow \quad \pi \text{ concentrates on lowest energy configurations ,} \\ \beta \to 0 \quad & \text{very high temperature} \quad \Rightarrow \quad \pi \text{ becomes uniform (energy irrelevant) .} \end{aligned}$$

The Hamiltonian for the Ising model is given by

$$H(\sigma) = -\sum_{x,y \in \Lambda} J_{xy} \sigma(x)\, \sigma(y) \tag{2.24}$$

where the *coupling constants* J_{xy} determine the interaction of the spins on the graph Λ. The classical situation is to consider a regular lattice $\Lambda \subseteq \mathbb{Z}^d$ with nearest neighbour interaction

$$J_{xy} = \begin{cases} 1 \ , \ x \sim y \\ 0 \ , \ \text{otherwise} \end{cases} . \tag{2.25}$$

Since $J_{xy} \geq 0$ neighbouring spins favour to lign up and tend towards a *ferromagnetic* behaviour.

Limit (accumulation) points of the sequence $\pi(\sigma) = \pi_\Lambda(\sigma)$ (2.23) as $\Lambda \nearrow \mathbb{Z}^d$ are called **Gibbs measures**. They are probability distributions on $\{-1, 1\}^{\mathbb{Z}^d}$, and by general compactness arguments there exists at least one such measure for all $\beta \geq 0$. The Ising model exhibits a *spin-flip symmetry*, i.e. since $H(\sigma) = H(-\sigma)$ flipped spin configurations have the same probability. For all $\beta > 0$ the model has two *ground states* $\sigma(x) \equiv \pm 1$ for all $x \in \Lambda$ which have lowest energy (highest probability). The set of Gibbs measures has to exhibit the same symmetry, and does this in two different ways:

In dimensions $d \geq 2$ there exists a *critical Temperature* $T_c = 1/\beta_c$ such that

- for $\beta < \beta_c$ (high temperature) there exists only one Gibbs measure π with $\mathbb{E}_\pi(\sigma(x)) = 0$;
- for $\beta > \beta_c$ (small temperature) there exist two Gibbs measures π^+ and π^-, such that $\mathbb{E}_{\pi^\pm}(\sigma(x)) = \pm m_* \neq 0$. The system is said to exhibit *spontaneous symmetry breaking* with *spontaneous magnetization* $m^* \neq 0$. In general, non-uniqueness of the Gibbs measure is called a **phase transition**.

Picture.

The physical validity of Gibbs measures can be postulated from general principles of equipartition of energy in equilibrium systems without referring to any dynamics. The goal is now to compute expected values w.r.t. (2.23) for finite large lattices, which is not a simple task since the $\sigma(x)$ are correlated due to the interaction in the Hamiltonian, and the the size of the state space $|S| = 2^L$ is huge even for moderate lattice sizes $L = 100$. The trick is to invent an artificial Markov chain $(\sigma_t : t \geq 0)$ (often also done in discrete time) with stationary distribution π, and use the Ergodic Theorem to sample from π. This method is called **Markov chain Monte Carlo**, and the conditions on $(\sigma_t : t \geq 0)$ usually imposed are the following:

- it should be ergodic, i.e. irreducible on S (which is large, but finite)
- it should be reversible w.r.t. π, i.e. the rates $c(\sigma, \sigma')$ fulfill the detailed balance relations

$$c(\sigma, \sigma')e^{-\beta H(\sigma)} = c(\sigma', \sigma)e^{-\beta H(\sigma')} \quad \text{for all } \sigma, \sigma' \in S\ . \tag{2.26}$$

  In fact, stationarity would be enough, but reversibility is easier to implement via detailed balance.

To ensure both conditions, one usually restricts to local (*spin flip*) dyanmics analogous to stochastic particle systems, where only single spins are flipped $\sigma \to \sigma^x$ with rates $c(\sigma, \sigma^x)$. The most basic choices are

- the **heat bath algorithm**, with

$$c(\sigma, \sigma^x) = \frac{e^{\beta H(\sigma)}}{e^{\beta H(\sigma)} + e^{\beta H(\sigma^x)}} \in (0, 1)\ , \tag{2.27}$$

  where the system is sampled at rate 1 and proposed flips are accepted with probability $c(\sigma, \sigma^x)$. This can be motivated on physical grounds (coupling to a heat bath), but has the disadvantage that the sampling rate is higher than necessary and acceptance probabilities are typically < 1.

- the **Metropolis algorithm**, with

$$c(\sigma, \sigma^x) = \begin{cases} e^{-\beta(H(\sigma^x) - H(\sigma))} & , \text{ if } H(\sigma^x) > H(\sigma) \\ 1 & , \text{ if } H(\sigma^x) \leq H(\sigma) \end{cases} \in (0, 1]\ , \tag{2.28}$$

  i.e. the system is sampled at rate 1 and whenever a proposed flip does not increase the energy it is accepted with probability 1, otherwise with probability < 1. Therefore, implementations of this algorithm are in general faster than the heat bath algorithm.

There are more involved non-local dynamics where whole clusters of spins are flipped such as the *Swendsen-Wang algorithm*,which are much faster than either of the above when the system is close to a phase transition.
Due to the phase transition and spontaneous symmetry breaking for $\beta > \beta_c$ the MCMC dynamics will exhibit ergodicity breaking in the limit $L \to \infty$ and converge to a mixture of the two Gibbs measures which depends on the boundary conditions imposed or the initial condition.

3 Processes with continuous state space

3.1 General properties and Brownian motion

This chapter is about processes with continuous state space $S = \mathbb{R}$ or $\mathbb{R}^d$ and continuous time $\mathbb{T} = [0, \infty)$. This is mathematically more complicated than Markov chains, and we will discuss some of the technical issues below. On the other hand, the sample paths are now real valued functions, our state space has an analytic structure and we will be able to use concepts from usual calculus.

For example we will often integrate over sets $A \in \mathbb{R}$ of possible values with respect to the distribution function $F(x) = \mathbb{P}(X \leq x)$ of a random variable X, e.g.

$$\mathbb{P}(X \in A) = \int_A dF(x) = \int_A f(x)\, dx \quad \text{where } f = F' \text{ is the pdf (if it exists)} . \tag{3.1}$$

Technical side remark: This cannot be done for all sets $A \subseteq \mathbb{R}$ but only for $A \in \mathcal{A}$, where $\mathcal{A} \subsetneq \mathcal{P}(\mathbb{R})$ is a so-called *σ-algebra*. This is a set of *measurable sets* where the measure $dF(x)$ or Lebesgue measure dx can be consistently defined on.

Example. ('Delta-function')

Characterization of processes on $\mathbb{R}$.
As for Markov chains, the **distributional properties** of a general stochastic process are determined by fixing all *finite-dimensional distributions (fdds)*

$$\mathbb{P}\big(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n\big) \quad \text{for all } t_1 < \ldots < t_n,\ x_1, \ldots, x_n \in \mathbb{R} \text{ and } n \in \mathbb{N} . \tag{3.2}$$

We focus here on the state space $S = \mathbb{R}$ where the fdds are given by joint distribution functions F as above, in principle this can be extended to more general state spaces.

In contrast to Markov chains, for continuous state space the fdds do not determine the process uniquely. Two processes with the same fdds are called *versions* of each other, and their **sample paths** can have very different properties. This fact cannot be ignored, since it is very important when studying properties such as first-passage times (first random time to enter a given set). One is usually interested in the most regular version of the process (in order to avoid complications) and there are basically two classes of processes that are usually considered.

- *Diffusion processes* have continuous sample paths, i.e. almost surely the function $t \mapsto X_t$ is continuous or $\mathbb{P}\big(\{\omega : t \mapsto X_t(\omega) \text{ is continuous}\}\big) = 1$.
  The most basic example is Brownian motion introduced below.
- More generally, if one wants to allow for discontinuities in the sample paths (such as Lévy processes or as a special case also continuous time Markov chains), one restricts to the following class of paths:

Definition 3.1 A real-valued, continuous-time process X is called *càdlàg* if almost surely, its sample paths are right continuous (continue à droite) and have left limits (limite à gauche), i.e.

$$\lim_{s \searrow t} X_s = X_t \quad \text{and} \quad \lim_{s \nearrow t} X_s \text{ exists}, \quad \text{for all } t \in [0, \infty) . \tag{3.3}$$

Example. (Lévy/Poisson process)

Description of the dynamics.
The transition probabilities of a Markov chain can also be generalized.

