Stochastic Processes Lecture Notes (2025/2026)

Griffin Reimerink

Contents

1 Simple random walks

1.1 Random walks on graphs

Definition *Simple random walk*

Let $G = (V, E)$ be a graph where each vertex has finite degree.
A **simple random walk** (**SRW**) is a sequence of vertices $(W_t)_{t\geq 0}$ such that for all $x \in V$ and integers $t \geq 0$, if $W_t = x$ then for all $y \in V$ satisfying $xy \in E$,

$$\mathbb{P}(W_{t+1} = y) = \frac{1}{\deg(x)}$$

independently of the previous steps of the walk.

Definition *Hitting time and first visit time*

$$\textbf{Hitting time}: \tau_x := \inf\{t \geq 0 : W_t = x\}$$
$$\textbf{First visit time}: \tau_x^+ := \inf\{t \geq 1 : W_t = x\}$$

Conditioning on starting point

$$\mathbb{P}_x(\,\cdot\,) = \mathbb{P}(\,\cdot \mid W_0 = x)$$

Definition *Escape probability*

For a graph with starting point x and endpoint y, the **escape probability** is $p_{\text{esc}} := \mathbb{P}_x(\tau_y < \tau_x)$

1.2 Random walks on $\mathbb{Z}$

Definition *Random walk on $\mathbb{Z}$*

On $\mathbb{Z}$, $(W_t)_{t\geq 0}$ can be described as a sum of Bernoulli distributed random variables.

$$W_t = \sum_{i=1}^{t} X_i \qquad X_i = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$$

Proposition

Let $a < 0 < b$ and $x \in [a, b]$ be integers and consider the simple random walk on $\mathbb{Z}$.

$$\mathbb{P}_x(\tau_b < \tau_a) = \frac{x-a}{b-a} \qquad \mathbb{P}_0(\tau_b < \tau_a) = -\frac{a}{b-a}$$

Theorem *Weak law of large numbers*

If $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables with finite variance, then

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{|\sum_{i=1}^{n} X_i - \mu n|}{n} > \varepsilon\right) = 0$$

Lemma

W_t has expectation 0 and variance t.

Lemma

$$\lim_{t\to\infty} \mathbb{P}(-\varepsilon t < W_t < \varepsilon t) = 1$$

Theorem *Central limit theorem*

If If $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables with finite expectation and variance, then

$$\frac{\sum_{i=1}^n X_i - \mu n}{\sqrt{n}\sigma} \xrightarrow{d} \mathcal{N}(0,1)$$

Proposition

$$W_t \xrightarrow{d} \mathcal{N}(0,t)$$

Proposition *Reflection principle*

For any random walk on $\mathbb{Z}$, denote

$$N(a,n) := \text{number of paths of length } n \text{ starting at } 0 \text{ and ending at } a$$
$$M(a,b,n) := \text{number of paths of length } n \text{ starting at } 0 \text{ and ending at a that visit } b$$

Then for any $0 < a < b$ we have

$$M(a,b,n) = N(2b-a,n)$$

1.3 Random walks on $\mathbb{Z}^d$

Definition *Random walk on $\mathbb{Z}^d$*

For the simple random walk $(W_t)_{t\geq 0}$ on $\mathbb{Z}^d$, we start at the origin, and

$$\mathbb{P}(W_{t+1} = W_t \pm e_i) = \frac{1}{2d} \qquad i = i, \ldots, d$$

where e_i are the basis vectors of $\mathbb{Z}^d$.

Definition *Recurrent random walk*

$(W_t)_{t\geq 0}$ is **recurrent** if $\mathbb{P}(\text{return to } \underline{0}) = 1$ and **transient** otherwise.

Lemma

A simple random walk on $\mathbb{Z}^d$ is recurrent if and only if $\mathbb{E}(\text{number of returns to } \underline{0}) = \infty$

Definition *Little-o notation*

$f_n = o_n(g_n)$ if $\dfrac{f_n}{g_n} \to 0$ as $n \to \infty$.

Theorem *Stirling's formula*

$$n! = (1 + o_n(1))\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Lemma

$$\lim_{k\to\infty} \mathbb{P}\left(X \sim \text{Bin}\left(2k, \frac{1}{2}\right) = k\right) = \frac{1}{\pi}$$

Theorem *Pòlya's theorem*

$(W_t)_{t\geq 0}$ on $\mathbb{Z}^d$ is recurrent if and only if $d \in \{1,2\}$.

