1. Probability

- **probability space** \( \Omega \)  (e.g. \{H, T\} , \{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. \( \mathbb{P}[\bigcup_i A_i] = \sum_i \mathbb{P}[A_i] \),
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. \( \mathbb{P}[\text{even number on a die}] = \mathbb{P}[2] + \mathbb{P}[4] + \mathbb{P}[6] = 1/2 \)
- For **continuous** \( \Omega \) (e.g. \( [0, 1] \)): \( \mathcal{F} \subseteq \mathcal{P}(\Omega) \)
1. Independence and conditional probability

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

  **Example.** Rolling a die repeatedly

- If $\Pr[B] > 0$ then the **conditional probability** of $A$ given $B$ is
  $$\Pr[A|B] := \frac{\Pr[A \cap B]}{\Pr[B]}.$$  
  
  If $A$ and $B$ are independent, then $\Pr[A|B] = \Pr[A]$.

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

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

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

Note that also $\Pr[A|C] = \sum_{i=1}^n \Pr[A \cap C \cap B_i] \Pr[B_i|C]$.

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

$$F(x) = \Pr[X \leq x] = \Pr[\{\omega : X(\omega) \leq x\}].$$

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

$$\pi(x_k) := \Pr[X = x_k], \quad k = 1, 2, \ldots.$$  

$X$ 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'

$$f(x) = F'(x) = \sum_k \pi(x_k) \delta(x - x_k).$$

- The expected value of $X$ is given by

$$\mathbb{E}[X] = \left\{ \sum_\omega X(\omega) \Pr[\omega] = \sum_k x_k \pi(x_k) \right\} = \int_{\Omega} X(\omega) \, \Pr(\omega) = \int_{\mathbb{R}} x f(x) \, dx.$$  

- The variance is given by $\text{var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

  - 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**

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

where $f^X(x) = \int_{\mathbb{R}} f(x, y) \, dy$ and $\pi^X(x_k) = \sum_l \pi(x_k, x_l)$ are the marginals.

**Example.** (successive) coin tosses with $\Omega = \{H, T\}$ and $X(H) = -1, X(T) = 1$
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 (\textit{iidrv}'s) with

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

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

- for a single \textit{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 \textit{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 \( \mu := \mathbb{E}[X_1] < \infty \) and \( \mathbb{E}[|X_1|] < \infty \). Then

\[
\frac{1}{n} Y_n = \frac{1}{n} \sum_{k=1}^{n} X_k \to \mu \quad \text{as } n \to \infty
\]

in distribution (i.e. the distr. fact. of \( Y_n \) converges to \( \mathbb{I}_{(\mu, \infty)}(x) \) 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 \( \mu := \mathbb{E}[X_1] < \infty \) and \( \sigma^2 := \text{var}[X_1] < \infty \). Then

\[
\frac{Y_n - n\mu}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu) \to \xi \quad \text{as } n \to \infty
\]

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

Expansion. as \( n \to \infty \), \( \sum_{k=1}^{n} X_k = n\mu + \sqrt{n}\sigma\xi + o(\sqrt{n}) \), \( \xi \sim N(0, 1) \)

1. Discrete-time Markov processes

Definition 1.4

A \textit{discrete-time stochastic process} with \textit{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 \textit{Markov}, if for all \( A \subseteq S, n \in \mathbb{N}_0 \) and \( s_0, \ldots, s_n \in S \)

\[
P(Y_{n+1} \in A|Y_n = s_{n}, \ldots, Y_0 = s_0) = P(Y_{n+1} \in A|Y_n = s_0).
\]

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

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

If \( S \) is discrete, the MP is called a \textit{Markov chain (MC)}.

The generic probability space \( \Omega \) is the \textit{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 \textit{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 \( N \), \( P \) is a distribution on the finite path space \( \Omega_N \) with

\[
\mathbb{P}(\omega) = \delta_{0,Y_0(\omega)} \prod_{n=1}^N \left( p \delta_{1,Y_n(\omega)} - Y_{n-1}(\omega) + q \delta_{-1,Y_n(\omega)} - Y_{n-1}(\omega) \right)
\]

\[
= \begin{cases} p \# \text{ of up-steps} & q \# \text{ of down-steps} \\
0 & \text{path not possible}
\end{cases}
\]

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 iid rv’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 \((X_n : n \in \mathbb{N}_0)\) by a homogeneous DTMC with state space \( S \). Then the transition function

\[ p_n(x, y) := \mathbb{P}[X_{n+1} = y | X_n = x] = \mathbb{P}[X_{n+k} = y | X_k = x] \quad \text{for all } k \geq 0 \]

is well defined and fulfills the **Chapman Kolmogorov equations**

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

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

\[
\mathbb{P}[X_{n+k} = y | X_0 = x] = \sum_{z \in S} \mathbb{P}[X_{n+k} = y | X_k = z, X_0 = x] \mathbb{P}[X_k = z | X_0 = x]
\]

\[ = \sum_{z \in S} \mathbb{P}[X_{n+k} = y | X_k = z] \mathbb{P}[X_k = z | X_0 = x]
\]

\[ = \sum_{z \in S} \mathbb{P}[X_n = y | X_0 = z] \mathbb{P}[X_k = z | X_0 = x]
\]

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 \( P_0 = I \), 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 \( A_1, \ldots, A_k \subseteq S \)

\[
\mathbb{P}[X_1 \in A_1, \ldots, X_k \in A_k] = \sum_{s_0 \in A_1} \cdots \sum_{s_{k-1} \in A_k} p(X_0, s_1)p(s_1, s_2) \cdots p(s_{k-1}, s_k).
\]

- 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

\[
\pi_n(x) = \sum_{y \in S} \sum_{s_0 \in S} \cdots \sum_{s_{n-1} \in S} \pi_0(y)p(y, s_1) \cdots p(s_{n-1}, x) \quad \text{or } \quad \pi_n = \pi_0 P^n.
\]
1. Transition matrices

The transition matrix \( P \) is stochastic, i.e.

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

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 \( p(L, 1) = p, \quad p(1, L) = q \),
- **absorbing** if \( p(L, L) = 1, \quad p(1, 1) = 1 \),
- **closed** if \( p(1, 1) = q, \quad p(L, L) = p \),
- **reflecting** if \( p(1, 2) = 1, \quad p(L, L-1) = 1 \).

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 \( y \in S \)

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

\( \pi \) 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

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

- Stationary distributions are left **eigenvectors** with **eigenvalue** \( 1 \).

