MA933 - Stochastic Modelling and Random Processes

MSc in Mathematics of Systems

Stefan Grosskinsky

Warwick, 2019

These notes and other information about the course are available on
www2.warwick.ac.uk/fac/sci/mathsys/courses/msc/ma933/
Contents

1 Basic probability, simple random walk, discrete-time Markov processes

2 Continuous time Markov chains

3 Processes with continuous state space

4 Stochastic Particle Systems

References

- C.W. Gardiner: Handbook of Stochastic Methods (3rd edition), Springer 2004
- G. Grimmett: Probability on Graphs, CUP 2010
  http://www.statslab.cam.ac.uk/~grg/books/ogs.html
1. Probability

- **sample space** $\Omega$ (e.g. \{H, T\}, \{H, T\}^N, \{paths of a stoch. process\})
- **events** $A \subseteq \Omega$ (measurable) subsets (e.g. odd numbers on a die)
  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is the set of all events (subset of the powerset)

**Definition 1.1**

A **probability distribution** $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ which is

(i) normalized, i.e. $\mathbb{P}[\emptyset] = 0$ and $\mathbb{P}[\Omega] = 1$

(ii) additive, i.e. $\mathbb{P} \left[ \bigcup_i A_i \right] = \sum_i \mathbb{P}[A_i]$, where $A_1, A_2, \ldots$ is a collection of disjoint events, i.e. $A_i \cap A_j = \emptyset$ for all $i, j$.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.

- For **discrete** $\Omega$: $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P}[A] = \sum_{\omega \in A} \mathbb{P}[\omega]$
  e.g. $\mathbb{P}[\text{even number on a die}] = \mathbb{P}[2] + \mathbb{P}[4] + \mathbb{P}[6] = \frac{1}{2}$

- For **continuous** $\Omega$ (e.g. [0, 1]): $\mathcal{F} \subsetneq \mathcal{P}(\Omega)$
1. Independence and conditional probability

- Two events $A, B \subseteq \Omega$ are called **independent** if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$.

**Example.** rolling a die repeatedly

- If $\mathbb{P}[B] > 0$ then the **conditional probability** of $A$ given $B$ is

  $$\mathbb{P}[A|B] := \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$ 

  If $A$ and $B$ are independent, then $\mathbb{P}[A|B] = \mathbb{P}[A]$.

**Lemma 1.1 (Law of total probability)**

Let $B_1, \ldots, B_n$ be a partition of $\Omega$ such that $\mathbb{P}[B_i] > 0$ for all $i$. Then

$$\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A \cap B_i] = \sum_{i=1}^{n} \mathbb{P}[A|B_i] \mathbb{P}[B_i].$$

Note that also

$$\mathbb{P}[A|C] = \sum_{i=1}^{n} \mathbb{P}[A|C \cap B_i] \mathbb{P}[B_i|C] \quad \text{provided } \mathbb{P}[C] > 0.$$
1. Random variables

Definition 1.2

A random variable $X$ is a (measurable) function $X : \Omega \to \mathbb{R}$.

The distribution function of the random variable is

$$F(x) = \mathbb{P}[X \leq x] = \mathbb{P}\left[\{\omega : X(\omega) \leq x\}\right].$$

$X$ is called discrete, if it only takes values in a countable subset $\Delta = \{x_1, x_2, \ldots\} \subseteq \mathbb{R}$, and its distribution is characterized by the probability mass function

$$\pi(x) := \mathbb{P}[X = x], \quad x \in \Delta.$$

$X$ is called continuous, if its distribution function is

$$F(x) = \int_{-\infty}^{x} f(y) \, dy \quad \text{for all } x \in \mathbb{R},$$

where $f : \mathbb{R} \to [0, \infty)$ is the probability density function (PDF) of $X$. 
1. Random variables

- In general, $f = F'$ is given by the derivative (exists for cont. rv’s).
  For discrete rv’s, $F$ is a step function with 'PDF'

$$f(x) = F'(x) = \sum_{y \in \Delta} \pi(y) \delta(x - y).$$

- The **expected value** of $X$ is given by $\mathbb{E}[X] = \left\{ \begin{array}{l} \sum_{x \in \Delta} x \pi(x) \\ \int_{\mathbb{R}} x f(x) \, dx \end{array} \right.$

- The **variance** is given by $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$,
  the **covariance** of two r.v.s by $\text{Cov}[X, Y] := \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]$.

- Two random variables $X, Y$ are independent if the events $\{X \leq x\}$ and $\{Y \leq y\}$
  are independent for all $x, y \in \mathbb{R}$. This implies for **joint distributions**

$$f(x, y) = f^X(x) f^Y(y) \quad \text{or} \quad \pi(x, y) = \pi^X(x) \pi^Y(y)$$

  with **marginals** $f^X(x) = \int_{\mathbb{R}} f(x, y) \, dy$ and $\pi^X(x) = \sum_{y \in \Delta_y} \pi(x, y)$.

- Independence implies $\text{Cov}[X, Y] = 0$, i.e. $X$ and $Y$ are **uncorrelated**.
  The inverse is in general false, but holds if $X$ and $Y$ are Gaussian.
1. Simple random walk

**Definition 1.3**

Let \( X_1, X_2, \ldots \in \{-1, 1\} \) be a sequence of independent, identically distributed random variables (iidrv's) with

\[
p = \mathbb{P}[X_i = 1] \quad \text{and} \quad q = \mathbb{P}[X_i = -1] = 1 - p.\]

The sequence \( Y_0, Y_1, \ldots \) defined as \( Y_0 = 0 \) and \( Y_n = \sum_{k=1}^{n} X_k \) is called the **simple random walk (SRW)** on \( \mathbb{Z} \).

- for a single increment \( X_k \) we have
  \[
  \mathbb{E}[X_k] = p - q = 2p - 1, \quad \text{var}[X_k] = p + q - (p - q)^2 = 4p(1 - p)
  \]

- \( \mathbb{E}[Y_n] = \mathbb{E}\left[ \sum_{k=1}^{n} X_k \right] = \sum_{k=1}^{n} \mathbb{E}[X_k] = n(2p - 1) \) (expectation is a linear operation)

- \( \text{var}[Y_n] = \text{var}\left[ \sum_{k=1}^{n} X_k \right] = \sum_{k=1}^{n} \text{var}[X_k] = 4np(1 - p) \) (for a sum of independent rv's the variance is additive)
1. LLN and CLT

**Theorem 1.2 (Weak law of large numbers (LLN))**

Let $X_1, X_2, \ldots \in \mathbb{R}$ be a sequence of iidrv’s with $\mu := \mathbb{E}[X_k] < \infty$ and $\mathbb{E}[|X_k|] < \infty$. Then

$$
\frac{1}{n} Y_n = \frac{1}{n} \sum_{k=1}^{n} X_k \rightarrow \mu \quad \text{as } n \rightarrow \infty
$$

in distribution (i.e. the distr. fct. of $Y_n$ converges to $\mathbb{1}_{[\mu, \infty)}(x)$ for $x \neq \mu$).

**Theorem 1.3 (Central limit theorem (CLT))**

Let $X_1, X_2, \ldots \in \mathbb{R}$ be a sequence of iidrv’s with $\mu := \mathbb{E}[X_k] < \infty$ and $\sigma^2 := \text{var}[X_k] < \infty$. Then

$$
\frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu) \rightarrow \xi \quad \text{as } n \rightarrow \infty
$$

in distr., where $\xi \sim N(0, 1)$ is a **standard Gaussian** with PDF $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

**Expansion.** as $n \rightarrow \infty$, $\sum_{k=1}^{n} X_k = n\mu + \sqrt{n}\sigma \xi + o(\sqrt{n})$, $\xi \sim N(0, 1)$.
1. Discrete-time Markov processes

**Definition 1.4**

A **discrete-time stochastic process** with **state space** $S$ is a sequence $Y_0, Y_1, \ldots = (Y_n : n \in \mathbb{N}_0)$ of random variables taking values in $S$.

The process is called **Markov**, if for all $A \subseteq S$, $n \in \mathbb{N}_0$ and $s_0, \ldots, s_n \in S$

$$
\mathbb{P}(Y_{n+1} \in A|Y_n = s_n, \ldots, Y_0 = s_0) = \mathbb{P}(Y_{n+1} \in A|Y_n = s_n) .
$$

A Markov process (MP) is called **homogeneous** if for all $A \subseteq S$, $n \in \mathbb{N}_0$ and $s \in S$

$$
\mathbb{P}(Y_{n+1} \in A|Y_n = s) = \mathbb{P}(Y_1 \in A|Y_0 = s) .
$$

If $S$ is discrete, the MP is called a **Markov chain (MC)**.

The generic probability space $\Omega$ is the **path space**

$$
\Omega = D(\mathbb{N}_0, S) := S^{\mathbb{N}_0} = S \times S \times \ldots
$$

which is uncountable even when $S$ is finite. For a given $\omega \in \Omega$ the function $n \mapsto Y_n(\omega)$ is called a **sample path**.