2 Networks

2.1 Electrical networks

Definition *Network*

A **network** is a graph $G = (V, E)$ with a **conductance** $c(e) > 0$ assigned to each edge $e \in E$. The **resistance** of an edge is $r(e) = \frac{1}{c(e)}$. The conductance of a vertex is $c(v) = \sum_{u:uv \in E} c(uv)$.

Definition *Harmonic function*

For a network $G = (V, E)$, a function $V \to \mathbb{R}$ is **harmonic** at x if

$$f(x) = \sum_{y:xy \in E} f(y) \cdot \frac{c(xy)}{c(x)} \quad \text{or equivalently} \quad \sum_{y:xy \in E} c(xy)(f(x) - f(y)) = 0$$

Definition *Voltage*

A **voltage** on a network G with $a \neq z \in V(G)$ is a function $V(G) \to \mathbb{R}$ which is harmonic at every $x \notin \{a, z\}$.

Lemma *Uniqueness principle*

If G is a connected finite network and f, g are voltages on G such that $f(a) = g(a)$ and $f(z) = g(z)$, then $f = g$.

Orientation of edges

We denote the orientation $\vec{E}$ of edges E by $\overrightarrow{xy}$ or $\overleftarrow{xy}$, where $\overrightarrow{xy} = \overleftarrow{yx}$.

Definition *Flow*

A **flow** $J : \overrightarrow{E} \to \mathbb{R}$ is an assignment of values to oriented edges such that

1. $J(\overrightarrow{xy}) = -J(\overrightarrow{yx})$
2. $J(x) := \sum_y J(\overrightarrow{xy}) = 0$ for all vertices $x \notin \{a, z\}$

We call $J(x)$ the **divergence** of x. We call $J(a)$ the **source** and $J(z)$ the **sink** of a network.

Lemma

$$\sum_{x \in V(G)} J(x) = J(a) + J(z) = 0$$

Definition *Flow strength*

The **strength** of a flow J is $\|J\| := J(a)$. A **unit flow** is a flow of strength 1.

Definition *Current flow*

Given a voltage W on a network G the **current flow** I is defined using Ohm's law:

$$I(\overrightarrow{xy}) := \frac{W(x) - W(y)}{r(xy)}$$

Note: I satisfies the definition of a flow.

Definition *Effective resistance*

The **effective resistance** for a voltage W and the corresponding current flow I is

$$R_{\text{eff}} := \frac{W(a) - W(z)}{\|I\|}$$

Lemma

R_{eff} does not depend on the choice of voltage.

2.2 Random walks on networks

Definition *Random walk on a network*

For a random walk $(W_t)_{t\geq 0}$ on a network, we have

$$\mathbb{P}(W_{t+1} = u \mid W_t = v) = \begin{cases} \frac{c(uv)}{c(v)} & \text{if } uv \in E \\ 0 & \text{otherwise} \end{cases}$$

Lemma

If G is a connected finite network, $a, z \in V(G)$ and $a \neq z$, then

$$f(x) = \mathbb{P}_x(\tau_z < \tau_a) \text{ is harmonic for all } x \notin \{a, z\} \qquad f(a) = 0 \qquad f(z) = 1$$

Theorem

On a network, the escape probability p_{esc} satisfies

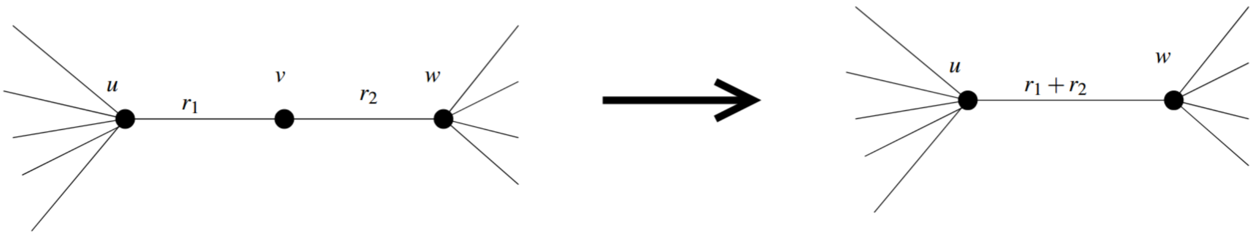
$$p_{\text{esc}} = \frac{1}{c(a) \cdot R_{\text{eff}}}$$