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 \( |u_i| \) and right (column) eigenvectors \( |v_i| \)

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

\[
A = \sum_{i=1}^{L} \lambda_i |v_i\rangle \langle u_i| \quad \text{and} \quad A^n = \sum_{i=1}^{L} \lambda_i^n |v_i\rangle \langle u_i|
\]

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

Since \( \pi_n = \pi_0 P^n \) we get

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

- The Gershgorin theorem implies that \( |\lambda_i| < 1 \) and the dependence on the initial condition \( \pi_0 \) decays exponentially in directions where \( |\lambda_i| < 1 \).
- \( \lambda_1 = 1 \) corresponds to the stationary distribution and \( |v_1\rangle = (1, \ldots, 1)^T \).
- Other eigenvalues with \( |\lambda_i| = 1 \) and \( \lambda_i \neq 1 \) correspond to persistent oscillations.
1. Lazy Markov chains

**Definition 1.6**

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

\[
p'(x, y) = \epsilon \delta_{x,y} + (1 - \epsilon) p(x, y), \quad \epsilon \in (0, 1)
\]

is called a **lazy version** of the original chain.

- Since all diagonal elements are bounded below by \( \epsilon > 0 \), the Gershgorin theorem now implies for the eigenvalues of \( P' \)

\[
|\lambda_i| = 1 \quad \Rightarrow \quad \lambda_i = 1.
\]

Such a matrix \( P' \) is called **aperiodic**, and there are no persistent oscillations.
- The stationary distribution is unique if and only if the eigenvalue \( \lambda = 1 \) has multiplicity 1, which is independent of lazyness and is discussed later.

1. Absorbing states

**Definition 1.7**

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 \( X_0 = k \in S = \{1, \ldots, L\} \),

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

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**

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

**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 \( \Rightarrow \quad h_k \equiv 1 \)

2. Continuous-time Markov chains

**Definition 2.1**

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**

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

For a given \( \omega \in \Omega \) the function \( t \mapsto X_t(\omega) \) is called a **sample path**.
2. Continuous-time Markov chains

Proposition 2.1

Let \((X_t : t \geq 0)\) by 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 \(P_{t+\Delta t} - P_t = P_t \Delta t^{-1} \Delta t = \Delta t P_t\), taking \(\Delta t \downarrow 0\) we get the so-called forward and backward equations

\[ \frac{d}{dt} P_t = P_G = G P_t , \quad \text{where} \quad G = \frac{dP_t}{dt} \bigg|_{t=0} \]

is called the generator of the process (sometimes also \(Q\)-matrix).

2. Continuous-time Markov chains

- The solution is given by the matrix exponential

\[ P_t = \exp(tG) = \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k = I + tG + \frac{t^2}{2} G^2 + \ldots \quad (2.1) \]

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

\[ \pi_t = \pi_0 \exp(tG) \quad \text{which solves} \quad \frac{d}{dt} \pi_t = \pi_t G . \quad (2.2) \]

- On a finite state space with \(\lambda_1, \ldots, \lambda_S \in \mathbb{C}\) being eigenvalues of \(G\), \(P_t\) has eigenvalues \(\exp(\lambda t)\) with the same eigenvectors \(\{v_i\}, \{u_i\}\).

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

\[ \langle \pi_0 \rangle \equiv \alpha_1 \langle v_1 \rangle + \ldots + \alpha_S \langle v_S \rangle \]

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

\[ \langle \pi_t \rangle = \alpha_1 \langle v_1 | e^{\lambda_1 t} \rangle + \ldots + \alpha_S \langle v_S | e^{\lambda_S t} \rangle \quad (2.3) \]

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

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

- 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.
2. Sample paths

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

\textbf{Proposition 2.2}

$W_x \sim \text{Exp}\left(\frac{1}{g(x,x)}\right)$, i.e. it is \textbf{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 $g(x,y)/|g(x,x)|$.

\textbf{Proof.} $W_x$ has the \textbf{memoryless property}, i.e. for all $t, u > 0$

$$
\mathbb{P}(W_x > t + u | W_x > t) = \mathbb{P}(W_x > t + u | X_t = x) = \mathbb{P}(W_x > u)
$$

where we used the Markov property and homogeneity. Therefore

$$
\mathbb{P}(W_x > t + u) = \mathbb{P}(W_x > u) \mathbb{P}(W_x > t) \Rightarrow \mathbb{P}(W_x > t) = e^{-\gamma t}
$$

where

$$
\gamma = \frac{d}{dt} \mathbb{P}(W_x > t) \bigg|_{t=0} = \lim_{\Delta t \searrow 0} \frac{p_{\Delta t}(x,x) + o(\Delta t) - 1}{\Delta t} = g(x,x) \leq 0.
$$

Conditioned on leaving the current state shortly, the probability to jump to $y$ is

$$
\lim_{\Delta t \searrow 0} \frac{p_{\Delta t}(x,y)}{1 - p_{\Delta t}(x,x)} = \lim_{\Delta t \searrow 0} \frac{\Delta t g(x,y)}{1 - \Delta t g(x,x)} = \frac{g(x,y)}{-g(x,x)}.
$$

2. Sample paths

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

- The \textbf{jump chain} $(Y_n : n \in \mathbb{N}_0)$ with $Y_0 := X_0$ is then a discrete-time Markov chain with transition matrix

$$
p^t(x,y) = \begin{cases} 
0 & \text{if } g(x,y) < 0 \\
g(x,y)/|g(x,x)| & \text{if } g(x,y) > 0
\end{cases}
$$

and

$$
p^t(x,y) = \delta_{x,y} \text{ if } g(x,x) = 0 \quad (\text{by convention}).
$$

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

2. Examples

- A \textbf{Poisson process} with \textbf{rate} $\lambda$ (short $\text{PP}(\lambda)$) is a CTMC with

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

The $\text{PP}(\lambda)$ has \textbf{stationary and independent increments} with

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

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

- A \textbf{birth-death chain} with \textbf{birth rates} $\alpha_x$ and \textbf{death rates} $\beta_x$ is a CTMC with

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

where $\beta_0 = 0$.

