Stochastic Processes

CO905 - Complexity Science

Stefan Grosskinsky

Warwick, 2009

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

17

Introduction

In this module we will cover the basics to study 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 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.

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.

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.

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

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

Example.

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.

$$\lim_{t \searrow 0} P(t) = P(0) = Id \quad \text{and} \quad G := \lim_{t \searrow 0} \frac{P(t) - Id}{t} \quad \text{exists}\,, \tag{1.9}$$

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

$$\frac{P(t+\Delta t) - P(t)}{\Delta t} = \frac{P(\Delta t) - Id}{\Delta t}\,P(t) = P(t)\frac{P(\Delta t) - Id}{\Delta t}\,, \tag{1.10}$$

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

$$\boxed{\frac{d}{dt}P(t) = G\,P(t) = P(t)\,G \quad \Rightarrow \quad P(t) = \exp(t\,G)\,.} \tag{1.11}$$

For finite state spaces $|S| < \infty$ the formal solution to this equation is given by a matrix exponential which is discussed in Section 1.3 in more detail. The distribution at time t is then given by

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\,\exp(t\,G)\,. \tag{1.12}$$

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.

Theorem 1.2 (Existence) *A Markov chain with finite state space S has at least one stationary distribution.*

Proof. Depends on discrete or continuous time, see later.

In Section 1.4 we will see a generalisation of this for inifinite state spaces. The question of uniqueness of stationary distributions is connected to the following definition.

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 1.3 (Uniqueness) *An irreducible Markov chain has at most one stationary distribution.*

Proof. Depends on discrete or continuous time, see later.

Examples.

1.2 Discrete time

Since $P(n) = P^n$ a discrete time Markov chain is uniquely determined by the transition matrix $P = P(1)$ and its initial distribution. So with Definition 1.2, $\boldsymbol{\pi}^*$ is a stationary distribution if and only if $\boldsymbol{\pi}^* P = \boldsymbol{\pi}^*$, i.e. $\boldsymbol{\pi}^*$ is a left eigenvector with eigenvalue $\lambda = 1$.

By definition, P is a *stochastic matrix*, i.e.

$$p_{ij} \in [0,1] \quad \text{and} \quad \sum_{j \in S} p_{ij} = 1 \,, \tag{1.13}$$

since the chain starting in i has to jump somewhere in S. So $P\,\mathbf{1} = \mathbf{1}$ and $\lambda = 1$ is an eigenvalue of P with right eigenvector $\mathbf{1} = (\ldots, 1, 1, \ldots)^T$. Therefore if S is finite, there exists at least one left eigenvector $\boldsymbol{\pi}^*$, whose entries can be shown to be non-negative. If the chain is irreducible this eigenvector is unique under the condition $\sum_{i \in S} \pi_i^* = 1$.

Theorem 1.4 *A discrete time, finite state irreducible Markov chain has a unique stationary distribution $\boldsymbol{\pi}^*$, where*

$$\pi_i^* = 1/\mu_i \quad \textit{with} \quad \mu_i := \mathbb{E}(T_i | X_0 = i) \quad \textit{and} \quad T_i := \min\{n \geq 1 : X_n = i\} \,, \tag{1.14}$$

*so that $\boldsymbol{\pi}^*$ is determined by the inverse of the **mean recurrence times** μ_i (in particular all $\pi_i^* > 0$).*

Proof. see [GS] pp 229 - 230 for recurrence times, for uniqueness see proof of Theorem 1.7.

Example.

Proposition 1.5 *Let $X = \big(X_n : n \in \{0, \ldots, N\}\big)$ be a finite state irreducible Markov chain with transition matrix P^X. Suppose further that X is stationary, i.e. $X_n \sim \boldsymbol{\pi}^*$ for all n. Then the **reversed chain** $Y = \big(Y_n : n \in \{0, \ldots, N\}\big)$ with $Y_n = X_{N-n}$ is a Markov chain with*

$$\textit{transition matrix} \qquad p_{ij}^Y = \frac{\pi_j^*}{\pi_i^*}\, p_{ji}^X \quad \textit{for all } i, j \in S \,. \tag{1.15}$$

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

$$\begin{aligned} &\mathbb{P}\big(Y_{n+1} = i_{n+1} \,\big|\, Y_n = i_n, \ldots, Y_0 = i_0\big) = \\ &\quad = \frac{\mathbb{P}(Y_k = i_k, 0 \leq k \leq n+1)}{\mathbb{P}(Y_k = i_k, 0 \leq k \leq n)} = \frac{\mathbb{P}(X_{N-k} = i_k, 0 \leq k \leq n+1)}{\mathbb{P}(X_{N-k} = i_k, 0 \leq k \leq n)} = \\ &\quad = \frac{\pi_{i_{n+1}}^* p_{i_{n+1} i_n} \cdots p_{i_1 i_0}}{\pi_{i_n}^* p_{i_n i_{n-1}} \cdots p_{i_1 i_0}} = \frac{\pi_{i_{n+1}}^* p_{i_{n+1} i_n}}{\pi_{i_n}^*} \end{aligned} \tag{1.16}$$

as required. □

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

Definition 1.4 Let $\boldsymbol{\pi}$ be a probability distribution on S. A discrete time Markov chain X with transition matrix P is called *reversible (w.r.t. $\boldsymbol{\pi}$)*, if it fulfilles the *detailed balance* conditions

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all } i, j \in S \, . \tag{1.17}$$

Proposition 1.6 *Suppose a discrete time Markov chain X is reversible w.r.t. $\boldsymbol{\pi}$. Then $\boldsymbol{\pi}$ is a stationary distribution of X.*

Proof. From (1.17) we deduce $\quad (\boldsymbol{\pi} P)_j = \sum_{i \in S} \pi_i p_{ij} = \sum_{i \in S} \pi_j p_{ji} = \pi_j \, .$ □

Note that Proposition 1.5 together with (1.17) implies that a reversible Markov chain and its time-reversal are indistinguishable, i.e. they have the same transition probabilities, since

$$p^Y_{ij} = \frac{\pi^*_j}{\pi^*_i} \, p^X_{ji} = \frac{\pi^*_i}{\pi^*_i} \, p^X_{ij} = p^X_{ij} \, . \tag{1.18}$$

The detailed balance relations (1.17) can be a useful tool to find stationary distributions of certain Markov chains 'without loops'.

Example.

Definition 1.5 A discrete time Markov chain is called *aperiodic* if for all $i \in S$, $p_{ii}(n)$ is eventually positive, i.e.

$$\text{there exists } N_i \in \mathbb{N} \text{ such that } p_{ii}(n) > 0 \text{ for all } n \geq N_i \, . \tag{1.19}$$

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

Example.

Theorem 1.7 *An irreducible, aperiodic Markov chain with finite state space is* ***ergodic****, i.e.*

$$p_{ij}(n) = \mathbb{P}(X_n = j \,|\, X_0 = i) \to \pi_j^* \quad \textit{as } t \to \infty\,, \quad \textit{for all } i, j \in S\,. \tag{1.20}$$

Proof. The statement 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 1.4.
Suppose further that all the eigenvalues are distinct, then P can be diagonalized, i.e.

$$B\,P\,B^{-1} = \Lambda = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_{|S|} \end{pmatrix} \tag{1.21}$$

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

$$P^n = \left(B^{-1}\Lambda B\right)^n = B^{-1} \begin{pmatrix} \lambda_1^n & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_{|S|}^n \end{pmatrix} B \to B^{-1} \begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix} B \tag{1.22}$$

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

$$\begin{pmatrix} 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \ldots & 0 \end{pmatrix} B = \begin{pmatrix} \pi_1^* \ldots \pi_{|S|}^* \\ \vdots \\ \pi_1^* \ldots \pi_{|S|}^* \end{pmatrix} \quad \text{which implies the statement.} \tag{1.23}$$

The proof can be extended to more general cases. □

Remark. Theorem 1.7 implies that for every initial distribution $\boldsymbol{\pi}(0)$,

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\,P^n \to \boldsymbol{\pi}^* \quad \text{as } n \to \infty\,. \tag{1.24}$$

Example.

