Stochastic Processes

CO905 - Complexity Science

Stefan Grosskinsky

Warwick, 2008

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

Contents

Introduction **2**

1 Markov chains **4**
1.1 General properties 4
1.2 Discrete time 6
1.3 Continuous time 10
1.4 Countably infinite state spaces 15

2 Processes with continuous state space **19**
2.1 Brownian motion and the Central limit theorem 19
2.2 General facts 23
2.3 Brownian motion and the heat equation 26
2.4 Diffusion processes and Fokker-Planck equations 27

3 Martingales **34**
3.1 Filtrations and adapted processes 34
3.2 Properties of martingales 37

4 Stochastic calculus **40**
4.1 Diffusion processes and SDEs 40
4.2 Semimartingales 42
4.3 Stochastic integration and Itô calculus 46
4.4 Solutions to SDEs 50

References

[GS] G. Grimmett and D. Stirzaker: Probability and Random Processes (3rd edition), Oxford 2001

[Ga] C.W. Gardiner: Handbook of Stochastic Methods (3rd edition), Springer 2004

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 Grimmett and Stirzaker, 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 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'*

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

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

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

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

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

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 since is uniquely determined by the initial distribution $\boldsymbol{\pi}(0)$ and the transition matrix P.

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.

$$\lim_{t \searrow 0} P(t) = P(0) = Id \quad \text{and} \quad G := \lim_{t \searrow 0} \frac{P(t) - Id}{t} \quad \text{exists}\,. \tag{1.8}$$

Together with the Chapman-Kolmogorov equations this implies that

$$\frac{P(t+\Delta t) - P(t)}{\Delta t} = \frac{P(\Delta t) - Id}{\Delta t}\,P(t) = P(t)\frac{P(\Delta t) - Id}{\Delta t}\,, \tag{1.9}$$

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

$$\boxed{\frac{d}{dt}P(t) = G\,P(t) = P(t)\,G \quad \Rightarrow \quad P(t) = \exp(t\,G)\,.} \tag{1.10}$$

For finite state spaces $|S| < \infty$ the formal solution to this equation is given by a matrix exponential which is discussed in Section 2.3 in more detail.

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 1.5 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 $j \neq i$.

Examples.

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

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

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. $p_{ij} \in [0, 1]$ and $\sum_{j \in S} p_{ij} = 1$, since the chain starting in i has to jump somewhere in S. So $\lambda = 1$ 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*

$$\pi_i^* = 1/\mu_i \quad \textit{with} \quad \mu_i := \mathbb{E}(T_i | X_0 = i) \quad \textit{and} \quad T_i := \min\{t \geq 1 : X_t = i\}\,, \tag{1.11}$$

*so that $\boldsymbol{\pi}^*$ is determined by the inverse of the **mean recurrence times** μ_i.*
Furthermore,

$$\boldsymbol{\pi}^* = \mathbf{1}\,(Id - P + U)^{-1}\,, \quad \textit{where} \quad u_{ij} = 1 \textit{ for all } i,j \in S\,. \tag{1.12}$$

Proof. see Grimmett and Stirzaker pp 229 - 230 for recurrence times.
for uniqueness see proof of Theorem 1.8.

$$\begin{aligned} \boldsymbol{\pi}^* = \boldsymbol{\pi}^* P \quad &\Leftrightarrow \quad \boldsymbol{\pi}^*(Id - P) = \mathbf{0} \quad \Leftrightarrow \quad \boldsymbol{\pi}^*(Id + P) + \mathbf{1} = \mathbf{1} \\ &\Leftrightarrow \quad \boldsymbol{\pi}^*(Id - P + U) = \mathbf{1} \end{aligned} \tag{1.13}$$

By uniqueness of $\boldsymbol{\pi}^*$ we know that this linear system of equations has a unique solution, which is equivalent to $Id - P + U)$ being invertible. □

Example.

Proposition 1.5 *Let $X = \big(X_n : n \in \{0,\dots,N\}\big)$ 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 with $Y_n = X_{N-n}$ is a Markov chain with transition matrix*

$$p^Y_{ij} = \frac{\pi^*_j}{\pi^*_i}\, p^X_{ji} \quad \textit{for all } i,j \in S\,. \tag{1.14}$$

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

$$\begin{aligned} &\mathbb{P}\big(Y_{n+1} = i_{n+1} \,\big|\, Y_n = i_n, \dots, Y_0 = i_0\big) = \\ &\quad = \frac{\mathbb{P}(Y_k = i_k, 0 \le k \le n+1)}{\mathbb{P}(Y_k = i_k, 0 \le k \le n)} = \frac{\mathbb{P}(X_{N-k} = i_k, 0 \le k \le n+1)}{\mathbb{P}(X_{N-k} = i_k, 0 \le k \le n)} = \\ &\quad = \frac{\pi^*_{i_{n+1}} p_{i_{n+1} i_n} \cdots p_{i_1 i_0}}{\pi^*_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 i_0}} = \frac{\pi^*_{i_{n+1}} p_{i_{n+1} i_n}}{\pi^*_{i_n}} \end{aligned} \tag{1.15}$$

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

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all } i,j \in S\,. \tag{1.16}$$

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

Note that Proposition 1.5 together with (1.16) implies that a reversible Markov chain and its time-reversal are indistinguishable, i.e. they have the same transition probabilities, since

$$p^Y_{ij} = \frac{\pi^*_j}{\pi^*_i}\, p^X_{ji} = \frac{\pi^*_i}{\pi^*_i}\, p^X_{ij} = p^X_{ij} \ . \tag{1.17}$$

The detailed balance relations (1.16) can be a useful tool to find stationary distributions of certain Markov chains 'without loops'.

Example.

Definition 1.5 The *period* $d(i)$ of a state $i \in S$ is defined as

$$d(i) := gcd\big\{t \geq 1 \,:\, p_{ii}(t) > 0\big\} \ , \tag{1.18}$$

the greatest common divisor of the epochs at which return is possible.

Theorem 1.7 *For an irreducible Markov chain all states have the same period, and we call the chain aperiodic if $d(i) = 1$ for all $i \in S$.*

Proof. see p.224 in Grimmett and Stirzaker

Example.

Theorem 1.8 *An irreducible, aperiodic Markov chain with finite state space is* ***ergodic****, i.e.*

$$p_{ij}(n) \to \pi_j^* \quad \textit{as } t \to \infty\,, \quad \textit{for all } i,j \in S\,. \tag{1.19}$$

Proof. The statement follows from 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 eigenvalue of P,

$$\lambda_1 = w^0 = 1,\ \lambda_2 = w^1, \ldots, \lambda_d = w^{d-1} \quad \textit{where} \quad w = e^{2\pi i/d}\,, \tag{1.20}$$

(ii) and the remaining eigenvalues $\lambda_{d+1}, \ldots, \lambda_{|S|}$ satisfy $|\lambda_j| < 1$.
Note that this includes uniqueness of the stationary distribution claimed in Theorem 1.4.
Suppose further that the eigenvalues are all distinct, then $B\,P\,B^{-1} = (\lambda_i \delta_{ij})_{ij}$ is a diagonal matrix with entries $\lambda_1, \ldots, \lambda_{|S|}$, where the rows of B are the left eigenvectors of P. Thus

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

since $\lambda_1 = 1$ and $|\lambda_i| < 1$ for all $i > 1$ if the chain is aperiodic.
The proof can be extended to countably infinite S. □

This implies that for every initial distribution

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\, P^n \to \boldsymbol{\pi}^* = (1/\mu_1, \ldots, 1/\mu_{|S|}) \quad \text{as } n \to \infty\,. \tag{1.22}$$

Example.