2.3 Simplifying the network

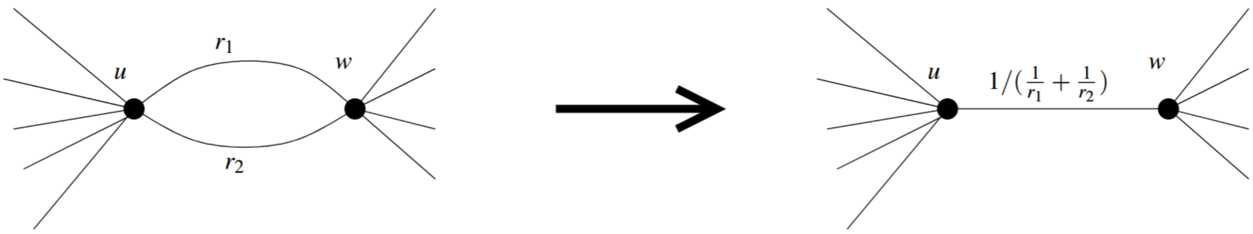
Simplification laws

The following three operations do not change the effective resistance R_{eff}:

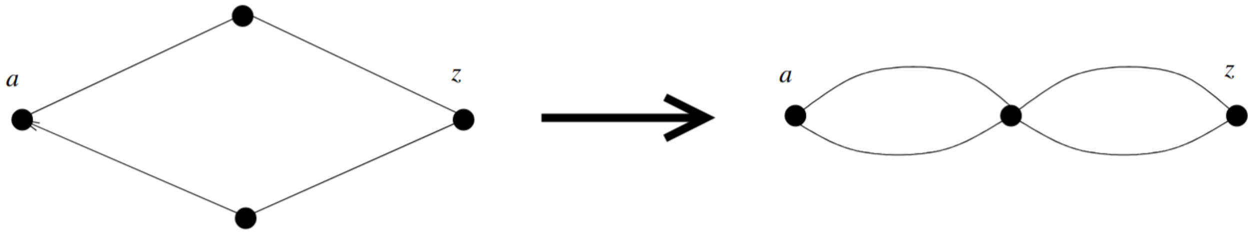
Series law



Parallel law



Gluing vertices of equal voltage



Note: Loops (edges from u to u) can be discarded without affecting R_{eff}.

Definition *Network automorphism*

An **automorphism** of a network G is a bijection $\varphi : V(G) \to V(G)$ such that uv is an edge if and only if $\varphi(u)\varphi(v)$ is an edge, and moreover $c(uv) = c(\varphi(u)\varphi(v))$ for all edges uv.

Lemma *Equal voltage criterion*

Let G be a network with distinguished nodes a, z and let φ be an automorphism such that $\varphi(a) = a$ and $\varphi(z) = z$. Then $W(u) = W(v)$ for all pairs of vertices u, v such that $\varphi(u) = v$.

2.4 Rayleigh's monotonicity law

Lemma

Let $f : V(G) \to \mathbb{R}$ be an arbitrary function and let J be a flow from a to z. Then

$$\sum_{x,y \in V(G)} (f(x) - f(y)) \cdot J(\overrightarrow{xy}) = 2(f(a) - f(z))\|J\|.$$

Definition *Energy of a flow*

The **energy** of a flow J is

$$\mathcal{E}(J) := \frac{1}{2} \sum_{x,y} \left(J(\overrightarrow{xy})\right)^2 \cdot r(xy)$$

Theorem *Thompson's principle*

$$R_{\text{eff}} = \inf\{\mathcal{E}(J) : J \text{ a flow from } a \text{ to } z \text{ with } \|J\| = 1\}$$

and the unique minimizer is the current flow of strength one.

Theorem *Rayleigh's monotonicity law*

Let $(r(e))_{e \in E}$ and $(r'(e))_{e \in E}$ be assignments of resistances such that $r(e) \leq r'(e)$, and fix $a, z \in V$. Then the corresponding effective resistance satisfies

$$R_{\text{eff}} \leq R'_{\text{eff}}$$

Corollary

- **Cutting law**: Removing an edge from G will not decrease R_{eff}.
- **Shorting law**: Gluing two vertices in G (regardless of voltage) will not increase R_{eff}.

Definition *Graph distance*

The **graph distance** of two vertices is the number of edges in the shortest path between them. We define

$$S_n = \{\text{vertices with graph distance } n \text{ from the origin}\}$$

Note

For any graph where S_n grows at most linearly, the simple random walk starting at the origin is recurrent. This can be proven analogously to the proof of Pòlya's theorem with $d = 2$ at the end of Lecture 4.

Transience of a simple random walk can be proven by embedding a k-regular tree into the graph.

3 Markov chains

3.1 Markov chains

Definition *Directed graph*

A **directed graph** $D = (V, A)$ consists of a set of vertices V and a set of **arrows** (or **arcs**) $A \subset V \times V$.