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

$$d(i) := gcd\big\{t \geq 1 \,:\, p_{ii}(t) > 0\big\}\,, \tag{1.25}$$

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. 1.5 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,

$$\lambda_1 = w^0 = 1,\ \lambda_2 = w^1, \ldots, \lambda_d = w^{d-1} \quad \textit{where} \quad w = e^{2\pi i/d}\,, \tag{1.26}$$

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

1.3 Continuous time

As derived in (1.11) the time evolution of a continuous time MC is governed by

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

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

$$P(t) = \exp(t\,G) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\,G^k\,. \tag{1.28}$$

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.

How does G look and what is the relation to the time evolution of X?

Assume that $X_t = i$. For small times Δt we have from (1.28)

$$p_{ij}(\Delta t) = g_{ij}\Delta t + o(\Delta t) \quad \text{for all } i \neq j \in S \,. \tag{1.29}$$

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

$$p_{ii}(\Delta t) = 1 + g_{ii}\Delta t + o(\Delta t) \quad \text{for all } i \in S \,, \tag{1.30}$$

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

$$1 = \sum_{j\in S} p_{ij}(\Delta t) = 1 + \Delta t \sum_{j\in S} g_{ij} \quad \Rightarrow \quad \sum_{j\in S} g_{ij} = 0 \quad \text{for all } i \in S \,. \tag{1.31}$$

Therefore the diagonal entries of G are

$$g_{ii} = -\sum_{j\neq i} g_{ij} \leq 0 \quad \text{for all } I \in S \,, \tag{1.32}$$

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

What does really 'happen' in a continuous time Markov chain?

Assume that $X_t = i$ and define the *holding time*

$$W_i := \inf\left\{t' \geq 0 \,:\, X_{t+t'} \neq i\right\} \,, \tag{1.33}$$

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 1.8 *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 $s, u > 0$

$$\begin{aligned} \mathbb{P}(W_i > s+u | W_i > s) &= \mathbb{P}(W_i > s+u | X_{t+s} = i) = \\ &= \mathbb{P}(W_i > u | X_t = i) = \mathbb{P}(W_i > u) \,, \end{aligned} \tag{1.34}$$

where we have used the Markov property and homogeneity. Therefore

$$\begin{aligned} \mathbb{P}(W_i > s+u) &= \mathbb{P}(W_i > s+u, W_i > s) = \mathbb{P}(W_i > s+u \,|\, W_i > s)\, \mathbb{P}(W_i > s) \\ &= \mathbb{P}(W_i > u)\, \mathbb{P}(W_i > s) \,. \end{aligned} \tag{1.35}$$

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

$$\bar{F}(s) = \mathbb{P}(W_i > s) = e^{\lambda s} \quad \text{where} \quad \lambda = \bar{F}'(0) \,. \tag{1.36}$$

Together with (1.30) we get

$$\bar{F}'(0) = \lim_{\Delta t \searrow 0} \frac{\mathbb{P}(W_i > \Delta t) - 1}{\Delta t} = \lim_{\Delta t \searrow 0} \frac{p_{ii}(\Delta t) + o(\Delta t) - 1}{\Delta t} = g_{ii} \leq 0 \,, \tag{1.37}$$

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

$$\mathbb{P}(X_{t+\Delta t} = j | X_t = i, W_i < \Delta t) = \frac{\mathbb{P}(X_{t+\Delta t} = j | X_t = i)}{\mathbb{P}(W_i < \Delta t | X_t = i)} \simeq$$
$$\simeq \frac{p_{ij}(\Delta t)}{1 - p_{ii}(\Delta t)} \to \frac{g_{ij}}{-g_{ii}} \quad \text{as } \Delta t \searrow 0 \,. \tag{1.38}$$

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

Picture.

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

$$p_{ij}^Y = \begin{cases} 0 & , \, i = j \\ g_{ij}/|g_{ii}| & , \, i \neq j \end{cases} \quad \text{if } g_{ii} > 0 \,, \quad \text{and} \quad p_{ij}^Y = \delta_{ij} \quad \text{if } g_{ii} = 0 \,. \tag{1.39}$$

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.

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

Proposition 1.9 *$X = (X_t : t \geq 0)$ is a Poisson process with rate λ if and only if it has stationary, independent increments, i.e.*

$$X_{t+t'} - X_{t'} \text{ is distributed like } X_t - X_0 \text{ and independent of } (X_s : s \leq t') \,, \tag{1.40}$$

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.

Using the forward equation (1.27) we can also get an evolution equation for the distribution,

$$\frac{d}{dt}\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, \frac{d}{dt} P(t) = \boldsymbol{\pi}(0)\, P(t)\, G = \boldsymbol{\pi}(t)\, G \,. \tag{1.41}$$

This is called the *Master equation* and using (1.32) the coordinate form is given by

$$\boxed{\frac{d}{dt}\pi_i(t) = \sum_{j\neq i}\Big(\pi_j(t)\, g_{ji} - \pi_i(t)\, g_{ij}\Big)\,.} \tag{1.42}$$

If $\boldsymbol{\pi} = \boldsymbol{\pi}^*$ is a stationary distribution, then both sides of the equation vanish.

Proposition 1.10 *Let G be the generator of a continuous time Markov chain, $P(t) = \exp(tG)$ and P^Y the transition matrix of the jump chain. Then*

$$\boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^* \quad \Leftrightarrow \quad \boldsymbol{\pi}^* G = \mathbf{0} \quad \Leftrightarrow \quad \bar{\boldsymbol{\pi}} P^Y = \bar{\boldsymbol{\pi}}\,, \tag{1.43}$$

where $\bar{\pi}_i = \pi_i^* |g_{ii}|$ *for all* $i \in S$.

Proof. Assume finite state space S and that all $|g_{ii}| > 0$.

$$\begin{aligned}
\boldsymbol{\pi}^* G = (0,\ldots,0) \quad &\Rightarrow \quad \boldsymbol{\pi}^* G^k = (0,\ldots,0) \quad \text{for all } k \geq 1 \\
&\Rightarrow \quad \sum_{k=1}^{\infty} \frac{t^k}{k!} \boldsymbol{\pi}^* G^k = (0,\ldots,0) \quad \text{for all } t \geq 0 \\
&\Rightarrow \quad \boldsymbol{\pi}^* \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k = \boldsymbol{\pi}^* \quad \text{for all } t \geq 0 \text{ since } G^0 = Id \\
&\Rightarrow \quad \boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^* \quad \text{for all } t \geq 0\,.
\end{aligned} \tag{1.44}$$

The backward direction follows directly from differentiating $\boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^*$ at $t = 0$.
By (1.39) we can write g_{ij} in terms of the entries of P^Y, $\quad g_{ij} = |g_{ii}|(p_{ij}^Y - \delta_{ij})$, and so

$$(\bar{\boldsymbol{\pi}} P^Y)_j - \bar{\pi}_j = \sum_{i\in S} \bar{\pi}_i (p_{ij}^Y - \delta_{ij}) = \sum_{i\in S} \pi_i^* g_{ij} = (\boldsymbol{\pi}^* G)_j\,, \tag{1.45}$$

and both sides vanish equivalently. □

Theorem 1.11 *A continuous time irreducible Markov chain with finite state space has a unique stationary distribution $\boldsymbol{\pi}^*$, where*

$$\pi_i^* = \frac{1}{\mu_i |g_{ii}|} \quad \textit{with} \quad \mu_i := \mathbb{E}(T_i | X_0 = i) \quad \textit{and} \quad T_i := \inf\{t \geq J_1 : X_t = i\}\,. \tag{1.46}$$

This follows immediately from Theorem 1.4 for discrete time by the tie-up with stationary measures of the jump chain (Proposition 1.10). Note that T_i is still the recurrence time for the jump chain Y. This forces a slightly different definition in terms of X, and T_i is often called a *first passage time*. Then π_i^* is determined by the average fraction of time the chain spends in state i,

$$\pi_i^* = \frac{1}{\mu_i |g_{ii}|} = \frac{\mathbb{E}(W_i)}{\mu_i} \quad \text{with the expected holding time } \mathbb{E}(W_i) = 1/|g_{ii}|\,. \tag{1.47}$$

The *detailed balance* conditions for a continuous-time Markov chain are

$$\pi_i g_{ij} = \pi_j g_{ji} \quad \text{for all } i, j \in S\,. \tag{1.48}$$

If they are fulfilled for a distribution $\boldsymbol{\pi}$, then $\boldsymbol{\pi}$ is stationary since every term in the right-hand side of (1.42) vanishes individually.