1.3 Continuous time

Subject to the boundary conditions $P(0) = Id$, equations (1.10) often have a unique solution

$$\frac{d}{dt}P(t) = G\,P(t) = P(t)\,G \quad \Rightarrow \quad P(t) = \exp(t\,G) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\,G^k\ . \tag{1.23}$$

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 (1.23)

$$p_{ij}(\Delta t) = g_{ij}\Delta t + o(\Delta t) \quad \text{for all } i \neq j \in S\ . \tag{1.24}$$

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

$$p_{ii}(\Delta t) = 1 + g_{ii}\Delta t + o(\Delta t) \quad \text{for all } i \in S\ , \tag{1.25}$$

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

$$1 = \sum_{j\in S} p_{ij}(\Delta t) = 1 + \Delta t \sum_{j\in S} g_{ij} \quad \Rightarrow \quad \sum_{j\in S} g_{ij} = 0 \quad \text{for all } i \in S\ . \tag{1.26}$$

Therefore the diagonal entries of G are

$$g_{ii} = -\sum_{j\neq i} g_{ij} \leq 0 \quad \text{for all } I \in S\ , \tag{1.27}$$

and $|g_{ii}|$ 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*

$$W_i := \inf\left\{t' \geq 0\ :\ X_{t+t'} \neq i\right\}\ , \tag{1.28}$$

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

Proposition 1.9 *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 $s, u > 0$

$$\begin{aligned}\mathbb{P}(W_i > s+u | W_i > s) \ &= \ \mathbb{P}(W_i > s+u | X_{t+s} = i) = \\ &= \ \mathbb{P}(W_i > u | X_t = i) = \mathbb{P}(W_i > u)\ ,\end{aligned} \tag{1.29}$$

where we have used the Markov property and homogeneity. Therefore we have for the tail

$$\bar{F}(s+u) = \mathbb{P}(W_i > s+u) = \mathbb{P}(W_i > s+u, W_i > s)\ , \tag{1.30}$$

and thus $\bar{F}(s+u) = \bar{F}(s)\,\bar{F}(u)$.
Analogous to the Chapman-Kolmogorov equations (1.5) this can be used to derive a differential equation for $\bar{F}$ which has an exponential solution

$$\bar{F}(s) = \mathbb{P}(W_i > s) = e^{\lambda s} \quad \text{where} \quad \lambda = \bar{F}'(0)\ . \tag{1.31}$$

Together with (1.25) we get

$$\bar{F}'(0) = \lim_{\Delta t \searrow 0} \frac{\mathbb{P}(W_i > \Delta t) - 1}{\Delta t} = \lim_{\Delta t \searrow 0} \frac{p_{ii}(\Delta t) + o(\Delta t) - 1}{\Delta t} = g_{ii} \leq 0\ , \tag{1.32}$$

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 $(s, s+\Delta t]$, is given by

$$\mathbb{P}(X_{t+s+\Delta t} = j | X_{t+s} = i, W_i < \Delta t) \simeq \frac{p_{ij}(\Delta t)}{1 - p_{ii}(\Delta t)} \to \frac{g_{ij}}{-g_{ii}} \quad \text{as } \Delta t \searrow 0\ . \tag{1.33}$$

Picture. □

The chain jumps at the *jump time* $J_n = \sum_{i=0}^{n-1} W_{Y_i}$ 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

$$p_{ij}^Y = \begin{cases} 0 & , i = j \\ g_{ij}/|g_{ii}| & , i \neq j \end{cases} \quad \text{if } g_{ii} > 0\ , \quad \text{and} \quad p_{ij}^Y = \delta_{ij} \quad \text{if } g_{ii} = 0\ . \tag{1.34}$$

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

$$X_{t+t'} - X_{t'} \text{ is distributed like } X_t - X_0 \text{ and independent of } (X_s : s \leq t') \,, \tag{1.35}$$

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 equation (1.23) we can also get an evolution equation for the distribution,

$$\frac{d}{dt}\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, \frac{d}{dt} P(t) = \boldsymbol{\pi}(0)\, P(t)\, G = \boldsymbol{\pi}(t)\, G \,. \tag{1.36}$$

This is called the *Master equation* and using (1.27) the coordinate form is given by

$$\boxed{\frac{d}{dt}\pi_i(t) = \sum_{j\neq i}\Big(\pi_j(t)\,g_{ji} - \pi_i(t)\,g_{ij}\Big)\,.} \tag{1.37}$$

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

Proposition 1.11 *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*

$$\boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^* \quad \Leftrightarrow \quad \boldsymbol{\pi}^* G = \mathbf{0} \quad \Leftrightarrow \quad \bar{\boldsymbol{\pi}} P^Y = \bar{\boldsymbol{\pi}}\,, \tag{1.38}$$

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

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

$$\begin{aligned}
\boldsymbol{\pi}^* G = (0,\ldots,0) \quad &\Leftrightarrow \quad \boldsymbol{\pi}^* G^k = (0,\ldots,0) \quad \text{for all } k \geq 1 \\
&\Leftrightarrow \quad \sum_{k=1}^{\infty} \frac{t^k}{k!}\boldsymbol{\pi}^* G^k = (0,\ldots,0) \quad \text{for all } t \geq 0 \\
&\Leftrightarrow \quad \boldsymbol{\pi}^* \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k = \boldsymbol{\pi}^* \quad \text{for all } t \geq 0 \text{ since } G^0 = Id \\
&\Leftrightarrow \quad \boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^* \quad \text{for all } t \geq 0\,.
\end{aligned} \tag{1.39}$$

By (1.34) 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

$$(\bar{\boldsymbol{\pi}} P^Y)_j - \bar{\pi}_j = \sum_{i\in S} \bar{\pi}_i (p_{ij}^Y - \delta_{ij}) = \sum_{i\in S} \pi_i^* g_{ij} = (\boldsymbol{\pi}^* G)_j\,, \tag{1.40}$$

and both sides vanish equivalently. □

Theorem 1.12 *A continuous time irreducible Markov chain with finite state space has a unique stationary distribution $\boldsymbol{\pi}^*$, where*

$$\pi_i^* = \frac{1}{\mu_i |g_{ii}|} \quad \textit{with} \quad \mu_i := \mathbb{E}(T_i | X_0 = i) \quad \textit{and} \quad T_i := \inf\{t \geq J_1 : X_t = i\}\,. \tag{1.41}$$

This follows immediately from Theorem 1.4 for discrete time by the tie-up with stationary measures of the jump chain (Proposition 1.11). 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,

$$\pi_i^* = \frac{1}{\mu_i |g_{ii}|} = \frac{\mathbb{E}(W_i)}{\mu_i} \quad \text{with the expected holding time } \mathbb{E}(W_i) = 1/|g_{ii}|\,. \tag{1.42}$$

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

$$\pi_i g_{ij} = \pi_j g_{ji} \quad \text{for all } i, j \in S\,. \tag{1.43}$$

If they are fulfilled for a distribution $\boldsymbol{\pi}$, then $\boldsymbol{\pi}$ is stationary since every term in the right-hand side of (1.37) vanishes individually.

Examples.

Theorem 1.13 *An irreducible Markov chain with finite state space is ergodic, i.e.*

$$p_{ij}(t) \to \pi_j^* \quad \textit{as } t \to \infty\,, \quad \textit{for all } i,j \in S\,. \tag{1.44}$$

Again, this follows directly by ergodicity of the jump chain (Theorem 1.8), and it implies

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, P(t) \to \boldsymbol{\pi}^* = \Big(\frac{\mathbb{E}(W_1)}{\mu_1}, \dots, \frac{\mathbb{E}(W_{|S|})}{\mu_{|S|}}\Big) \quad \text{as } t \to \infty\,, \tag{1.45}$$

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

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

$$\text{if } i \to j \quad \text{then} \quad p_{ij}(t) > 0 \quad \text{for all } t > 0\,. \tag{1.46}$$

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

$$g_{ii_1} g_{i_1 i_2} \cdots g_{i_{n-1} j} > 0 \quad \text{for some } i_1, \dots i_{n-1} \in S, n \in \mathbb{N}\,, \tag{1.47}$$

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

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.

Examples.

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

$$\begin{aligned} &\textit{recurrent,} \text{ if} \quad \mathbb{P}\big(\{t \geq 0 : X_t = i\} \text{ is unbounded} \,\big|\, X_0 = i\big) = 1 \quad \text{and} \\ &\textit{transient,} \text{ if} \quad \mathbb{P}\big(\{t \geq 0 : X_t = i\} \text{ is unbounded} \,\big|\, X_0 = i\big) = 0 \,. \end{aligned} \tag{1.48}$$

Proposition 1.14 *Let $i \in S$ be a non-absorbing state. Then*

$$\begin{aligned} &\mathbb{P}(T_i < \infty) \;=\; 1 \quad \Leftrightarrow \quad \sum_{n=0}^{\infty} p_{ii}(n) \textit{ or } \int_0^{\infty} p_{ii}(t) = \infty \quad \Rightarrow i \textit{ recurrent} \,, \\ &\mathbb{P}(T_i < \infty) \;<\; 1 \quad \Leftrightarrow \quad \sum_{n=0}^{\infty} p_{ii}(n) \textit{ or } \int_0^{\infty} p_{ii}(t) < \infty \quad \Rightarrow i \textit{ transient} \,. \end{aligned} \tag{1.49}$$

Each state is either recurrent or transient. If the chain is irreducible then the states are either all recurrent or all transient.