Definition 3.2 Let X be a stochastic process on $\mathbb{R}$. The conditional distribution function

$$F(t, x|s, y) = \mathbb{P}(X_t \leq x | X_s = y) , \tag{3.4}$$

is called the *transition kernel* of X. If it has a density we call this the *transition density*,

$$f(t, x|s, y) = \frac{\partial F}{\partial x}(t, x|s, y) . \tag{3.5}$$

Note that for a homogeneous process, the kernel is actually only a function of $t-s$ and it suffices to study $f(t, x|0, y)$. In analogy to Markov chains we then have the *Chapman-Kolmogorov equation*

$$f(t + s, x|0, y) = \int_{\mathbb{R}} f(t, z|0, y)\, f(s, x|0, z)\, dz \quad \text{for all } x, y \in \mathbb{R},\ t, s > 0 . \tag{3.6}$$

Again, this implies that the process can be generated by a single operator, the generator $\mathcal{L}$ which we will derive below for several processes.

Proposition 3.1 *The fdds of a Markov process are uniquely determined by the transition kernels and the initial distribution (density) $f(0, x)$. In particular, the distribution at time t is given by*

$$f(t, x) = \int_{\mathbb{R}} f(t, x|0, y)\, f(0, y)\, dy . \tag{3.7}$$

Proof. Sample calculation for $0 \le t_1 \le t_2$ with densities using the Markov property,

$$\mathbb{P}\big(X_{t_1} \le x_1, X_{t_2} \le x_2\big) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f(0,x)\, f(t_1, y|0, x)\, f(t_2, z|t_1, y)\, dx\, dy\, dz\ .$$

□

Example.

Definition 3.3 A real-valued stochastic process $B = (B_t : t \ge 0)$ is called *standard Brownian motion* (BM) or *Wiener process* if

- $B_0 = 0$;
- it has *stationary, independent increments*, i.e. for all $s, t > 0$

$$B_{s+t} - B_s \text{ is distributed like } B_t - B_0 \text{ and independent of } (B_u : u \le s)\ ;$$

- for all $t > 0$, $B(t) \sim N(0, t)$ is Gaussian;
- almost surely, sample paths $t \mapsto B(t)$ are continuous.

Theorem 3.2 (Wiener, 1923)
*Brownian motion exists, i.e. there exists a (unique) probability measure $\mathcal{W}$ on the space of continuous paths $C\big([0,\infty), \mathbb{R}\big)$ (called the **Wiener measure**), such that the process with sample paths distributed according to $\mathcal{W}$ has the properties of Brownian motion as defined above.*

Proof. see e.g. [MP], Section I.1

Remarks.

- If B is a standard BM then $(\sigma B_t + x; t \ge 0)$ is a BM starting in $x \in \mathbb{R}$ with *diffusion coefficient* $\sigma > 0$.
- Since increments are stationary, $B_t - B_s \sim B_{t-s} \sim N(0, t-s)$ for all $t > s \ge 0$, and since they are also independent, $B_t|_{B_s = y} \sim N(y, t-s)$, so the transition kernel for BM is

$$f(t, x|s, y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\Big(-\frac{(x-y)^2}{2(t-s)}\Big)\ . \tag{3.8}$$

- $f(t,x|0,y)$ is also known as the *heat kernel*, since it is the fundamental solution of the *heat equation* (or diffusion equation)

$$\frac{\partial}{\partial t} f(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2} f(t,x) \quad \text{with initial condition} \quad f(0,x) = \delta_y(x)\ . \tag{3.9}$$

The heat equation is therefore the analogue of the forward or master equation for Markov chains in the case of Brownian motion.

Proposition 3.3 *Let B be a standard BM. For all $t_1, \ldots, t_n$, $n \in \mathbb{N}$*

$$(B_{t_1}, \ldots, B_{t_n}) \sim N(\mathbf{0}, \Sigma) \quad \textit{with} \quad \sigma_{ij} = \min\{t_i, t_j\}\ , \tag{3.10}$$

i.e. the fdds for BM are multivariate Gaussian with zero mean and covariance matrix $\Sigma = (\sigma_{ij})_{i,j}$. Conversely, if X is a Gaussian process with continuous paths, mean 0 and covariance function $\sigma(s,t) = \min\{s,t\}$, then X is a standard BM.

Definition 3.4 A process X with Gaussian fdds is called *Gaussian processes*, and it is characterized by the mean $m(t) = E(X_t)$ and covariance function $\sigma(s,t) = \mathrm{cov}(X_s, X_t)$.

In general, multivariate Gaussians are uniquely defined by their mean and covariance, see Handout 3 for more details.

Proof. For standard BM $B_t \sim N(0,t)$ and it suffices to show that $\mathrm{cov}(B_s, B_t) = \min\{s,t\}$. Take $s < t$, then

$$\mathbb{E}(B_s B_t) = \mathbb{E}\big(B_s^2 + B_s(B_t - B_s)\big) = \mathbb{E}(B_s^2) + 0\ , \tag{3.11}$$

since B has independent increments and $\mathbb{E}(B_s) = 0$. Thus $\mathrm{cov}(B_s, B_t) = \mathrm{var}(B_s) = s$. On the other hand, for a Gaussian process X with covariances $\sigma(s,t) = \min\{s,t\}$ we have for $t > s$ and all $u \le s$

$$\mathbb{E}\big((X_t - X_s)X_u\big) = \mathbb{E}(X_t X_u) - \mathbb{E}(X_s X_u) = u - u = 0\ , \tag{3.12}$$

so the increments $X_t - X_s$ are uncorrelated with $(X_u : u \le s)$. So the joint distribution of $X_t - X_s$ and X_u is multivariate Gaussian with vanishing covariances, so they are in fact independent.
(Note that in general uncorrelated rvs could still be dependent, only for Gaussian rvs these concepts are equivalent, since they are fully characterized by their means and covariances.)
Finally, $X_t - X_s$ is Gaussian with mean 0 and variance (for $s < t$)

$$\mathrm{var}(X_t - X_s) = \mathrm{var}(X_t) + \mathrm{var}(X_s) - 2\mathrm{cov}(X_t, X_s) = t + s - 2s = t - s\ , \tag{3.13}$$

so the increments are also stationary. □

3.2 Brownian motion and random walk

From (3.9) we can identify the Laplacian as the generator of a standard BM $(B_t : t \geq 0)$. Let $g \in C^2(\mathbb{R})$ be an observable, such as $g(x) = x$. Then the expected value at time t is given by

$$\mathbb{E}\big(g(B_t)\big) = \int_{\mathbb{R}} g(x)\, f(t,x)\, dx \;, \tag{3.14}$$

where $f(t,x)$ is the Gaussian pdf of B_t solving (3.9) with $f(0,x) = \delta_0(x)$. Using this we get

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\big(g(B_t)\big) \;&=\; \int_{\mathbb{R}} g(x)\, \frac{\partial f(t,x)}{\partial t}\, dx = \frac{1}{2}\int_{\mathbb{R}} g(x)\, \frac{\partial^2 f(t,x)}{\partial x^2}\, dx = \\ &=\; \frac{1}{2}\int_{\mathbb{R}} \frac{\partial^2 g(x)}{\partial x^2}\, f(t,x)\, dx = \frac{1}{2}\mathbb{E}\big(g''(B_t)\big) = \mathbb{E}\big((\mathcal{L}g)(B_t)\big) \;. \end{aligned} \tag{3.15}$$

This follows from twice ingetrating by parts, using that for $k = 0, 1, 2$ products of derivatives

$$\partial_x^{2-k} g(x)\, \partial_x^k f(t,x) \to 0 \quad \text{as } |x| \to \infty \quad \text{fast enough} \;, \tag{3.16}$$

so that we can ignore boundary terms. This is true for all observables g that have finite expectation $\mathbb{E}\big(g(B_t)\big)$ since the $f(t,x)$ have of course Gaussian tail in x for fixed t. Therefore standard BM can be *characterized* as a process on $\mathbb{R}$ with generator

$$\mathcal{L} = \frac{1}{2}\frac{\partial^2}{\partial x^2} \quad \text{i.e.} \quad (\mathcal{L}g)(x) = \frac{1}{2}g''(x) \quad \text{for all } x \in \mathbb{R} \;, \tag{3.17}$$

and $B_0 = 0$. Note that as for IPS, the generator is an operator with a certain domain of definition, which for the Laplacian are the twice differentiable functions $C^2(\mathbb{R})$. Knowing the time evolution of those, one can define it for more general observables $g : \mathbb{R} \to \mathbb{R}$ through approximation.

Definition 3.5 A *jump process* $(X_t : t \geq 0)$ with jump rate (density) $q(x,y)$ is a real-valued process defined by the generator

$$(\mathcal{L}g)(x) = \int_{-\infty}^{\infty} q(x,y)\big(g(y) - g(x)\big)\, dy \;,$$

for all $g : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} |g(x)|\, dx < \infty$. The total jump rate has to be uniformly bounded, i.e.

$$\int_{-\infty}^{\infty} q(x,y)\, dy \leq C < \infty \quad \text{for all} \quad x \in \mathbb{R} \;, \tag{3.18}$$

for the dynamnics to be well defined.