Examples.

Theorem 1.12 *An irreducible Markov chain with finite state space is ergodic, i.e.*

$$p_{ij}(t) \to \pi_j^* \quad \textit{as } t \to \infty\,, \quad \textit{for all } i,j \in S\,. \tag{1.49}$$

Again, this follows directly by ergodicity of the jump chain (Theorem 1.7), and it implies

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, P(t) \to \boldsymbol{\pi}^* \quad \text{as } t \to \infty\,, \tag{1.50}$$

for every initial distribution $\boldsymbol{\pi}(0)$.

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

$$\text{if } i \to j \quad \text{then} \quad p_{ij}(t) > 0 \quad \text{for all } t > 0\,. \tag{1.51}$$

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

$$g_{ii_1} g_{i_1 i_2} \cdots g_{i_{n-1} j} > 0 \quad \text{for some } i_1, \ldots i_{n-1} \in S, n \in \mathbb{N}\,, \tag{1.52}$$

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

1.4 Remarks on 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 defined in (1.46) for continuous and in (1.14) for discrete time.

Definition 1.6 A state $i \in S$ is called

$$\begin{aligned} \textit{transient}, \text{ if} \quad & \mathbb{P}(T_i = \infty) > 0 \,, \\ \textit{null recurrent}, \text{ if} \quad & \mathbb{P}(T_i < \infty) = 1 \quad \text{and} \quad \mathbb{E}(T_i) = \infty \,, \\ \textit{positiv recurrent}, \text{ if} \quad & \mathbb{P}(T_i < \infty) = 1 \quad \text{and} \quad \mathbb{E}(T_i) < \infty \,. \end{aligned} \tag{1.53}$$

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.

1.5 The ergodic theorem

Theorem 1.14 *Let $X = (X_t, t \in \mathbb{T})$ be an ergodic Markov chain with unique stationary distribution π^*. Then for every $f : S \to \mathbb{R}$ (observable)*

$$\frac{1}{t}\int_0^t f(X_s)\,ds \quad or \quad \frac{1}{N}\sum_{n=0}^{N} f(X_n) \quad \longrightarrow \quad \mathbb{E}_{\pi^*}(f) \quad as\ t, N \to \infty\,. \tag{1.54}$$

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 $f = \mathbb{1}_i$ the right-hand side of (1.54) 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).

2 Processes with continuous state space

2.1 Brownian motion and the Central limit theorem

Let $Y_1, Y_2, \ldots \in \mathbb{R}$ be iidrvs with mean $\mathbb{E}(Y_i) = 0$ and variance $\mathrm{var}(Y_i) = \sigma^2 > 0$. Then define the discrete-time process

$$X_n := \sum_{i=1}^{n} Y_i \quad \text{with} \quad X_0 = 0 \,. \tag{2.1}$$

For example if $Y_i \sim U\big(\{-1, 1\}\big)$ then X is a simple symmetric random walk. Then by the **Central Limit Theorem** (CLT) as $n \to \infty$

$$\frac{X_n}{\sqrt{n}} \to \xi \sim N(0, \sigma^2) \qquad \text{(Gaussian rv with mean 0 and variance } \sigma^2) \,, \tag{2.2}$$

or, equivalently, for all $y \in \mathbb{R}$

$$\mathbb{P}\Big(\frac{X_n}{\sqrt{n}} \le y\Big) \to \int_{-\infty}^{y} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-x^2/(2\sigma^2)}}_{f_X(x)}\, dx \quad \text{with Gaussian pdf} \quad f_X(x) \,. \tag{2.3}$$

We can use the CLT to look at the process X_n in rescaled time $t = n\Delta t$. According to (2.2), $X_{[t/\Delta t]}/\sqrt{1/\Delta t}$ should converge to a t-dependent random variable as $\Delta t \to 0$, and we define

$$B_t := \lim_{\Delta t \to 0} \sqrt{\Delta t}\, X_{[t/\Delta t]} = \lim_{\Delta t \to 0} \frac{\sqrt{t}}{\sqrt{t/\Delta t}} \sum_{i=1}^{[t/\Delta t]} Y_i = \sqrt{t}\, \xi_t \sim N(0, t\sigma^2) \,. \tag{2.4}$$

Here the $\xi_t \sim N(0,1)$ are different for each t, but they are certainly not independent. Note that by the CLT the time rescaling induces a space rescaling

$$\boxed{t = \Delta t\, n \,, \quad b = (\Delta t)^\alpha x \quad \text{with} \quad \alpha = 1/2} \,, \tag{2.5}$$

and on all other spatial scales, the limiting process does either not exist or is degenerate,

$$B_t = 0 \quad \text{for } \alpha > 1/2 \,, \quad B_t \text{ is not well defined for } \alpha < 1/2 \,. \tag{2.6}$$

Properties of the process $B = (B_t : t \geq 0)$:

$$B_0 \;=\; 0\,, \quad B_t \sim N(0, t\sigma^2) \quad \text{and analogously to (2.4)}$$
$$B_t - B_s \;=\; \lim_{\Delta t \to 0} \sqrt{\Delta t} \sum_{i=[s/\Delta t]}^{[t/\Delta t]} Y_i \sim N(0, (t-s)\sigma^2) \tag{2.7}$$

for all $t \geq s \geq 0$. So B has *stationary increments*, i.e. $B_t - B_s \sim B_{t-s} - B_0$, and by independence of the Y_i, B has *independent increments*, i.e.

$$B_t - B_s \quad \text{is independent of} \quad \{B_u : u \leq s\} \quad \text{for all } t \geq s \geq 0\,. \tag{2.8}$$

So far B is only the result of an informal derivation, an important question is wether it actually exists as a mathematical object.

Theorem 2.1 Existence of Brownian motion (Wiener)
There exists a process $B = (B_t : t \geq 0)$ with stationary independent increments, such that $B_0 = 0$ and $B_t \sim N(0, t)$.

Proof. see e.g. Rogers and Williams, Section I.6

What are the regularity properties of a sample path?

From (2.4) we expect for Brownian motion

$$B_{t+h} - B_t = \sqrt{h}\,\xi \sim N(0, h\sigma^2) \to 0 \quad \text{as } h \to 0\,. \tag{2.9}$$

Therefore Brownian sample paths are continuous (and more precisely, Hölder continuous with index $1/2$). But they are nowhere differentiable, since

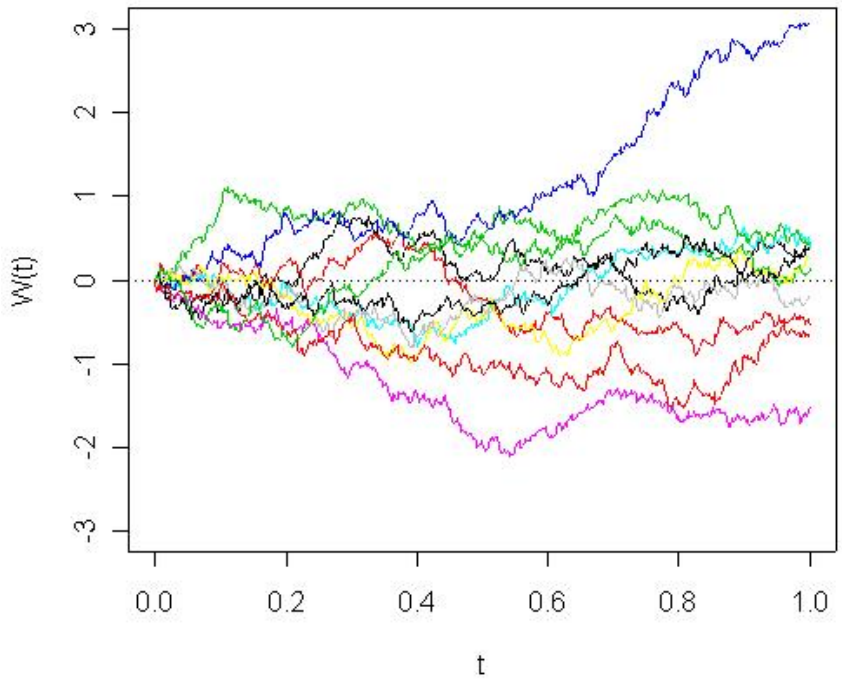
$$\frac{B_{t+h} - B_t}{h} = \frac{\sigma}{\sqrt{h}}\,\xi \quad \text{has no limit as } h \to 0\,. \tag{2.10}$$