Up to finite time $N$ and with finite $S$, $\Omega_N = S^{N+1}$ is finite.
1. Discrete-time Markov processes

Examples.

- For the simple random walk we have state space $S = \mathbb{Z}$ and $Y_0 = 0$. Up to time $N$, $\mathbb{P}$ is a distribution on the finite path space $\Omega_N$ with

  \[ \mathbb{P}(\omega) = \begin{cases} p & \text{# of up-steps} \\ q & \text{# of down-steps} \\ 0 & \text{path } \omega \text{ not possible} \end{cases}, \text{ path } \omega \text{ possible} \]

  There are only $2^N$ paths in $\Omega_N$ with non-zero probability.
  For $p = q = 1/2$ they all have the same probability $(1/2)^N$.

- For the generalized random walk with $Y_0 = 0$ and increments $Y_{n+1} - Y_n \in \mathbb{R}$, we have $S = \mathbb{R}$ and $\Omega_N = \mathbb{R}^N$ with an uncountable number of possible paths.

- A sequence $Y_0, Y_1, \ldots \in S$ of iidrv’s is also a Markov process with state space $S$.

- Let $S = \{1, \ldots, 52\}$ be a deck of cards, and $Y_1, \ldots, Y_{52}$ be the cards drawn at random without replacement. Is this a Markov process?
1. Discrete-time Markov chains

Proposition 1.4

Let \((X_n : n \in \mathbb{N}_0)\) be a homogeneous DTMC with discrete state space \(S\). Then the transition function

\[ p_n(x, y) := P[X_n = y|X_0 = x] = P[X_{k+n} = y|X_k = x] \text{ for all } k \geq 0 \]

is well defined and fulfills the Chapman Kolmogorov equations

\[ p_{k+n}(x, y) = \sum_{z \in S} p_k(x, z) p_n(z, y) \text{ for all } k, n \geq 0, \ x, y \in S. \]

Proof. We use the law of total probability, the Markov property and homogeneity
1. Markov chains

- In matrix form with \( P_n = (p_n(x, y) : x, y \in S) \) the Chapman Kolmogorov equations read

\[
P_{n+k} = P_n P_k \quad \text{and in particular} \quad P_{n+1} = P_n P_1.
\]

With \( P_0 = \mathbb{I} \), the obvious solution to this recursion is

\[
P_n = P^n \quad \text{where we write} \quad P_1 = P = (p(x, y) : x, y \in S) .
\]

- The transition matrix \( P \) and the initial condition \( X_0 \in S \) completely determine a homogeneous DTMC, since for all \( k \geq 1 \) and all events \( A_1, \ldots, A_k \subseteq S \)

\[
\mathbb{P}[X_1 \in A_1, \ldots, X_k \in A_k] = \sum_{s_1 \in A_1} \cdots \sum_{s_k \in A_k} p(X_0, s_1)p(s_1, s_2) \cdots p(s_{k-1}, s_k) .
\]

- Fixed \( X_0 \) can be replaced by an initial distribution \( \pi_0(x) := \mathbb{P}[X_0 = x] . \)

The distribution at time \( n \) is then

\[
\pi_n(x) = \sum_{y \in S} \sum_{s_1 \in S} \cdots \sum_{s_{n-1} \in S} \pi_0(y)p(y, s_1) \cdots p(s_{n-1}, x) \quad \text{or} \quad \langle \pi_n \rangle = \langle \pi_0 \rangle P^n .
\]
1. Transition matrices
The transition matrix $P$ is **stochastic**, i.e.

$$p(x, y) \in [0, 1] \quad \text{and} \quad \sum_y p(x, y) = 1,$$

or equivalently, the column vector $|1\rangle = (1, \ldots, 1)^T$ is **eigenvector** with **eigenvalue** 1: $P|1\rangle = |1\rangle$

**Example 1 (Random walk with boundaries)**

Let $(X_n : n \in \mathbb{N}_0)$ be a SRW on $S = \{1, \ldots, L\}$ with $p(x, y) = p\delta_{y,x+1} + q\delta_{y,x-1}$. The boundary conditions are

- **periodic** if $p(L, 1) = p$, $p(1, L) = q$,
- **absorbing** if $p(L, L) = 1$, $p(1, 1) = 1$,
- **closed** if $p(1, 1) = q$, $p(L, L) = p$,
- **reflecting** if $p(1, 2) = 1$, $p(L, L - 1) = 1$. 
1. Stationary distributions

**Definition 1.5**

Let \((X_n : n \in \mathbb{N}_0)\) be a homogeneous DTMC with state space \(S\). The distribution \(\pi(x), x \in S\) is called **stationary** if for all \(y \in S\)

\[
\sum_{x \in S} \pi(x)p(x, y) = \pi(y) \quad \text{or} \quad \langle \pi \mid P = \langle \pi \mid .
\]

\(\pi\) is called **reversible** if it fulfills the **detailed balance** conditions

\[
\pi(x)p(x, y) = \pi(y)p(y, x) \quad \text{for all } x, y \in S .
\]

- Reversibility implies stationarity, since

\[
\sum_{x \in S} \pi(x)p(x, y) = \sum_{x \in S} \pi(y)p(y, x) = \pi(y) .
\]

- Stationary distributions as row vectors \(\langle \pi \mid = (\pi(x) : x \in S)\) are **left eigenvectors** with **eigenvalue** 1: \(\langle \pi \mid = \langle \pi \mid P .\)
1. Absorbing states

**Definition 1.6**

A state \( s \in S \) is called **absorbing** for a DTMC with transition matrix \( p(x, y) \), if

\[
p(s, y) = \delta_{s,y} \quad \text{for all } y \in S.
\]

**RW with absorbing BC.**

Let \( h_k \) be the **absorption probability** for \( X_0 = k \in S = \{1, \ldots, L\} \),

\[
h_k = \mathbb{P}[\text{absorption}|X_0 = k] = \mathbb{P}[X_n \in \{1, L\} \text{ for some } n \geq 0|X_0 = k].
\]

Conditioning on the first jump and using Markov, we have the recursion

\[
h_k = ph_{k+1} + qh_{k-1} \quad \text{for } k = 2, \ldots, L - 1; \quad h_1 = h_L = 1.
\]

**Ansatz for solution** \( h_k = \lambda^k, \quad \lambda \in \mathbb{C} \):

\[
\lambda = p\lambda^2 + q \quad \Rightarrow \quad \lambda_1 = 1, \quad \lambda_2 = q/p
\]

**General solution** of 2nd order linear recursion

\[
h_k = a\lambda_1^k + b\lambda_2^k = a + b(q/p)^k, \quad a, b \in \mathbb{R}.
\]

Determine coefficients from boundary condition \( \Rightarrow \quad h_k \equiv 1 \).
1. Distribution at time $n$
Consider a DTMC on a finite state space with $|S| = L$, and let $\lambda_1, \ldots, \lambda_L \in \mathbb{C}$ be the eigenvalues of the transition matrix $P$ with corresponding

left (row) eigenvectors $\langle u_i |$ and right (column) eigenvectors $| v_i \rangle$

in bra-ket notation. Assuming that all eigenvalues are distinct we have

$$P = \sum_{i=1}^{L} \lambda_i | v_i \rangle \langle u_i | \quad \text{and} \quad P^n = \sum_{i=1}^{L} \lambda_i^n | v_i \rangle \langle u_i |$$

since eigenvectors can be chosen orthonormal $\langle u_i | v_j \rangle = \delta_{i,j}$.

Since $\langle \pi_n \rangle = \langle \pi_0 | P^n$ we get

$$\langle \pi_n \rangle = \langle \pi_0 | v_1 \rangle \lambda_1^n \langle u_1 | + \ldots + \langle \pi_0 | v_L \rangle \lambda_L^n \langle u_L | .$$

- The Gershgorin theorem implies that $|\lambda_i| \leq 1$ and contributions with $|\lambda_i| < 1$ decay exponentially (see hand-out 1).
- $\lambda_1 = 1$ corresponds to the stationary distribution $\langle \pi | = \langle u_1 |$ and $| v_1 \rangle = |1 \rangle$.
- Other $\mathbb{C} \ni \lambda_i \neq 1$ with $|\lambda_i| = 1$ correspond to persistent oscillations.
1. Lazy Markov chains

Definition 1.7

Let $(X_n : n \in \mathbb{N}_0)$ be a DTMC with transition matrix $p(x, y)$. The DTMC with transition matrix

$$p^\epsilon(x, y) = \epsilon \delta_{x,y} + (1 - \epsilon) p(x, y), \quad \epsilon \in (0, 1)$$

is called a lazy version of the original chain.