Definition *Markov chain*

A **Markov chain** M is a sequence of random variables $X_0, X_1, X_2, \dots$ on a (finite) **state space** $S = \{1, \dots, n\}$. We define **transition probabilities** for all $i, j \in S$:

$$\mathbb{P}(X_{t+1} = j \mid X_t = i) = P_{ij} \in [0, 1]$$

The transition probability is independent of t and for all $i \in S$ we have $\sum_{j \in S} P_{ij} = 1$
Markov chains have the following properties:

1. **Markov property**: the state at time t + 1 depends only on the state at time t.

   $$\text{for all } t \geq 0,\, i_0, \dots, i_{t+1} \qquad \mathbb{P}(x_{t+1} = i_{t+1} \mid X_0 = i_0, \dots, X_t = i_t) = \mathbb{P}(X_{t+1} = i_{t+1} \mid X_t = i_t)$$

2. **Time homogeneity**:

   $$\mathbb{P}(x_{t+1} = j \mid X_t = i) = \mathbb{P}(x_1 = j \mid X_0 = i)$$

Any Markov chain is equivalent to a random walk on a weighted directed graph D, where $(i, j) \in A \iff P_{ij} > 0$.

Notation

Let $i \in S$, A an event and v a probability distribution. Then we denote

$$\mathbb{P}_i(A) = \mathbb{P}(A \mid X_0 = i) \qquad \mathbb{P}_v(A) = \mathbb{P}(A \mid X_0 \overset{d}{=} v)$$

We also define the hitting time and first visit time in the same way as before.

Definition *Stochastic matrix*

A **stochastic matrix** is a square matrix with nonnegative elements such that its rows sum up to 1.
We can collect the values of P_{ij} of a Markov chain in a stochastic matrix P called a **transition matrix**.

Lemma

For every $t \geq 0$ and $i, j \in S$, we have $\mathbb{P}_i(X_t = j) = P^t_{ij}$

3.2 Stationary distributions

Definition *Stationary distribution*

A distribution π on S is **stationary** if $\pi = \pi P$.

Lemma

If there exists a state i such that $\mathbb{E}_i(\tau_i^+) < \infty$, then $\pi = (\pi_1, \dots, \pi_n)$ with

$$\pi_j = \frac{\mathbb{E}_i|\{\text{visits to } j \text{ before } \tau_i^+\}|}{\mathbb{E}_i(\tau_i^+)}$$

Definition *Irreducible Markov chain*

A Markov chain is **irreducible** if from every $i \in S$ we can reach every $j \neq i \in S$ in one or more steps.

Lemma

If a Markov chain is irreducible, then $\mathbb{E}_i(\tau_i^+) < \infty$ for all $i \in S$.

Lemma

If a Markov chain is irreducible, then there is precisely one stationary distribution.

Corollary

If a Markov chain is irreducible, then its unique stationary distribution satisfies

$$\pi_i = \frac{1}{\mathbb{E}_i(\tau_i^+)}$$

3.3 Periodicity

Theorem

For all $a_1, \dots, a_n \in \mathbb{N}$, there exist $x_1, \dots, x_n \in \mathbb{Z}$ such that $\gcd(a_1, \dots, a_n) = a_1 x_1 + \dots + a_n x_n$.

Definition

$$A + A := \{a + b : a, b \in A\}$$

Theorem

If $A \subset \mathbb{N}$ is nonempty, $A + A \subseteq A$, and $\gcd(A) = 1$, then there exists $N \in \mathbb{N}$ such that

$$\{N, N+1, N+2, \dots\} \subset A$$

Lemma

Consider a Markov chain with state space S and transition matrix P, and define

$$A_n := \{t \geq 1 : P_{ii}^t > 0\}$$

If the Markov chain is irreducible, then for all $i, j \in S$

$$\gcd(A_i) = \gcd(A_j)$$

Definition *Period*

The **period** of an irreducible Markov chain is $\gcd(A_1)$ where $A_1 = \{t \geq 1 : P_{11}^t > 0\}$
An irreducible Markov chain is **aperiodic** if its period is 1.

Theorem

If a Markov chain is irreducible and aperiodic, then there exists t_0 such that

$$P_{ij}^t > 0 \quad \text{for all } t \geq t_0,\ i, j \in S$$

3.4 Convergence

Definition *Total variational distance*

Let X, Y be random variables on a finite state space S. The **total variational distance** of X and Y is

$$d_{TV}(X, Y) = \max_{A \subseteq S} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$$

Theorem *Markov chain convergence theorem*

Consider an irreducible and aperiodic Markov chain on a state space S with stationary distribution π.
Then there exists $0 \leq \alpha < 1$ such that for all initial distributions μ on S,

$$d_{TV}(\pi, \mu P^t) \leq \alpha^t$$

Lemma

$$d_{TV}(X,Y) = \frac{1}{2}\sum_{x\in S}|\mathbb{P}(X=x)-\mathbb{P}(Y=x)|$$

Definition *Coupling of random variables*

Let μ,ν be probability distributions on S.
A **coupling** of μ,ν is a random vector $(X,Y)\in S\times S$ such that X has distribution μ and Y has distribution ν.