Proof. see Section 6.2 in Grimmett and Stirzaker

Examples.

Apparently recurrence and transience of a Markov chain is not enough to characterize the existence of a stationary distribution $\boldsymbol{\pi}^*$ completely. If it exists, the entries $\pi_i^* \propto 1/\mu_i$ are inversely proportional to the mean recurrence times $\mu_i = \mathbb{E}(T_i|X_0 = i)$. If the chain is transient, then certainly $\mu_i = \infty$ for all i, $\boldsymbol{\pi}^*$ does not exist and in fact $\pi_i(t) \to 0$.

Definition 1.7 A recurrent state $i \in S$ is called *positive recurrent* if $\mu_i < \infty$ and *null recurrent* if $\mu_i = \infty$.

Theorem 1.15 *Let X be an irreducible (non-explosive[1]) Markov chain. Then X has a unique stationary distribution if and only if it is positive recurrent.*

So the positive recurrent (non-explosive) Markov chains behave exactly like chains with finite state space concerning their stationary distributions. This also holds for dynamic properties and the convergence to equilibrium.

[1] explained at the end of this section

Examples.

Therefore, X can be positive recurrent while the corresponding jump chain Y is null recurrent. However, we have the following connection.

Proposition 1.16 *Let X be a (non-explosive) Markov chain with jump chain Y. Then*

$$\begin{aligned} &i \in S \text{ is transient for } X \quad \Leftrightarrow \quad i \text{ is transient for } Y\ , \\ &i \in S \text{ is recurrent for } X \quad \Leftrightarrow \quad i \text{ is recurrent for } Y\ . \end{aligned} \tag{1.50}$$

Transient continuous time chains can get lost at infinity even in finite time. This phenomenon is called *explosion*. Define the *explosion time*

$$J_\infty := \lim_{n\to\infty} J_n = \sum_{i=1}^{\infty} W_i \in (0,\infty]\ . \tag{1.51}$$

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.

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 $n \to \infty$

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

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 $B = (B_t : t \geq 0)$:

$$B_0 \;=\; 0\,, \quad B_t \sim N(0, t\sigma^2) \quad \text{and analogously to (2.4)}$$

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

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 $B_t \sim N(0, t)$. B is called a **standard Brownian motion (BM)** or **Wiener process**.*

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

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

$$(B_{t_1}, \ldots, B_{t_n}) \sim N(\mathbf{0}, \Gamma) \quad \textit{with} \quad \gamma_{ij} = \min\{t_i, t_j\}\,, \tag{2.9}$$

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 $\operatorname{cov}(B_s, B_t) = \min\{s, t\}$. Take $s < t$, then

$$\mathbb{E}(B_s B_t) = \mathbb{E}\big(B_s^2 + B_s(B_t - B_s)\big) = \mathbb{E}(B_s^2) + 0\,, \tag{2.10}$$

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

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

$$f_{t_1,..,t_n}(\mathbf{x}) = \frac{1}{(2\pi \det \Gamma)^{n/2}} \exp\big(-\tfrac{1}{2}\,\mathbf{x}\,\Gamma^{-1}\,\mathbf{x}^T\big) \quad \text{with} \quad \mathbf{x} = (x_1, \ldots, x_n)\,. \tag{2.11}$$

What are the regularity properties of a Brownian sample path?

From (2.4) we expect for Brownian motion

$$B_{t+h} - B_t = \sqrt{h}\,\xi \sim N(0, h\sigma^2) \to 0 \; a.s. \quad \text{as } h \to 0\,. \tag{2.12}$$

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

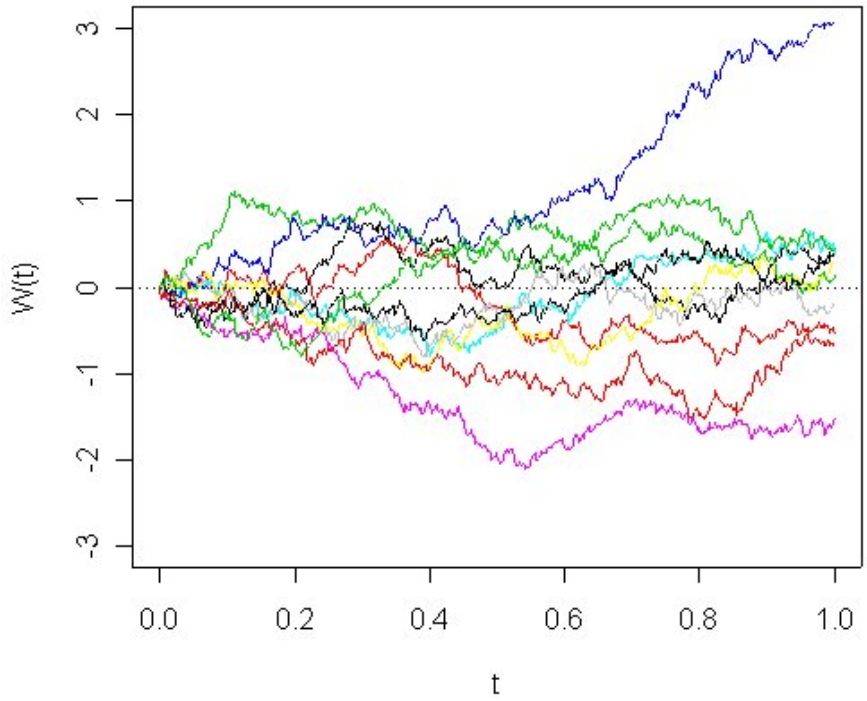
$$\frac{B_{t+h} - B_t}{h} = \frac{\sigma}{\sqrt{h}}\,\xi \quad \text{has no limit as } h \to 0\,. \tag{2.13}$$

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 2.3 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 $\mathcal{W}$ is a Brownian motion as defined in Theorem 2.1*

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

Examples of sample paths.



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 limiting process for a very general class of models.

Definition 2.1 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 defined in Theorem 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 *(i) If $d = 1$ BM is **point-recurrent**, i.e.*

$$\mathbb{P}\big(\{t \geq 0 : B_t = 0\} \text{ is unbounded} \big| \, B_0 = 0\big) = 1 \, . \tag{2.15}$$

*(ii) If $d = 2$, BM is **neighbourhood-recurrent**, i.e. for every $\epsilon > 0$*

$$\mathbb{P}\big(\{t \geq 0 : |\mathbf{B}_t| < \epsilon\} \text{ is unbounded} \big| \, \mathbf{B}_0 = 0\big) = 1 \, . \tag{2.16}$$

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

$$\mathbb{P}(T_x = \infty) = 1 \, , \quad \text{where} \quad T_x = \inf\{t > 0 : \mathbf{B}_t = x\} \, . \tag{2.17}$$

(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'. see class

2.2 General facts

In this section we discuss 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 of a random variable X, e.g.

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

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 socalled *σ-algebra*. This is a set of *measurable sets* where the measure $dF(x)$ can be consistently defined on.

Remember also that a random variable is actually a *measurable function* $X : \Omega \to \mathbb{R}$, i.e. for each measurable $A \subseteq \mathbb{R}$ the preimage $X^{-1}(A) = \{\omega : X(\omega) \in A\}$ is measurable with respect to $\mathbb{P}$ on Ω. We always us the shorthand

$$\mathbb{P}(X \in A) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\} , \tag{2.19}$$

and the probability space Ω is hidden from the discussion.

Definition 2.2 Let $X, Y : \Omega \to \mathbb{R}$ be random variables and $A \subseteq \mathbb{R}$ measurable. Then we say

$$\begin{aligned}
&X \in A \quad \textit{almost surely (a.s.)}, \quad \text{if} \quad \mathbb{P}(X \in A) = 1 \quad \text{and} \\
&X = Y \quad a.s., \quad \text{if} \quad \mathbb{P}(X = Y) = \mathbb{P}\big(\{\omega : X(\omega) = Y(\omega)\}\big) = 1 \quad \text{and} \\
&X \sim Y \quad \text{if} \quad F_X(x) = \mathbb{P}(X \le x) = F_Y(x) \quad \text{for all } x \in \mathbb{R} .
\end{aligned} \tag{2.20}$$

Example.