The jump process jumps from x to the interval $[y, y+dy)$ with rate $q(x,y)\, dy \geq 0$. The evolution equation corresponding to the master equation (sometimes called *Kolmogorov-Feller equation*) is

$$\frac{\partial}{\partial t}\, f(t,x) = \int_{-\infty}^{\infty} \Big(q(y,x)\, f(t,y) - q(x,y)\, f(t,x)\Big)\, dy \;. \tag{3.19}$$

A special case is the **compound Poisson process** where $q(x,y)$ is only a function of $y-x$, and we write $q(y-x) = \lambda p(y-x)$. Here $\lambda > 0$ is the rate of a Poisson process $(N(t) : t \geq 0)$ setting the time scale, $p : \mathbb{R} \to [0,1]$ is the distribution of i.i.d. increments $Y_1, Y_2, \ldots \in \mathbb{R}$ and the process can be written as the sum

$$X(t) = \sum_{i=1}^{N(t)} Y_i \,. \tag{3.20}$$

Note that continuous-time Markov chains are included in this class as a limiting (degenerate) case, e.g. for the usual Poisson process we have $q(x,y) = \lambda \delta_1(y-x)$ and all increments are $Y_i = +1$.

Proposition 3.4 Scaling to BM
Let $(X_t : t \geq 0)$ be a jump process with translation invariant rates $q(y-x)$ which have

$$\begin{aligned} &\textit{mean zero, i.e.} && \int_{\mathbb{R}} q(y-x)(y-x)\,dy = \int_{\mathbb{R}} q(z)\,z\,dz = 0 \,, \\ &\textit{finite variance, i.e.} && \sigma^2 := \int_{\mathbb{R}} q(z)\,z^2\,dz < \infty \,. \end{aligned} \tag{3.21}$$

Then the rescaled process $(X^\epsilon_t : t \in [0,T])$ with $X^\epsilon_t = \epsilon X_{t/\epsilon^2}$ converges in distribution as $\epsilon \to 0$ to a BM $(B_t : t \in [0,T])$ with generator $\mathcal{L} = \frac{1}{2}\sigma^2 \partial_x^2$, for all $T > 0$.

Proof. Since the process X^ϵ is speeded up by a factor $1/\epsilon^2$, its generator is given by

$$\begin{aligned} \big(\mathcal{L}^\epsilon g\big)(x) \;&=\; \frac{1}{\epsilon^2}\int_{\mathbb{R}} q(z)\big(g(x+\epsilon z) - g(x)\big)\,dz = \\ &=\; \frac{1}{\epsilon^2}\int_{\mathbb{R}} q(z)\Big(\epsilon z\, g'(x) + \tfrac{1}{2}\epsilon^2 z^2\, g''(x) + O(\epsilon^3)\Big)dz \to \frac{\sigma^2}{2}\, g''(x) \,, \end{aligned} \tag{3.22}$$

where we used (3.21). This holds for all $g \in C^2(\mathbb{R})$ with $\int_{\mathbb{R}} |g(x)|\,dx < \infty$.
An additional argument is required to show that limiting sample paths are a.s. continuous (tightness on pathspace), which requires restriction to finite time intervals $[0,T]$. □

If the mean $\mu := \int_{\mathbb{R}} z\, q(z)\,dz \neq 0$, the same holds for

$$X^\epsilon_t = \epsilon X_{t/\epsilon^2} - \mu\, t/\epsilon \quad \text{with} \quad \sigma^2 := \int_{\mathbb{R}} (z-\mu)^2 q(z)\,dz \,. \tag{3.23}$$

Example. Let $(X_t : t \geq 0)$ be a continuous-time symmetric random walk on $\mathbb{Z}$ with generator

$$G = \begin{pmatrix} \ddots & \ddots & \ddots & & \\ & \frac{1}{2} & -1 & \frac{1}{2} & \\ & & \ddots & \ddots & \ddots \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \ddots & \ddots & \ddots & & \\ & 1 & -2 & 1 & \\ & & \ddots & \ddots & \ddots \end{pmatrix} = \frac{1}{2}\Delta \,, \tag{3.24}$$

which is proportional to the discrete Laplacian Δ. Then it is not surprising that the limit is given by BM with generator $\frac{1}{2}\frac{\partial^2}{\partial x^2}$.
However, BM is the unversal scaling limit for a huge class of processes with finite variance of jumps, and plays a similar role as the Gaussian distribution in the central limit theorem (CLT). Only if the jumps have heavy tails with $\sigma^2 = \infty$ or the process is non-Markovian with non-exponential waiting time distributions $\big($ $(N(t) : t \geq 0)$ is a so-called *renewal process*$\big)$ then different scaling limits are possible which are discussed in a later section. There is a famous convergence result analogous to the above also for discrete-time random walks.

Theorem 3.5 Donsker's invariance principle
Let $(X_n : n \in \mathbb{N})$ be a sequence of iid random variables with $\mathbb{E}(X_n) = 0$ and $\mathrm{var}(X_n) = 1$ and consider the random walk

$$(S_n : n \in \mathbb{N}) \quad \textit{with} \quad S_n = \sum_{k=0}^{n} X_k \,. \tag{3.25}$$

Then the rescaled, continuous-time extention

$$(S_t^n : t \in [0,1]) \quad \textit{with} \quad S_t^n = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} \tag{3.26}$$

converges in distribution as $n \to \infty$ to a standard BM $(B_t : t \in [0,1])$.

Proof. As before, convergence to continuous paths can be shown and requires restriction to times in $[0,1]$ or any other finite time interval (for details see e.g. [MP], Section 5.3.3).
We will identify BM in the limit as a Gaussian process with covariances $\mathrm{cov}(B_s, B_t) = \min\{s,t\}$. First we can write

$$S_t^n = \sqrt{\frac{\lfloor nt \rfloor}{n}} \frac{1}{\sqrt{\lfloor nt \rfloor}} \sum_{k=0}^{\lfloor nt \rfloor} X_k \to \sqrt{t}\, N(0,1) \quad \text{as } n \to \infty \tag{3.27}$$

by the usual CLT. Therefore, $B_t := \lim_{n\to\infty} S_t^n \sim N(0,t)$ for all $t \in [0,1]$. Now we compute the covariances for $0 < s < t \leq 1$:

$$\begin{aligned} \mathbb{E}(S_s^n S_t^n) \;&=\; \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\Big(\sum_{k=0}^{\lfloor ns \rfloor} X_k \Big(\sum_{l=0}^{\lfloor ns \rfloor} X_l + \sum_{l=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} X_l \Big) \Big) = \\ &=\; \lim_{n\to\infty} \frac{1}{n} \Big(\sum_{k,l=0}^{\lfloor ns \rfloor} \underbrace{\mathbb{E}(X_k X_l)}_{=\delta_{k,l}} + \mathbb{E}\Big(\sum_{k=0}^{\lfloor ns \rfloor} X_k \sum_{l=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} X_l \Big) \Big) = s \,, \end{aligned} \tag{3.28}$$

where the second expectation vanishes since all increments X_k are independent with mean 0. $\square$.

Obviously, the above focus on standardized increments X_k is no restriction, if they have mean μ and variance σ^2 the statement holds with

$$S_t^n = \frac{1}{\sqrt{n}} \sum_{k=0}^{\lfloor nt \rfloor} \frac{X_k - \mu}{\sigma} \,. \tag{3.29}$$

In summary, processes with mean zero and finite variance increments universally scale to Brownian motion under the **diffusive scaling**

$$\Delta t \propto \epsilon,\, 1/n\,, \quad \Delta x \propto \epsilon^H,\, 1/n^H \quad \text{or simply} \quad \boxed{\Delta x \propto (\Delta t)^H \quad \text{with} \quad H = 1/2}\,. \tag{3.30}$$

Regularity properties of BM sample paths.
For the increments of Brownian motion we have

$$B_{t+h} - B_t = \sqrt{h}\, N(0,1) \sim N(0,h) \to 0 \quad \text{as } h \to 0\,. \tag{3.31}$$

This is consistent with Brownian sample paths being continuous (and more precisely, Hölder continuous with index $H = 1/2$). But they are nowhere differentiable, since

$$\frac{B_{t+h} - B_t}{h} = \frac{\sigma}{\sqrt{h}}\,\xi \quad \text{has no limit as } h \to 0\,. \tag{3.32}$$

Note that Brownian motion is invariant under diffusive rescaling and therefore *self-similar* with *Hurst exponent* $H = 1/2$, i.e.

$$(B_{\lambda t} : t \geq 0) \sim \lambda^H (B_t : t \geq 0) \quad \text{for all } \lambda > 0\,, \text{ with } H = 1/2\,. \tag{3.33}$$

To formally differentiate Browian motion, define for each fixed $h > 0$

$$\xi_t^h := \frac{B_{t+h} - B_t}{h} \sim N(0, 1/h) \quad \text{for all } t \geq 0\,, \tag{3.34}$$

which is a stationary Gaussian process with mean 0. The covariances are, e.g. for $s < t$

$$\begin{aligned}
\mathbb{E}(\xi_s^h \xi_t^h) &= \frac{1}{h^2}\mathbb{E}\Big(B_{s+h}B_{t+h} + B_s B_t - B_s B_{t+h} - B_{s+h}B_t\Big) = \\
&= \frac{1}{h^2}\left(s + h + s - s - \left\{ \begin{array}{cc} s+h & ,\, t \geq s+h \\ t & ,\, t < s+h \end{array} \right\}\right) = \left\{ \begin{array}{cc} 0 & ,\, t \geq s+h \\ \frac{s+h-t}{h^2} & ,\, t < s+h \end{array} \right. .
\end{aligned}$$

The same can be done for $s > t$ and leads to the following picture and definition.