Lemma

If μ,ν are distributions on S, then

$$d_{TV}(\mu,\nu) = \min\{\mathbb{P}(X\neq Y) : (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\}$$

3.5 Some additional tricks

Definition *Lazy Markov chain*

For a periodic Markov chain with transition matrix P, we define the **lazy Markov chain** with transition matrix

$$Q = \frac{1}{2}(I+P)$$

Definition *Essential class*

We write $i\to j$ if it is possible to move from i to j in zero or more steps.
States i and j **communicate**, denoted $i\leftrightarrow j$, if $i\to j$ and $j\to i$.
A state i is an **essential state** if $i\leftrightarrow j$ for every j such that $i\to j$. A state that is not essential is **inessential**.
An **essential class** is an equivalence class under $\leftrightarrow$ of which every state is essential.

Definition *Detailed balance equations*

A distribution μ on S satisfies the **detailed balance equations** if for all $i,j\in S$,

$$\mu_i P_{ij} = \mu_j P_{ji}$$

Lemma

If μ satisfies the detailed balance equations then μ is stationary.

4 Branching processes

4.1 Branching processes

Definition *Branching process*

A **branching process** or **Galton-Watson process** is defined as follows:

- In generation $t = 0$, there is a single individual.
- This individual has a random number X of children, where X is a random variable taking values in $\mathbb{N}_0$
- This process repeats: each individual k in generation t has $X_{t,k}$ children.
- The process ends (dies out) if all individuals in a certain generation t have zero children.
- We say the process **survives** if it goes on indefinitely and it goes **extinct** if the process dies out at any point.

Definition *Number of individuals*

$$Z_0 = 1 \qquad Z_t = \sum_{k=1}^{Z_{t-1}} X_{t-1,k}$$

Regimes

Let $\mu = \mathbb{E}[X]$.

- If $\mu > 1$ (**supercritical regime**) then the expected number of offspring in generation t grows arbitrarily large, exponentially fast with t.
- If $\mu < 1$ (**subcritical regime**) then it decreases exponentially fast to zero.
- We call the case $\mu = 1$ the **critical regime**.

4.2 Probability generating functions

Definition *Probability generating function*

The **probability generating function** (**PGF**) of a random variable X in $\mathbb{N}_0$ is

$$G_x(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k)$$

Lemma *Properties of the PGF*

1. $G(0) = \mathbb{P}(X = 0)$ and $G(1) = 1$
2. $\mathbb{P}(X = k) = \dfrac{G^{(k)}(0)}{k!}$
3. $G'(s) = \sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}(X = k) \geq 0$ and $G'(1) = \mathbb{E}[X]$
4. $G''(s) \geq 0$ and $G''(1) = \mathbb{E}[X(X-1)] = \mathrm{Var}(X) - \mathbb{E}[X] + (\mathbb{E}[X])^2$

Theorem

Provided that $0 < \mathbb{P}(X = 0) < 1$, we have that $\mathbb{P}(\text{extinction})$ is the least nonnegative root of $q = G(q)$.
Moreover, we have that $\mathbb{P}(\text{extinction}) = 1$ if and only if $\mu = \mathbb{E}[X] \leq 1$

Proposition *Properties of Z_t*

Given $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$, we have

- $G_{Z_t}(s) = G_X(G_{Z_{t-1}}(s)) = (G_X \circ \cdots \circ G_X)(s)$ (where G_X is composed t times)
- $\mathbb{E}[Z_t] = \mu^t$
- $\mathrm{Var}[Z_t] = \begin{cases} \sigma^2 \cdot t & \text{if } \mu = 1 \\ \sigma^2 \cdot \mu^{t-1} \cdot \left(\frac{\mu^t - 1}{\mu - 1}\right) & \text{if } \mu \neq 1 \end{cases}$

4.3 Duality

Definition *Dual branching process*

Consider a branching process with distribution X such that $\mu = \mathbb{E}[X] > 1$.
For the **dual branching process**, we condition on extinction, and we have random variables $\widetilde{X}$ and $\widetilde{Z}_t$.