These properties do not follow from (2.7) and (2.8), which can be fulfilled also by discontinuous processes. But under the restriction that $t \mapsto B_t(\omega)$ is a continous function of t for all $\omega \in \Omega$, BM is unique. So we restrict ourselves to the *path space* $C\big([0,\infty), \mathbb{R}\big)$ of continuous functions, and the process can then be described by a probability measure on that space.

Theorem 2.2 Uniqueness of Brownian motion (Wiener)
*There exists a unique probability measure $\mathcal{W}$ on the path space $C\big([0,\infty), \mathbb{R}\big)$ (called the **Wiener measure**), such that the process with sample paths distributed according to $\mathcal{W}$ has the properties of 2.1.*

Proof. see e.g. Rogers and Williams, Section I.6

Examples of sample paths.



Definition 2.1 A real-valued process $B = (B_t : t \geq 0)$ with continuous paths and stationary independent increments, such that $B_0 = 0$ and $B_t \sim N(0, t)$, is called a *Brownian motion* or *Wiener process*.

It suffices to look at *standard* BMs B with $\sigma^2 = 1$ and $B_0 = 0$, then $\sqrt{\sigma}B + x_0$ is a BM with variance σ^2 starting in x_0. All distributional properties of BM are characterized by the finite dimensional distributions.

Proposition 2.3 *Let B be a standard BM. For all $t_1, \ldots, t_n$, $n \in \mathbb{N}$ the vector*

$$(B_{t_1}, \ldots, B_{t_n}) \sim N(\mathbf{0}, \Gamma) \quad \textit{with} \quad \gamma_{ij} = \min\{t_i, t_j\} \,, \tag{2.11}$$

has multivariate Gaussian distribution with zero means and covariance matrix $\Gamma = (\gamma_{ij})_{i,j}$.

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

$$\mathbb{E}(B_s B_t) = \mathbb{E}\big(B_s^2 + B_s(B_t - B_s)\big) = \mathbb{E}(B_s^2) + 0 \,, \tag{2.12}$$

since B has independent increments and $\mathbb{E}(B_s) = 0$. Thus $\mathrm{cov}(B_s, B_t) = \mathrm{var}(B_s) = s$. □

Reminder. The pdf of the multivariat Gaussian $(B_{t_1}, \ldots, B_{t_n})$ is given by

$$f_{t_1,..,t_n}(\mathbf{x}) = \frac{1}{(2\pi \det \Gamma)^{n/2}} \exp\big(-\tfrac{1}{2}\,\mathbf{x}\,\Gamma^{-1}\,\mathbf{x}^T\big) \quad \text{with} \quad \mathbf{x} = (x_1, \ldots, x_n) \,. \tag{2.13}$$

Remark. Note that if the increments Y_i in (2.2) are not identically distributed or independent, the CLT still holds under more general conditions (see e.g. Gardiner, Section 2.8.2). So Brownian motion is the natural scaling limit for a very general class of models.

Definition 2.2 A *d-dimensional standard Brownian motion* $\mathbf{B} = (\mathbf{B}_t : t \geq 0)$ is a collection of d independent one-dimensional BMs $B^1, \ldots, B^d$ as given in Def. 2.1, i.e.

$$\mathbf{B}_t = (B_t^1, \ldots, B_t^d) \quad \text{for all } t \geq 0 \,. \tag{2.14}$$

So the pdf of the increments $\mathbf{B}_t - \mathbf{B}_s$ is $\quad f_{t-s}(\mathbf{x}) = (2\pi(t-s))^{-d/2} \exp\Big(- \frac{\|\mathbf{x}\|_2^2}{2(t-s)} \Big)$.

Analogous to the random walk, one can study recurrence and transience for BM depending on the space dimension.

Theorem 2.4 *Let $T_A = \inf\{t > 0 : \mathbf{B}_t \in A,\, B_s \notin A \text{ for some } s < t\}$ be the first passage time for the set $A \subseteq \mathbb{R}^d$, analogous to the times T_x for points $x \in \mathbb{R}^d$.*

*(i) If $d = 1$ BM is **point-recurrent**, i.e. for all $x \in \mathbb{R}$*

$$\mathbb{P}\big(\{T_x < \infty \big|\, B_0 = x\big) = 1 \,. \tag{2.15}$$

*(ii) If $d = 2$, BM is **neighbourhood-recurrent**, i.e. for every $\epsilon > 0$, $x \in \mathbb{R}^2$*

$$\mathbb{P}\big(T_{B(x,\epsilon)} < \infty \big|\, \mathbf{B}_0 = x\big) = 1 \quad \textit{where} \quad B(x,\epsilon) = \{y \in \mathbb{R}^d : \|y\| < \epsilon\} \,. \tag{2.16}$$

*However, points are **polar**, i.e. for all $x \in \mathbb{R}^2$*

$$\mathbb{P}(T_x = \infty) = 1 \,, \quad \textit{independent of the starting point} \,. \tag{2.17}$$

(iii) If $d \geq 3$, BM is transient, i.e. $|\mathbf{B}_t| \to \infty$ as $t \to \infty$ with probability one.

Proof. see e.g. Rogers and Williams, Section I.18

Proposition 2.5 *For dimension $d \geq 2$, the image $\{\mathbf{B}_t : t \geq 0\} \subseteq \mathbb{R}^d$ of the sample path of a BM $\mathbf{B}$ has **Hausdorff** (or **fractal**) **dimension** 2.*

'Proof'.

2.2 General facts

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.

$$\mathbb{P}(X \in A) = \int_A dF(x) = \int_A f(x)\, dx \quad \text{where } f = F' \text{ is the pdf (if it exists)} . \tag{2.18}$$

Technical side remark (non-examinable): 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 $dF(x)$ can be consistently defined on.

As for Markov chains, the distributional properties of a general stochastic process are determined by fixing all *finite-dimensional distributions (fdds)*

$$F_{\mathbf{t}}(\mathbf{x}) = \mathbb{P}\big(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n\big) , \tag{2.19}$$

for all $\mathbf{t} = (t_1, \ldots, t_n) \in [0, \infty)^n$, $t_i \neq t_j$, $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $n \in \mathbb{N}$. We focus here on the state space $S = \mathbb{R}$ where the fdds are given by joint distribution functions $F_{\mathbf{t}}$ as above. In principle this can be extended to more general state spaces.

Theorem 2.6 *If a collection $\{F_{\mathbf{t}}\}$ of fdds fulfills the* ***Kolmogorov consistency relations***

$$\begin{aligned} F_{\mathbf{t},t_{n+1}}(\mathbf{x}, x_{n+1}) &\to F_{\mathbf{t}}(\mathbf{x}) \quad \textit{as} \quad x_{n+1} \to \infty\,, \quad \textit{and} \\ F_{\Pi\mathbf{t}}(\Pi\mathbf{x}) &= F_{\mathbf{t}}(\mathbf{x}) \quad \textit{for all permutations } \Pi \textit{ of } (1,\dots,n)\,, \end{aligned} \tag{2.20}$$

then there exists a stochastic process $X = (X_t : t \geq 0)$ (on some probability space Ω) that has fdds $\{F_{\mathbf{t}}\}$.