Definition 3.6 The **non-existing** limit process $(\xi_t : t \geq 0)$ with $\xi_t = \lim_{h\to 0} \xi_t^h$ is called *white noise*, and can be interpreted as a stationary Gaussian process with mean 0 and covariance 'function' $\mathrm{cov}(\xi_s \xi_t) = \delta(t-s)$.

3.3 Diffusion processes and Fokker-Planck equations

Definition 3.7 A Markov process X is called a *diffusion process*, if it has a.s. continuous sample paths and as $h \searrow 0$

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t = x\big) &= a(t,x)\,h + o(h)\,, \\ \mathbb{E}\big((X_{t+h} - X_t)^2 \,\big|\, X_t = x\big) &= b(t,x)\,h + o(h)\,, \end{aligned} \tag{3.35}$$

for some functions $a(t,x) \in \mathbb{R}$ (*drift coefficient*) and $b(t,x) \geq 0$ (*diffusion coefficient*).

The distributional properties are uniquely characterized by the drift and the diffusion coefficient.

Theorem 3.6 *Let X be a diffusion process with drift $a(t,x)$ and diffusion coefficient $b(t,x)$. Then the transition density $f = f(t,x|s,y)$ exists and satisfies the* ***(forward) Fokker-Planck equation***

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}\big(a(t,x)\,f\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(t,x)\,f\big) \tag{3.36}$$

for all $0 \leq s \leq t$, $x, y \in \mathbb{R}$. In particular, this equation holds also for the density

$$f(t,x) = \int_{\mathbb{R}} f(t,x|0,y)\,f(0,y)\,dy \tag{3.37}$$

with general initial conditions $f(0,y)$.

Proof. by Taylor expansion for observables (see hand-out 6 if you are interested).

Examples. (BM with drift, OU process)

Stationary pdfs $f^*(x)$ of a **time-homogeneous diffusion process** with constant drift $a(x)$ and diffusion $b(x)$ are given by stationary solutions to (3.36), i.e.

$$0 = -\frac{\partial}{\partial x}\big(a(x)\,f^*(x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(x)\,f^*(x)\big)\,. \tag{3.38}$$

Integrating (3.38) and denoting the derivative by $'$ we get under mild regularity assumptions on $a(x)$ and $b(x)$ (see also later remarks)

$$\begin{aligned} 0 \;&=\; -\int_{-\infty}^{x} \big(a(y)\, f^*(y)\big)' dy + \frac{1}{2}\int_{-\infty}^{x} \big(b(y)\, f^*(y)\big)'' dy = \\ &=\; -a(x)\, f^*(x) + \frac{1}{2}\big(b(x)\, f^*(x)\big)' \quad (+const.) \end{aligned} \tag{3.39}$$

Under the assumption that $f^*(x)$ vanishes sufficiently fast as $x \to -\infty$ we can ignore the boundary terms and the above constant vanishes. If we consider processes on other state spaces, such as $S = [0,1]$ with boundary conditions, this has to be taken into account in the derivation. The above is a first order linear differential equation and differentiating with the product rule we get

$$f^{*\prime}(x) = \frac{2a(x) - b'(x)}{b(x)}\, f^*(x)\ . \tag{3.40}$$

So the solution is

$$f^*(x) = f^*(0) \exp\Big(\int_0^x \frac{2a(y) - b'(y)}{b(y)}\, dy \Big) \tag{3.41}$$

where $f^*(0)$ is fixed by normalization $\int_{\mathbb{R}} f^*(x)\, dx = 1$.

Examples. (BM with drift, OU process)

For solutions of Fokker-Planck equations in various cases (including time-dependent ones) see [Ga], Chapter 5.

Time evolution of observables.
Let $(X_t : t \geq 0)$ be a diffusion process on $S = \mathbb{R}$, the Fokker-Planck equation is often written in the short form $\frac{\partial}{\partial t} f = \mathcal{L}^* f$, with the (adjoint) *generator*

$$\big(\mathcal{L}^* f\big)(t,x) = -\frac{\partial}{\partial x}\big(a(t,x)\, f\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(t,x)\, f\big) \ . \tag{3.42}$$

Let $g : S \to \mathbb{R}$ be an observable, such as $g(x) = x$. Then the expected value

$$\mathbb{E}\big(g(X_t)\big) = \int_{\mathbb{R}} g(x)\, f(t,x)\, dx \tag{3.43}$$

obeys the following evolution equation (using the notation (3.42)),

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\big(g(X_t)\big) \ &= \ \int_{\mathbb{R}} g(x)\, \frac{\partial f(t,x)}{\partial t}\, dx = \int_{\mathbb{R}} g(x)\, (\mathcal{L}^* f)(t,x)\, dx = \\ &= \ \int_{\mathbb{R}} (\mathcal{L} g)(t,x)\, f(t,x)\, dx = \mathbb{E}\big((\mathcal{L} g)(X_t)\big) \ . \end{aligned} \tag{3.44}$$

This follows from integration by parts, since

$$\int_{\mathbb{R}} g(x) \frac{\partial}{\partial x_i}\big(a_i(t,x)\, f(t,\mathbf{x})\big) dx = -\int_{\mathbb{R}} \Big(\frac{\partial}{\partial x_i} g(x)\Big) a_i(t,x)\, f(t,x) dx \ , \tag{3.45}$$

where we again assume that $f(t,x) \to 0$ as $|x| \to \infty$ sufficiently fast, so that we can ignore the boundary terms. For the diffusion part this can be done twice and leads to the *generator*

$$\mathcal{L} = a(t,x)\frac{\partial}{\partial x} + \frac{1}{2} b(t,x) \frac{\partial^2}{\partial x^2} \ . \tag{3.46}$$

Examples.
The time evolution of the mean $m_t = \mathbb{E}(X_t)$ is given by choosing $g(x) = x$ and computing $\mathcal{L} x = a(t,x)$, which gives

$$\frac{d}{dt}\mathbb{E}(X_t) = \mathbb{E}\big(a(t,X_t)\big) \ . \tag{3.47}$$

This is not very surprising when looking at the definition 3.7 of diffusion processes. For the second moment with $g(x) = x^2$ we get $\mathcal{L} x^2 = 2x a(t,x) + b(t,x)$ and thus

$$\frac{d}{dt}\mathbb{E}(X_t^2) = 2\mathbb{E}\big(X_t\, a(t,X_t)\big) + \mathbb{E}\big(b(t,X_t)\big) \ , \tag{3.48}$$

so the drift also influences the time evolution of higher moments.

(OU process)

3.4 Beyond diffusion

Lévy processes

Definition 3.8 A *Lévy process* $(X_t : t \geq 0)$ is a càdlàg Markov process with stationary, independent increments.

Since each increment $X_- X_0 = \sum_{k=1}^{n} \left(X_{\frac{kt}{n}} - X_{\frac{(k-1)t}{n}}\right)$ can be written as an arbitrary sum of i.i.d. increments, their distribution has to be *infinitely divisible*, i.e. a stable distribution such as Gaussian or Poisson.
By the **Lévy-Khintchine** representation, a Lévy process can be characterized by a drift $a \in \mathbb{R}$, a diffusion coefficient $b \geq 0$ and a measure ν on $\mathbb{R}$ which describes the jumps of the process. The generator is given by

$$(\mathcal{L}g)(x) = a\frac{\partial g(x)}{\partial x} + \frac{b}{2}\frac{\partial^2 g(x)}{\partial x^2} + \int_{z\neq 0} \left(g(x+z) - g(x) - z\frac{\partial g(x)}{\partial x}\mathbb{1}_{(0,1)}(z)\right)\nu(dz) \tag{3.49}$$

$\nu(dz)$ is the expected number of jumps of $X(t)$ per unit time of size $[z, z+dz)$ and it has to satisfy

$$\int_{|z|\geq 1} \nu(dz) < \infty \quad \text{and} \quad \int_{0<|z|<1} z^2\,\nu(dz) < \infty\,, \tag{3.50}$$

i.e. there is a finite number of jumps of size ≥ 1, and a possibly infinite number of small jumps but with finite variance.