Theorem *Duality principle*

$$\mathbb{P}\big(Z_1 = z_1, Z_2 = z_2, \ldots, Z_t = z_t \mid \text{extinction}\big) = \mathbb{P}\big(\tilde{Z}_1 = z_1, \tilde{Z}_2 = z_2, \ldots, \tilde{Z}_t = z_t\big)$$

for all $t \in \mathbb{N}$ and nonnegative integers $z_1, z_2, \ldots, z_t$.

Theorem *Duality principle, version 2*

For all rooted ordered trees T,

$$\mathbb{P}(T = \tau \mid \text{extinction}) = \mathbb{P}(\widetilde{T} = \tau)$$

where T is the tree of $(Z_t)_{t\geq 0}$ and $\widetilde{T}$ is the tree of $(\widetilde{Z}_t)_{t\geq 0}$

4.4 Relation to random walks

Notation

$$Z_{tot} = Z_0 + Z_1 + \cdots \qquad S_n = \sum_{i=1}^{n}(X_i - 1)$$

where X_i are i.i.d. distributed like X.

Lemma

$$\mathbb{P}(Z_{tot} = n) = \mathbb{P}(S_n = -1,\, S_1, \ldots, S_{n-1} > -1) \qquad \text{for all } n \in \mathbb{N}$$

Theorem *Otter-Dwass formula*

$$\mathbb{P}(Z_{tot} = n) = \frac{1}{n} \cdot \mathbb{P}(S_n = 1)$$

5 Martingales

5.1 Conditional expectation

Definition *Conditional expectation*

$$\mathbb{E}[Y \mid X = x] = \sum_y y \cdot \mathbb{P}(Y = y | X = x)$$

$\mathbb{E}[Y \mid X_1, \dots, X_n] = \varphi(X_1, \dots, X_n)$ is a random variable, satisfying

$$\varphi(x_1, \dots, x_n) = \mathbb{E}[Y \mid X_1 = x_1, \dots, X_n = x_n] \qquad \text{for all } x_1, \dots, x_n$$

Lemma *Tower rule*

$$\mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(Y)$$

Proposition *Properties of conditional expectation*

1. If λ, μ are constants, then $\mathbb{E}(\lambda Y + \mu Z \mid X_1, \dots, X_n) = \lambda \mathbb{E}(Y \mid X_1, \dots, X_n) + \mu \mathbb{E}(Z \mid X_1, \dots, X_n)$
2. $\mathbb{E}[g(X_1, \dots, X_n) \cdot Y \mid X_1, \dots, X_n = g(X_1, \dots, X_n)\mathbb{E}(Y \mid X_1, \dots, X_n)$
3. If $h : \mathbb{R}^n \to \mathbb{R}^n$ is a bijection, then $\mathbb{E}[Y \mid X_1, \dots, X_n] = \mathbb{E}[Y \mid h(X_1, \dots, X_n)]$
4. $\mathbb{E}(\mathbb{E}(Y \mid X_1, \dots, X_{n+m}) \mid X_1, \dots, X_n) = \mathbb{E}(Y \mid X_1, \dots, X_n)$

5.2 Martingales

Definition *Martingale*

A sequence of random variables (Y_n) is a **martingale** with respect to the sequence of random variables (X_n) if

$$\mathbb{E}[Y_{n+1} \mid X_1, \dots, X_n] = Y_n$$

Lemma

If (Y_n) is a martingale w.r.t. (X_n), then

$$\mathbb{E}Y_1 = \mathbb{E}Y_2 = \dots$$

and for all $n, m \in \mathbb{N}_0$,

$$\mathbb{E}(Y_{n+m} \mid X_1, \dots, X_n) = \mathbb{E}(Y_n \mid X_1, \dots, X_{n+m}) = Y_n$$

Theorem *Chebyshev inequality*

$$\mathbb{P}\big(|Y - \mathbb{E}Y| \geq \lambda\big) \leq \frac{\mathrm{Var}(Y)}{\lambda^2}$$

Theorem *Chernoff bound*

If $Y \sim \mathrm{Bin}(n, p)$, then

$$\mathbb{P}(|Y - \mathbb{E}Y| > \lambda) \leq 2e^{-\frac{\lambda^2}{2n}}$$

Lemma

Suppose X is a random variable with $\mathbb{E}X = 0$ and $|X| \leq 1$ almost surely. Then

$$\mathbb{E}[e^{tX}] \leq e^{t^2/2}$$

Theorem *Azuma-Hoeffding inequality*

Suppose (Y_n) is a martingale w.r.t. (X_n), and (c_n) is a sequence of constants satisfying

$$|Y_n - Y_{n-1}| \leq c_n \qquad \text{almost surely for all } n$$

Then

$$\mathbb{P}(|Y_n - Y_0| \geq \lambda) \leq 2\exp\left(-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}\right)$$