- $P^\epsilon$ has the same eigenvectors as $P$ with eigenvalues $\lambda_i^\epsilon = \lambda_i (1 - \epsilon) + \epsilon$ since

$$\langle u_i | P^\epsilon = \epsilon \langle u_i | + \lambda_i (1 - \epsilon) \langle u_i |$$

(analogously for $|v_i\rangle$)

- This implies $|\lambda_i^\epsilon| < |\lambda_i| \leq 1$ unless $\lambda_i = 1$. Such a matrix $P^\epsilon$ is called aperiodic, and there are no persistent oscillations.

- The stationary distribution is unique if and only if the eigenvalue $\lambda = 1$ is unique (has multiplicity 1), which is independent of lazyness (discussed later).
A continuous-time stochastic process with state space $S$ is a family $(X_t : t \geq 0)$ of random variables taking values in $S$. The process is called **Markov**, if for all $A \subseteq S$, $n \in \mathbb{N}$, $t_1 < \ldots < t_{n+1} \in [0, \infty)$ and $s_1, \ldots, s_n \in S$

$$\mathbb{P}(X_{t_{n+1}} \in A | X_{t_n} = s_n, \ldots, X_{t_1} = s_1) = \mathbb{P}(X_{t_{n+1}} \in A | X_{t_n} = s_n).$$

A Markov process (MP) is called **homogeneous** if for all $A \subseteq S$, $t, u > 0$ and $s \in S$

$$\mathbb{P}(X_{t+u} \in A | X_u = s) = \mathbb{P}(X_t \in A | X_0 = s).$$

If $S$ is discrete, the MP is called a continuous-time **Markov chain (CTMC)**.

The generic probability space $\Omega$ of a CTMC is the space of **right-continuous paths**

$$\Omega = D([0, \infty), S) := \{X : [0, \infty) \to S \mid X_t = \lim_{u \searrow t} X_u\}$$

$\mathbb{P}$ is a probability distribution on $\Omega$, which by **Kolmogorov’s extension theorem** is fully specified by its **finite dimensional distributions (FDDs)** of the form

$$\mathbb{P}[X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n], \quad n \in \mathbb{N}, \ t_i \in [0, \infty), \ A_i \subseteq S.$$
2. Continuous-time Markov chains

Proposition 2.1

Let \((X_t : t \geq 0)\) be a homogeneous CTMC with state space \(S\). Then for all \(t \geq 0\) the transition function

\[
p_t(x, y) := \mathbb{P}[X_t = y | X_0 = x] = \mathbb{P}[X_{t+u} = y | X_u = x] \quad \text{for all } u \geq 0
\]

is well defined and fulfills the Chapman Kolmogorov equations

\[
p_{t+u}(x, y) = \sum_{z \in S} p_t(x, z) p_u(z, y) \quad \text{for all } t, u \geq 0, \ x, y \in S.
\]

In matrix notation \(P_t = (p_t(x, y) : x, y \in S)\) we get

\[
P_{t+u} = P_t P_u \quad \text{with} \quad P_0 = I.
\]

In particular

\[
\frac{P_{t+\Delta t} - P_t}{\Delta t} = P_t \frac{P_{\Delta t} - I}{\Delta t} = \frac{P_{\Delta t} - I}{\Delta t} P_t,
\]

taking \(\Delta t \downarrow 0\) we get the so-called forward and backward equations

\[
\frac{d}{dt} P_t = P_t G = GP_t,
\]

where

\[
G = \left. \frac{dP_t}{dt} \right|_{t=0}
\]

is called the generator of the process (sometimes also \(Q\)-matrix).
2. Continuous-time Markov chains

- The solution is given by the matrix exponential

\[ P_t = \exp(tG) = \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k = \mathbb{I} + tG + \frac{t^2}{2} G^2 + \ldots \]  

(2.1)

- The distribution \( \pi_t \) at time \( t > 0 \) is then given by

\[ \langle \pi_t \rangle = \langle \pi_0 \rangle \exp(tG) \text{ which solves } \frac{d}{dt} \langle \pi_t \rangle = \langle \pi_t \rangle G. \]  

(2.2)

- On a finite state space with \( \lambda_1, \ldots, \lambda_L \in \mathbb{C} \) eigenvalues of \( G \), \( P_t \) has eigenvalues \( \exp(t\lambda_i) \) with the same eigenvectors \( \langle v_i|, \ |u_i\rangle \).

If the \( \lambda_i \) are distinct, we can expand the initial condition in the eigenvector basis

\[ \langle \pi_0 \rangle = \alpha_1 \langle v_1 | + \ldots + \alpha_L \langle v_L |, \]  

where \( \alpha_i = \langle \pi_0 | u_i \rangle \). This leads to

\[ \langle \pi_t \rangle = \alpha_1 \langle v_1 | e^{\lambda_1 t} + \ldots + \alpha_L \langle v_L | e^{\lambda_L t}. \]  

(2.3)
2. Continuous-time Markov chains

- Using (2.1) we have for $G = \{g(x, y) : x, y \in S\}$

  \[ p_{\Delta t}(x, y) = g(x, y) \Delta t + o(\Delta t) \quad \text{for all } x \neq y \in S. \]

  So $g(x, y) \geq 0$ can be interpreted as transition rates.

\[ p_{\Delta t}(x, x) = 1 + g(x, x) \Delta t + o(\Delta t) \quad \text{for all } x \in S, \]

  and since $\sum_y p_{\Delta t}(x, y) = 1$ this implies that

  \[ g(x, x) = -\sum_{y \neq x} g(x, y) \leq 0 \quad \text{for all } x \in S. \]

- (2.2) can then be written intuitively as the Master equation

\[ \frac{d}{dt} \pi_t(x) = \sum_{y \neq x} \pi_t(y) g(y, x) - \sum_{y \neq x} \pi_t(x) g(x, y) \quad \text{for all } x \in S. \]

  \[ \underbrace{\sum_{y \neq x} \pi_t(y) g(y, x)}_{\text{gain term}} - \underbrace{\sum_{y \neq x} \pi_t(x) g(x, y)}_{\text{loss term}} \]

- The Gershgorin theorem now implies that either $\lambda_i = 0$ or $\text{Re}(\lambda_i) < 0$ for the eigenvalues of $G$, so there are no persistent oscillations for CTMCs.
2. Stationary distributions

**Definition 2.2**

Let \((X_t : t \geq 0)\) be a homogeneous CTMC with state space \(S\). The distribution \(\pi(x), x \in S\) is called **stationary** if \(\langle \pi | G = \langle 0 \rangle\), or for all \(y \in S\)

\[
\sum_{x \in S} \pi(x)g(x, y) = \sum_{x \neq y} (\pi(x)g(x, y) - \pi(y)g(y, x)) = 0 .
\]  
(2.4)

\(\pi\) is called **reversible** if it fulfills the **detailed balance conditions**

\[
\pi(x)g(x, y) = \pi(y)g(y, x) \quad \text{for all } x, y \in S .
\]  
(2.5)

- again, **reversibility implies stationarity**, since with (2.5) every single term in the sum (2.4) vanishes

- Stationary distributions are left **eigenvectors** of \(G\) with **eigenvalue** 0.

- \(\langle \pi | G = \langle 0 \rangle\) implies \(\langle \pi | P_t = \langle \pi | (I + \sum_{k \geq 1} t^k G^k / k!) = \langle \pi | \) for all \(t \geq 0\).
2. Stationary distributions

**Proposition 2.2 (Existence)**

A DTMC or CTMC with **finite** state space $S$ has **at least one** stationary distribution.

**Proof.** Since $P$ and $G$ have row sum 1 and 0 we have $P \rho = \rho$ and $G \rho = \mathbf{0}$.

So 1 and 0 are eigenvalues, and left eigenvectors can be shown to have non-negative entries and thus can be normalized to be stationary distributions $\langle \pi |$. □

**Remark.** If $S$ is countably infinite, stationary distributions may not exist, as for example for the SRW on $\mathbb{Z}$ or the Poisson process on $\mathbb{N}$ (see later).

**Definition 2.3**

A CTMC (or DTMC) is called **irreducible**, if for all $x, y \in S$

$$p_t(x, y) > 0 \text{ for some } t > 0 \quad (p_n(x, y) > 0 \text{ for some } n \in \mathbb{N}) .$$

**Remark.** For continuous time irreducibility implies $p_t(x, y) > 0$ for **all** $t > 0$. 
2. Stationary distributions

Proposition 2.3 (Uniqueness)

An irreducible Markov chain has at most one stationary distribution.

Proof. Follows from the Perron Frobenius theorem:
Let $P$ be a stochastic matrix ($P = P_t$ for any $t \geq 0$ for CTMCs). Then

1. $\lambda_1 = 1$ is an eigenvalue of $P$, it is singular if and only if the chain is irreducible. Corresponding left and right eigenvectors have non-negative entries.