The simplest examples are Brownian motion with $a = 0$, $b < 0$, $\nu \equiv 0$ and Gaussian increments which is purely diffusive, and the Poisson process with $a = b = 0$, $\nu(dz) = \lambda\delta_1(z)\,dz$, which is a jump process (see above) with Poisson increments $X_t - X_0 \sim Poi(\lambda t)$.
If $\nu(dz) = \frac{C\,dz}{|z|^{1+\alpha}}$ with $C > 0$ and $0 < \alpha < 2$, large jumps with infinite mean $(0 < \alpha < 1)$ or infinite variance $(1 < \alpha < 2)$ lead to superdiffusive behaviour. If in addition $a = b = 0$ the process is called **α-stable symmetric Lévy process** or **Lévy flight**. The evolution for the density can then be written as space-fractional diffusion equation

$$\frac{\partial f(t,x)}{\partial t} = D_\alpha\,\frac{\partial^\alpha f(t,x)}{\partial |x|^\alpha} \tag{3.51}$$

with an anomalous diffusion coefficient $\quad D_\alpha = 2C\alpha^{-1}\Gamma(1-\alpha)\,\cos(\pi\alpha/2)\,.$
The (symmetric) *Riesz fractional derivative* of order α is defined by its Fourier transform

$$\mathcal{F}\Big[\frac{\partial^\alpha f}{\partial |x|^\alpha}\Big](t,k) = -|k|^\alpha \mathcal{F}(\rho)(t,k)\,. \tag{3.52}$$

It also has an integral representation analogous to (3.49), for details see [MFH], Section 3.3. This process is self-similar with *Hurst exponent* $H = 1/\alpha$, i.e.

$$\big(X_{\lambda t} : t \geq 0\big) \sim \lambda^{1/\alpha}\big(X_t : t \geq 0\big) \quad \text{for all } \lambda > 0\,. \tag{3.53}$$

This implies in particular that $X_t \sim t^{1/\alpha} X_1$ and therefore $\mathbb{E}(X_t^2) \propto t^{2/\alpha}$, , and since $\alpha < 2$ the α-stable symmetric Lévy process is superdiffusive. Recall that Brownian motion is self similar with $H = 1/2$.

Fractional Brownian motion

Definition 3.9 A *fractional Brownian motion (fBM)* $(B_t^H : t \geq 0)$ with Hurst index $H \in (0,1)$ is a mean zero Gaussian process with continuous paths and covariances

$$\mathbb{E}\big(B_t^H\, B_s^H\big) = \frac{1}{2}\big(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\big)\,, \quad t, s \geq 0\,. \tag{3.54}$$

The process has stationary increments, i.e. for all $t > s \geq 0$

$$B_t^H - B_s^H \sim B_{t-s}^H \sim N(0, (t-s)^{2H})\,, \tag{3.55}$$

but in general they are **not** independent, and the process is even **non-Markovian**!

Only for $H = 1/2$ it reduces to standard BM with covariances $\mathbb{E}\big(B_t^{1/2}\, B_s^{1/2}\big) = \min\{t, s\}$.
fBM is self similar with Hurst exponent H, i.e.

$$\big(B_{\lambda t}^H : t \geq 0\big) \sim \lambda^H \big(B_t^H : t \geq 0\big) \quad \text{for all } \lambda > 0\,, \tag{3.56}$$

and for $H > 1/2$ it is super-diffusive resulting from positively correlated increments, whereas for $H < 1/2$ it is sub-diffusive with negatively correlated increments.
Analogously to standard BM we can again define the discrete noise process

$$\xi_t^H(h) := \frac{B_{t+h}^H - B_t^H}{h} \sim N(0, h^{2H-2}) \quad \text{for all } t \geq 0\,, \tag{3.57}$$

which is a stationary Gausian process with mean 0. For the covariances we get now as $h \to 0$

$$\begin{aligned}
\mathbb{E}(\xi_s^H(h)\xi_t^H(h)) &= \frac{1}{h^2}\mathbb{E}\Big(B_{s+h}^H B_{t+h}^H + B_s^H B_t^H - B_s^H B_{t+h}^H - B_{s+h}^H B_t^H\Big) = \\
&= \frac{1}{h^2}\Big(|t-s+h|^{2H} + |t-s-h|^{2H} - 2|t-s|^{2H}\Big) = \\
&= \frac{2H(2H-1)}{h^2}\,|t-s|^{2H-2}\big(h^2 + O(h^3)\big) \to 2H(2H-1)|t-s|^{2(H-1)}\,.
\end{aligned}$$

Definition 3.10 The **non-existing** limit process $(\xi_t^H : t \geq 0)$ with $\xi_t^H = \lim_{h\to 0} \xi_t^H(h)$ is called *fractional Gaussian noise*, and can be interpreted as a stationary Gaussian process with mean 0 and covariance 'function' $\mathrm{cov}(\xi_s \xi_t) = 2H(2H-1)|t-s|^{2(H-1)}$, which exhibits a long-range power law decay.

4 Some stochastic calculus

4.1 Diffusion processes and SDEs

Diffusion processes can be described also by stochastic differential equations. Let X be a diffusion process in $\mathbb{R}$ with drift $a(t,x)$ and diffusion coefficient $b(t,x) = \sigma^2(t,x)$ given by

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t = x\big) &= a(t,x)\,h + o(h)\,, \\ \mathbb{E}\big((X_{t+h} - X_t)^2 \,\big|\, X_t = x\big) &= \sigma^2(t,x)\,h + o(h)\,. \end{aligned} \tag{4.1}$$

In general for a random variable Y with mean μ and variance σ^2 we can write

$$Y = \mu + \sigma\xi \quad \text{where} \quad \xi = \frac{Y-\mu}{\sigma} \quad \text{has mean } 0 \text{ and unit variance}\,. \tag{4.2}$$

Also the increments of the process X at time t are random variables with mean and variance depending on X_t and given by

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t\big) &= a(t,X_t)\,h + o(h)\,, \\ \mathrm{var}\big(X_{t+h} - X_t \,\big|\, X_t\big) &= \sigma^2(t,X_t)\,h - a(t,X_t)^2\,h^2 + o(h) = \sigma^2(t,X_t)\,h + o(h)\,. \end{aligned} \tag{4.3}$$

Therefore with $\xi_{t,t+h} = \big(X_{t+h} - X_t - a(t,X_t)\big)/\sqrt{\sigma^2(t,X_t)\,h}$ we get

$$X_{t+h} - X_t = a(t,X_t)\,h + \sigma(t,X_t)\,\sqrt{h}\,\xi_{t,t+h} + o(h)\,. \tag{4.4}$$

Then

$$\mathbb{E}(\sqrt{h}\,\xi_{t,t+h}) = 0 \quad \text{and} \quad \mathrm{var}(\sqrt{h}\,\xi_{t,t+h}) = h\,, \tag{4.5}$$

which looks like the increment of a Brownian motion. Indeed, if the process X has independent increments also the $\xi_{t,t+h}$ are independent and

$$\xi_{t,t+h} = \sum_{k=1}^{n} \xi_{t+h(k-1)/n,\,t+hk/n} \tag{4.6}$$

can be written as a sum of arbitrarily many independent random variables with mean 0 and variance $1/n$. Therefore $\sqrt{h}\,\xi_{t,t+h} \sim N(0,h)$ are Gaussian and can thus be interpreted as increments of a Brownian motion. Now we can write

$$X_{t+h} - X_t = a(t,X_t)\,h + \sigma(t,X_t)(B_{t+h} - B_t) + o(h) \quad \text{for a BM } B\,. \tag{4.7}$$

Sending $h \to 0$ we get a differential equation for each path of X, i.e. for fixed $\omega \in \Omega$. But since paths of a BM are not differentiable the differential equation is often written as

$$dX_t = a(t,X_t)\,dt + \sigma(t,X_t)\,dB_t\,. \tag{4.8}$$

Definition 4.1 (4.8) is called a *stochastic differential equation (SDE)* with *drift* $a(t,x)$ and *diffusion* $\sigma(t,x)$. Alternatively, in the physics literature you often find

$$\frac{dX_t}{dt} = a(t,X_t) + \sigma(t,X_t)\,\eta_t\,, \tag{4.9}$$

and call this a *Langevin equation*, where $\eta_t = dB_t/dt$ is *white noise*.

White noise $\eta_t = dB_t/dt$ can be understood as a normalized random force term on X uncorrelated in time. As we have seen before, it is formally given by a Gaussian process with mean 0 and covariance function $\mathrm{cov}(t,s) = \delta_0(t-s)$, which makes sense if integrated over time.