5.3 Convergence

Theorem *Doob-Kolmogorov inequality*

If (Y_n) is a martingale with respect to (X_n), then

$$\mathbb{P}\left(\max_{1\leq i\leq n} |Y_i| \geq \varepsilon\right) \leq \frac{\mathbb{E}Y_n^2}{\varepsilon^2}$$

Theorem *Martingale convergence theorem*

Let (Y_n) be a martingale with respect to (X_n) satisfying $\sup_n(\mathbb{E}[Y_n^2]) < \infty$.
Then there exists a random variable Y such that $Y_n \to Y$ almost surely as $n \to \infty$.

5.4 Applications of martingales

Definition *Travelling salesman problem*

Let $P_1, \ldots, P_n$ be points in the Euclidean plane. We define the **length of the shortest tour** as

$$L(P_1, \ldots, P_n) := \min_\pi \sum_{i=1}^{n-1} \|P_{\pi(i+1)} - P_{\pi(i)}\| + \|P_{\pi(n)} - P_{\pi(1)}\|$$

where π ranges to all permutations of $1, \ldots, n$.

Theorem *Beardwood-Halton-Hammersley*

There exists a constant β, such that if $P_1, \ldots, P_n$ are i.i.d uniform in $[0,1]^2$,

$$\frac{L(P_1, \ldots, P_n)}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathbb{P}} \beta$$

6 Poisson processes

6.1 Poisson processes

Definition *Poisson process*

A **Poisson process** with intensity $\lambda > 0$ is a random function $t \to N(t)$ with domain $[0, \infty)$ and taking values in $Z_{\geq 0}$ such that

1. $N(0) = 0$
2. If $s \leq t$ then $N(s) \leq N(t)$
3. For every $t \geq 0$, as $h \searrow 0$, we have

$$\mathbb{P}(N(t+h) = j \mid N(t) = i) = \begin{cases} 0 & \text{if } j < i \\ 1 - \lambda h + o(h) & \text{if } j = i \\ \lambda h + o(h) & \text{if } j = i + 1 \\ o(h) & \text{if } j > i + 1 \end{cases}$$

4. If $s < t$, then $N(s)$ (the number of arrivals in $[0, s]$) and $N(t) - N(s)$ (the number of arrivals in $(s, t]$), are independent.

Lemma

If $N(t)$ satisfies the definition of a Poisson process with intensity λ, then

$$N(t) \overset{d}{=} \text{Poi}(\lambda t)$$

for all $t \in [0, \infty)$.

Corollary

For every $0 \leq s < t$ we have

$$N(t) - N(s) \overset{d}{=} \text{Poi}(\lambda(t - s))$$

Theorem

A Poisson process with intensity $\lambda > 0$ satisfies

1. $N(A) \overset{d}{=} \text{Poi}(\lambda \cdot \text{length}(A))$ for every finite interval $A \subseteq [0, \infty)$
2. If $A_1, \ldots, A_n$ are disjoint intervals then $N(A_1), \ldots, N(A_n)$ are independent random variables.

Note: these properties form an alternative definition for a Poisson process.
This definition can be extended to $\mathbb{R}^d$ by replacing intervals with boxes and length with volume.

6.2 Interarrival times

Definition *Interarrival times*

We can alternatively describe the Poisson process by a strictly increasing sequence (T_n) of **arrival times**.

$$T_i := \inf\{t : N(t) \geq i\}$$

We denote the **interarrival times** by

$$X_i := T_i - T_{i-1}$$

where $T_0 = 0$.

Theorem

The interarrival times are i.i.d. $\text{Exp}(\lambda)$ distributed.

6.3 Transformations

Theorem *Thinning theorem*

If $\mathscr{P}$ is a Poisson process of intensity λ and $\mathscr{Q} \subseteq \mathscr{P}$ is constructed by keeping each point of $\mathscr{P}$ with probability p, independently of all other points, then $\mathscr{Q}$ is a Poisson process of intensity $p\lambda$.

Theorem *Superposition theorem*

If $\mathscr{P}_1, \mathscr{P}_2$ are independent Poisson processes with intensities λ_1, λ_2 respectively,
then $\mathscr{P} := \mathscr{P}_1 \cup \mathscr{P}_2$ is a Poisson process of intensity $\lambda_1 + \lambda_2$.