2. if the chain is continuous-time or discrete-time aperiodic, all remaining eigenvalues $\lambda_i \in \mathbb{C}, i \neq 1$ satisfy $\text{Re}(\lambda_i) < 0$ or $|\lambda_i| < 1$, respectively.

The second part of the Perron Frobenius theorem also implies convergence of the transition functions to the stationary distribution, since

$$ p_t(x, y) = \sum_{i=1}^{|S|} \langle \delta_x | u_i \rangle \langle v_i | e^{\lambda_i t} \rightarrow \langle v_1 | = \langle \pi | \quad \text{as } t \to \infty. $$
2. Sample paths

Sample paths \( t \mapsto X_t(\omega) \) are piecewise constant and right-continuous by convention. For \( X_0 = x \), define the **holding time** \( W_x := \inf \{ t > 0 : X_t \neq x \} \).

**Proposition 2.4**

\( W_x \sim \text{Exp}(|g(x, x)|) \), i.e. it is **exponentially distributed** with mean \( 1/|g(x, x)| \), and if \( |g(x, x)| > 0 \) the chain jumps to \( y \neq x \) after time \( W_x \) with probability \( g(x, y)/|g(x, x)| \).

**Proof.** \( W_x \) has the **memoryless property**, i.e. for all \( t, u > 0 \)

\[
P(W_x > t + u | W_x > t) = P(W_x > t + u | X_t = x) = P(W_x > u)
\]

where we used the Markov property and homogeneity. Therefore

\[
P(W_x > t + u) = P(W_x > u)P(W_x > t) \Rightarrow P(W_x > t) = e^{\gamma t}
\]

where

\[
\gamma = \left. \frac{d}{dt} P(W_x > t) \right|_{t=0} = \lim_{\Delta t \downarrow 0} \frac{p_{\Delta t}(x, x) + o(\Delta t) - 1}{\Delta t} = g(x, x) \leq 0 .
\]

Conditioned on leaving the current state shortly, the probability to jump to \( y \) is

\[
\lim_{\Delta t \downarrow 0} \frac{p_{\Delta t}(x, y)}{1 - p_{\Delta t}(x, x)} = \lim_{\Delta t \downarrow 0} \frac{\Delta t g(x, y)}{1 - 1 - \Delta t g(x, x)} = \frac{g(x, y)}{-g(x, x)} .
\]
2. Sample paths

- the **jump times** $J_0, J_1, \ldots$ are defined recursively as
  \[ J_0 = 0 \quad \text{and} \quad J_{n+1} = \inf\{ t > J_n : X_t \neq X_{J_n} \} . \]

- due to right-continuous paths, jump times are **stopping times**, i.e. for all $t \geq 0$, the event $\{J_n \leq t\}$ depends only on $(X_s : 0 \leq s \leq t)$.

- By the **strong Markov property** (allows conditioning on state at stopping time), subsequent holding times and jump probabilities are all independent.

- The **jump chain** $(Y_n : n \in \mathbb{N}_0)$ with $Y_n := X_{J_n}$ is then a discrete-time Markov chain with transition matrix
  \[ p^Y(x, y) = \begin{cases} 0 & , x = y \quad \text{if } g(x, x) < 0 \quad \text{and} \\ g(x, y)/|g(x, x)| & , x \neq y \quad \text{if } g(x, x) > 0 \\ \delta_{x, y} & \text{if } g(x, x) = 0 \end{cases} \]

- A **sample path** is constructed by simulating the jump chain $(Y_n : n \in \mathbb{N}_0)$ together with independent **holding times** $(W_{Y_n} : n \in \mathbb{N}_0)$, so that $J_n = \sum_{k=0}^{n-1} W_{Y_k}$
2. Examples

- A **Poisson process** with **rate** $\lambda$ (short PP($\lambda$)) is a CTMC with

  $$S = \mathbb{N}_0, \ X_0 = 0 \ \text{and} \ g(x, y) = \lambda \delta_{x+1,y} - \lambda \delta_{x,y}.$$  

  The PP($\lambda$) has **stationary and independent increments** with

  $$\mathbb{P}[X_{t+u} = n + k | X_u = n] = p_t(0, k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \text{for all} \ u, t > 0, \ k, n \in \mathbb{N}_0$$

  since $\pi_t(k) = p_t(0, k)$ solves the Master equation $\frac{d}{dt}\pi_t(k) = (\pi_t G)(k)$.

- A **birth-death chain** with **birth rates** $\alpha_x$ and **death rates** $\beta_x$ is a CTMC with

  $$S = \mathbb{N}_0 \ \text{and} \ g(x, y) = \alpha_x \delta_{x+1,y} + \beta_x \delta_{x-1,y} - (\alpha_x + \beta_x) \delta_{x,y},$$

  where $\beta_0 = 0$.

  Special cases include
  - **M/M/1 server queues**: $\alpha_x \equiv \alpha > 0, \ \beta_x \equiv \beta > 0$ for $x > 1$
  - **M/M/$\infty$ server queues**: $\alpha_x \equiv \alpha > 0, \ \beta_x = x\beta$
  - **population growth model**: $\alpha_x = x\alpha, \ \beta_x = x\beta$
2. Ergodicity

**Definition 2.4**

A Markov process is called **ergodic** if it has a unique stationary distribution $\pi$ and

$$p_t(x, y) = \mathbb{P}[X_t = y | X_0 = x] \to \pi(y) \text{ as } t \to \infty , \text{ for all } x, y \in S .$$

**Theorem 2.5**

An **irreducible** (aperiodic) MC with finite state space is **ergodic**.

**Theorem 2.6 (Ergodic Theorem)**

Consider an **ergodic Markov chain** with unique stationary distribution $\pi$. Then for every bounded function $f : S \to \mathbb{R}$ we have with probability 1

$$\frac{1}{T} \int_0^T f(X_t) \, dt \quad \text{or} \quad \frac{1}{N} \sum_{n=1}^N f(X_n) \to \mathbb{E}_{\pi}[f] \quad \text{as } T, N \to \infty .$$

- for a proof see e.g. [GS], chapter 9.5
- in practice, use relaxation/burn-in time before computing time averages
2. Markov Chain Monte Carlo (MCMC)

Typical problems related to sampling from $\pi$ on a very large state space $S$

- Compute expectations $\mathbb{E}_\pi[f] = \sum_{x \in S} f(x) \pi(x)$
- for Gibbs measures $\pi(x) = \frac{1}{Z(\beta)} e^{-\beta H(x)}$ (stach. mech. problems), compute partition function $Z(\beta) = \sum_{x \in S} e^{-\beta H(x)}$

Use the ergodic theorem to estimate expectations by time averages

- assume $\pi(x) > 0$ for all $x \in S$ (otherwise restrict $S$)
- invent CTMC/DTMC such that $\pi$ is stationary, e.g. via detailed balance

$$\pi(x) g(x, y) = \pi(y) g(y, x) \quad \text{or} \quad \pi(x) p(x, y) = \pi(y) p(y, x)$$

for Gibbs measures $e^{-\beta H(x)} g(x, y) = e^{-\beta H(y)} g(y, x)$

Typically $g(x, y) = q(x, y) a(x, y)$, i.e. propose move from $x$ to $y$ with rate $q(x, y) = q(y, x)$ (irreducible on $S$ but ’local’), accept with probability $a(x, y)$

- **Heat bath algorithm:** $a(x, y) = \frac{e^{-\beta H(y)}}{e^{-\beta H(x)} + e^{-\beta H(y)}}$

- **Metropolis-Hastings:** $a(x, y) = \begin{cases} 1, & \text{if } H(y) \leq H(x) \\ e^{\beta (H(x) - H(y))}, & \text{if } H(y) > H(x) \end{cases}$
2. Reversibility

Proposition 2.7 (Time reversal)

Let $(X_t : t \in [0, T])$ be a finite state, irreducible CTMC with generator $G^X$ on a compact time interval which is \textit{stationary}, i.e. $X_t \sim \pi$ for $t \in [0, T]$. Then the \textbf{time reversed chain}

$$(Y_t : t \in [0, T]) \quad \text{with} \quad Y_t := X_{T-t}$$

is a stationary CTMC with generator $g^Y(x, y) = \frac{\pi(y)}{\pi(x)} g^X(y, x)$ and stat. prob. $\pi$.

- An analogous statement holds for stationary, finite state, irreducible DTMCs with $p^Y(x, y) = \frac{\pi(y)}{\pi(x)} p^X(y, x)$.
- Stationary chains with reversible $\pi$ are \textbf{time-reversible}, $g^Y(x, y) = g^X(x, y)$.
- The definition of stationary chains can be extended to negative times, $(X_t : t \in \mathbb{R})$, with the time reversed chain given by $Y_t := X_{-t}$.
- The time reversal of non-stationary MCs is in general \textbf{not} a homogeneous MC, for DTMCs using Bayes’ Theorem we get $p^Y(x, y; n) = \frac{\pi_{N-n-1}(y)}{\pi_{N-n}(x)} p^X(y, x)$.
2. Countably infinite state space

For infinite state space, Markov chains can get ’lost at infinity’ and have no stationary distribution. Let \( T_x := \inf \{ t > J_1 : X_t = x \} \) be the first return time to a state \( x \).