Proof. Is related to the Skorohod representation theorem. Basically one takes Ω to be the path space of the process. Some hints are given in [GS] Section 8.6

Example.

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

Definition 2.3 Let X be a stochastic process on $\mathbb{R}$. The conditional distribution function

$$F(t, x|s, y) = \mathbb{P}(X_t \leq x | X_s = y)\,, \tag{2.21}$$

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

$$f(t, x|s, y) = \frac{\partial F}{\partial x}(t, x|s, y)\,. \tag{2.22}$$

Note that for a homogeneous process, the kernel is actually only a function of $t - s$.

Proposition 2.7 *The fdds of a Markov process are uniquely determined by the transition kernels and the initial distribution.*

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

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

□

Example.

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).

In the previous section we saw that the sample paths of BM are continuous. Many interesting phenomena cannot be modeled with continuous processes alone, but one usually concentrates on the following class of processes.

Definition 2.4 A real-valued, continuous-time process X is called *càdlàg* if its sample paths are right continuous (continue à droite) and have left limits (limite à gauche), i.e.

$$\lim_{s \searrow t} X_s(\omega) = X_t(\omega) \quad \text{and} \quad \lim_{s \nearrow t} X_s(\omega) \text{ exists}, \quad \text{for all } \omega \in \Omega,\ t \in [0, \infty) \,. \tag{2.23}$$

For example continuous-time Markov chains (e.g. the Poisson process) are defined as càdlàg.

2.3 Brownian motion and the heat equation

We are looking for an evolution equation for the transition densities, analogous to the forward equation (or master equation) for Markov chains. First we will derive it for Brownian motion as scaling limit from the simple random walk.

Let $(X_n : n \in \mathbb{N})$ be a simple random walk. Then the distribution at time n is given by $\boldsymbol{\pi}(n+1) = \boldsymbol{\pi}(n)\, P$, which can be written in the following incremental form

$$\boldsymbol{\pi}(n+1) - \boldsymbol{\pi}(n) = \boldsymbol{\pi}(n)(P - Id) \,, \tag{2.24}$$

where $P - Id$ is proportional to the discrete Laplacian Δ,

$$P - Id = \begin{pmatrix} \ddots & \ddots & \ddots & & \\ & \frac{1}{2} & -1 & \frac{1}{2} & \\ & & \ddots & \ddots & \ddots \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \ddots & \ddots & \ddots & & \\ & 1 & -2 & 1 & \\ & & \ddots & \ddots & \ddots \end{pmatrix} = \frac{1}{2}\Delta \,. \tag{2.25}$$

In the previous section we saw that under the scaling $\boxed{t = \Delta t\, n,\ x = (\Delta t)^\alpha k \text{ with } \alpha = 1/2}$, $(\Delta t)^\alpha X_{[t/\Delta t]} \to B_t$ converges to Brownian motion as $\Delta t \to 0$. Therefore the mass function

$\pi_k(n)$ should converge to the pdf $f(t,x)$ of B_t, i.e.

$$\lim_{\Delta t\to 0}\ \frac{1}{(\Delta t)^\alpha}\pi_{x/(\Delta t)^\alpha}(t/\Delta t) = f(t,x) = (2\pi t)^{-1/2}\exp\big(-x^2/(2t)\big)\ . \tag{2.26}$$

where $1/(\Delta t)^\alpha$ is the volume element to turn the probability $\pi_k(n)$ into a density.

Plugging the scaling into the discrete-time Master equation (2.24), we can derive a differential equation for f. We assume that for large n,k (i.e. small Δt), $\pi_k(n)$ is approximately given by

$$\pi_k(n)/(\Delta t)^\alpha \simeq f\big(n\Delta t, k(\Delta t)^\alpha\big) = f(t,x)\ . \tag{2.27}$$

Then we get by Taylor expansion

$$\begin{aligned}\pi_{k\pm1}(n)/(\Delta t)^\alpha\ &\simeq\ f(t,x) \pm (\Delta t)^\alpha\frac{\partial}{\partial x}f(t,x) + \frac{(\Delta t)^{2\alpha}}{2}\frac{\partial^2}{\partial x^2}f(t,x) + O\big((\Delta t)^{3\alpha}\big)\\ \pi_k(n+1)/(\Delta t)^\alpha\ &\simeq\ f(t,x) + \Delta t\frac{\partial}{\partial t}f(t,x) + O\big((\Delta t)^2\big)\ .\end{aligned} \tag{2.28}$$

Thus if $\alpha = 1/2$ (otherwise the limit is again degenrate),

$$\begin{aligned}\frac{\partial}{\partial t}f(t,x)\ &=\ \lim_{\Delta t\to 0}\frac{\pi_k(n+1)-\pi_k(n)}{(\Delta t)^{1+\alpha}} = \lim_{\Delta t\to 0}\frac{1}{2(\Delta t)^{1+\alpha}}\big(\pi_{k-1}(n) - 2\pi_k(n) + \pi_{k+1}(n)\big) =\\ &=\ \lim_{\Delta t\to 0}\frac{(\Delta t)^{2\alpha}}{2\Delta t}\frac{\partial^2}{\partial x^2}f(t,x) + O\big((\Delta t)^{3\alpha-1}\big) = \frac{1}{2}\frac{\partial^2}{\partial x^2}f(t,x)\ .\end{aligned} \tag{2.29}$$

So since standard BM starts in the origin, its pdf fulfills

$$\frac{\partial}{\partial t}f(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}f(t,x) \quad \text{with initial condition} \quad f(0,x) = \delta_0(x)\ . \tag{2.30}$$

This PDE is the so-called *heat equation* which has been well studied and indeed (2.26) is its unique solution.

Note that with the implicit initial condition $f(t,x) = f(t,x|0,0)$ in terms of transition densities. An analogous derivation conditioned on $B_s = y$ gives the same equation for $f(t,x|s,y)$ with the more general initial condition $f(s,x|s,y) = \delta_y(x)$.

Indeed, as we have seen before $B_t \sim N(0, t-s)$ for $t \geq s$, and therefore the transition density is given by

$$f(t,x|s,y) = \big(2\pi(t-s)\big)^{-1/2} \exp\Big(-\frac{(x-y)^2}{2(t-s)}\Big) \,. \tag{2.31}$$

$f(t,x|s,y)$ is also called the *heat kernel*, since it is the fundamental solution to that PDE (2.30). That means that for every intial distribution $f(0,y)$ we have

$$f(t,x) = \int_{\mathbb{R}} f(t,x|0,y)\, f(0,y)\, dy \,. \tag{2.32}$$

We can also derive (2.30) from the forward equation $\frac{d}{dt}P(t) = P(t)\,G$ or the master equation (1.42) of a continuous-time Markov chain, by rescaling only space as $x = \epsilon k$ with $\epsilon \to 0$. In these derivations the exact structure of the generator G or $P - Id$ is not important and this equation holds for a whole class of processes, including e.g. symmetric jumps with finite range.

2.4 Diffusion processes and Fokker-Planck equations

Definition 2.5 A Markov process X is called a *diffusion process*, if it has continuous sample paths and as $h \searrow 0$

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t = x\big) &= a(t,x)\,h + o(h) \,, \\ \mathbb{E}\big((X_{t+h} - X_t)^2 \,\big|\, X_t = x\big) &= b(t,x)\,h + o(h) \,, \end{aligned} \tag{2.33}$$

for some functions $a(t,x)$ (*drift coefficient*) and $b(t,x)$ (*diffusion coefficient*).

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

Theorem 2.8 *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** (or **forward equation**)*

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}\big(a(t,x)\,f\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(t,x)\,f\big) \tag{2.34}$$

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

$$f(t,x) = \int_{\mathbb{R}} f(t,x|0,y)\, f(0,y)\, dy \tag{2.35}$$

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

Proof. by Taylor expansion similar to Section 2.3

Examples.