So in general we have $X = Y\ a.s. \quad \Rightarrow \quad X \sim Y$.

Similar concepts exist for convergence of random variables,

$$\begin{aligned}
&X_n \to X \quad \text{means } \textit{convergence in distribution}, \text{ i.e. } \quad F_{X_n}(x) \to F_X(x) , \\
&X_n \to X\ a.s. \quad \text{means } \textit{almost sure convergence}, \text{ i.e. } \quad \mathbb{P}(X_n \to X) = 1 .
\end{aligned} \tag{2.21}$$

(Note that convergence in distribution is required only for all $x \in \mathbb{R}$ at which F is continuous due to technical reasons, see e.g. Grimmett and Stirzaker, Section 7.2.)
In general, 'almost sure' is stronger then 'in distribution' and we have

$$X = Y \ a.s. \quad \Rightarrow \quad X \sim Y \ , \quad X_n \to X \ a.s. \quad \Rightarrow \quad X_n \to X \ . \tag{2.22}$$

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

$$F_{\mathbf{t}}(\mathbf{x}) = \mathbb{P}\big(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n\big) \ , \tag{2.23}$$

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 $n \in \mathbb{N}$. For state space $S = \mathbb{R}$ these are characterized by joint distribution functions $F_{\mathbf{t}}$.

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

$$\begin{aligned} F_{\mathbf{t},t_{n+1}}(\mathbf{x}, x_{n+1}) \ &\to \ F_{\mathbf{t}}(\mathbf{x}) \quad \textit{as} \quad x_{n+1} \to \infty \ , \quad \textit{and} \\ F_{\Pi\mathbf{t}}(\Pi\mathbf{x}) \ &= \ F_{\mathbf{t}}(\mathbf{x}) \quad \textit{for all permutations } \Pi \textit{ of } (1, \ldots, n) \ , \end{aligned} \tag{2.24}$$

then there exists a prob. space Ω and a process $X = (X_t : t \geq 0)$ on 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 Grimmett and Stirzaker, Section 8.6

Example.

So Brownian motion is an example of a *Gaussian process*.

Definition 2.3 A real-valued, continuous-time process X is called a *Gaussian process* if each finite-dimensional vector $(X_{t_1}, \ldots, X_{t_n}) \sim N\big(\boldsymbol{\mu}(t), V(t)\big)$ is Gaussian with mean vector $\boldsymbol{\mu}$ and covariance matrix V, which may depend on t.

Analogous to Proposition 2.2, all finite dimensional distributions of a Gaussian process are uniquely specified by its mean and covariance matrix (which has to be positive definite). But note that Gaussian processes are not necessarily Markov. Standard Brownian motion is a Gaussian process with zero mean and covariance $\min\{s, t\}$.

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

Definition 2.4 Let X be a stochastic process. The conditional distribution function

$$F(t,x|s,y) = \mathbb{P}(X_t \leq x | X_s = y) \;, \tag{2.25}$$

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

$$f(t,x|s,y) = \frac{\partial F}{\partial x}(t,x|s,y) \;. \tag{2.26}$$

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.

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 2.5 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.

$$\lim_{s \searrow t} X_s(\omega) = X_t(\omega) \quad \text{and} \quad \lim_{s \nearrow t} X_s(\omega) \text{ exists} \;, \quad \text{for all } \omega \in \Omega,\ t \in [0,\infty) \;. \tag{2.27}$$

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

$$\boldsymbol{\pi}(n+1) - \boldsymbol{\pi}(n) = \boldsymbol{\pi}(n)(P - Id)\,, \tag{2.28}$$

where $P - Id$ is proportional to the discrete Laplacian $\Delta : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$,

$$P - Id = \begin{pmatrix} \ddots & \ddots & \ddots & & \\ & \frac{1}{2} & -1 & \frac{1}{2} & \\ & & \ddots & \ddots & \ddots \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \ddots & \ddots & \ddots & & \\ & 1 & -2 & 1 & \\ & & \ddots & \ddots & \ddots \end{pmatrix} = \frac{1}{2}\Delta\,. \tag{2.29}$$

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.

$$\lim_{\Delta t \to 0} \pi_{x/(\Delta t)^{\alpha}}(t/\Delta t) = f(t,x) = (2\pi t)^{-1/2} \exp\big(-x^2/(2t)\big)\,. \tag{2.30}$$

Plugging the scaling into the discrete-time Master equation (2.28), 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

$$\pi_k(n) \simeq f\big(k(\Delta t)^{\alpha}, n\Delta t\big) = f(t,x)\,. \tag{2.31}$$

Then we get by Taylor expansion

$$\begin{aligned} \pi_{k\pm1}(n) \;&\simeq\; f(t,x) \pm (\Delta t)^{\alpha} \frac{\partial}{\partial x} f(t,x) + \frac{(\Delta t)^{2\alpha}}{2} \frac{\partial^2}{\partial x^2} f(t,x) + O\big((\Delta t)^{3\alpha}\big) \\ \pi_k(n+1) \;&\simeq\; f(t,x) + \Delta t \frac{\partial}{\partial t} f(t,x) + O\big((\Delta t)^2\big)\,. \end{aligned} \tag{2.32}$$

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

$$\begin{aligned} \frac{\partial}{\partial t} f(t,x) \;&=\; \lim_{\Delta t \to 0} \frac{\pi_k(n+1) - \pi_k(n)}{\Delta t} = \lim_{\Delta t \to 0} \frac{1}{2\Delta t} \big(\pi_{k-1}(n) - 2\pi_k(n) + \pi_{k+1}(n)\big) = \\ &=\; \lim_{\Delta t \to 0} \frac{(\Delta t)^{2\alpha}}{2\Delta t} \frac{\partial^2}{\partial x^2} f(t,x) + O\big((\Delta t)^{3\alpha - 1}\big) = \frac{1}{2} \frac{\partial^2}{\partial x^2} f(t,x)\,. \end{aligned} \tag{2.33}$$

So since standard BM starts in the origin, its pdf should fulfill

$$\frac{\partial}{\partial t} f(t,x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} f(t,x) \quad \text{with initial condition} \quad f(0,x) = \delta_0(x) \,. \tag{2.34}$$

This PDE is the socalled *heat equation* which has been well studied and indeed (2.30) 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 kernel $F(t,x|s,y)$ has density function

$$f(t,x|s,y) = \frac{\partial}{\partial x} F(t,x|s,y) = \big(2\pi(t-s)\big)^{-1/2} \exp\Big(-\frac{(x-y)^2}{2(t-s)}\Big) \,. \tag{2.35}$$

$f(t,x|s,y)$ is also called the *heat kernel*, since it is the fundamental solution to that PDE (2.34), i.e. for every intial distribution $f(0,y)$ we have

$$f(t,x) = \int_{\mathbb{R}} f(t,x|0,y)\, f(0,y)\, dy \,. \tag{2.36}$$

We can also derive (2.34) from the forward equation $\frac{d}{dt} P(t) = P(t)\, G$ or the master equation (1.37) 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.

2.4 Diffusion processes and Fokker-Planck equations

Definition 2.6 A Markov process X is called a *diffusion process*, if

$$\begin{aligned} \mathbb{P}\big(|X_{t+h} - X_t| > \epsilon \,\big|\, X_t = x\big) &= o(h) \quad \text{for all } \epsilon > 0,\, x \in \mathbb{R} \,, \\ \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t = x\big) &= a(t,x)\, h + o(h) \,, \\ \mathbb{E}\big((X_{t+h} - X_t)^2 \,\big|\, X_t = x\big) &= b(t,x)\, h + o(h) \,, \end{aligned} \tag{2.37}$$

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

By the first property diffusion processes have continuous sample paths. Their 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****)*

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}\big(a(t,x)\, f\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(t,x)\, f\big) \tag{2.38}$$

for all $0 \le s \le t$, $x, y \in \mathbb{R}$.

Proof. by Taylor expansion similar to Section 2.3

Examples.

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

$$\frac{\partial f}{\partial s} = -a(s,y)\,\frac{\partial f}{\partial y} - \frac{1}{2}\,b(s,y)\,\frac{\partial^2 f}{\partial y^2} \tag{2.39}$$

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

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 (2.38), i.e.

$$0 = -\frac{\partial}{\partial x}\big(a(x)\, f^*(x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(x)\, f^*(x)\big)\ . \tag{2.40}$$

Examples.