Special cases include

- $M/M/1$ server queues: $\alpha_\infty \equiv \alpha > 0, \beta_x \equiv \beta > 0$ for $x > 1$

- $M/M/\infty$ server queues: $\alpha_\infty \equiv \alpha > 0, \beta_x = x \beta$

- population growth model: $\alpha_x = x \alpha, \beta_x = x \beta$
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 \(\langle \pi \mid G \rangle = \langle 0 \mid\), or for all \(y \in S\)

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

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

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

- reversibility implies stationarity, since with (2.5) every single summand in (2.4) vanishes.
- Stationary distributions are left **eigenvectors** of \(G\) with **eigenvalue** 0.
- \(\langle \pi \mid G = \langle 0 \mid\) implies \(\langle \pi \mid P_t = \langle \pi \mid (\mathbb{I} + \sum_{k \geq 1} t^k G^k / k!) = \langle \pi \mid\) for all \(t \geq 0\).

**Proposition 2.3 (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 \(P \mathbb{1} = \mathbb{1}\) and \(G \mathbb{1} = \mathbb{0}\). 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 \(\langle \pi \mid\).

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

**Definition 2.3**

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

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

**Remark.** For continuous time irreducibility implies \(p_t(x, y) > 0\) for all \(t > 0\).

**Proposition 2.4 (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

- \(\lambda_1 = 1\) 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.
- if the chain is continuous-time or discrete-time aperiodic, all remaining eigenvalues \(\lambda_i \in \mathbb{C}\) 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, which is usually called ergodicity.
2. Ergodicity

**Theorem 2.5 (Ergodicity)**

An irreducible (aperiodic) MC with finite state space is ergodic, i.e. it has a unique stationary distribution \( \pi \) and

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

**Theorem 2.6 (Ergodic Theorem)**

Consider an ergodic Markov chain with unique stationary distribution \( \pi \). Then for every observable \( f : S \to \mathbb{R} \) we have

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

- for a proof see e.g. [GS], chapter 9.5
- Stationary expectations can be approximated by time averages, which is the basis for a technique called Markov chain Monte Carlo.
- for example, choosing the indicator function \( f = 1_x \) we get \( \mathbb{E}_\pi[f] = \pi(x) \)

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 stationary, i.e. \( X_t \sim \pi \) for \( t \in [0, T] \). Then the time reversed chain

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

is a stationary CTMC with generator

\[ g^Y(x, y) = \frac{\pi(y)}{\pi(x)} g^X(y, x) . \]

- 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 time-reversible,

\[ g^Y(x, y) = g^X(y, x) . \]

- 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 chains is in general not a homogeneous MP.

**Example.** SRW on finite state space

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 2.4**

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 \( \mathbb{E}[T_x | X_0 = x] < \infty \)

- For an irreducible MC all states are either transient, null or positive recurrent.
- The MC has a unique stationary distribution if and only if it is positive recurrent.
- A transient CTMC can exhibit explosion. Define the explosion time

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

The chain is called **non-explosive** if \( \mathbb{P}[J_\infty = \infty] = 1 \).

This is always the case if \( S \) is finite or \( \sup_{x \in S} |g(x, x)| < \infty \).
2. Stochastic particle systems

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

Only local transitions are allowed with rates
\[
\begin{align*}
\eta \to \eta' & \quad \text{with rate } c(\eta, \eta') \quad (\text{reaction}) \\
\eta \to \eta^0 & \quad \text{with rate } c(\eta, \eta^0) \quad (\text{transport})
\end{align*}
\]
where $\eta'(k) = \begin{cases} 
\eta(k) & k \neq i \\
1 - \eta(k) & k = i
\end{cases}$ and $\eta^0(k) = \begin{cases} 
\eta(k) & k \neq i, j \\
\eta(j) & k = i \\
\eta(i) & k = j
\end{cases}$

**Definition 2.5**

An **stochastic particle system** is a CTMC with state space $S = \{0, 1\}^\Lambda$ and generator $G$ with off-diagonal elements
\[
g(\eta, \zeta) = \sum_{i \in \Lambda} \mathbb{I}_{\eta'}(\zeta) c(\eta, \eta') \quad \text{(or } \sum_{i,j \in \Lambda} \mathbb{I}_{\eta'}(\zeta) c(\eta, \eta^0)) .
\]

2. 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 2.6**

The **contact process (CP)** $\eta_t : t \geq 0$ is an IPS with rates
\[
c(\eta, \eta') = 1 \cdot \delta_{\eta(0), 1} + \delta_{\eta(0), 0} \sum_{j \neq i} q(j, i) \delta_{\eta(j), 1} \quad \text{for all } i \in \Lambda .
\]

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 $q(i, j)$)
  \[
  \mathbb{E}[T|\eta_0 \equiv 1] \propto \log L \quad \text{for } \lambda < \lambda_c \quad \text{and} \quad \mathbb{E}[T|\eta_0 \equiv 1] \propto e^{CL} \quad \text{for } \lambda > \lambda_c .
  \]

3. Markov processes with $S = \mathbb{R}$

**Proposition 3.1**

Let $(X_t : t \geq 0)$ by 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)
\[
\mathbb{P}[X_0 \leq x_1, \ldots, X_n \leq x_n] = \int_{-\infty}^{x_1} dz_0 p_0(z_0) \int_{-\infty}^{x_2} dz_1 p_t(z_0, z_1) \cdots \int_{-\infty}^{x_n} dz_n p_{t_{n-1}}(z_{n-1}, z_n)
\]
for all $n \in \mathbb{N}, 0 < t_1 < \cdots < t_n$ and $x_1, \ldots, x_n \in \mathbb{R}$. 


32 / 63
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 \( R(x) = \int_{\mathbb{R}} r(x, y) \, dy < R < \infty \) 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

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

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

to get the **Kolmogorov-Feller equation**

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

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

### 3. Gaussian processes

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

\[
f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp \left( -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right),
\]

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 \( n \in \mathbb{N} \), and all \( t_1, \ldots, t_n \geq 0 \) the vector \( (X_{t_1}, \ldots, X_{t_n}) \) has a multivariate Gaussian distribution.

**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 \( 0 \leq t_1 < \cdots < t_n \)

\[
\{X_{t_{k+1}} - X_{t_k} : 1 \leq k < n\} \quad \text{are independent.}
\]

**Example.** The Poisson process \( (N_t : t \geq 0) \sim \text{PP}(\lambda) \) 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 increments 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.
\[
\mathbb{P} \left[ \{ \omega : B_t(\omega) \text{ is continuous in } t \geq 0 \} \right] = 1.
\]

**Theorem 3.4 (Wiener)**

There exists a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) 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 corresponding stochastic process on \(\Omega\), i.e. a 'unique' law \(\mathbb{P}\).