(For DTMCs return times are defined as \( T_x := \inf \{ n \geq 1 : X_n = x \} \))

**Definition 2.5**

A state \( x \in S \) is called

- **transient**, if \( \mathbb{P}[T_x = \infty | X_0 = x] > 0 \)
- **null recurrent**, if \( \mathbb{P}[T_x < \infty | X_0 = x] = 1 \) and \( \mathbb{E}[T_x | X_0 = x] = \infty \)
- **positive recurrent**, if \( \mathbb{P}[T_x < \infty | X_0 = x] = 1 \) and \( \mathbb{E}[T_x | X_0 = x] < \infty \)

and these properties partition \( S \) into **communicating classes**.

- For an irreducible MC all states are either transient, null or positive recurrent.
- A MC has a unique stationary distribution if and only if it is positive recurrent

and in this case \( \pi(x) = \frac{1}{\mathbb{E}[T_x | X_0 = x]} \mathbb{E} \left[ \int_0^{T_x} \mathbf{1}_x(X_s) ds | X_0 = x \right] \).
2. Countably infinite state space

A CTMC with an infinite transient component in $S$ can exhibit explosion.

**Definition 2.6**

For a CTMC define the explosion time

$$J_\infty := \lim_{n \to \infty} J_n \in (0, \infty] \quad \text{where } J_n \text{ are the jump times of the chain}.$$ 

The chain is called **non-explosive** if $\mathbb{P}[J_\infty = \infty] = 1$, otherwise it is **explosive**.

- If the exit rates are uniformly bounded, i.e. $\sup_{x \in S} |g(x, x)| < \infty$, then the chain is non-explosive, which is always the case if $S$ is finite.
- As an example, consider a pure birth chain with $X_0 = 1$ and rates

$$g(x, y) = \alpha_x \delta_{y,x+1} - \alpha_x \delta_{y,x}, \quad x, y \in S = \mathbb{N}_0.$$ 

If $\alpha_x \to \infty$ fast enough (e.g. $\alpha_x = x^2$) we get

$$\mathbb{E}[J_\infty] = \sum_{x=1}^{\infty} \mathbb{E}[W_x] = \sum_{x=1}^{\infty} \frac{1}{\alpha_x} < \infty$$

since holding times $W_x \sim \text{Exp}(\alpha_x)$. This implies $\mathbb{P}[J_\infty = \infty] = 0 < 1$. 
3. Markov processes with $S = \mathbb{R}$

**Proposition 3.1**

Let $(X_t : t \geq 0)$ be a homogeneous MP as in Definition 18 with state space $S = \mathbb{R}$. Then for all $t \geq 0$ the **transition kernel** for all $x, y \in \mathbb{R}$

$$P_t(x, dy) := \mathbb{P}[X_t \in dy | X_0 = x] = \mathbb{P}[X_{t+u} \in dy | X_u = x] \quad \text{for all } u \geq 0$$

is well defined. If it is absolutely continuous the **transition density** $p_t$ with

$$P_t(x, dy) = p_t(x, y) dy$$

exists and fulfills the **Chapman Kolmogorov equations**

$$p_{t+u}(x, y) = \int_{\mathbb{R}} p_t(x, z) p_u(z, y) dz \quad \text{for all } t, u \geq 0, \, x, y \in \mathbb{R}.$$ 

As for CTMCs, the transition densities and the initial distribution $p_0(x)$ describe all **finite dimensional distributions (fdds)**

$$\mathbb{P}[X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n] = \int_{\mathbb{R}} dz_0 p_0(z_0) \int_{-\infty}^{x_1} dz_1 p_{t_1}(z_0, z_1) \cdots \int_{-\infty}^{x_n} dz_n p_{t_n-t_{n-1}}(z_{n-1}, z_n)$$

for all $n \in \mathbb{N}$, $0 < t_1 < \ldots < t_n$ and $x_1, \ldots, x_n \in \mathbb{R}$. 
3. Jump processes

\((X_t : t \geq 0)\) is a **jump process** with state space \(S = \mathbb{R}\) characterized by a

**jump rate density** \(r(x, y) \geq 0\) with a uniformly bounded

**total exit rate** \(R(x) = \int_{\mathbb{R}} r(x, y) \, dy < \bar{R} < \infty\) for all \(x \in \mathbb{R}\).

**Ansatz** for transition function as \(\Delta t \to 0\):

\[ p_{\Delta t}(z, y) = r(z, y) \Delta t + (1 - R(z) \Delta t) \delta(y - z) \]

Then use the Chapman Kolmogorov equations

\[
p_{t+\Delta t}(x, y) - p_t(x, y) = \int_{\mathbb{R}} p_t(x, z) p_{\Delta t}(z, y) \, dz - p_t(x, y) = \\
= \int_{\mathbb{R}} p_t(x, z) r(z, y) \Delta t \, dz + \int_{\mathbb{R}} (1 - R(z) \Delta t - 1) p_t(x, z) \delta(y - z) \, dz
\]

to get the **Kolmogorov-Feller equation** \((x\) is a fixed initial condition)

\[
\frac{\partial}{\partial t} p_t(x, y) = \int_{\mathbb{R}} \left( p_t(x, z) r(z, y) - p_t(x, y) r(y, z) \right) \, dz.
\]

As for CTMC sample paths \(t \mapsto X_t(\omega)\) are piecewise constant and right-continuous.
3. Gaussian processes

$\mathbf{X} = (X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \Sigma)$ is a **multivariate Gaussian** in $\mathbb{R}^n$ if it has PDF

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp \left( -\frac{1}{2} \langle \mathbf{x} - \mu \mid \Sigma^{-1} \mid \mathbf{x} - \mu \rangle \right),$$

with **mean** $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$ and **covariance matrix**

$$\Sigma = (\sigma_{ij} : i, j = 1, \ldots, n), \quad \sigma_{ij} = \text{Cov}[X_i, X_j] = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)].$$

**Definition 3.1**

A stochastic process $(X_t : t \geq 0)$ with state space $S = \mathbb{R}$ is a **Gaussian process** if for all $n \in \mathbb{N}$, $0 \leq t_1 < \ldots < t_n$ the vector $(X_{t_1}, \ldots, X_{t_n})$ is a multivariate Gaussian.

**Proposition 3.2**

All fdds of a Gaussian process $(X_t : t \geq 0)$ are fully characterized by the **mean** and the **covariance function**

$$m(t) := \mathbb{E}[X_t] \quad \text{and} \quad \sigma(s, t) := \text{Cov}[X_s, X_t].$$
3. Stationary independent increments

**Definition 3.2**

A stochastic process \((X_t : t \geq 0)\) has **stationary increments** if

\[
X_t - X_s \sim X_{t-s} - X_0 \quad \text{for all } 0 \leq s \leq t.
\]

It has **independent increments** if for all \(n \geq 1\) and \(0 \leq t_1 < \cdots < t_n\)

\[
\{X_{t_{k+1}} - X_{t_k} : 1 \leq k < n\}
\]

are independent.

**Example.** The Poisson process \((N_t : t \geq 0) \sim PP(\lambda)\) has stationary independent increments with \(N_t - N_s \sim \text{Poi}(\lambda(t - s))\).

**Proposition 3.3**

The following two statements are equivalent for a stochastic process \((X_t : t \geq 0)\):

- \(X_t\) has stationary independent increments and \(X_t \sim \mathcal{N}(0, t)\) for all \(t \geq 0\).
- \(X_t\) is a Gaussian process with \(m(t) = 0\) and \(\sigma(s, t) = \min\{s, t\}\).

Stationary independent incr. have **stable distributions** such as Gaussian or Poisson.
3. Brownian motion

**Definition 3.3**

**Standard Brownian motion** \((B_t : t \geq 0)\) is a stochastic process that satisfies either of the two equivalent properties in Proposition 3.3 and has **continuous paths**, i.e.

\[
P\left[\{\omega : t \mapsto B_t(\omega) \text{ is continuous in } t \geq 0\}\right] = 1.
\]

**Theorem 3.4 (Wiener 1923)**

There exists a probability space \((\Omega, \mathcal{F}, P)\) on which standard Brownian motion exists.

**Proof idea.** Construction on \(\Omega = \mathbb{R}^{[0, \infty)}\), using Kolmogorov's extension theorem: For every 'consistent' description of finite dimensional distributions (fdds) there exists a 'canonical' process \(X_t[\omega] = \omega(t)\) characterized by a law \(P\) on \(\Omega\). The main problem is to show that there exists a 'version' of the process that has continuous paths, i.e. \(P\) can be chosen to concentrate on continuous paths \(\omega\).