In general, integrating (2.40) and denoting the derivative by $'$ we get

$$\begin{aligned} 0 \;&=\; -\int_{-\infty}^{x} \big(a(y)\, f^*(y)\big)' dy + \frac{1}{2}\int_{-\infty}^{x} \big(b(y)\, f^*(y)\big)'' dy = \\ &=\; -a(x)\, f^*(x) + \frac{1}{2}\big(b(x)\, f^*(x)\big)' \quad (+const.) \end{aligned} \tag{2.41}$$

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

$$f^{*\prime}(x) = \frac{2a(x) - b'(x)}{b(x)}\, f^*(x)\,. \tag{2.42}$$

So the solution is

$$f^*(x) = f^*(0) \exp\Big(\int_0^x \frac{2a(y) - b'(y)}{b(y)}\, dy\Big) \tag{2.43}$$

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

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 2.6

$$\begin{aligned} \mathbb{E}\big(\mathbf{X}_{t+h} - \mathbf{X}_t \,\big|\, \mathbf{X}_t = \mathbf{x}\big) \;&=\; \mathbf{a}(t,\mathbf{x})\, h + o(h)\,, \\ \mathbb{E}\big((\mathbf{X}_{t+h} - \mathbf{X}_t) \otimes (\mathbf{X}_{t+h} - \mathbf{X}_t) \,\big|\, \mathbf{X}_t = \mathbf{x}\big) \;&=\; b(t,\mathbf{x})\, h + o(h)\,, \end{aligned} \tag{2.44}$$

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

$$b_{ij} = \mathbb{E}\big((X^i_{t+h} - X^i_t)(X^j_{t+h} - X^j_t) \,\big|\, \mathbf{X}_t = \mathbf{x}\big)\,. \tag{2.45}$$

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

$$\frac{\partial f}{\partial t} \;=\; -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}\big(a_i(t,\mathbf{x})\, f\big) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}\big(b_{ij}(t,\mathbf{x})\, f\big) = \mathcal{L}^* f \;. \tag{2.46}$$

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.

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

$$\bar{g}(t) := \mathbb{E}\big(g(\mathbf{X}_t)\big) = \int_{\mathbb{R}^d} g(\mathbf{x})\, f(t,\mathbf{x})\, d^d x \tag{2.47}$$

obeys the following evolution equation,

$$\begin{aligned} \frac{d}{dt}\bar{g}(t) \;&=\; \int_{\mathbb{R}^d} g(\mathbf{x})\, \frac{\partial f(t,\mathbf{x})}{\partial t}\, d^d x = \int_{\mathbb{R}^d} g(\mathbf{x})\, (\mathcal{L}^* f)(t,\mathbf{x})\, d^d x = \\ &=\; \int_{\mathbb{R}^d} (\mathcal{L} g)(\mathbf{x})\, f(t,\mathbf{x})\, d^d x = \overline{\mathcal{L} g}(t) = \mathbb{E}\big((\mathcal{L} g)(\mathbf{X}_t)\big) \;. \end{aligned} \tag{2.48}$$

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

$$\int_{\mathbb{R}^d} g(\mathbf{x}) \frac{\partial}{\partial x_i}\big(a_i(t,\mathbf{x})\, f(t,\mathbf{x})\big) d^d x = -\int_{\mathbb{R}^d} \Big(\frac{\partial}{\partial x_i} g(\mathbf{x})\Big) a_i(t,\mathbf{x})\, f(t,\mathbf{x}) d^d x \;, \tag{2.49}$$

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

$$\mathcal{L} = \sum_{i=1}^{d} a_i(t,\mathbf{x}) \frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} b_{ij}(t,\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} \;. \tag{2.50}$$

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 (2.39). 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*

$$\frac{\partial f(t,\mathbf{x})}{\partial t} = (\mathcal{L}^* f)(t,\mathbf{x}) \;, \quad f(0,\mathbf{x}) = f_0(\mathbf{x}) \;, \tag{2.51}$$

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

$$\xi\, b(\mathbf{x})\, \xi^T = \sum_{i,j=1}^{d} b_{ij}(\mathbf{x})\, \xi_i\, \xi_j \geq \alpha \|\xi\|_2^2 \quad \text{for some } \alpha > 0 \text{ and all } \xi \in \mathbb{R}^d \;. \tag{2.52}$$

Theorem 2.9 *Under the assumption (2.52) and the growth conditions*

$$f_0(\mathbf{x}) \leq C e^{\alpha \|\mathbf{x}\|_2^2}, \quad \Big|\frac{\partial a_i(\mathbf{x})}{\partial x_i}\Big| \leq C_1\big(1 + \|\mathbf{x}\|_2^2\big), \quad \Big|\frac{\partial^2 b_{ij}(\mathbf{x})}{\partial x_i \partial x_j}\Big| \leq C_2\big(1 + \|\mathbf{x}\|_2^2\big) \tag{2.53}$$

for some constants $C, C_1, C_2 > 0$, the initial value problem (2.51) has a unique classical $C^{1,2}\big((0,\infty), \mathbb{R}^d\big)$ solution.

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

$$J_i(\mathbf{x}, f) := a_i(\mathbf{x})\, f - \frac{1}{2} \sum_{j=1}^{d} \frac{\partial}{\partial x_j} \big(b_{ij}(\mathbf{x})\, f \big) \,, \tag{2.54}$$

the Fokker-Planck equation (2.46) can be written as a *continuity equation*

$$\frac{\partial f(t, \mathbf{x})}{\partial t} + \nabla_{\mathbf{x}} \cdot \mathbf{J}\big(\mathbf{x}, f(t, \mathbf{x})\big) = 0. \tag{2.55}$$

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

$$\frac{\partial}{\partial t} \int_A f(t, \mathbf{x})\, d^d x = - \int_A \nabla_{\mathbf{x}} \cdot \mathbf{J}\big(\mathbf{x}, f(t, \mathbf{x})\big)\, d^d x = - \int_{\partial A} \mathbf{J}\big(\mathbf{x}, f(t, \mathbf{x})\big) \cdot d\mathbf{S} \,. \tag{2.56}$$

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 (2.56) vanishes and the total probability is conserved, i.e.

$$\mathbb{P}(\mathbf{X}_t \in A) = \int_A f(t, \mathbf{x})\, d^d x = 1 \,. \tag{2.57}$$

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

Definition 2.7 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

$$\frac{\partial f(t,\mathbf{x})}{\partial t} = \nabla \cdot \big((\nabla V(\mathbf{x}))\, f(t,\mathbf{x}) \big) + \frac{b}{2}\, \Delta f(t,\mathbf{x}) \tag{2.58}$$

and the generator is

$$\mathcal{L} = -\big(\nabla V(\mathbf{x})\big) \cdot \nabla + \frac{b}{2}\, \Delta \,. \tag{2.59}$$

Examples.

Proposition 2.10 *Assume that $V : \mathbb{R}^d \to \mathbb{R}$ is smooth and that*

$$Z := \int_{\mathbb{R}^d} e^{-2V(\mathbf{x})/b}\, d^d x < \infty \,. \tag{2.60}$$

*Then the diffusion process $\mathbf{X}$ with generator (2.59) is ergodic. The unique stationary distribution is the **Gibbs distribution** with density*

$$f^*(\mathbf{x}) = \frac{1}{Z}\, e^{-2V(\mathbf{x})/b} \,, \tag{2.61}$$

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

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

$$\frac{b}{2}\Delta f^* = \frac{b}{2}\nabla \cdot (\nabla f^*) = -\nabla \cdot \big((\nabla V)\, f^*\big)\,. \tag{2.62}$$

Substituting this in (2.58) the right-hand side vanishes and f^* is stationary.
Uniqueness and ergodicity follow from the fact gradient flows fulfill the conditions such that the Fokker-Planck equation (2.58) has a unique (time-dependent) solution. □

3 Martingales

Martingales are an important class of processes in the study of stochastic differential equations. They can be discrete or continuous both in time and space.

3.1 Filtrations and adapted processes

Definition 3.1 Let $Y = \{Y_n : n \in \mathbb{N}\}$ and $X = \{X_n : n \in \mathbb{N}\}$ be sequences of real-valued random variables. Y is a *martingale* with respect to X if for all $n \in \mathbb{N}$

$$\mathbb{E}\big(|Y_n|\big) < \infty \quad \text{and} \quad \mathbb{E}\big(Y_{n+1}\big|X_0, \ldots, X_n\big) = Y_n \,. \tag{3.1}$$

Examples.