The main problem is to show that there exists a 'version' of the process that has continuous paths, i.e. \(\mathbb{P}\) can be chosen to concentrate on continuous paths \(\omega\).

**Remark.** Construction of \((N_t : t \geq 0) \sim PP(\lambda)\) is
\[
N_t := \max \left\{ k \geq 1 : \tau_1 + \cdots + \tau_k \leq t \right\}, \quad \tau_1, \tau_2, \cdots \sim \text{Exp}(\lambda) \text{idrvs}
\]

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.
  \[
  (B_t : t \geq 0) \sim \xi^H(B_{t/c} : t \geq 0) \quad \text{for all } c > 0.
  \]
- \(t \mapsto B_t\) 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 \(\sigma(s, t) = \begin{cases} 0 & |t - s| > h \\ (h - |t - s|)/h^2 & |t - s| < h \end{cases}\).
  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 state space \(S\) we have for \(f : S \to \mathbb{R}\)
\[
\frac{d}{dt} \mathbb{E}[f(X_t)] = \frac{d}{dt} \langle \pi_t | f \rangle = \frac{d}{dt} \langle \pi_0 | P_t | f \rangle = \langle \pi_0 | P_t G | f \rangle = \langle \pi_t | G | f \rangle.
\]

The **generator** \(G\) can be defined as an operator \(G : C(S) \to C(S)\)
\[
G(f)(x) = \langle G(f) | x \rangle = \sum_{y \neq y' \in S} g(x, y) \left[ f(y) - f(x) \right].
\]

For **Brownian motion** we have
\[
\frac{d}{dt} \langle \pi_t | f \rangle = \langle \pi_t | \mathcal{L} | f \rangle, \quad \text{where for } f \in C^2(\mathbb{R})
\]
the **generator of BM** is
\[
(\mathcal{L} f)(x) = \frac{\sigma^2}{2} \Delta f(x) \quad \text{or } \frac{\sigma^2}{2} f''(x).
\]

For **jump processes** with \(S = \mathbb{R}\) and rate density \(r(x, y)\) the generator is
\[
(\mathcal{L} f)(x) = \int_{\mathbb{R}} r(x, y) \left[ f(y) - f(x) \right] dy.
\]
3. Brownian motion as scaling limit

**Proposition 3.5**

Let $(X_t : t \geq 0)$ be a jump process with translation invariant rates $r(x, y) = q(y - x)$ which have

- mean zero $\int_R q(z) z \, dz = 0$ and
- finite second moment $\sigma^2 := \int_R q(z) z^2 \, dz < \infty$.

Then for all $T > 0$ the rescaled process

$$(\epsilon X_{t\epsilon^2} : t \in [0, T]) \Rightarrow (B_t : t \in [0, T])$$

as $\epsilon \to 0$ converges in distribution to a BM with generator $L = \frac{1}{2} \sigma^2 \Delta$ 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 $\sigma^2(x, t) > 0$ is a real-valued process with continuous paths and generator

$$(L f)(x) = a(x, t)f'(x) + \frac{1}{2} \sigma^2(x, t)f''(x).$$

**Examples.**
- The Ornstein-Uhlenbeck process is a diffusion process with generator
  $$(L f)(x) = -\alpha x f'(x) + \frac{1}{2} \sigma^2 f''(x), \quad \alpha, \sigma^2 > 0.$$
  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
  $$(L f)(x) = -\frac{x}{1-t} f'(x) + \frac{1}{2} f''(x).$$
  Equivalently, it can be characterized as a SBM conditioned on $B_1 = 0$.

4. Graphs - definition

**Definition 4.1**

A graph (or network) $G = (V, E)$ consists of a finite set $V = \{1, \ldots, N\}$ of vertices (or nodes, points), and a set $E \subseteq V \times V$ of edges (or links, lines). The graph is called undirected if $(i, j) \in E$ implies $(j, i) \in E$, otherwise directed. The structure of the graph is encoded in the adjacency (or connectivity) matrix

$$A = (a_{ij} : i, j \in V) \quad \text{where} \quad a_{ij} = \begin{cases} 1, & (i, j) \in E \\ 0, & (i, j) \notin E \end{cases}.$$ 

We denote the number of edges by $K = |E|$ for directed, or $K = |E|/2$ for undirected graphs.

- Graphs we consider do not have self edges, i.e. $(i, i) \notin E$ for all $i \in V$, or multiple edges, since edges $(i, j)$ are unique elements of $E$.
- Weighted graphs with edge weights $w_{ij} \in \mathbb{R}$ can be used to represent continuous- or discrete-time Markov chains.
- In general graphs can also be infinite, but we will focus on finite graphs. Many of the following graph characteristics only make sense in the finite case.
4. Graphs - paths and connectivity

**Definition 4.2**

A **path** $\gamma_{ij}$ of length $l = |\gamma_{ij}|$ from vertex $i$ to $j$ is sequence of vertices

$$\gamma_{ij} = (v_1 = i, v_2, \ldots, v_{l+1} = j) \quad \text{with} \quad (v_k, v_{k+1}) \in E \text{ for all } k = 1, \ldots, l,$$

and $v_k \neq v_{k'}$ for all $k \neq k' \in \{1, \ldots, l\}$ (i.e. each vertex is visited only once). If such a path exists, we say that vertex $i$ is **connected to** $j$ (write $i \rightarrow j$).

**Shortest paths** between vertices $i, j$ are called **geodesics** (not necessarily unique) and their length $d_{ij}$ is called the **distance** from $i$ to $j$. If $i \not\rightarrow j$ we set $d_{ij} = \infty$.

A graph is **connected** if $d_{ij} < \infty$ for all $i, j \in V$.