As for ordinary differential equations, it is often better to look at the integrated version of (4.8), since it requires less regularity assumptions.

Definition 4.2 A continuous process $X = (X_t : t \geq 0)$ is a *(weak) solution* of the SDE (4.8) with initial distribution μ if (there exists a probability space $(\Omega, \mathbb{P})$ such that)

$$X_t = X_0 + \int_0^t a(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \tag{4.10}$$

where $(B_t : t \geq 0)$ is a BM and $X_0 \sim \mu$. Furthermore, we impose the regularity condition

$$\int_0^t \left(\sigma(s, X_s)^2 + |b(s, X_s)|\right)^2 < \infty \quad \mathbb{P} - a.s. \quad \text{for all } t \geq 0\,, \tag{4.11}$$

for the stochastic integrals to be well defined (see later). The solution is called *unique (in law)*, if for any two solutions X and X' with the same initial distribution μ we have

$$(X_t : t \geq 0) \sim (X'_t : t \geq 0)\,. \tag{4.12}$$

So in order to solve SDEs we have to make sense of the two *stochastic integrals* in (4.10). There is also the concept of 'strong' solutions and uniqueness for SDEs defined on a given probability space, which we do not discuss here.
Let $X = (X_t : t \geq 0)$ and $Y = (Y_t : t \geq 0)$ be two continuous processes.

We partition the time interval $[0, t]$ such that

$$0 = t_0 < t_1 < \ldots < t_n = t \quad \text{with} \quad t_k - t_{k-1} \to 0 \text{ for all } k = 1, \ldots, n, \text{ as } n \to \infty \tag{4.13}$$

Then we would like to approximate the stochastic integral $I = (I_t : t \geq 0)$ by

$$I_t^n = \sum_{k=1}^n Y_{t_{k-1}}(X_{t_k} - X_{t_{k-1}}) \to I_t = \int_0^t Y_s\, dX_s \quad \text{as } n \to \infty\,. \tag{4.14}$$

This is a (particular) Riemann sum approximation of the integrals in (4.10), the simple choice $X_t = t$ yields the first, and $X_t = B_t$ the second. The general question we investigate in the following is, for which processes X and Y the approximations converge and in what sense they converge. Answers turn out to be quite different for the two integrals. But it turns out that the choice of the time partition is not crucial, so we can arbitrarily choose $t_k = tk/n$ to fix ideas.

4.2 Stochastic integration and Itô calculus

Proposition 4.1 *For the integrator $X_t = t$ and continuous integrand Y the limit in (4.14) exists pathwise for each $\omega \in \Omega$ and we can define*

$$I_t(\omega) = \int_0^t Y_s(\omega)\, ds := \lim_{n\to\infty} I_t^n(\omega) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_{t_{k-1}}(\omega)\,. \tag{4.15}$$

Proof. Usual convergence for the Riemann sum approximations holds for each fixed ω since the total variation of $X_t = t$ is finite, i.e.

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left| X_{t_k}(\omega) - X_{t_{k-1}}(\omega) \right| = t < \infty \quad \text{for all } t \geq 0\,. \tag{4.16}$$

Examples.

Theorem 4.2 Itô integral
Let Y be a continuous process and $X = B$ a standard BM. If

$$\mathbb{E}\Big(\int_0^t Y_s^2 ds \Big) < \infty \tag{4.17}$$

for some $t \geq 0$, then

$$I_t = \int_0^t Y_s\, dB_s := \lim_{n\to\infty} I_t^n = \lim_{n\to\infty} \sum_{k=1}^{n} Y_{t_{k-1}} (B_{t_k} - B_{t_{k-1}}) \tag{4.18}$$

exists in the L^2-sense, i.e. for all $s \leq t$, $\quad \mathbb{E}\big((I_s^n - I_s)^2\big) \to 0\,.$
*If (4.17) holds for all $t \geq 0$, then $I = (I_t : t \geq 0)$ is a continuous process with $\mathbb{E}(I_t) = 0$ and is called the **(stochastic) Itô integral** of Y w.r.t. B.*

Proof. see e.g. [MP] Chapter 7

How do we calculate Itô integrals? Let's start with a simple example.

We see that for the Itô integral with $\alpha = 0$ we get

$$\int_{t_0}^{t} B_s \, dB_s = \frac{1}{2}\big((B_t^2 - B_{t_0}^2) - (t - t_0)\big) \, . \tag{4.19}$$

Another common choice are centred intermediate points with $\alpha = 1/2$. Here we get

$$S - \int_{t_0}^{t} B_s \, dB_s = \frac{1}{2}(B_t^2 - B_{t_0}^2) \, , \tag{4.20}$$

and this integral is called the *Stratonovich integral*. The advantage of this choice is that it obeys the usual rules of calculus and arises naturally in approximations of Brownian motion by smooth processes. But now dependence of Y_{τ_k} and the increment $X_{t_k} - X_{t_{k-1}}$ is more complicated, leading to several technical difficulties compared to Itô. Therefore the preferred choice is usually the Itô integral, and from this one can recover the Stratonovich version by a simple transformation. The unexpected term $(t - t_0)$ in (4.19) has to be there, since the result should have vanishing expectation. These additional terms can be easily understood by the rules of *Itô calculus*, introduced below.

It is often convenient to use the following intuitive differential notation,

$$I_t = I_{t_0} + \int_{t_0}^{t} Y_s \, dX_s \quad \Leftrightarrow \quad dI_t = Y_t \, dX_t \, . \tag{4.21}$$

For example for the integral (4.19) this gives $B_t \, dB_t = \frac{1}{2}(dB_t^2 - dt)$, leading to the rule

$$\boxed{d(B_t^2) = 2B_t \, dB_t + dt} \, . \tag{4.22}$$

From the above derivation we see that the origin of this chain rule is

$$\boxed{(dB_t)^2 = dt} \qquad \text{or more generally} \quad (\sigma dB_t)^2 = \sigma^2 dt \, . \tag{4.23}$$

For a consistent calculus all terms up to order dt have to be taken into account. For usual calculus this involves only first order expansions, but in stochastic calculus this means going up to second order.

To summarize, a diffusion process $(X_t : t \geq 0)$ with generator

$$\mathcal{L}f(x) = a(t,x)\,\partial_x f(x) + \frac{1}{2}\sigma^2(t,x)\,\partial_x^2 f(x) \tag{4.24}$$

and initial condition X_0 solves the SDE

$$dX_t = a(t,X_t)\,dt + \sigma(t,X_t)\,dB_t\ , \tag{4.25}$$

and is defined implicitly as a solution to the integral equation

$$X_t - X_0 = \int_0^t a(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dB_s\ . \tag{4.26}$$

Proposition 4.3 *Let $(X_t : t \geq 0)$ be a diffusion process with $\sigma(t,x) \equiv 0$. Then*

$$X_0 = \mathbb{E}(X_0) \quad \Rightarrow \quad X_t = \mathbb{E}(X_t) \quad \textit{for all } t \geq 0\ , \tag{4.27}$$

i.e. the process is deterministic and solves the ODE $\frac{dX_t}{dt} = a(t,X_t)$.

Proof. Follows directly from the SDE becoming an ODE and uniqueness of the solution for given initial conditions. □

4.3 Martingales

The theory of martingales is usually presented more generally involving filtrations, a time-dependent family of σ-algebras that represents the knowledge about the process with increasing time. To avoid such technicalities in the following we choose a not entirely rigorous presentation, which captures the basic ideas and is based on an intuitive understanding of conditional expectations (which in complete generality are in fact rather complicated objects).

Definition 4.3 A stochastic process $(X_t : t \geq 0)$ on $\mathbb{R}$ is called a *martingale* if

$$\mathbb{E}\big(X_t | (X_u : 0 \leq u \leq s)\big) = X_s \quad \text{for all } t \geq s \geq 0\ , \tag{4.28}$$

and $\mathbb{E}(|X_t|) < \infty$ for all $t \geq 0$.

In particular, for a martingale $(X_t : t \geq 0)$ we have $\mathbb{E}(X_t) = \mathbb{E}(X_0)$ for all $t \geq 0$.
Martingales can be interpreted as the capital when playing a 'fair game', where the future expected capital is equal to the present amount.

Examples. (BM,RW,jump process, fBM)

Proposition 4.4 *The Itô integral $I_t = \int_0^t Y_s \, dB_s$ of a diffusion process $(Y_t : t \geq 0)$ w.r.t. Brownian motion is a martingale.*

Proof.