Alternatively, the conditional expectation can be defined with respect to the σ-algebra generated by X. This approach can be extended to continuous time.
A *σ-algebra* $\mathcal{F}$ on the probability space Ω is a set of *measurable* sets, that fulfills the following consistency conditions: For all $A, A_1, A_2, ... \in \mathcal{F}$

$$\emptyset \in \mathcal{F}\,, \quad A^c = \Omega \setminus A \in \mathcal{F}\,, \quad \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}\,. \tag{3.2}$$

Examples.

A *filtration* is a collection $(\mathcal{F}_t, t \in \mathbb{T})$ of sub-σ-algebras of $\mathcal{F}$ which is increasing, i.e. $s \leq t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t$. A process $X = (X_t : t \in \mathbb{T})$ is *adapted* to the filtration $(\mathcal{F}_t : t \in \mathbb{T})$ if X_t is $\mathcal{F}_t$-measurable for every t.

Example.

Interpretation: $\mathcal{F}_t$ is the total information available up to time t.

As in the above example, this can be generated by a process $X = (X_t : t \in \mathbb{T})$. The *natural filtration* for the process X is given by

$$\mathcal{F}_t^X = \sigma\big(\{X_s : s \leq t,\ s \in \mathbb{T}\}\big) \quad \text{for all } t \in \mathbb{T}\ , \tag{3.3}$$

i.e. the information generated by the process X. This is the smallest filtration X is adapted to. Then we can write for a random variable Z

$$\mathbb{E}(Z|X_0,\ldots,X_t) = \mathbb{E}(Z|\mathcal{F}_t^X)\ . \tag{3.4}$$

In general we have the following monotone behaviour with respect to the available information.

Proposition 3.1 Tower property
Let $\mathcal{F}_1 \subseteq \mathcal{F}_2$ be σ-algebras. Then for every random variable Z with $\mathbb{E}\big(|Z|\big) < \infty$,

$$\mathbb{E}\big(\mathbb{E}(Z|\mathcal{F}_2)\big|\mathcal{F}_1\big) = \mathbb{E}\big(\mathbb{E}(Z|\mathcal{F}_1)\big|\mathcal{F}_2\big) = \mathbb{E}(Z|\mathcal{F}_1)\ . \tag{3.5}$$

With this notation continuous-time and discrete-time martingales can be treated together, in the first case $\mathbb{T} = [0,\infty)$, in the second $\mathbb{T} = \mathbb{N}$.

Definition 3.2 A real-valued process $X = (X_t : t \in \mathbb{T})$ is called a *martingale* with respect to the filtration $(\mathcal{F}_t : t \in \mathbb{T})$, if it is adapted to $(\mathcal{F}_t : t \in \mathbb{T})$ and for all $t \in \mathbb{T}$

$$\mathbb{E}\big(|X_t|\big) < \infty \quad \text{and} \quad \mathbb{E}(X_t|\mathcal{F}_s) = X_s \quad \text{for all } s \le t,\ s \in \mathbb{T}\ . \tag{3.6}$$

X is just called a martingale if it is a martingale w.r.t. its natural filtration $(\mathcal{F}_t^X : t \in \mathbb{T})$.

In particular this implies that $\mathbb{E}(X_t) = X_0$ for all $t \in \mathbb{T}$.

Example.

Proposition 3.2 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 with continuous sample paths and $B_0 = 0$.

3.2 Properties of martingales

As mentioned before, for general continuous-time processes we concentrate on càdlàg processes, i.e. processes with right-continuous sample paths that have left limits.

Theorem 3.3 Martingale convergence theorem
Let $X = (X_t : t \in \mathbb{T})$ be a (càdlàg) martingale. If $\mathbb{E}\big(|X_t|\big) \leq M$ for some $M > 0$ and all $t \in \mathbb{T}$, then $X_\infty = \lim_{t\to\infty} X_t$ exists $a.s.$ and $\mathbb{E}\big(X_\infty\big) < \infty$.

This is basically the same as saying that X converges to a stationary distribution and X_∞ is a random variable with that distribution. In particular this implies $X_t = \mathbb{E}(X_\infty|\mathcal{F}_t)$.

Examples.

Definition 3.3 A random variable $T \in \mathbb{T} \cup \{\infty\}$ is called a *stopping time* (with respect to the filtration $(\mathcal{F}_t : t \in \mathbb{T})$), if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \in \mathbb{T}$.

Examples.

Interpretation: If T is a stopping time then at each time $t \in \mathbb{T}$ there is always enough information (in $\mathcal{F}_t$) to decide wether $T = t$ or not.

Proposition 3.4 *Let $X = (X_t : t \in \mathbb{T})$ be a (càdlàg) martingale and $T \in \mathbb{T}$ a stopping time, both w.r.t. $(\mathcal{F}_t : t \in \mathbb{T})$. Then the* ***stopped process***

$$X^T := \begin{cases} X_t & , \, t < T \\ X_T & , \, t \geq T \end{cases} \quad \textit{is a martingale w.r.t. } (\mathcal{F}_t : t \in \mathbb{T}) \, . \tag{3.7}$$

Proof. We focus on discrete time $\mathbb{T} = \mathbb{N}$. We can write

$$X_n^T = \sum_{i=0}^{n-1} X_i \, \mathbb{1}_{T=i} + X_n \, \mathbb{1}_{T \geq n} \tag{3.8}$$

so X_n^T is adapted to $\mathcal{F}_n$ and $\mathbb{E}\big(|X_n^T|\big) \leq \sum_{i=0}^n \mathbb{E}\big(|X_i|\big) < \infty$.
Also $X_{n+1}^T - X_n^T = (X_{n+1} - X_n) \, \mathbb{1}_{T>n}$, so we have

$$\begin{aligned} \mathbb{E}(X_{n+1}^T | \mathcal{F}_n) - X_n^T \; &= \; \mathbb{E}(X_{n+1}^T - X_n^T | \mathcal{F}_n) = \mathbb{E}\big((X_{n+1} - X_n) \, \mathbb{1}_{T>n} \big| \mathcal{F}_n\big) = \\ &= \; \big(\mathbb{E}(X_{n+1} | \mathcal{F}_n) - X_n\big) \, \mathbb{1}_{T>n} = 0 \, , \end{aligned} \tag{3.9}$$

and X^T is a martingale w.r.t. $(\mathcal{F}_n : n \in \mathbb{N})$. □

Example.

Theorem 3.5 Optional stopping theorem
Let $X = (X_t : t \in \mathbb{T})$ be a (càdlàg) martingale with $\mathbb{E}\big(|X_t|\big) \leq M$ for some $M > 0$ and all $t \in \mathbb{T}$. Then for all stopping times $S \leq T$ with $\mathbb{P}(S < \infty) = 1$ we have

$$\mathbb{E}(X_T|\mathcal{F}_S) = X_S \; a.s. \; . \tag{3.10}$$

Proof. By the convergence theorem 3.3 $X_t \to X_\infty \; a.s.$ and we have $X_T = X_\infty^T$ even if $T = \infty$. Then again for discrete time $\mathbb{T} = \mathbb{N}$ we have

$$\mathbb{E}(X_T|\mathcal{F}_S) = \sum_{s\in\mathbb{N}} \mathbb{E}(X_\infty^T|\mathcal{F}_s)\, \mathbb{1}_{S=s} = \sum_{s\in\mathbb{N}} X_s^T \, \mathbb{1}_{S=s} = X_S^T = X_S \tag{3.11}$$

using Proposition 3.4 and that $S \leq T$. □

In particular with $S = 0$ this implies that $\mathbb{E}(X_T) = X_0$ for all stopping times T.

Example.

4 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

$$Y = \mu + \sigma\xi \quad \text{where} \quad \xi = \frac{X-\mu}{\sigma}\,. \tag{4.2}$$

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 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

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

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

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

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

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

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 (4.8) is often written as

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

This is called a *stochastic differential equation (SDE)*.

The (non-existing) derivative $\eta_t = dB_t/dt$ is called *white noise*, and can be understood as a normalized random force term on X uncorrelated in time. Formally it is given by a Gaussian process with covariance $\delta_{t,s}$. Physicists often write

$$\frac{dX_t}{dt} = a(t, X_t) + \sigma(t, X_t)\, \eta_t \ , \tag{4.10}$$

instead of (4.9) and call this a *Langevin equation*.