The **diameter** and the **characteristic path length** of the graph $G$ are given by

$$\text{diam}(G) := \max\{d_{ij} : i, j \in V\} \in \mathbb{N}_0 \cup \{\infty\},$$

$$L = L(G) := \frac{1}{N(N-1)} \sum_{i,j \in V} d_{ij} \in [0, \infty].$$

For undirected graphs we have $d_{ij} = d_{ji}$ which is finite if $i \leftrightarrow j$, and they can be decomposed into **connected components**, where we write

$$C_i = \{j \in V : j \leftrightarrow i\} \quad \text{for the component containing vertex } i.$$

4. Graphs - degrees

**Definition 4.3**

The **in-** and **out-degree** of a node $i \in V$ is defined as

$$k_i^\text{in} = \sum_{j \in V} a_{ji} \quad \text{and} \quad k_i^\text{out} = \sum_{j \in V} a_{ij}.$$

$k_1^\text{in}, \ldots, k_N^\text{in}$ is called the **in-degree sequence** and the **in-degree distribution** is

$$\{p_i^\text{in}(k) : k \in \{0, \ldots, K\}\} \quad \text{with} \quad p_i^\text{in}(k) = \frac{1}{N} \sum_{i \in V} \delta_{k, k_i^\text{in}},$$

giving the fraction of vertices with in-degree $k$. The same holds for out-degrees, and in undirected networks we simply write $k_i = k_i^\text{in} = k_i^\text{out}$ and $p(k)$.

- Note that $\sum_{i \in V} k_i = \sum_{i,j \in V} a_{ij} = |E|$ (also for directed), average and variance are

  $$\langle k \rangle = \frac{1}{N} \sum_{i \in V} k_i = |E|/N = \sum_k k p(k), \quad \sigma^2 = \langle k^2 \rangle - \langle k \rangle^2.$$

- In a **regular graph** (usually undirected) all vertices have equal degree $k_i \equiv k$.
- Graphs where the degree distribution shows a power law decay, i.e. $p(k) \propto k^{-\alpha}$ for large $k$, are often called **scale-free**.

Real-world networks are often scale-free with exponent around $\alpha \approx 3$.

4. Graphs - first examples

**Example 2 (Some graphs)**

The **complete graph** $K_N$ with $N$ vertices is an undirected graph where all $N(N-1)/2$ vertices $E = \{(i, j) : i \neq j \in V\}$ are present.

**Regular lattices** $\mathbb{Z}^d$ with edges between nearest neighbours are examples of regular graphs with degree $k = 2d$.

**Definition 4.4**

A **tree** is an undirected graph where any two vertices are connected by exactly one path. Vertices with degree 1 are called **leaves**.

In a **rooted tree** one vertex $i \in V$ is the designated **root**, and the graph can be directed, where all vertices point towards or away from the root.

A **cycle** is a closed path $\gamma_{ij}$ of length $|\gamma_{ij}| > 2$. $G$ is a tree if and only if

- it is connected and has no cycles;
- it is connected but is not connected if a single edge is removed;
- it has no cycles but a cycle is formed if any edge is added.
4. Graphs - degree correlations

**Definition 4.5**

For undirected graphs, the *joint degree distribution* of nodes linked by an edge is

\[
q(k, k') = \frac{1}{|E|} \sum_{i,j \in E} \delta_{k_i, k} \delta_{k_j, k'} = \frac{\sum_{i,j \in V} a_{ij} \delta_{k_i, k} \delta_{k_j, k'}}{\sum_{i,j \in V} a_{ij}} = q(k, k') .
\]

With the marginal \(q(k') = \sum_k q(k, k')\) we have the *conditional degree distribution*

\[
q(k|k') = q(k, k')/q(k') \quad \text{with average} \quad k_{nn}(k') := \sum_k kq(k|k') .
\]

The network is called *uncorrelated* if \(k_{nn}(k') = 0\) in \(k'\) and *assortative* if \(k_{nn}(k') > 0\) in \(k'\) and *disassortative* if \(k_{nn}(k') < 0\) in \(k'\).

- The marginal \(q(k)\) corresponds to *edge biased degree sampling*, i.e.

\[
q(k) = \sum_{k'} q(k, k') = \frac{1}{|E|} \sum_{i,j \in V} a_{ij} \delta_{k_i, k} = \frac{N}{|E|} \sum_{i,j \in V} k\delta_{k_i, k} = \frac{kp(k)}{\langle k \rangle} .
\]

For uncorrelated networks \(q(k|k') = q(k)\) and thus \(k_{nn}(k') = \langle k^2 \rangle / \langle k \rangle .

**4. Subgraphs**

- The degree of correlation can be quantified by the *correlation coefficient*

\[
\chi := \frac{(\langle kk' \rangle - \langle k \rangle^2 q)}{(\langle k^2 \rangle - \langle k \rangle^2 q)} = \frac{\sum_{k,k'} kk' (q(k, k') - q(k)q(k'))}{\sum_k k^2 q(k) - (\sum_k kq(k))^2} \in [-1, 1] .
\]

**Definition 4.6**

A subgraph \(G' = (V', E')\) of \(G = (V, E)\) is a graph such that \(V' \subseteq V\) and \(E' \subseteq E\).

- Small connected subgraphs are also called *motifs*, the simplest non-trivial examples in undirected graphs are connected triples and triangles.
- Fully connected (complete) subgraphs which are maximal with respect to connectedness are called *cliques*.
- A *spanning tree* is a tree subgraph that contains all vertices of the graph.
- A subgraph \(G'\) is called a *community*, if (for example)

\[
\sum_{i,j \in V'} a_{ij} > \sum_{i \in V', j \in V'} a_{ij} \quad \text{(there are also other definitions)} .
\]

**4. Clustering**

Clustering aims to quantify the probability that two neighbours of a given vertex are themselves neighbours. Two different definitions are used in the literature.

**Definition 4.7**

The *global clustering coefficient* for an undirected graph is defined as

\[
C = \frac{3 \times \# \text{ of (connected) triangles}}{\# \text{ of (connected) triples}} = \frac{3 \sum_{i<j<k} a_{ij}a_{jk}a_{ki}}{\sum_{i<j<k} (a_{ij}a_{jk} + a_{jk}a_{ki} + a_{ki}a_{ij})} \in [0, 1] .
\]

Alternatively, one can define a *local clustering coefficient*

\[
C_i = \frac{\# \text{ of triangles containing vertex } i}{\# \text{ of triples centered on vertex } i} = \frac{\sum_{j<k} a_{ij}a_{jk}a_{ki}}{\sum_{j<k} a_{ij}a_{jk}a_{kl}} \in [0, 1] ,
\]

and use the average \(\langle C_i \rangle = \frac{1}{k} \sum_i C_i\) to quantify clustering.

- For a tree we have \(C = \langle C_i \rangle = 0\) and for the complete graph \(C = \langle C_i \rangle = 1\).
- Higher-order clustering coefficients can be defined similarly, using different subgraphs as basis.
4. Graph spectra

**Definition 4.8**

The spectral density of a graph \( G = (V, E) \) is

\[
\rho(\lambda) := \frac{1}{N} \sum_{i \in V} \delta(\lambda - \lambda_i)
\]

where \( \lambda_1, \ldots, \lambda_N \in \mathbb{C} \) are the eigenvalues of the adjacency matrix \( A \).