Theorem *Scaling theorem*

If $\mathscr{P}$ is a Poisson proces of intensity λ and $\mathscr{Q} := \varphi[\mathscr{P}]$ where $\varphi(x) := ax$ with $a > 0$,
then $\mathscr{Q}$ is a Poisson process of intensity λ/a.

7 Brownian motion

Definition *Brownian motion*

Brownian motion is a random process $(B(t))_{t\geq 0}$ satisfying

1. The process has independent increments, that is, for all $0 < t_1 < \cdots < t_n$, the random variables

   $$B(t_1) - B(0), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$$

   are independent.
2. For every $0 \leq s < t$ the increment $B(t) - B(s)$ is $\mathscr{N}(0, t-s)$ distributed.
3. Almost surely, the function $t \mapsto B(t)$ is continuous.

If $B(0) = 0$ we speak of **standard Brownian motion**.

Theorem

Standard Brownian motion exists.

Proposition

For every $0 < t_1 < \cdots < t_n$ the random vector

$$[B(t_1), \ldots, B(t_n)]^T \overset{d}{=} \mathscr{N}(\mathbf{0}, \Sigma)$$

follows the multivariate normal distribution with covariance matrix Σ given by $\Sigma_{ij} = \min(t_i, t_j)$.

7.1 Transformations

Theorem *Translation invariance*

If $a \geq 0$ is fixed and $B(t)$ denotes a Brownian motion then the process given by

$$X(t) := B(a+t) - B(a))$$

is a standard Brownian motion.

Theorem *Scale invariance*

If $a \neq 0$ is fixed and $B(t)$ denotes a standard Brownian motion, then the process given by

$$X(t) := a^{-1} \cdot B(a^2 t)$$

is also a standard Brownian motion.

Theorem *Time inversion*

If $B(t)$ denotes a standard Brownian motion, then the process given by

$$X(t) := \begin{cases} t \cdot B(1/t) & \text{if } t > 0 \\ 0 & \text{if } t = 0 \end{cases}$$

is also a standard Brownian motion.

7.2 Properties

Theorem

Almost surely, for every $0 \leq a < b$, the function $t \mapsto B(t)$ is non-monotone on the interval $[a, b]$.

Proposition

For every fixed s, almost surely, $t \mapsto B(t)$ is not differentiable at s.

Theorem

Almost surely, $t \mapsto B(t)$ is non-differentiable at every $t \in [0, \infty)$.

Theorem *Second arcsine law*

Let $B(t)$ be standard Brownian motion and denote $L := \max\{t \in [0,1] : B(t) = 0\}$. Then for all $t \in [0, 1]$

$$\mathbb{P}(L \leq t) = \frac{2}{\pi} \arcsin(\sqrt{t})$$

Theorem *Third arcsine law*

Let $B(t)$ be standard Brownian motion and denote $T := \underset{t\in[0,1]}{\operatorname{argmax}} B(t)$. Then for all $t \in [0, 1]$

$$\mathbb{P}(T \leq t) = \frac{2}{\pi} \arcsin(\sqrt{t})$$

7.3 Brownian motion as a limit

Definition B_n

Consider the symmetric random walk on $\mathbb{Z}$, where $X_1, X_2, \dots$ are i.i.d with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ and let S_n be the partial sums of X_n. We turn the partial sums into a continuous function by linear interpolation:

$$S^*(t) := S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor) \cdot (S_{\lceil t \rceil} - S_{\lfloor t \rfloor})$$

Now we set

$$B_n(t) := \frac{S^*(nt)}{\sqrt{n}}$$

Definition *Metric for random functions*

$$\|f\|_\infty := \sup_{t\in[0,1]} |f(t)| \qquad \operatorname{dist}(f, g) := \|f - g\|_\infty$$

Proposition

If B_n is as above and B is standard Brownian motion, then there exist couplings of B_n, B such that

$$\mathbb{P}(\operatorname{dist}(B_n, B) > \varepsilon) \xrightarrow{n\to\infty} 0$$

Theorem *Skorokhod embedding theorem*

Let X be a random variable with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] < \infty$.
Then there exists a random time T such that

1. For each constant $t \in [0, \infty)$ the event $\{T = t\}$ depends only on $(B(s))_{S \leq t}$
2. $B(T) \overset{d}{=} X$

Theorem *Donsker's invariance theorem*

Suppose that $X_1, X_2, \ldots$ are i.i.d with $\mathbb{E}[X_1] = 0$, $\mathrm{Var}(X_1) = 1$ and define B_n as above.
Then there is a series of couplings such that

$$\mathbb{P}(\mathrm{dist}(B_n, B) > \varepsilon) \xrightarrow{n \to \infty} 0$$

for all $\varepsilon > 0$.

Index