As for ordinary differential equations, it is often better to look at the integrated version of (4.9), since it requires less regularity assumptions,

$$X_t - X_0 = \int_0^t a(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \ . \tag{4.11}$$

In general, a *solution* of the SDE with initial condition $X_0 = x_0$ consists of

- a probability space and a filtration $\big(\Omega, \mathcal{F}, (\mathcal{F}_t : t \geq 0), \mathbb{P}\big)$,
- a BM $B = (B_t : t \geq 0)$ adapted to $(\mathcal{F}_t : t \geq 0)$,
- a continuous process $X = (X_t : t \geq 0)$ adapted to $(\mathcal{F}_t : t \geq 0)$ that fulfilles (4.11) with $X_0 = x_0$. (4.12)

As usual, the probability space is often not mentioned explicitly, X is just given as some function of B and $(\mathcal{F}_t : t \geq 0)$ is the natural filtration for B. But in any case, we have to make sense of the two *stochastic integrals* in (4.11).

From now on let us fix some probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t : t \geq 0), \mathbb{P}\big)$. Let $X = (X_t : t \geq 0)$ and $Y = (Y_t : t \geq 0)$ be two càdlàg adapted 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 define the stochastic integral $I = (I_t : t \geq 0)$ by

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

The question is, for which X and Y is this limit well defined, does it depend on the choice of $\tau_k \in [t_{k-1}, t_k]$ and in what sense does the limit hold?

Integrands: The most general integrands Y we will consider are continuous adapted processes. This includes all diffusion processes but could be further generalized to socalled *previsible processes*.

Integrators: The most general integrators X for which (4.14) can be defined are socalled *semimartingales*, which we introduce in the following.

4.2 Semimartingales

Definition 4.1 Let X be a càdlàg adapted process. For each $\omega \in \Omega$ define

$$V_t(\omega) := \lim_{n\to\infty} \sum_{k=1}^{n} \left| X_{t_k}(\omega) - X_{t_{k-1}}(\omega) \right| \in [0, \infty] \,, \tag{4.15}$$

which is non-decreasing in t. $V = (V_t : t \geq 0)$ is called the *total variation process* of X and X is *of finite variation* if $V_t < \infty$ for all $t \geq 0$.

$V_t(\omega)$ corresponds to the length of the path $\big(X_s(\omega) : 0 \leq s \leq t\big)$.

Examples.

Proposition 4.1 *Let X be a continuous martingale. Then there exists a unique adapted increasing process $[X] = \big([X]_t : t \geq 0\big)$ with $[X]_0 = 0$, such that*

$$X^2 - [X] = \big(X_t^2 - [X]_t : t \geq 0\big) \quad \textit{is a continuous martingale} \,. \tag{4.16}$$

*$[X]$ is called the **quadratic variation** of X and for*

$$[X]_t^n := \sum_{k=1}^{n} (X_{t_k} - X_{t_{k-1}})^2 \quad \textit{we have} \quad [X]_t^n \to [X]_t \quad \textit{as } n \to \infty \,, \tag{4.17}$$

in the sense that for all $\epsilon > 0$ *and* $t \geq 0$, $\quad \mathbb{P}\Big(\sup_{s \leq t} \big| [X]^n_s - [X]_s \big| > \epsilon \Big) \to 0$.

Note that if we define the quadratic variation pathwise like the total variation, it does not have the nice properties stated above.

Examples.

Proposition 4.2 *Let* X *be a continuous process of finite variation. Then* $[X] \equiv 0$. *In particular, if* X *is a continuous martingale of finite variation, then* $X \equiv X_0$.

Proof. By the definitions (4.17) and 4.1 we have

$$[X]^n_t = \sum_{k=1}^{n} (X_{t_k} - X_{t_{k-1}})^2 \;\leq\; \underbrace{\sup_{k=1,..,n} \big| X_{t_k} - X_{t_{k-1}} \big|}_{\to 0 \text{ as } n \to \infty} \underbrace{\sum_{k=1}^{n} \big| X_{t_k} - X_{t_{k-1}} \big|}_{\leq V_t} \;. \tag{4.18}$$

by continuity by finite variation

Therefore $\quad [X]_t = \lim_{n \to \infty} [X]^n_t = 0 \quad$ for all $t \geq 0$.
If X is a continuous martingale with say $X_0 = 0$ we have

$$\mathbb{E}(X_{t_k} X_{t_{k-1}}) = \mathbb{E}\big(\mathbb{E}(X_{t_k} X_{t_{k-1}} | \mathcal{F}_{t_{k-1}})\big) = \mathbb{E}\big(X_{t_{k-1}}\, \mathbb{E}(X_{t_k} | \mathcal{F}_{t_{k-1}})\big) = \mathbb{E}(X^2_{t_{k-1}}) \tag{4.19}$$

Therefore we have for a time partition of arbitrary size n,

$$\mathbb{E}(X_t^2) = \mathbb{E}\bigg(\sum_{k=1}^{n} \big(X^2_{t_k} - X^2_{t_{k-1}} \big) \bigg) = \mathbb{E}\bigg(\sum_{k=1}^{n} \big(X_{t_k} - X_{t_{k-1}} \big)^2 \bigg) \to 0 \tag{4.20}$$

as $n \to \infty$ by the first statement. Thus $X_t = 0$ for all $t \geq 0$. □

Definition 4.2 A *semimartingale* X is a càdlàg adapted process which may be written as

$$X = X_0 + M + A \quad \text{with} \quad M_0 = A_0 = 0 \,, \tag{4.21}$$

where M is a martingale and A is a process of finite variation.

Examples.

Corollary 4.3 *For continuous semimartingales X the decomposition (4.21) is unique and it is also called the* ***Doob-Meyer decomposition*** *of X.*

Proof.

Definition 4.3 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

$$Y = (Y_t : t \geq 0) \quad \text{with} \quad Y_t = \sum_{k=1}^{N_t} Z_k \tag{4.22}$$

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

$$X_t = X_0 + a\,t + \sigma B_t + Y_t \quad \text{with stationary, independent increments} \tag{4.23}$$

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

Proposition 4.4 *A Lévy process X is a càdlàg semimartingale.*

Proof.

It can also be shown that every càdlàg semimartingale with stationary independent increments has to be a Lévy process, so they are quite general.

4.3 Stochastic integration and Itô calculus

Theorem 4.5 Itô integral
Let X be a continuous semimartingale with Doob-Meyer decomposition $X = X_0 + M + A$ (4.21) and Y be a continuous adapted process. If

$$\sup_{0\le s\le t} \mathbb{E}(X_s^2) < \infty \quad \textit{and} \quad \mathbb{E}\bigg(\int_0^t Y_s^2 d[X]_s \bigg) < \infty \tag{4.24}$$

for some $t \geq 0$, then

$$I_t = \int_0^t Y_s \, dX_s := \lim_{n\to\infty} \sum_{k=1}^{n} Y_{t_{k-1}} (X_{t_k} - X_{t_{k-1}}) \tag{4.25}$$

exists in the sense of (4.17). If (4.24) holds for all $t \geq 0$, then $I = (I_t : t \geq 0)$ is a continuous semimartingale with decomposition

$$I_t = 0 + \int_0^t Y_s \, dM_s + \int_0^t Y_s \, dA_s \tag{4.26}$$

and is called the (stochastic) Itô integral of Y w.r.t. X.

Examples.

With considerable technical effort, the Itô integral can be generalized to non-continuous processes. (4.26) implies that

$$\boxed{\text{if } X \text{ is a martingale, then } \textstyle\int_0^t Y_s \, dX_s \text{ is a martingale}} \, . \tag{4.27}$$

So for example Itô integrals w.r.t. BM with $X = B$ are martingales. Surprisingly, also a converse statement holds.

Proposition 4.6 *Let X be a martingale. Then there exists an adapted process Y such that*

$$X_t = X_0 + \int_0^t Y_s \, dB_s \quad \textit{for all } t \geq 0 \,, \quad \textit{where } B \textit{ is a standard BM .} \tag{4.28}$$

Moreover, if X is a continuous martingale with $X_0 = 0$ and $[X]_t \to \infty$ as $t \to \infty$, we have

$$X_t = B_{[X]_t} \quad \textit{for all } t \geq 0 \,, \quad \textit{where } B \textit{ is a standard BM .} \tag{4.29}$$

Proof.