- Perron-Frobenius: A has a real eigenvalue \( \lambda_1 > 0 \) with maximal modulus and real, non-negative eigenvector(s). If the graph is connected, it has multiplicity 1 and \( |\lambda_i| < \lambda_1 \) for all other eigenvalues with \( j \neq 1 \).
- For undirected graphs, \( (A^n)_{ij} \) is equal to the number of walks (paths which allow repeated vertices) from \( i \) to \( j \) of length \( n \). We also have
  \[
  \text{Tr}(A^n) = \sum_{i,j} (A^n)_{ij}
  \]
  which can be used to derive statements like:
  \[
  \sum_{i < j} \lambda_i \lambda_j = -|E|, \quad \sum_{i < j < l} \lambda_i \lambda_j \lambda_l = 2 \cdot \# \text{ of triangles in } G.
  \]

4. Graph Laplacian

**Definition 4.9**

The Graph Laplacian for a graph \( (V, E) \) with adjacency matrix \( A \) is defined as

\[
Q := A - D \quad \text{where} \quad D = \left( \delta_{ij} \sum_{l \neq i} a_{il} : i, j \in V \right).
\]

- \( Q \) has eigenvalues in \( \mathbb{C} \) with real part \( \text{Re}(\lambda) < 0 \) except for \( \lambda_1 = 0 \), which follows directly from the Gershgorin theorem and vanishing row sums.
- The multiplicity of \( \lambda_1 \) equals the number of connected components in undirected graphs. Properly chosen orthogonal eigenvectors to \( \lambda_1 \) have non-zero entries on the individual connected components.
- The smaller the second largest real part of an eigenvalue, the harder it is to cut \( G \) into separated components by removing edges.
- \( Q \) defines a generator matrix of a continuous-time random walk on \( V \) with transition rates \( a_{ij} \). Using weighted graphs, any finite state CTMC can be represented in this way.
- The Laplacian determines the first order linearized dynamics of many complex processes on graphs and is therefore of particular importance.

4. More general graphs and networks

- For multigraphs, multiple edges between nodes and loops \( (a_{ii} > 0) \) are allowed.
- Hypergraphs \( (V, E) \) are generalizations in which an edge can connect any number of vertices. Formally, the set of hyperedges \( E \subseteq P(V) \) is a set of non-empty subsets of \( V \).
- In bipartite graphs the edge set can be partitioned into two sets \( V_1, V_2 \subseteq V \) each non-empty, with no connections within themselves, i.e. \( a_{ij} = a_{ji} = 0 \) for all \( i, j \in V_1 \) and also for all \( i, j \in V_2 \).
- Simple undirected examples include regular lattices \( \mathbb{Z}^d \) for \( d \geq 1 \) which are partitioned into sites with even and odd parity. Feed-forward neural networks are examples of directed graphs with bipartite or multi-partite structure.
- Multilayer networks \( M = (G, C) \) consist of a family of \( m \) (weighted or unweighted) graphs \( G_\alpha = (V_\alpha, E_\alpha) \) (called layers of \( M \)), and the set of interconnections between nodes of different layers

\[
C = \{ c_{\alpha,\beta} \subseteq V_\alpha \times V_\beta : \alpha, \beta \in \{1, \ldots, m\}, \alpha \neq \beta \}.
\]

Real examples include transportation networks or social networks with different types of connections.
5. E-R Random graphs

Definition 5.1

An \( \text{(Erdős-Rényi) random graph} \ G \sim \mathcal{G}_{N,K} \) has uniform distribution on the set of all undirected graphs with \( N \) vertices and \( K = |E|/2 \) edges, i.e.

\[
\mathbb{P}_{N,K}[G = (V, E)] = \delta_{|V|,N} \delta_{|E|,2K} / \binom{N(N-1)/2}{K}.
\]

An \( \text{(E-R) random graph} \ G \sim \mathcal{G}_{N,p} \) has \( N \) vertices and each (undirected) edge is present independently with probability \( p \in [0,1] \), i.e.

\[
\mathbb{P}_{N,p}[G = (V, E)] = \delta_{|V|,N} p^{|E|/2}(1-p)^{N(N-1)/2-|E|/2}.
\]

- The ensemble \( \mathcal{G}_{N,p} \) is easier to work with and is mostly used in practice, and for \( N, K \) large, \( \mathcal{G}_{N,K} \) is largely equivalent to \( \mathcal{G}_{N,p} \) with \( p = 2K/(N(N-1)) \).
- Since edges are present independently, graphs \( G \in \mathcal{G}_{N,K} \) should typically be uncorrelated. Indeed, one can show that \( \chi(G) \), \( \mathbb{E}[\chi] \to 0 \) as \( N \to \infty \).

5. E-R Random graphs - properties

- The number of undirected edges for \( G \sim \mathcal{G}_{N,p} \) is \textit{random}, \( K \sim \text{Bin} \left( \frac{N(N-1)}{2}, p \right) \).
  Thus, the average degree is \textit{random}, \( \langle k \rangle = 2K/N \) with \( \mathbb{E}[\langle k \rangle] = (N-1)p \).
- The expected number of triangles in a \( \mathcal{G}_{N,p} \) graph is \( \binom{N}{3}p^3 \), and the number of triples is \( \binom{N}{3} 3p^2 \).
  Since fluctuations are of lower order, this implies for all \( G_N \sim \mathcal{G}_{N,p} \)

\[
C(G_N) = \frac{3\binom{N}{3}p^3(1+o(1))}{\binom{N}{3} 3p^2(1+o(1))} \to p \quad \text{as } N \to \infty.
\]

- The \textit{expected} degree distribution for \( G_N \sim \mathcal{G}_{N,p} \) is binomial. In the limit \( N \to \infty \) with \( p = p_N = z/(N-1) \) keeping \( z = \mathbb{E}[\langle k \rangle] \) fixed we have

\[
\mathbb{E}[p_k] = \mathbb{P}[k = k] = \binom{N-1}{k} p_N^k (1-p_N)^{N-1-k} \to \frac{k^z}{k!} e^{-z}.
\]

Therefore, E-R \( \mathcal{G}_{N,p} \) graphs are sometimes called \textbf{Poisson random graphs}.
- In this scaling limit E-R graphs are \textit{locally tree-like}, i.e. finite connected components

\[
C_n := \{ j \in V : j \leftrightarrow i, d_j \leq n \}, \quad n \text{ fixed}
\]

are tree subgraphs as \( N \to \infty \) with probability 1.