**Remark.** Construction of \((N_t : t \geq 0) \sim PP(\lambda)\) is

\[
N_t := \max \left\{ k \geq 1 : \tau_1 + \cdots + \tau_k \leq t \right\}, \quad \tau_1, \tau_2, \cdots \sim \text{Exp}(\lambda) \text{ iids}
\]
3. Properties of Brownian motion

- SBM is a time-homogeneous MP with \( B_0 = 0 \).
- \( \sigma B_t + x \) with \( \sigma > 0 \) is a (general) BM with \( B_t \sim \mathcal{N}(x, \sigma^2 t) \).
- The transition density is given by a Gaussian PDF
  \[
  p_t(x, y) = \frac{1}{\sqrt{2\pi \sigma^2 t}} \exp \left( -\frac{(y - x)^2}{2\sigma^2 t} \right)
  \]
  This is also called the heat kernel, since it solves the heat/diffusion equation
  \[
  \frac{\partial}{\partial t} p_t(x, y) = \frac{\sigma^2}{2} \frac{\partial^2}{\partial y^2} p_t(x, y) \quad \text{with} \quad p_0(x, y) = \delta(y - x).
  \]
- SBM is self-similar with Hurst exponent \( H = 1/2 \), i.e.
  \[
  (B_{\lambda t} : t \geq 0) \sim \lambda^H (B_t : t \geq 0) \quad \text{for all } \lambda > 0.
  \]
- \( t \mapsto B_t \) is \( \mathbb{P} \)–a.s. not differentiable at \( t \) for all \( t \geq 0 \).
  For fixed \( h > 0 \) define \( \xi^h_t := (B_{t+h} - B_t) / h \sim \mathcal{N}(0, 1/h) \), which is a mean-0 Gaussian process with covariance \( \sigma(s, t) = \begin{cases} 0 & \text{if } |t - s| > h \\ (h - |t - s|) / h^2 & \text{if } |t - s| < h \end{cases} \).
  The (non-existent) derivative \( \xi_t := \lim_{h \to 0} \xi^h_t \) is called white noise and is formally a mean-0 Gaussian process with covariance \( \sigma(s, t) = \delta(t - s) \).
3. Generators as operators

For a CTMC \((X_t: t \geq 0)\) with discrete state space \(S\) we have for \(f: S \rightarrow \mathbb{R}\)

\[
\mathbb{E}[f(X_t)] = \sum_{x \in S} \pi_t(x)f(x) = \langle \pi_t | f \rangle \quad \text{and} \quad \frac{d}{dt} \langle \pi_t | = \langle \pi_t | G \quad \text{master equation}
\]

Therefore
\[
\frac{d}{dt} \mathbb{E}[f(X_t)] = \frac{d}{dt} \langle \pi_t | f \rangle = \langle \pi_t | Gf \rangle = \mathbb{E}[(Gf)(X_t)]
\]

The generator \(G\) can be defined as an operator \(G\) acting on functions \(f: S \rightarrow \mathbb{R}\)

\[
G(f)(x) = (Gf)(x) = \sum_{y \neq x} g(x, y) [f(y) - f(x)]
\]

For Brownian motion use the heat eq. and integration by parts for \(f \in C^2(\mathbb{R})\)

\[
\frac{d}{dt} \mathbb{E}_x[f(X_t)] = \int_{\mathbb{R}} \partial_t p_t(x, y)f(y)dy = \frac{\sigma^2}{2} \int_{\mathbb{R}} \partial^2_y p_t(x, y)f(y)dy = \mathbb{E}_x[(Lf)(X_t)]
\]

where the generator of BM is \((Lf)(x) = \frac{\sigma^2}{2} \Delta f(x) \quad \text{or} \quad \frac{\sigma^2}{2} f''(x)\)

For jump processes with \(S = \mathbb{R}\) and rate density \(r(x, y)\) the generator is

\[
(Lf)(x) = \int_{\mathbb{R}} r(x, y) [f(y) - f(x)] dy
\]
3. Brownian motion as scaling limit

**Proposition 3.5**

Let \((X_t : t \geq 0)\) be a jump process on \(\mathbb{R}\) with **translation invariant rates** \(r(x, y) = q(y - x)\) which have

- **mean zero** \(\int_{\mathbb{R}} q(z) z \, dz = 0\) and
- **finite second moment** \(\sigma^2 := \int_{\mathbb{R}} q(z) z^2 \, dz < \infty\).

Then for all \(T > 0\) the rescaled process

\[(\epsilon X_{t/\epsilon^2} : t \in [0, T]) \Rightarrow (B_t : t \in [0, T]) \quad \text{as} \quad \epsilon \to 0\]

converges in distribution to a BM with generator \(\mathcal{L} = \frac{1}{2} \sigma^2 \Delta\) for all \(T > 0\).

**Proof.** Taylor expansion of the generator for test functions \(f \in C^3(\mathbb{R})\), and tightness argument for continuity of paths (requires fixed interval \([0, T]\)).
3. Diffusion processes

**Definition 3.4**

A **diffusion process** with drift \( a(x, t) \in \mathbb{R} \) and diffusion \( \sigma(x, t) > 0 \) is a real-valued process with continuous paths and generator

\[
(Lf)(x) = a(x, t)f'(x) + \frac{1}{2} \sigma^2(x, t)f''(x).
\]

**Examples.**

- The **Ornstein-Uhlenbeck process** is a diffusion process with generator

\[
(Lf)(x) = -\alpha xf'(x) + \frac{1}{2} \sigma^2 f''(x), \quad \alpha, \sigma^2 > 0.
\]

It has a Gaussian stationary distribution \( \mathcal{N}(0, \sigma^2/(2\alpha)) \).

If the initial distribution \( \pi_0 \) is Gaussian, this is a **Gaussian process**.

- The **Brownian bridge** is a Gaussian diffusion with \( X_0 = 0 \) and generator

\[
(Lf)(x) = -\frac{x}{1 - t} f'(x) + \frac{1}{2} f''(x).
\]

Equivalently, it can be characterized as a SBM conditioned on \( B_1 = 0 \).
3. Diffusion processes

Time evolution of the mean. Use \[ \frac{d}{dt} \mathbb{E}[f(X_t)] = \mathbb{E}[(Lf)(x_t)] \] with \( f(x) = x \)

\[ \frac{d}{dt} \mathbb{E}[X_t] = \mathbb{E}[a(X_t, t)] \]

Time evolution of the transition density. With \( X_0 = x \) we have for \( p_t(x, y) \)

\[ \int_{\mathbb{R}} \frac{\partial}{\partial t} p_t(x, y) f(y) dy = \frac{d}{dt} \mathbb{E}[f(X_t)] = \int_{\mathbb{R}} p_t(x, y) Lf(y) dy \quad \text{for any } f. \]

Use integration by parts to get the Fokker-Planck equation

\[ \frac{\partial}{\partial t} p_t(x, y) = -\frac{\partial}{\partial y} \left( a(y, t) p_t(x, y) \right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left( \sigma^2(y, t) p_t(x, y) \right). \]

Stationary distributions for time-independent \( a(y) \in \mathbb{R} \) and \( \sigma^2(y) > 0 \)

\[ \frac{d}{dy} \left( a(y) p^*(y) \right) = \frac{1}{2} \frac{d^2}{dy^2} \left( \sigma^2(y) p^*(y) \right), \]

leads to a stationary density (modulo normalization fixing \( p^*(0) \))

\[ p^*(x) = p^*(0) \exp \left( \int_0^x \frac{2a(y) - (\sigma^2)'(y)}{\sigma^2(y)} dy \right). \]
3. Beyond diffusion

Definition 3.5

A \textbf{Lévy process} \((X_t : t \geq 0)\) is a real-valued process with right-continuous paths and stationary, independent increments.

The generator has a part with \textbf{constant drift} \(a \in \mathbb{R}\) and \textbf{diffusion} \(\sigma^2 \geq 0\)

\[
\mathcal{L}f(x) = af'(x) + \frac{\sigma^2}{2}f''(x) + \int_{\mathbb{R}} \left( f(x+z) - f(x) - zf'(x) \mathbb{1}_{(0,1)}(|z|) \right) q(z)dz,
\]

and a translation invariant \textbf{jump part} with density \(q(z)\) (or measure \(\nu(dz)\)) and fulfills \(\int_{|z|>1} q(z)dz < \infty\) and \(\int_{0<|z|<1} z^2 q(z)dz < \infty\).

- Diffusion processes, in particular \textbf{BM} with \(a = 0, \sigma^2 > 0\) and \(q(z) \equiv 0\), or jump processes, in particular \textbf{Poisson} with \(a = \sigma = 0\) and \(q(z) = \lambda \delta(z - 1)\).
- For \(a = \sigma = 0\) and heavy-tailed jump distribution

\[
q(z) = \frac{C}{|z|^{1+\alpha}} \quad \text{with} \quad C > 0 \text{ and } \alpha \in (0, 2]
\]

the process is called \textbf{\(\alpha\)-stable symmetric Lévy process} or \textbf{Lévy flight}.