So every martingale X is an integral w.r.t. standard BM, and if it is continuous, it is actually a (time-changed) standard BM on the time scale $[X]_t$ rather than t. Note that of course consistently $[B]_t = t$.

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

$$\int_{t_0}^{t} B_s \, dB_s = \frac{1}{2}\big((B_t^2 - B_{t_0}^2) - (t - t_0)\big) \, . \tag{4.30}$$

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

$$S\int_{t_0}^{t} B_s \, dB_s = \frac{1}{2}(B_t^2 - B_{t_0}^2) \, , \tag{4.31}$$

and this integral is called the *Stratonovich integral*. The advantage of this choice is that it obeys the usual rules of calculus. But now dependence of Y_{τ_k} and the increment $X_{t_k} - X_{t_{k-1}}$ is more complicated, leading to several disadvantages compared to Itô:

- $S\int Y_s \, dX_s$ can only be defined if also Y is a semimartingale .
- Even if X is a martingale, $S\int Y_s \, dX_s$ is in general NOT a martingale .

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 (4.30) has to be there, since the result should be a martingale. 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,

$$I_t = I_{t_0} + \int_{t_0}^{t} Y_s \, dX_s \quad \Leftrightarrow \quad dI_t = Y_t \, dX_t \, . \tag{4.32}$$

For a continuous martingale M we get analogous to our above computation

$$\int_{t_0}^{t} M_s \, dM_s = \frac{1}{2}\Big((M_t^2 - M_{t_0}^2) - \big([M]_t - [M]_{t_0}\big)\Big) \, . \tag{4.33}$$

This is equivalent to

$$M_t \, dM_t = \frac{1}{2}\big(d(M_t^2) - d[M]_t\big) \quad \text{or} \quad d(M_t^2) = 2M_t \, dM_t + d[M]_t \, . \tag{4.34}$$

This is basically an application of the chain rule for Itô calculus. The meaning of the quadratic variation term becomes clear if we compute an increment by hand,

$$\begin{aligned} M_{t+h}^2 - M_t^2 &= (M_{t+h} - M_t)(M_{t+h} + M_t) = (M_{t+h} - M_t)(M_{t+h} - M_t + 2M_t) = \\ &= (M_{t+h} - M_t)^2 + 2M_t(M_{t+h} - M_t) \, . \end{aligned} \tag{4.35}$$

Taking $h \to 0$ we get $\quad d(M_t^2) = 2M_t \, dM_t + (dM_t)^2 \, , \quad$ and comparing with (4.34),

$$\boxed{d[M]_t = (dM_t)^2 = O(dt)} \, . \tag{4.36}$$

In usual calculus these terms are of negligible order $o(dt)$, but for martingales they have to be taken into account, for example for BM $\quad d[B]_t = (dB_t)^2 = dt$.
If $X = X_0 + M + A$ is a semimartingale, by Proposition 4.2, $[A] \equiv 0$. Thus $[X] = [M]$ and the general version of (4.34) is

$$\boxed{d(X_t^2) = 2X_t \, dX_t + d[X]_t} \quad \big(= 2(M_t + A_t) \, d(M_t + A_t) + d[M]_t \big) \, . \tag{4.37}$$

Using (4.36) we can easily see that the quadratic variation of the Itô integral

$$I_t = \int_{t_0}^{t} Y_s \, dX_s \quad \text{is given by} \quad [I]_t = \int_{t_0}^{t} Y_s^2 \, d[X]_s \quad \text{for all } t \geq t_0 \, . \tag{4.38}$$

This follows directly from

$$d[I]_t = (dI_t)^2 = (Y_t \, dX_t)^2 = Y_t^2 \, (dX_t)^2 = Y_t^2 \, d[X]_t \, . \tag{4.39}$$

This should clarify condition (4.24) which insures that the integral has a finite quadratic variation. These findings are summarized in the following very usefull result.

Theorem 4.7 Itô's formula
Let X be a continuous semimartingale and $g \in C^2(\mathbb{R}, \mathbb{R})$. Then

$$dg(X_t) = g'(X_t) \, dX_t + \frac{1}{2} g''(X_t) \, d[X]_t \, , \tag{4.40}$$

or in the integrated version

$$g(X_t) = g(X_0) + \int_0^t g'(X_t) \, dX_t + \frac{1}{2} \int_0^t g''(X_t) \, d[X]_t \, . \tag{4.41}$$

Proof. Taylor expansion with terms up to order dt.

In particular, we see that $f(X_t)$ is again a semimartingale with decomposition

$$g(X_t) = g(X_0) + \underbrace{\int_0^t g'(X_t) \, dM_t}_{\text{cont. martingale}} + \underbrace{\int_0^t g'(X_t) \, dA_t + \frac{1}{2} \int_0^t g''(X_t) \, d[X]_t}_{\text{finite variation}} \, . \tag{4.42}$$

Examples.

4.4 Solutions to SDEs

Let X be a solution of the SDE

$$dX_t = a(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \; . \tag{4.43}$$

For the quadratic variation of X we get

$$d[X]_t = (dX_t)^2 = \sigma^2(t, X_t)\, dt + o(dt) \; , \tag{4.44}$$

and using Itô's formula, the SDE for an observable $g(X_t)$ with $g \in C^2(\mathbb{R}, \mathbb{R})$ is

$$\begin{aligned} dg(X_t) \;&=\; g'(X_t)\, dX_t + \frac{1}{2}\, g''(X_t)\, (dX_t)^2 = \\ &=\; g'(X_t)\big(a(t, X_t)\, dt + \sigma(t, X_t)\, dB_t\big) + \frac{1}{2}\, g''(X_t)\sigma^2(t, X_t)\, dt \; . \end{aligned} \tag{4.45}$$

Taking the expectation on both sides, we get with PDF $f(t, x)$ by partial integration

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\big(g(X_t)\big) \;&=\; \int_{\mathbb{R}} g(x)\, \frac{\partial}{\partial t} f(t, x)\, dx = \\ &=\; \int_{\mathbb{R}} \Big(g'(x)\, a(t, x) + \frac{1}{2} g''(x)\sigma^2(t, x)\Big)\, f(t, x)\, dx = \\ &=\; \int_{\mathbb{R}} g(x) \bigg(\frac{\partial}{\partial x}\big(a(t, x)\, f(t, x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(\sigma^2(t, x)\, f(t, x)\big)\bigg) dx \; , \end{aligned} \tag{4.46}$$

since the expected value of the martingale part vanishes. This holds for arbitrary functions g, and therefore we must have

$$\frac{\partial}{\partial t} f(t, x) = \frac{\partial}{\partial x}\big(a(t, x)\, f(t, x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(\sigma^2(t, x)\, f(t, x)\big) \; . \tag{4.47}$$

Thus $f(t, x)$ fulfilles the Fokker-Planck equation and X is a diffusion process with drift $a(t, x)$ and diffusion $\sigma^2(t, x)$.

How many solutions to the SDE (4.43) are there?

Definition 4.4 We say that a SDE has a *weak solution* if there exists a solution for all initial values $X_0 = x_0 \in \mathbb{R}$. The solution is *unique in law*, if all solutions started from x_0 have the same distribution. The solution is *pathwise unique*, if for a fixed probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_t : t \geq 0), \mathbb{P}\big)$ and a fixed BM B, any two solutions X and X' fulfill

$$X_0 = X_0' \; a.s. \quad \Rightarrow \quad \mathbb{P}(X_t = X_t' \text{ for all } t \geq 0) = 1 \; . \tag{4.48}$$

If a solution X is adapted to the natural filtration of B it is called a *strong solution*.

For time-independent drift and diffusion there is a general theorem about existence and uniqueness for SDEs.

Theorem 4.8 *Suppose that $a : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are Lipschitz-continuous, i.e.*

$$\big|a(x) - a(y)\big| \leq K|x - y| \quad \text{for some } K > 0 \text{ and all } x, y \in \mathbb{R}\,. \tag{4.49}$$

Then for each $\big(\Omega, \mathcal{F}, (\mathcal{F}_t : t \geq 0), \mathbb{P}\big)$ and each BM B adapted to $(\mathcal{F}_t : t \geq 0)$ solutions to

$$dX_t = a(X_t)\,dt + \sigma(X_t)\,dB_t \tag{4.50}$$

are pathwise unique and there exists a strong solution for any starting point $x_0 \in \mathbb{R}$.

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

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.