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 (2.34), i.e.

$$0 = -\frac{\partial}{\partial x}\big(a(x)\, f^*(x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(x)\, f^*(x)\big)\ . \tag{2.36}$$

Integrating (2.36) and denoting the derivative by $'$ we get under mild regularity assumptions on $a(x)$ and $b(x)$ (e.g. boundedness)

$$\begin{aligned} 0 \ &= \ -\int_{-\infty}^{x} \big(a(y)\, f^*(y)\big)' dy + \frac{1}{2}\int_{-\infty}^{x} \big(b(y)\, f^*(y)\big)'' dy = \\ &= \ -a(x)\, f^*(x) + \frac{1}{2}\big(b(x)\, f^*(x)\big)' \quad (+const.) \end{aligned} \tag{2.37}$$

This is a first order linear differential equation and differentiating with the product rule we get

$$f^{*\prime}(x) = \frac{2a(x) - b'(x)}{b(x)}\, f^*(x)\ . \tag{2.38}$$

So the solution is

$$f^*(x) = f^*(0) \exp\Big(\int_0^x \frac{2a(y) - b'(y)}{b(y)}\, dy \Big) \tag{2.39}$$

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

Examples.

Diffusion processes can be generalized to higher dimensions. $\mathbf{X}$ in $\mathbb{R}^d$ is called a diffusion process if in addition to the continuity property analogous to Definition 2.5

$$\begin{aligned} \mathbb{E}\big(\mathbf{X}_{t+h} - \mathbf{X}_t \,\big|\, \mathbf{X}_t = \mathbf{x}\big) &= \mathbf{a}(t,\mathbf{x})\, h + o(h)\ , \\ \mathbb{E}\big((\mathbf{X}_{t+h} - \mathbf{X}_t)(\mathbf{X}_{t+h} - \mathbf{X}_t)^T \,\big|\, \mathbf{X}_t = \mathbf{x}\big) &= b(t,\mathbf{x})\, h + o(h)\ , \end{aligned} \tag{2.40}$$

with drift vector $\mathbf{a}(t,x) \in \mathbb{R}^d$ and diffusion matrix $b \in \mathbb{R}^{d\times d}$, where

$$b_{ij} = \mathbb{E}\big((X^i_{t+h} - X^i_t)(X^j_{t+h} - X^j_t) \,\big|\, \mathbf{X}_t = \mathbf{x}\big)\ . \tag{2.41}$$

This is the covariance matrix of the increments of the process. The Fokker-Planck equation for $f = f(t, \mathbf{x})$ is now given by

$$\frac{\partial f}{\partial t} \;=\; -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}\big(a_i(t,\mathbf{x})\, f\big) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}\big(b_{ij}(t,\mathbf{x})\, f\big) = \mathcal{L}^* f \;. \tag{2.42}$$

where the right-hand side defines a linear operator $\mathcal{L}^*$ on the set of functions $f : \mathbb{R}^d \to \mathbb{R}$. $\mathcal{L}^*$ is called the *(adjoint) generator* of the process $\mathbf{X}$ and is the analogous quantity of the generator of a continuous-time Markov chain.

Definition 2.6 Let $\mathbf{X}$ be a diffusion process with time-independent drift $\mathbf{a}(\mathbf{x})$ and diffusion $b(\mathbf{x})$. $V : \mathbb{R}^d \to \mathbb{R}$ is called a *potential* for $\mathbf{X}$, if $\mathbf{a}(\mathbf{x}) = -\nabla V(\mathbf{x})$. If $b_{ij}(\mathbf{x}) = b\,\delta_{ij}$ we call $\mathbf{X}$ a (noise-perturbed) *gradient flow*.

The Fokker-Planck equation of a gradient flow is given by

$$\frac{\partial f(t,\mathbf{x})}{\partial t} = \nabla \cdot \big((\nabla V(\mathbf{x}))\, f(t,\mathbf{x})\big) + \frac{b}{2}\,\Delta f(t,\mathbf{x}) \;. \tag{2.43}$$

Examples.

Proposition 2.9 *Assume that $V : \mathbb{R}^d \to \mathbb{R}$ is smooth and that*

$$Z := \int_{\mathbb{R}^d} e^{-2V(\mathbf{x})/b}\, d^d x < \infty \;. \tag{2.44}$$

*Then the diffusion process $\mathbf{X}$ with (2.42) is ergodic. The unique stationary distribution is the **Gibbs distribution** with density*

$$f^*(\mathbf{x}) = \frac{1}{Z}\, e^{-2V(\mathbf{x})/b} \;, \tag{2.45}$$

*and the normalization factor Z is called **partition function**.*

Proof. We have from (2.45) $\nabla f^* = -\frac{2}{b}(\nabla V)\, f^*$ and thus

$$\frac{b}{2}\Delta f^* = \frac{b}{2}\nabla \cdot (\nabla f^*) = -\nabla \cdot \big((\nabla V)\, f^*\big) \;. \tag{2.46}$$

Substituting this in (2.43) the right-hand side vanishes $\mathcal{L}^* f^* = 0$, and f^* is stationary.
Uniqueness and ergodicity follow from the fact gradient flows fulfill general conditions such that the Fokker-Planck equation (2.43) has a unique (time-dependent) solution. □

Examples.

2.5 Further remarks (non-examinable)

- Defining the *probability current* $\mathbf{J}(f)$ with i-th component

$$J_i(\mathbf{x}, f) := a_i(\mathbf{x})\, f - \frac{1}{2} \sum_{j=1}^{d} \frac{\partial}{\partial x_j} \big(b_{ij}(\mathbf{x})\, f\big)\,, \tag{2.47}$$

  the Fokker-Planck equation (2.42) can be written as a *continuity equation*

$$\frac{\partial f(t,\mathbf{x})}{\partial t} + \nabla_{\mathbf{x}} \cdot \mathbf{J}\big(\mathbf{x}, f(t,\mathbf{x})\big) = 0. \tag{2.48}$$

  Integrating this equation over a domain $A \subseteq \mathbb{R}^d$ and using integration by parts like above we get

$$\frac{\partial}{\partial t} \int_A f(t,\mathbf{x})\, d^d x = -\int_A \nabla_{\mathbf{x}} \cdot \mathbf{J}\big(\mathbf{x}, f(t,\mathbf{x})\big)\, d^d x = -\int_{\partial A} \mathbf{J}\big(\mathbf{x}, f(t,\mathbf{x})\big) \cdot d\mathbf{S} \tag{2.49}$$

The second identity follows from *Stokes' theorem* (also called *Gauss' integration theorem*). If $A = \mathbb{R}^d$ or the system is *closed* in A then $\mathbf{J}\big(\mathbf{x}, f(t,\mathbf{x})\big) = \mathbf{0}$ for all $\mathbf{x} \in \partial A$. So the right-hand side of (2.49) vanishes and the total probability is conserved, i.e.

$$\mathbb{P}(\mathbf{X}_t \in A) = \int_A f(t,\mathbf{x})\, d^d x = 1\ . \tag{2.50}$$

An important class of diffusion processes with direct connections to statistical mechanics are noise-perturbed gradient flows.

- $f = f(t,x|s,y)$ is also the solution to the so-called *backward Fokker-Planck equation*

$$\frac{\partial f}{\partial s} = -a(s,y)\,\frac{\partial f}{\partial y} - \frac{1}{2}\,b(s,y)\,\frac{\partial^2 f}{\partial y^2} \tag{2.51}$$

which can be derived from the backward equation of a continuous time MC.