\textbf{self-similar} \((X_{\lambda t} : t \geq 0) \sim \lambda^H(X_t : t \geq 0)\), \(\lambda > 0\) with \(H = 1/\alpha\)

\(\Rightarrow\) \textbf{super-diffusive behaviour} with \(\mathbb{E}[X_t^2] \propto t^{2/\alpha}\)
3. Beyond diffusion

In general, a process \((X_t : t \geq 0)\) is said to exhibit **anomalous diffusion** if

\[
\text{Var}[X_t] / t \to \begin{cases} 
0 & \text{, sub-diffusive} \\
\infty & \text{, super-diffusive} 
\end{cases} \text{ as } t \to \infty.
\]

**Definition 3.6**

A fractional Brownian motion (fBM) \((B^H_t : t \geq 0)\) with **Hurst index** \(H \in (0, 1)\) is a **mean-zero Gaussian process** with continuous paths, \(B^H_0 = 0\) and **covariances**

\[
\mathbb{E}[B^H_t B^H_s] = \frac{1}{2} \left( t^{2H} + s^{2H} - |t - s|^{2H} \right) \text{ for all } s, t \geq 0.
\]

- For \(H = 1/2\), fBM is standard Brownian motion.
- fBM has stationary Gaussian increments where for all \(t > s \geq 0\)

\[
B^H_t - B^H_s \sim B^H_{t-s} \sim \mathcal{N}(0, (t-s)^{2H}),
\]

which for \(H \neq 1/2\) are **not** independent and the process is **non-Markov**.
- fBM is **self-similar**, i.e. \((B^H_{\lambda t} : t \geq 0) \sim \lambda^H (B^H_t : t \geq 0)\) for all \(\lambda > 0\).
3. Fractional BM and noise

- fBM exhibits **anomalous diffusion** with \( \text{Var}[B_t^H] = t^{2H} \)
- \( H > 1/2 \): super-diffusive with positively correlated increments
- \( H < 1/2 \): sub-diffusive with negatively correlated increments

\[
\mathbb{E}[B_1^H(B_{t+1}^H - B_t^H)] = \frac{(t+1)^{2H} - 2t^{2H} + (t-1)^{2H}}{2} \xrightarrow{t \to \infty} H(2H-1)t^{2(H-1)}
\]

For a **stationary process** \((X_t : t \geq 0)\) on \(\mathbb{R}\) define the **autocorrelation/covariance fct**

\[
c(t) := \text{Cov}[X_s, X_{s+t}] \quad \text{for all } s, t \in \mathbb{R}.
\]

Its Fourier transform is the **spectral density** \( S(\omega) := \int_{\mathbb{R}} c(t)e^{-i\omega t}dt \)

- **white noise** \((\xi_t : t \geq 0)\), stationary GP with mean zero and \( c(t) = \delta(t) \Rightarrow S(\omega) \equiv 1 \).

- **fractional** or **1/f noise** \((\xi_t^H : t \geq 0)\), stationary GP with mean zero and

\[
c(t) = \frac{2H(2H-1)}{|t|^{2(1-H)}} \Rightarrow S(\omega) \propto |\omega|^{2(1-H)-1} = \frac{1}{\omega^{2H-1}}
\]
3. SDEs and Itô’s formula

Let \((B_t : t \geq 0)\) be a standard BM. Then a diffusion process with drift \(a(x, t)\) and diffusion \(\sigma(x, t)\) solves the **Stochastic differential equation (SDE)**

\[
dX_t = a(X_t, t)dt + \sigma(X_t, t)dB_t.
\]

Here \(dB_t\) is white noise, interpreted in integrated form as

\[
X_t - X_0 = \int_0^t a(X_s, s)ds + \int_0^t \sigma(X_s, s)dB_s.
\]

**Theorem 3.6 (Itô’s formula for diffusions)**

Let \((X_t : t \geq 0)\) be a diffusion with generator \(\mathcal{L}\) and \(f : \mathbb{R} \to \mathbb{R}\) a smooth. Then

\[
f(X_t) - f(X_0) = \int_0^t (\mathcal{L}f)(X_s)ds + \int_0^t \sigma(X_s, s)f'(X_s)dB_s.
\]

or, equivalently in terms of SDEs

\[
df(X_t) = a(X_t, t)f'(X_t)dt + \frac{1}{2}\sigma^2(X_t, t)f''(X_t)dt + \sigma(X_t, t)f'(X_t)dB_t.
\]
3. SDEs and Itô’s formula

Itô’s formula for diffusions implies the following.

**Proposition 3.7**

Let \((X_t : t \geq 0)\) be a diffusion process with drift \(a(x, t)\) and diffusion \(\sigma(x, t)\), and \(f : \mathbb{R} \to \mathbb{R}\) a smooth invertible function. Then \((Y_t : t \geq 0)\) with \(Y_t = f(X_t)\) is a diffusion process with \((x = f^{-1}(y))\)

\[
\text{drift} \quad a(x, t)f'(x) + \frac{1}{2} \sigma^2(x, t)f''(x) \quad \text{and diffusion} \quad \sigma(x, t)f'(x).
\]

**Geometric BM.** \(Y_t := e^{\theta B_t}\), so \(f(x) = e^{\theta x}\) with \(f'(x) = \theta f(x)\) and \(f''(x) = \theta^2 f(x)\), where \((B_t : t \geq 0)\) is standard BM with \(a \equiv 0\), \(\sigma^2 \equiv 1\) and \(\theta \in \mathbb{R}\).

Then \((Y_t : t \geq 0)\) is a diffusion process with SDE

\[
dY_t = \frac{\theta}{2} Y_t dt + \theta Y_t dB_t.
\]

**Exponential martingale.**

\(Z_t := e^{\theta B_t - \frac{\theta^2}{2} t}\), so \(f(x, t) = e^{\theta x - \theta^2 t/2}\) and \(\partial_t f(x, t) = -\frac{\theta^2}{2} f(x, t)\)

Then

\[
dZ_t = \frac{\theta}{2} Z_t dt - \frac{\theta}{2} Z_t dt + \theta Z_t dB_t = \theta Z_t dB_t
\]

and \((Z_t : t \geq 0)\) is a martingale (see next slide) with \(\mathbb{E}[Z_t] \equiv Z_0 = 1\).
3. Fluctuations and martingales

**Definition 3.7**

A real-valued stochastic process \((M_t : t \geq 0)\) is a **martingale** w.r.t. the process \((X_t : t \geq 0)\) if for all \(t \geq 0\) we have \(\mathbb{E}[|M_t|] < \infty\) and

\[
\mathbb{E}[M_t \mid \{X_u : 0 \leq u \leq s\}] = M_s \quad \text{a.s. for all } s \leq t.
\]

If in addition \(\mathbb{E}[M_t^2] < \infty\), there exists a unique increasing process \([M]_t : t \geq 0\) called the **quadratic variation**, with \([M]_0 = 0\) and such that \(M_t^2 - [M]_t\) is martingale.

**Theorem 3.8 (Itô’s formula)**

Let \((X_t : t \geq 0)\) be a Markov process on state space \(S\) with generator \(\mathcal{L}\). Then for any smooth enough \(f : S \times [0, \infty) \to \mathbb{R}\)

\[
f(X_t, t) - f(X_0, 0) = \int_0^t \mathcal{L}f(X_s, s) ds + \int_0^t \partial_s f(X_s, s) ds + M^f_t,
\]

where \((M^f_t : t \geq 0)\) is a martingale w.r.t. \((X_t : t \geq 0)\) with \(M^f_0 = 0\) and

quadratic variation \([M^f]_t = \int_0^t \left((\mathcal{L}f^2)(X_s, s) - 2(f \mathcal{L}f)(X_s, s)\right) ds\).
3. Fluctuations and martingales

- For a **Poisson process** \((N_t : t \geq 0)\) with rate \(\lambda > 0\) Itô’s formula implies that
  \[ M_t := N_t - \lambda t \quad \text{is a martingale with quadr. variation} \quad [M]_t = \lambda t. \]

- **Watanabe’s characterization of PP**: Let \((N_t : t \geq 0)\) be a **counting process**, i.e. a jump process on \(S = \mathbb{N}\) with jump size \(+1\) only. If \(M_t = N_t - \lambda t\) is a martingale, then \((N_t : t \geq 0) \sim PP(\lambda)\).

- For a **diffusion process**, choosing \(f(X_t, t) = X_t\) in Itô’s formula leads to
  \[ X_t - X_0 = \int_0^t a(X_s, s)ds + M_t \quad \text{with} \quad [M]_t = \int_0^t \sigma^2(X_s, s)ds. \]

  In particular for BM with \(a(x, t) \equiv 0\) and \(\sigma^2(x, t) \equiv \sigma^2\) we have
  \((B_t : t \geq 0)\) is a martingale with quadratic variation \([B]_t = t\).