We know from before that for a stochastic process $(X_t : t \geq 0)$ with generator $\mathcal{L}$

$$\frac{d}{dt}\,\mathbb{E}\big(f(X_t)\big) = \mathbb{E}\big(\mathcal{L}f(X_t)\big)\,, \tag{4.29}$$

and therefore for all $t \geq s \geq 0$

$$\mathbb{E}\Big(f(X_t) - f(X_s)\Big) = \int_s^t \mathbb{E}(\mathcal{L}f(X_u))\,du = \mathbb{E}\bigg(\int_s^t \mathcal{L}f(X_u)\,du\bigg)\,, \tag{4.30}$$

where now we can make sense also of the stochastic integral in the last expression. This motivates a very general result connecting Markov processes with martingales.

Theorem 4.5 Martingale problem
$(X_t : t \geq 0)$ *is a Markov process with generator $\mathcal{L}$ on $\mathbb{R}$ if and only if*

$$M_t^f := f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)\,ds \tag{4.31}$$

is a martingale for all bounded functions $f : \mathbb{R} \to \mathbb{R}$ in the domain of the generator. (I.e. $(X_t : t \geq 0)$ solves the martingale problem with operator $\mathcal{L}$.)

Proof. Let $(X_t : t \geq 0)$ be a MP with generator $\mathcal{L}$. Then for all $s \leq t$ we can write

$$M_t^f = \underbrace{f(X_s) - f(X_0) + \int_0^s \mathcal{L}f(X_u)\,du}_{M_s^f} + f(X_t) - f(X_s) + \int_s^t \mathcal{L}f(X_u)\,du\,. \tag{4.32}$$

Therefore, using (4.30)

$$\mathbb{E}\Big(M_t^f \Big| (M_u^f : 0 \leq u \leq s)\Big) = M_s^f + \mathbb{E}\bigg(f(X_t) - f(X_s) + \int_s^t \mathcal{L}f(X_u)\,du\bigg) = M_s^f\,. \tag{4.33}$$

Also, for bounded functions it can be shown that $\mathbb{E}(|M_t^f|) < \infty$ for all $t \geq 0$.
On the other hand, if (4.31) is a martingale, this implies

$$\frac{d}{dt}\mathbb{E}(M_t^f) = \frac{d}{dt}\mathbb{E}\big(f(X_t)\big) - \mathbb{E}\big(\mathcal{L}f(X_s)\big) = 0 \tag{4.34}$$

for all f, which uniquely identifies $\mathcal{L}$ as the generator of the process. □

For a diffusion process $(X_t : t \geq 0)$, choosing $f(x) = x$ in Theorem 4.5,

$$M_t = X_t - X_0 - \int_0^t a(s, X_s)\, ds = \int_0^t \sigma(s, X_s)\, dB_s \tag{4.35}$$

is a martingale, and it fulfills the SDE $\quad dM_t = \sigma(t, X_t)\, dB_t \quad$ which depends on X_t.
Therefore, $(X_t : t \geq 0)$ itself is a martingale if and only if $a(t, x) \equiv 0$ and it solves the SDE

$$dX_t = \sigma(t, X_t)\, dB_t \ .$$

Note that the integral representation of the SDE

$$X_t - X_0 = \int_0^t a(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \tag{4.36}$$

provides a decomposition in a (deterministic) drift part and a (fluctuating) martingale part.

Examples. Let $(N_t : t \geq 0)$ be a Poisson process $PP(\lambda)$ with generator $\mathcal{L}f(x) = \lambda\big(f(x+1) - f(x)\big)$. Then we have with Theorem 4.5

$$f(x) = x \quad \Rightarrow \quad N_t - \lambda t \quad \text{is a martingale .} \tag{4.37}$$

For standard Brownian motion with generator $\mathcal{L} = \frac{1}{2}\partial_x^2$ we have with Theorem 4.5

$$\begin{aligned} f(x) = x \quad &\Rightarrow \quad B_t \quad \text{itself is obviously a martingale} \\ f(x) = x^2 \quad &\Rightarrow \quad B_t^2 - t \quad \text{is a martingale .} \end{aligned} \tag{4.38}$$

In fact, the reverse is also true and leads to another characterization of BM.

Theorem 4.6 Lévy's characterization of Brownian motion
A continuous-time process B is a standard Brownian motion if and only if B and $(B_t^2 - t : t \geq 0)$ are martingales, and it has continuous sample paths with $B_0 = 0$.

Proof. One direction follows from Theorem 4.5, for the other see e.g. [MP], Chapter 2.

Proposition 4.7 *Let $(X_t : t \geq 0)$ be a martingale on $\mathbb{R}$ with discrete absorbing set $A \subseteq \mathbb{R}$. Let $T_A = \inf_{t \geq 0}\{X_t \in A\}$ be the hitting time of A and assume that $\mathbb{P}(T_A < \infty) = 1$ and $\mathbb{E}(T_A) < \infty$. Then for*

$$\mu(a) := \lim_{t \to \infty} \mathbb{P}(X_t = a) \quad \textit{for all } a \in A \ , \tag{4.39}$$

*(this is called the **harmonic measure**) we have* $\quad \mathbb{E}(X_0) = \sum_{a \in A} a\mu(a)$.

Proof. Follows directly from $\mathbb{E}(X_t) = \mathbb{E}(X_0)$ for all $t \geq 0$ and absorption in A with prob. 1. □

Remarks.

- There is a more general version of this result called *optional stopping/sampling*, involving the concept of *stopping times* T, which are certain random times such as the time of absorption in a set A. See [K] for more details.
- The above result also holds for continuous sets A where μ has to be replaced by a density.
- For continuous martingales defined on an interval $[a, b]$ with absorbing boundary conditions, this can be used to completely characterize the harmonic measure, i.e. $a\mu(a) + b\mu(b) = X_0$ and $\mu(a) + \mu(b) = 1$ (see homework).

4.4 Diffusion processes and Itô's formula

Let X be a solution of the SDE

$$dX_t = a(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \,. \tag{4.40}$$

The following very useful result summarizes our findings in Section 4.2 and gives an explicit formula for time evolution of an observable $g(t, X_t)$, which can also have an explicit time dependence.

Theorem 4.8 Itô's formula
Let X be a solution of (4.40) and $g \in C^2(\mathbb{R} \times \mathbb{R}, \mathbb{R})$. Then

$$dg(t, X_t) = \partial_t g(t, X_t)\, dt + \partial_x g(t, X_t)\, dX_t + \frac{\sigma^2(t, X_t)}{2} \partial_x^2 g(t, X_t)\, dt \,, \tag{4.41}$$

or in the (extended) integrated version

$$\begin{aligned} g(t, X_t) \;=\; & g(0, X_0) + \int_0^t \partial_t g(s, X_s)\, ds + \int_0^t \partial_x g(s, X_s)\, \sigma(s, X_s) dB_s \\ & + \int_0^t \Big(\partial_x g(s, X_s)\, a(s, X_s) + \tfrac{1}{2} \partial_x^2 g(s, X_s)\, \sigma^2(s, X_s) \Big) ds \,. \end{aligned} \tag{4.42}$$

Proof. Taylor expansion with terms up to order dt, using (4.23) and the Itô chain rule (4.22). In incremental form this gives (omitting the arguments of a and σ)

$$\begin{aligned} dg(t, X_t) &= \partial_t g(t, X_t)\, dt + \partial_x g(t, X_t)\, dX_t + \frac{1}{2}\, \partial_x^2 g(t, X_t)\, (dX_t)^2 + o(dt) = \\ &= \partial_t g(t, X_t)\, dt + \partial_x g(t, X_t) \big(a\, dt + \sigma\, dB_t \big) + \frac{\sigma^2}{2}\, \partial_x^2 g(t, X_t)\, dt + o(dt) \,. \end{aligned} \tag{4.43}$$

□

Let $f(t, x)$ be the pdf of the process X that solves the SDE (4.40). Taking the expectation on both sides, we get from Itô's formula for an observable $g(X_t)$ using partial integration

$$\begin{aligned} \frac{d}{dt} \mathbb{E}\big(g(X_t)\big) \;=\; & \int_{\mathbb{R}} g(x)\, \frac{\partial}{\partial t} f(t, x)\, dx = \int_{\mathbb{R}} (\mathcal{L} g)(x)\, f(t, x)\, dx = \\ = \; & \int_{\mathbb{R}} \Big(g'(x)\, a(t, x) + \frac{1}{2} g''(x) \sigma^2(t, x) \Big)\, f(t, x)\, dx = \\ = \; & \int_{\mathbb{R}} g(x) \bigg(- \frac{\partial}{\partial x} \big(a(t, x)\, f(t, x) \big) + \frac{1}{2} \frac{\partial^2}{\partial x^2} \big(\sigma^2(t, x)\, f(t, x) \big) \bigg) dx \,, \end{aligned} \tag{4.44}$$

since boundary terms and the expected value of the Itô integral vanishes (it is a martingale). This holds for arbitrary functions g, and therefore we must have

$$\frac{\partial}{\partial t} f(t, x) = - \frac{\partial}{\partial x} \big(a(t, x)\, f(t, x) \big) + \frac{1}{2} \frac{\partial^2}{\partial x^2} \big(\sigma^2(t, x)\, f(t, x) \big) = \mathcal{L}^* f(t, x) \,. \tag{4.45}$$

Thus $f(t, x)$ fulfilles the Fokker-Planck equation with adjoint generator $\mathcal{L}^*$ introduced in (3.42) and X is a diffusion process with drift $a(t, x)$ and diffusion $\sigma^2(t, x)$. In Section 4.1 we have derived the opposite statement, so let us summarize both of them.