Vertex degrees are \( k_j \sim \text{Poi}(z) \) and iid \( k_j \sim \text{Poi}(z) + 1 \).

5. Percolation

Percolation studies robustness of connectivity properties of graphs under deletion of edges or vertices (e.g. random attacks or immunization).

Definition 5.2

Consider a connected, undirected graph \( G = (V, E) \). \textbf{Bond percolation} is a static probabilistic model with state space

\[
S = \Omega = \{0,1\}^E = \{ e_{ij} \in \{0,1\} : (i,j) \in E \},
\]

and distribution \( p = P[e_{ij} = 1] = 1 - P[e_{ij} = 0] \), i.e. \( e_{ij} \sim \text{Be}(p) \) iid with \( p \in [0,1] \).

Edges \( (i,j) \in E \) are called \textit{open} if \( e_{ij} = 1 \) and \textit{closed} if \( e_{ij} = 0 \), and we denote by

\[
G^o = (V, E^o) \quad \text{with} \quad E^o = \{ (i,j) \in E : e_{ij} = 1 \} \subseteq E
\]

the (random) subgraph containing only open edges. A sequence of connected graphs \( G_N \) of increasing size \( |V_N| = N \) exhibits \textbf{percolation with parameter} \( p \) if

\[
|C_N|/N \geq c > 0 \quad \text{as } N \to \infty \quad \text{with probability } 1,
\]

where \( |C_N| = \max_{v \in V_N} |C_v| \) is the size of the largest connected component \( C_N \) of \( G_N \).
5. Percolation and E-R graphs

- Alternatively, percolation can be defined on an infinite graph \( G (\text{e.g. } \mathbb{Z}^d) \) with

  \[
  \text{percolation probability } \theta(p) = \mathbb{P}[|C_0| = \infty] = \left\{ \begin{array}{ll}
  0 & , \text{ for } p < p_c \\
  > 0 & , \text{ for } p > p_c
  \end{array} \right. ,
  \]

  changing behaviour at a critical value \( p_c \in [0, 1] \).

- In site percolation vertices and their adjacent edges are deleted.

- E-R random graphs \( G_{N,p} \) have the same distribution as open subgraphs \( G_{N}^c \subseteq K_N \) under bond percolation with parameter \( p \) on the complete graph \( K_N \).

**Theorem 5.1 (Giant component for E-R graphs)**

Consider \( G_{N,p} \sim G_{N,p} \) with \( p = z/N \) and maximal connected component \( \tilde{C}_{N,p} \). Then

\[
|\tilde{C}_{N,p}| = \left\{ \begin{array}{ll}
\Theta(\log N) , & \text{for } z < 1 \\
\Theta(N^{2/3}) , & \text{for } z = 1 \\
\Theta(N) , & \text{for } z > 1
\end{array} \right.
\]

\( c(z) := \lim_{N \to \infty} |\tilde{C}_{N,p}|/N \) is a continuous function of \( z \). For \( z > 1 \), \( \tilde{C}_{N,p} \) is the only giant component of size \( O(N) \), and the second largest is of order \( O(\log N) \).

Local trees with 1 + Poisson degrees die out with probability 1 if and only if \( z \leq 1 \).

5. The Wigner semi-circle law

**Theorem 5.2 (Wigner semi-circle law)**

Let \( A = (a_{ij} : i,j = 1, \ldots, N) \) be a real, symmetric matrix with iid entries \( a_{ij} \) for \( i \leq j \) with finite moments, and \( \mathbb{E}[a_{ij}] = 0 \), \( \text{var}[a_{ij}] = \sigma^2 \) (called a Wigner matrix). Then the spectral density \( \rho_N \) of the matrix \( A/N \) converges with probability one to

\[
\rho_N(\lambda) \to \rho_{sc}(\lambda) \equiv \left\{ \begin{array}{ll}
(2\pi \sigma^2)^{-1/2} \sqrt{4\sigma^2 - \lambda^2} & , \text{if } |\lambda| < 2\sigma \\
0 & , \text{otherwise}
\end{array} \right.
\]

The bulk of eigenvalues of unscaled Wigner matrices typically lies in the interval \([-2\sqrt{N}\sigma, 2\sqrt{N}\sigma] \).

Adjacency matrices \( A \) of \( G_{N,p} \) random graphs are symmetric with iid \( \text{Be}(p) \) entries with \( \mathbb{E}[a_{ij}] = p \) and \( \text{var}[a_{ij}] = p(1-p) \), so are not Wigner matrices.

\( A \) has a maximal Perron-Frobenius eigenvalue of order \( pN \), but all other eigenvalues have modulus of order \( \sqrt{N} \).

For fixed \( p > 0 \) the Wigner semi-circle law holds for \( N \to \infty \) as stated above.

For scaled \( p = p_N \gg p_c = 1/N \) the width of the support reduces to \( 4\sqrt{N}\sigma_{p_N} \) with \( \sigma_{p_N} = \sqrt{p_N} \) and a modified version holds.

For \( p = p_N \ll p_c = 1/N \) the asymptotic spectral density deviates from \( \rho_{sc} \).

5. Preferential attachment

The prevalence of power-law degree distributions in real complex networks can be attributed to growth mechanisms subject to preferential attachment.

**Definition 5.3**

Starting with a complete graph \( (V_0, E_0) \) of \( |V_0| = m_0 \) nodes, at each time step \( t = 1, \ldots, N - m_0 \) a new node \( j = t + m_0 \) is added. It forms \( m \leq m_0 \) undirected edges with existing nodes \( i \in V_{t-1} \) with a probability proportional to their degree \( \pi_{r+1} = k_i/\sum_{i \in V_{t-1}} k_i \) (preferential attachment).

The resulting, undirected graph with \( N \) nodes and \( K = m_0(m_0 - 1)/2 + m(N - m_0) \) is called a Barabási-Albert random graph, denoted by \( G_{N,K}^\text{BA} \).

- As \( N \to \infty \), the average degree is \( \langle k \rangle = 2m \) and the degree distribution \( p_N(k) \) converges to a distribution \( p(k) \) with power law tail, i.e. \( p(k) = Ck^{-\alpha} \) as \( k \to \infty \) where \( \alpha = 3 \), which is close to exponents observed for real-world networks.