- **Lévy’s characterization of BM**: Any continuous martingale \((M_t : t \geq 0)\) on \(\mathbb{R}\) with \(M_0 = 0\) and quadratic variation \([M]_t = t\) is standard Brownian motion.

  Furthermore, **any continuous martingale** \((M_t : t \geq 0)\) on \(\mathbb{R}\) with \(M_0 = 0\) is a continuous (random) time-change of a standard BM, i.e.
  \((M_t : t \geq 0) \sim (B_{[M]} : t \geq 0)\) for SBM \((B_t : t \geq 0)\).
3. Fluctuations and martingales

- For a diffusion process \((X_t : t \geq 0)\) we have

  \[
  X_t - X_0 = \int_0^t a(X_s, s)ds + M_t \quad \text{with} \quad [M]_t = \int_0^t \sigma^2(X_s, s)ds .
  \]

  with \(M_t\) a continuous martingale \(\Rightarrow\) \((M_t : t \geq 0) \sim (B_{[M]}_t : t \geq 0)\)

- Related time-changed BMs can be written as stochastic Itô integrals

  \[
  M_t = \int_0^t \sigma(X_s, s)dB_s := B_{[M]}_t .
  \]

  Therefore \(\sigma \equiv 0\) implies deterministic dynamics with \(M_t \equiv 0\),
  and the corresponding SDE is an ODE \(dX_t/dt = a(X_t, t)\).

  Vanishing drift \(a \equiv 0\) implies \(X_t - X_0 = M_t\) or \(dX_t = \sigma(X_t, t)dB_t\)
  and the process \((X_t : t \geq 0)\) is a martingale.

- Recall the exponential martingale \(e^{\theta B_t - \theta^2 t/2}\) as a non-trivial example.
3. Martingales and conservation laws

Consider a CTMP \((X_t : t \geq 0)\) on state space \(S\) with generator \(\mathcal{L}\), and an observable \(f : S \to \mathbb{R}\) such that \(\mathcal{L}f : S \to \mathbb{R}\) is well defined (e.g. \(f \in C^2(S, \mathbb{R})\) for diffusions).

**Proposition 3.9**

If \(\mathcal{L}f(x) = 0\) for all \(x \in S\), then \(f(X_t)\) is a martingale, and is conserved in expectation, i.e. (for any initial condition \(X_0\))

\[
E[f(X_t)] = E[f(X_0)] \quad \text{for all } t \geq 0.
\]

If in addition \(\mathcal{L}f^2(x) = 0\) for all \(x \in S\), then \(f(X_t)\) is conserved (or a conserved quantity), i.e. (for any initial condition \(X_0\))

\[
f(X_t) = f(X_0) \quad \text{almost surely for all } t \geq 0.
\]

**Proof.** The first claim follows directly from Itô’s formula (Theorem 3.8). For the second claim, we have \(f(X_t) = f(X_0) + M_f^t\) and \(M_f^t\) has quadratic variation

\[
[M_f^t]_t = \int_0^t \left( (\mathcal{L}f^2)(X_s, s) - 2(f \mathcal{L}f)(X_s, s) \right) ds = 0,
\]

for all \(t \geq 0\), which implies \(M_f^t \equiv 0\) almost surely.
4. Stochastic particle systems

- **lattice/population**: $\Lambda = \{1, \ldots, L\}$, finite set of points
- **state space** $S$ is given by the set of all **configurations**
  \[ \eta = (\eta(i) : i \in \Lambda) \in S = \{0, 1\}^L \] (often also written $\{0, 1\}^\Lambda$).

$\eta(i) \in \{0, 1\}$ signifies the presence of a particle/infection at site/individual $i$.

- Only local transitions are allowed with rates
  \[
  \begin{align*}
  \eta \to \eta^i & \quad \text{with rate } c(\eta, \eta^i) \quad \text{(reaction)} \\
  \eta \to \eta^{ij} & \quad \text{with rate } c(\eta, \eta^{ij}) \quad \text{(transport)}
  \end{align*}
  \]

where \[ \eta^i(k) = \begin{cases} 
\eta(k) & k \neq i \\
1 - \eta(k) & k = i 
\end{cases} \]
and \[ \eta^{ij}(k) = \begin{cases} 
\eta(k) & k \neq i, j \\
\eta(j) & k = i \\
\eta(i) & k = j 
\end{cases} \]

**Definition 4.1**

A **stochastic particle system** is a CTMC with state space $S = \{0, 1\}^\Lambda$ and generator

\[
\mathcal{L}f(\eta) = \sum_{i \in \Lambda} c(\eta, \eta^i) [f(\eta^i) - f(\eta)] \quad \text{or} \quad \mathcal{L}f(\eta) = \sum_{i, j \in \Lambda} c(\eta, \eta^{ij}) [f(\eta^{ij}) - f(\eta)].
\]
4. Contact process

The contact process is a simple stochastic model for the SI epidemic with infection rates \( q(i, j) \geq 0 \) and uniform recovery rate 1.

**Definition 4.2**

The contact process (CP) \((\eta_t : t \geq 0)\) is an IPS with rates

\[
c(\eta, \eta^i) = 1 \cdot \delta_{\eta(i), 1} + \delta_{\eta(i), 0} \sum_{j \neq i} q(j, i) \delta_{\eta(j), 1} \quad \text{for all } i \in \Lambda.
\]

Usually, \( q(i, j) = q(j, i) \in \{0, \lambda\} \), i.e. connected individuals infect each other with fixed rate \( \lambda > 0 \).

- The CP has one absorbing state \( \eta(i) = 0 \) for all \( i \in \Lambda \), which can be reached from every initial configuration. Therefore the process is ergodic and the infection eventually gets extinct with probability 1.

- Let \( T := \inf\{t > 0 : \eta_t \equiv 0\} \) be the extinction time. Then there exists a critical value (epidemic threshold) \( \lambda_c > 0 \) such that (for irreducible \( q(i, j) \))

\[
\mathbb{E}[T|\eta_0 \equiv 1] \propto \log L \quad \text{for } \lambda < \lambda_c \quad \text{and} \quad \mathbb{E}[T|\eta_0 \equiv 1] \propto e^{CL} \quad \text{for } \lambda > \lambda_c.
\]
4. Voter model

The voter model describes opinion dynamics with **influence rates** \( q(i, j) \geq 0 \) at which individual \( i \) persuades \( j \) to switch to her/his opinion.

**Definition 4.3**

The **linear voter model (VM)** \((\eta_t : t \geq 0)\) is an IPS with rates

\[
c(\eta, \eta^i) = \sum_{j \neq i} q(j, i) \left( \delta_{\eta(i), 1} \delta_{\eta(j), 0} + \delta_{\eta(i), 0} \delta_{\eta(j), 1} \right) \quad \text{for all } i \in \Lambda.
\]

In non-linear versions the rates can be replaced by general (symmetric) functions.

- The VM is **symmetric** under relabelling opinions \( 0 \leftrightarrow 1 \).
- If \( q(i, j) \) is irreducible there are two absorbing states, \( \eta \equiv 0, 1 \), both of which can be reached from every initial condition. Therefore the VM is not ergodic, and **stationary measures** are

  \[
  \alpha \delta_0 + (1 - \alpha) \delta_1 \quad \text{with } \alpha \in [0, 1] \text{ depending on the initial condition}.
  \]

- **Coexistence** of both opinions can occur on infinite lattices (e.g. \( \mathbb{Z}^d \) for \( d \geq 3 \)).
4. Exclusion process

The exclusion process describes transport of a conserved quantity (e.g. mass or energy) with transport rates \( q(i, j) \geq 0 \) site \( i \) to \( j \).

**Definition 4.4**

The exclusion process (EP) \( (\eta_t : t \geq 0) \) is an IPS with rates

\[
c(\eta, \eta^{ij}) = q(i, j)\delta_{\eta(i), 1}\delta_{\eta(j), 0} \quad \text{for all} \ i, j \in \Lambda.
\]

The EP is called **simple (SEP)** if jumps occur only between nearest neighbours on \( \Lambda \). The SEP is **symmetric (SSEP)** if \( q(i, j) = q(j, i) \), otherwise **asymmetric (ASEP)**.

- The SEP is mostly studied in a 1D geometry with periodic or open boundaries.
- For periodic boundary conditions the total number of particles \( N = \sum_i \eta(i) \) is **conserved**. The process is ergodic on the sub-state space
  \[
  S_N = \left\{ \eta \in \{0, 1\}^L : \sum_i \eta(i) = N \right\}
  \]
  for each value \( N = 0, \ldots, L \), and has a unique stationary distribution.
- For open boundaries particles can be created and destroyed at the boundary, the system is ergodic on \( S \) and has a unique stationary distribution.