- Let $g : \mathbb{R}^d \to \mathbb{R}$ be an observable, such as $g(\mathbf{X}_t) = \|\mathbf{X}_t\|_2^2$. Then the expected value

$$\bar{g}(t) := \mathbb{E}\big(g(\mathbf{X}_t)\big) = \int_{\mathbb{R}^d} g(\mathbf{x})\, f(t,\mathbf{x})\, d^d x \tag{2.52}$$

obeys the following evolution equation,

$$\begin{aligned} \frac{d}{dt}\bar{g}(t) \;&=\; \int_{\mathbb{R}^d} g(\mathbf{x})\,\frac{\partial f(t,\mathbf{x})}{\partial t}\, d^d x = \int_{\mathbb{R}^d} g(\mathbf{x})\,(\mathcal{L}^* f)(t,\mathbf{x})\, d^d x = \\ &=\; \int_{\mathbb{R}^d} (\mathcal{L} g)(\mathbf{x})\, f(t,\mathbf{x})\, d^d x = \overline{\mathcal{L} g}(t) = \mathbb{E}\big((\mathcal{L} g)(\mathbf{X}_t)\big)\ . \end{aligned} \tag{2.53}$$

This follows by partial integration, since for each $i = 1, \ldots, d$

$$\int_{\mathbb{R}^d} g(\mathbf{x}) \frac{\partial}{\partial x_i}\big(a_i(t,\mathbf{x})\, f(t,\mathbf{x})\big) d^d x = -\int_{\mathbb{R}^d} \Big(\frac{\partial}{\partial x_i} g(\mathbf{x})\Big) a_i(t,\mathbf{x})\, f(t,\mathbf{x}) d^d x\ , \tag{2.54}$$

because $f(t,\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$ so there are no boundary terms. For the diffusion part this can be done twice and leads to

$$\mathcal{L} = \sum_{i=1}^d a_i(t,\mathbf{x}) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d b_{ij}(t,\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j}\ . \tag{2.55}$$

This operator is called the *generator* of the process $\mathbf{X}$ and describes the expected time evolution of observables. Note that this also determines the right-hand side of the backward Fokker-Planck equation (2.51). It is technically more convenient than $\mathcal{L}^*$ and therefore diffusion processes are often characterized by defining their generator.

- For time-independent drift $\mathbf{a}(\mathbf{x})$ and diffusion $b(\mathbf{x})$ existence and uniqueness of the *initial value problem*

$$\frac{\partial f(t,\mathbf{x})}{\partial t} = (\mathcal{L}^* f)(t,\mathbf{x})\ , \quad f(0,\mathbf{x}) = f_0(\mathbf{x})\ , \tag{2.56}$$

is well understood under the assumption of *uniform ellipticity*, i.e.

$$\xi\, b(\mathbf{x})\, \xi^T = \sum_{i,j=1}^{d} b_{ij}(\mathbf{x})\, \xi_i\, \xi_j \geq \alpha \|\xi\|_2^2 \quad \text{for some } \alpha > 0 \text{ and all } \xi \in \mathbb{R}^d\,. \tag{2.57}$$

Theorem 2.10 *Under the assumption (2.57) and the growth conditions*

$$f_0(\mathbf{x}) \leq C e^{\alpha \|\mathbf{x}\|_2^2}, \quad \Big|\frac{\partial a_i(\mathbf{x})}{\partial x_i}\Big| \leq C_1\big(1 + \|\mathbf{x}\|_2^2\big), \quad \Big|\frac{\partial^2 b_{ij}(\mathbf{x})}{\partial x_i \partial x_j}\Big| \leq C_2\big(1 + \|\mathbf{x}\|_2^2\big) \tag{2.58}$$

for some constants $C, C_1, C_2 > 0$, the initial value problem (2.56) has a unique ***classical*** *$C^{1,2}\big((0,\infty), \mathbb{R}^d\big)$ solution.*

3 Some stochastic calculus

3.1 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

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t = x\big) &= a(t,x)\,h + o(h)\,, \\ \mathbb{E}\big((X_{t+h} - X_t)^2 \,\big|\, X_t = x\big) &= \sigma^2(t,x)\,h + o(h)\,. \end{aligned} \tag{3.1}$$

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

$$Y = \mu + \sigma\xi \quad \text{where} \quad \xi = \frac{X-\mu}{\sigma}\,. \tag{3.2}$$

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

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t\big) &= a(t,X_t)\,h + o(h)\,, \\ \mathrm{var}\big(X_{t+h} - X_t \,\big|\, X_t\big) &= \sigma^2(t,X_t)\,h - a(t,X_t)^2\,h^2 + o(h) = \sigma^2(t,X_t)\,h + o(h)\,. \end{aligned} \tag{3.3}$$

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

$$X_{t+h} - X_t = a(t,X_t)\,h + \sigma(t,X_t)\,\sqrt{h}\,\xi_{t,t+h} + o(h)\,. \tag{3.4}$$

Then

$$\mathbb{E}(\sqrt{h}\,\xi_{t,t+h}) = 0 \quad \text{and} \quad \mathrm{var}(\sqrt{h}\,\xi_{t,t+h}) = h\,, \tag{3.5}$$

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

$$\xi_{t,t+h} = \sum_{k=1}^{n} \xi_{t+(k-1)/n,\,t+k/n} \tag{3.6}$$

can be written as a sum of arbitrarily many independent random variables with mean 0 and variance 1. 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

$$X_{t+h} - X_t = a(t,X_t)\,h + \sigma(t,X_t)(B_{t+h} - B_t) + o(h) \quad \text{for a BM } B\,. \tag{3.7}$$

Deviding by h we get in the limit $h \to 0$

$$\frac{dX_t}{dt} = a(t,X_t) + \sigma(t,X_t)\,\frac{dB_t}{dt}\,. \tag{3.8}$$

This is a differential equation for each path of X, i.e. for fixed $\omega \in \Omega$. But paths of a BM are not differentiable and therefore (3.8) is often written as

$$dX_t = a(t,X_t)\,dt + \sigma(t,X_t)\,dB_t\,. \tag{3.9}$$

Definition 3.1 (3.9) is called a *stochastic differential equation (SDE)* with *drift* $a(t,x)$ and *diffusion* $\sigma(t,x)$. Alternatively, physicists often write

$$\frac{dX_t}{dt} = a(t,X_t) + \sigma(t,X_t)\,\eta_t\,, \tag{3.10}$$

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

The (non-existing) derivative $\eta_t = dB_t/dt$ can be understood as a normalized random force term on X uncorrelated in time. Formally it is given by a Gaussian process with mean 0 and covariance $\delta(t-s)\,Id$, which makes sense if integrated over time.

As for ordinary differential equations, it is often better to look at the integrated version of (3.9), since it requires less regularity assumptions.

Definition 3.2 A continuous process $X = (X_t : t \geq 0)$ is a *solution* of the SDE (3.9) with initial condition $X_0 = x_0$ if

$$X_t = x_0 + \int_0^t a(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \tag{3.11}$$

holds with probability 1 for all $t \geq 0$ (or $t \in [0,T]$). The solution is called *unique*, if any two solutions X^1 and X^2 satisfy

$$\mathbb{P}\big(X_t^1 = X_t^2 \text{ for all } t \geq 0\big) = 1 \qquad \big(\text{ or } t \in [0,T] \text{ resp.}\big)\,. \tag{3.12}$$

So in order to solve SDEs we have to make sense of the two *stochastic integrals* in (3.11).
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

$$0 = t_0 < t_1 < \ldots < t_n = t \quad \text{with} \quad t_k - t_{k-1} \to 0 \text{ for all } k = 1, \ldots, n, \text{ as } n \to \infty \tag{3.13}$$

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

$$I_t^n = \sum_{k=1}^{n} Y_{t_{k-1}} (X_{t_k} - X_{t_{k-1}}) \to I_t = \int_0^t Y_s\, dX_s \quad \text{as } n \to \infty\,. \tag{3.14}$$

This is a (particular) Riemann sum approximation of the integrals in (3.11), 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 $t_k = k/n$ to fix ideas.

3.2 Stochastic integration and Itô calculus

Proposition 3.1 *For the integrator $X_t = t$ and continuous integrand Y the limit in (3.14) exists pointwise and we can define*

$$I_t(\omega) = \int_0^t Y_s(\omega)\, ds := \lim_{n\to\infty} I_t^n(\omega) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_{t_{k-1}}(\omega)\,. \tag{3.15}$$

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.

$$\sum_{k=1}^{n} \big|X_{t_k}(\omega) - X_{t_{k-1}}(\omega)\big| = t < \infty \quad \text{for all } t \geq 0\,. \tag{3.16}$$

Examples.