Proposition 4.9 *X is a diffusion process as defined in Def. 3.7 if and only if it is a solution of the SDE (4.40) with $\sigma^2(t,x) = b(t,x)$.*

Remember that the generator $\mathcal{L}$ of a diffusion process introduced in (3.46) is given by

$$\mathcal{L} = a(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2}{\partial x^2} \,, \tag{4.46}$$

and describes the time evolution of expected values of observables (3.44). Using $\mathcal{L}$ we can rewrite Itô's formula

$$g(X_t) = g(X_0) + \int_0^t (\mathcal{L}g)(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dB_s \,. \tag{4.47}$$

So we see that the time evolution of $g(X_t)$ is given by the generator part plus fluctuations in terms of an Itô integral, which is a martingale with vanishing expected value.

Scaling. Let $X = (X_t : t \geq 0)$ be a diffusion process with generator $\mathcal{L}_X$ given by (4.46).

- The **space-rescaled** process Y with $Y_t = \lambda X_t$, $\lambda \in \mathbb{R}$ is a diffusion process with generator

$$\mathcal{L}_Y f(y) = \lambda\, a(t, y/\lambda)\partial_y f(y) + \lambda^2 \frac{\sigma^2(t,y/\lambda)}{2}\partial_y^2 f(y) \,. \tag{4.48}$$

  Using Itô's formula for $Y_t = g(X_t) = \lambda X_t$ we get immediately

$$dY_t = \lambda dX_t = \lambda a(t, Y_t/\lambda)dt + \lambda\sigma(t, Y_t/\lambda)dB_t \,. \tag{4.49}$$

- For the **time-rescaled** process Y with $Y_t = X_{\lambda t}$ we have $d(\lambda t) = \lambda dt$ and $dB_{\lambda t} = \sqrt{\lambda}dB_t$, and therefore

$$dY_t = dX_{\lambda t} = \lambda a(\lambda t, Y_t)\,dt + \sqrt{\lambda}\sigma(\lambda t, Y_t)\,dB_t \,. \tag{4.50}$$

  So it is a diffusion process with drift $\lambda a(\lambda t, y)$ and diffusion $\sqrt{\lambda}\sigma(\lambda t, y)$.

- For two diffusion processes X and Y we have for the **sum** $Z_t = X_t + Y_t$,

$$dZ_t = dX_t + dY_t = \big(a_X(t,X_t) + a_Y(t,Y_t)\big)dt + \big(\sigma_X(t,X_t) + \sigma_Y(t,Y_t)\big)dB_t \,. \tag{4.51}$$

  This is in general NOT a diffusion process, since drift and diffusion depend on X and Y individually, and cannot necessarily be written as a function of Z alone.

4.5 Geometric Brownian motion and the Black-Scholes formula

Definition 4.4 Let $(B_t : t \geq 0)$ be a standard BM. Then the process

$$(Z_t : t \geq 0) \quad \text{with} \quad Z_t = Z_0 \exp(\mu t + \sigma B_t) \,, \quad Z_0 > 0 \,, \tag{4.52}$$

is a *geometric Brownian motion* with *drift parameter* $\mu \in \mathbb{R}$ and *variance parameter* $\sigma^2 \geq 0$.

Properties.

- Using Itô's formula with $g(t,x) = \exp(\mu t + \sigma x)$, and using $\partial_t g = \mu g$, $\partial_x g = \sigma g$ we get

$$dZ_t = dg(t, B_t) = \Big(\mu + \frac{\sigma^2}{2}\Big) Z_t dt + \sigma Z_t dB_t \,. \tag{4.53}$$

  So Z is a diffusion process with drift $\Big(\mu + \frac{\sigma^2}{2}\Big) z$ and diffusivity σz.

- Therefore, $\mathbb{E}(Z_t) = Z_0 e^{(\mu + \frac{\sigma^2}{2})t}$ and Z is a martingale if and only if $\mu = -\sigma^2/2$.

- If $\mu > 0$ ($\mu < 0$), $Z_t \to \infty$ (0) with probability 1, since the drift dominates the BM in the exponent. So if $\mu = -\sigma^2/2$, $Z_t \to 0$ $a.s.$ but $\mathbb{E}(Z_t) \equiv Z_0$.
  Even better, if $-\sigma^2/2 < \mu < 0$ then $Z_t \to 0$ $a.s.$ but $\mathbb{E}(Z_t) \to \infty$ diverges exponentially! This is due to very rare trajectories that contribute a huge amount to the mean.

- The **log-returns** $\log \frac{Z_{t+h}}{Z_t} \sim \mu h + \sigma(B_{t+h} - B_t)$ are independent of $(Z_s : 0 \leq s \leq t)$ and have a Gaussian distribution $N(\mu h, \sigma^2 h)$.
  Log-returns of stock/commodity prices are indeed often uncorrelated, but are usually not Gaussian and have heavy tails. Still geometric BM is often used as a standard model...

Definition 4.5 A *geometric random walk* is a process $(S_n : n \in \mathbb{N})$ where

$$S_n = S_{n-1} R_n \quad \text{for } n = 1, 2, \ldots \quad \text{where} \quad R_1, R_2, \ldots \geq 0 \quad \text{are iidrv} \,, \tag{4.54}$$

representing the relative change of S_n (multiplicative noise).

Note that $\log(S_n/S_0) = \sum_{k=1}^n \log R_k$ is a standard RW with iid increments $\log R_k$. Multiplication with iid relative returns R_k is used as a simple model for the formation of stock prices. In an appropriate scaling limit this leads to geometric Brownian motion independent of the actual distribution of the R_k.

Proposition 4.10 *Assume that* $R_k = 1 + \xi_k/\sqrt{n}$ *where the* ξ_k *are iid with* $\mathbb{E}(\xi_k) = 0$ *and* $\mathbb{E}(\xi_k^2) = \sigma^2$. *Then as* $n \to \infty$

$$S_{\lfloor nt \rfloor} \to Z_t = e^{\sigma B_t} \quad \textit{for all } t \in [0, T] \,, \quad T > 0 \,, \tag{4.55}$$

converges to a geometric BM with $\mu = 0$ *and variance parameter* σ^2.

Proof. As $n \to \infty$ we have $\log R_k \simeq \xi_k/\sqrt{n}$ and thus

$$\log(S_{\lfloor nt \rfloor}/S_0) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \xi_k \to \sigma B_t \tag{4.56}$$

converges to BM following Donsker's invariance principle (Thm 3.5). Then for all $t \in [0,T]$ also $S_{\lfloor nt \rfloor} \to S_0 e^{\sigma B_t}$. □

The Black-Scholes formula.
A *call option* with *strike prize* K and *expiration time* t provides the right to buy one unit of a *security* at time t at prize K. Let $(Z_t : t \geq 0)$ be the price of the security, which is modelled as a geometric Brownian motion with drift parameter μ and variance parameter σ^2.
The *no arbitrage* dogma dictates that $e^{-rt} Z_t$ is a martingale, where r is the *base rate* of interest when putting money on a bank account. Therefore, $\mu = r - \sigma^2/2$.

Theorem 4.11 Black-Scholes option pricing formula
The 'fair' price C, i.e. the price with no arbitrage, of the above call option is given by

$$C = e^{-rt}\mathbb{E}\big((Z_t - K)^+\big) = Z_0\Phi(\beta) - K\, e^{-rt}\Phi(\beta - \sigma\sqrt{t})\,, \tag{4.57}$$

where $\beta = \frac{1}{\sigma\sqrt{t}}\Big(rt + \frac{\sigma^2 t}{2} - \log\frac{K}{Z_0}\Big)$ and Φ is the standard normal distribution function.

Proof. The expected profit under exercising the call option at time t is

$$\mathbb{E}\big((Z_t - K)^+\big) \quad \text{where} \quad (Z_t - K)^+ = \max\big\{0, (Z_t - K)\big\}\,, \tag{4.58}$$

since for $Z_t < K$ the option is just not used. Discounted against the base rate, this should be the price of one call option. Then use that $W = e^{-rt}\frac{Z_t}{Z_0}$ is a log-normal random variable with $\log W \sim N(-\sigma^2 t/2, \sigma^2 t)$ and do the computation... □