  This is independent of the parameters \( m_0 \) and \( m \).

- Characteristic path length and clustering coefficient typically behave like \( L = O(\log N) \) and \( C = O(N^{-0.75}) \) for \( G_{N,K}^\text{BA} \) graphs.

- The expected degree of nodes increases with their age.
5. Preferential attachment

Variations of the model connecting to vertices $i$ with probability proportional to $k_i + k_0$ lead to power law degree distributions with $\alpha = 3 + k_0/m$.

For non-linear preferential attachment proportional to $k_i^{\gamma}$ the degree distribution converges to a stretched exponential $\exp(-Ck^{1-\gamma})$ for $\gamma \in [0, 1)$, and for $\gamma > 1$ a single vertex connects to almost all other vertices.

There are many other and also older models that develop the idea of preferential attachment, for some of which one can do exact computations.

(see e.g. [M.E.J. Newman, Siam Review 45(2), 167256 (2003)]).

5. Small-world networks

**Definition 5.4**

A sequence of connected graphs $G_N$ with increasing size $|V_N| = N$ exhibits the small-world property, if the characteristic path length $L(G_N) = O(\log N)$.

Examples include trees with degrees $k_i \geq 3$ and also the giant or largest component in E-R random graphs. In most graph models small-worldness is paired with low clustering coefficients, e.g. 0 for trees and $p$ for $G_{N,p}$ graphs. However, many real examples of small world networks exhibit also large clustering coefficients, such as networks of social contacts.

**Definition 5.5**

Consider a $2m$-regular ring graph with adjacency matrix $a_{ij} = \begin{cases} 1; & |i-j| \leq m \\ 0; & \text{otherwise} \end{cases}$ of size $N$ with a total number of $K = mN$ undirected edges.

For all $i$, each edge $(i,j)$ with a clockwise neighbour with $j > i$ is rewired with probability $p \in [0, 1]$, i.e. replaced by an edge $(i,l)$ where $l$ is chosen uniformly among vertices not adjacent to $i$. The resulting graph is a Watts-Strogatz random graph, denoted by $G_{N,K}^{WS}$.

5. Watts-Strogatz model

- W-S random graphs interpolate between a regular lattice for $p = 0$ and a $G_{N,K}$ E-R random graph conditioned on the event that all vertices have degree $k_i \geq m$.

- Expected clustering coefficient $E[C(p)]$ and characteristic path length $E[L(p)]$ are monotone decreasing functions of $p$ and show the following behaviour.

$N = 1000$ and $m = 5$, taken from [D.J. Watts, S.H. Strogatz, Nature 393, 440-442 (1998)].
5. Configuration model

**Definition 5.6**

The configuration model $G_{N,D}^{\text{conf}}$ is defined as the uniform distribution among all undirected graphs with $N$ vertices with a given degree sequence $D = \{k_1, \ldots, k_N\}$, such that $\sum_{i \in V} k_i = 2K$.

- Not all sequences $D$ that sum to an even number are graphical.
- Sampling is usually done by attaching $k_i$ half edges to each vertex $i$ and matching them randomly. This can lead to self loops and rejections.
- General randomized graphs with given degree distribution $p(k)$ can be sampled in the same way. If $k_{\max} = \max_i k_i$ is bounded, one can show that these graphs exhibit a giant (connected) component of size $O(N)$ if
  \[ Q := \sum_{k \geq 0} k(k-2)p(k) > 0 , \]
  and if $Q < 0$ the largest component is of size $O(k_{\max}^2 \log N)$.
- For directed versions with $D^{\text{in}}$ and $D^{\text{out}}$ we need $\sum_{i \in V} k_i^{\text{in}} = \sum_{i \in V} k_i^{\text{out}}$.

6. Planar graphs and spatial point processes

**Definition 6.1**

A planar graph is an undirected graph that can be embedded in the plane, i.e. it can be drawn in such a way that no edges cross each other. The edges of a particular embedding partition the plane into faces. A connected planar graph $G$ has a dual graph $G^*$, which has one vertex in each face of $G$, and a unique edge crossing each edge of $G$. $G^*$ may be a multigraph with self-loops.

A maximal planar graph is called a triangulation.

- Every planar graph is 4-partite or 4-colourable.
- In a triangulation each face is bounded by three edges. By induction, every triangulation with $N>2$ nodes has $K=3N-6$ undirected edges and $2N-4$ faces.

**Definition 6.2**

A random countable set $\Pi \subset \mathbb{R}^d$ is called a spatial point process. $\Pi \subset \mathbb{R}^d$ is called a homogeneous Poisson point process $\text{PPP}(\lambda)$ with rate $\lambda > 0$ if

- for all $A \subset \mathbb{R}^d$ we have $N(A) := |\Pi \cap A| \sim \text{Poi}(\lambda |A|)$.
- for all disjoint $A_1, \ldots, A_n \subset \mathbb{R}^d$, $N(A_1), \ldots, N(A_n)$ are independent.

6. Planar graphs and spatial point processes

- To sample from a $\text{PPP}(\lambda)$ e.g. in a box $A = [0,L]^d$, pick $N(A) \sim \text{Poi}(\lambda L^d)$, then place $N(A)$ particles independently in $A$ each with uniform distribution, i.e. pick the $d$ coordinates uniformly in $[0,L]$.
- A Poisson process $\text{PP}(\lambda)$ is equivalent to a $\text{PPP}(\lambda)$ on $[0, \infty)$.

**Definition 6.3**

Let $\Pi = \{x_1, x_2, \ldots\}$ be a countable subset of $\mathbb{R}^d$, endowed with a distance function $d(x,y)$. A Voronoi tessellation (or diagram) is given by the family of Voronoi cells $\{A_1, A_2, \ldots\} \subset \mathbb{R}^d$ where

\[ A_i = \{ x \in \mathbb{R}^d : d(x,x_i) \leq d(x,x_j) \text{ for all } j \neq i \} \]

is the set of points closest to $x_i$.

**Properties in 2 dimensions.**

- The shape of Voronoi cells depends on the distance function, for Euclidean distance $d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ they are convex polygons, and boundaries between adjacent cells are straight lines.
- The dual graph of a Voronoi diagram of a set $\Pi$ is called Delaunay triangulation, which is not unique if 4 or more cells intersect in a point.