Theorem 3.2 Itô integral
Let Y be a continuous process and $X = B$ a standard BM. If

$$\mathbb{E}\bigg(\int_0^t Y_s^2 ds\bigg) < \infty \tag{3.17}$$

for some $t \geq 0$, then

$$I_t = \int_0^t Y_s\, dB_s := \lim_{n\to\infty} I_t^n = \lim_{n\to\infty} \sum_{k=1}^{n} Y_{t_{k-1}}(B_{t_k} - B_{t_{k-1}}) \tag{3.18}$$

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 (3.17) 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. Rogers and Williams

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

$$\int_{t_0}^{t} B_s \, dB_s = \frac{1}{2}\big((B_t^2 - B_{t_0}^2) - (t - t_0)\big) \, . \tag{3.19}$$

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

$$S\int_{t_0}^{t} B_s \, dB_s = \frac{1}{2}(B_t^2 - B_{t_0}^2) \, , \tag{3.20}$$

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 (see later). 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 (3.19) 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,

$$I_t = I_{t_0} + \int_{t_0}^{t} Y_s \, dX_s \quad \Leftrightarrow \quad dI_t = Y_t \, dX_t \, . \tag{3.21}$$

For example for the integral (3.19) this gives $B_t \, dB_t = \frac{1}{2}(dB_t^2 - dt)$, leading to the rule

$$\boxed{d(B_t^2) = 2B_t \, dB_t + dt} \, . \tag{3.22}$$

This is basically a modified *chain rule* for Itô calculus. The meaning of the term dt becomes clear if we compute an increment by hand:

Comparing with the chain rule (3.22) we get

$$\boxed{(dB_t)^2 = dt} \qquad \text{or more generally} \quad (\sigma dB_t)^2 = \sigma^2 dt \; . \tag{3.23}$$

In usual calculus higher order terms are of negligible order $o(dt)$, but for BM they have to be taken into account.

Non-examinable remarks:
With considerable technical effort, the Itô integral can be generalized to non-continuous processes. But there is a particularly important class of integrators for which this can be done quite easily.

Def. Let $N = (N_t : t \geq 0) \sim PP(\lambda)$ and $Z_1, Z_2, \ldots$ a sequence of iidrv's with $\mathbb{E}\big(|Z_k|\big) < \infty$ and distribution function F. Then

$$Q = (Q_t : t \geq 0) \quad \text{with} \quad Q_t = \sum_{k=1}^{N_t} Z_k \tag{3.24}$$

is called a *compound Poisson process* or *jump process*. For $a, \sigma \in \mathbb{R}$ and B a standard BM a process of the form

$$X_t = X_0 + a\,t + \sigma B_t + Q_t \quad \text{with stationary, independent increments} \tag{3.25}$$

is called a *Lévy process*. X is completely determined by the *Lévy triple* $(a, \sigma^2, \lambda F)$.

Lévy processes are the simplest generalization of Brownian motion to non-continuous processes with jumps. By definition $t \mapsto Q_t(\omega)$ is piecewise constant for each ω, so the integral can be computed directly. The condition of stationary, independent increments implies that F cannot be arbitrary, but has to be an *infinitely divisible law*. Examples are Gaussians or α-stable laws.

3.3 Diffusion processes and Itô's formula

Let X be a solution of the SDE

$$dX_t = a(t, X_t)\, dt + \sigma(t, X_t)\, dB_t\ . \tag{3.26}$$

The following very useful result summarizes our findings in Section 3.2 and gives an explicit formula for time evolution of an observable $g(X_t)$.

Theorem 3.3 Itô's formula
Let X be a solution of (3.26) and $g \in C^2(\mathbb{R}, \mathbb{R})$. Then

$$dg(X_t) = g'(X_t)\, dX_t + \frac{1}{2} g''(X_t)\, \sigma^2\, dt\ , \tag{3.27}$$

or in the (extended) integrated version

$$\begin{aligned} g(X_t) \ &= \ g(X_0) + \int_0^t g'(X_s)\, \sigma(s, X_s) dB_s \\ &\quad + \int_0^t \Big(g'(X_s)\, a(s, X_s) + \tfrac{1}{2} g''(X_s)\, \sigma^2(s, X_s) \Big) ds\ . \end{aligned} \tag{3.28}$$

Proof. Taylor expansion with terms up to order dt, using (3.23) and the Itô chain rule (3.22).

Examples.

In detail Itô's formula says in incremental form

$$\begin{aligned} dg(X_t) \ &= \ g'(X_t)\, dX_t + \frac{1}{2}\, g''(X_t)\, (dX_t)^2 = \\ &= \ g'(X_t) \big(a(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \big) + \frac{1}{2}\, g''(X_t) \sigma^2(t, X_t)\, dt\ . \end{aligned} \tag{3.29}$$

Let $f(t,x)$ be the pdf of the process X introduced in Section 2. Taking the expectation on both sides, we get by partial integration

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\big(g(X_t)\big) &= \int_{\mathbb{R}} g(x)\,\frac{\partial}{\partial t}f(t,x)\,dx = \\ &= \int_{\mathbb{R}} \Big(g'(x)\,a(t,x) + \frac{1}{2}g''(x)\sigma^2(t,x)\Big)\,f(t,x)\,dx = \\ &= \int_{\mathbb{R}} g(x)\bigg(-\frac{\partial}{\partial x}\big(a(t,x)\,f(t,x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(\sigma^2(t,x)\,f(t,x)\big)\bigg)dx\ , \quad (3.30)\end{aligned}$$

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

$$\frac{\partial}{\partial t}f(t,x) = -\frac{\partial}{\partial x}\big(a(t,x)\,f(t,x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(\sigma^2(t,x)\,f(t,x)\big) = \mathcal{L}^*f(t,x)\ . \quad (3.31)$$

Thus $f(t,x)$ fulfilles the Fokker-Planck equation with adjoint generator $\mathcal{L}^*$ introduced in (2.42) and X is a diffusion process with drift $a(t,x)$ and diffusion $\sigma^2(t,x)$. In Section 3.1 we have derived the opposite statement, so let us summarize both of them.

Proposition 3.4 *X is a diffusion process as defined in Def. 2.5 if and only if it is a solution of the SDE (3.26) with $\sigma^2(t,x) = b(t,x)$.*

Non-examinable remarks.
Remember that the generator $\mathcal{L}$ introduced in (2.55) is given by

$$\mathcal{L} = a(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2}{\partial x^2} \,, \tag{3.32}$$

and describes the time evolution of expected values of observables (2.53). Using $\mathcal{L}$ we can rewrite Itô's formula

$$g(X_t) = g(X_0) + \int_0^t (\mathcal{L}g)(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dB_s \,. \tag{3.33}$$

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.
Another interesting question is, how many solutions to the SDE (3.26) there are. For time-independent drift and diffusion there is a general theorem about existence and uniqueness for SDEs.

Theorem 3.5 *Suppose that $a : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are Lipschitz-continuous, i.e.*

$$\big|a(x) - a(y)\big| \leq K|x-y| \quad \textit{for some } K > 0 \textit{ and all } x, y \in \mathbb{R} \,. \tag{3.34}$$

Then (for each probability space Ω and BM B) there exists a unique solution to the SDE

$$dX_t = a(X_t)\,dt + \sigma(X_t)\,dB_t \tag{3.35}$$

for any initial condition $X_0 = x_0 \in \mathbb{R}$.

Proof. analogous to ordinary differential equations using the contraction mapping theorem and Gronwall's Lemma.

There are more elaborate notions of 'solutions' to SDEs in a weak and strong sense, which we did not discuss. A proper understanding of these involves a substantial amount of measure theory. Itô's formula and the existence and uniqueness theorem can be extended to higher space dimensions. There is also a Stratonovich interpretation of SDEs which is directly connected to the Itô version given here. Both can be found in Gardiner, Section 4.3.