

# Lecture 1: Revision of probability concepts, and the simple random walk

Susana Gomes

October 7<sup>th</sup>, 2021

# Plan for today

1. Short intro to the topics in this module
2. Revision of some basic probability definitions
3. Basics of random variables
4. The simple random walk, LLN and CLT
5. (if there is time) Intro to stochastic processes

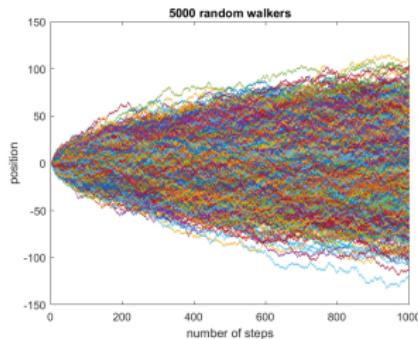
# Introduction

# Introduction

This module will be an overview of (some of) the most common concepts in modelling real world systems.

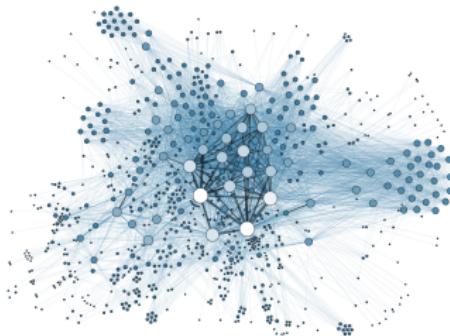
The first half of the module will broadly cover “stochastic modelling”, and it will focus on **Markov chains**, more general stochastic processes (including **diffusion processes**, Brownian motion, etc) and interacting particle systems.

Right: A lot of random walkers



After this, we will move on to **graphs** and **networks**. We will explore some basic graph definitions and properties, random graphs, percolation, and application of all this to networks.

Right: A network, taken from wikipedia



# Basic probability definitions

# Events and sets of events

In order to formulate probabilistic problems, we need to define some essential concepts, such as events, the sample space, probability spaces, etc. The next couple of slides are mostly to set notation up that we will use all the time.

- The **sample space**  $\Omega$  is the space where events we want to study live.  
*e.g., if you roll a die twice,  $\Omega$  is the set of all pairs of numbers between 1 and 6 ( $\{(1, 1), (1, 2), \dots, (6, 6)\}$ ).*
- An **event**  $\omega$  is a realisation of one of the members of the sample space.  
*e.g. (1, 6).*
- In practice, we work with **sets of events**  $A$  and we consider  $\omega \in A$ .  $A$  is always a (measurable) subset of  $\Omega$ .  
*e.g., "the two rolls add up to 8",  $A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$ .*
- The **set of all events**  $\mathcal{F} \subset \mathcal{P}(\Omega)$  is a subset of the powerset  $\mathcal{P}(\Omega)$ .  
*In order to be able to define probability, we need  $\mathcal{F}$  to be a closed system, i.e., it needs to be a  $\sigma$ -algebra.*

# $\sigma$ -algebras

Let  $\Omega$  be a set, and let  $\mathcal{P}(\Omega)$  be its power set. We say that  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  is a  **$\sigma$ -algebra** if it has the following properties:

1.  $\mathcal{F}$  is closed under the complement operation:

$$A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}.$$

2.  $\Omega \in \mathcal{F}$ .

*Note that together with 1., this means that the set of no events  $\emptyset$  is in  $\mathcal{F}$  since  $\emptyset = \Omega^C$ .*

3.  $\mathcal{F}$  is closed under *countable* union:

$$A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

# Probability

We can now define what is a probability distribution!

## Probability distributions

A **probability distribution**  $\mathbb{P}$  on  $(\Omega, \mathcal{F})$  is a function which satisfies the following properties:

(i) it is positive, i.e.  $\mathbb{P}(A) \geq 0$ .

(ii) it is normalised, i.e.  $\mathbb{P}(\emptyset) = 0$  and  $\mathbb{P}(\Omega) = 1$ .

(iii) it is additive, i.e.  $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ ,

where  $A_1, A_2, \dots$  is a collection of disjoint events, i.e.  $A_i \cap A_j = \emptyset, \forall i, j$ .

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

Note that (i) and (ii) mean that we have that  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ .

# Some relevant properties to note

- From the probability axioms + properties of  $\sigma$ -algebras, we have

$$\mathbb{P}(A^C) = 1 - \mathbb{P}(A), \quad \text{and} \quad \mathbb{P}(\emptyset) = 0.$$

- If  $\Omega$  is a **discrete set**, we always have

$$\mathcal{F} = \mathcal{P}(\Omega) \quad \text{and} \quad \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

e.g.  $\mathbb{P}(\text{the two rolls add up to } 8) =$

$$\mathbb{P}((2, 6)) + \mathbb{P}((3, 5)) + \mathbb{P}((4, 4)) + \mathbb{P}((5, 3)) + \mathbb{P}((6, 2)) = \frac{5}{36}.$$

- If  $\Omega$  is **continuous** (e.g.  $[0, 1]$ ), then we have

$$\mathcal{F} \subsetneq \mathcal{P}(\Omega)$$

# Independence and conditional probability

- We say that two events  $A, B \subset \Omega$  are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

*e.g. rolling a die repeatedly.*

- If  $\mathbb{P}(B) > 0$  then we define the **conditional probability** of  $A$  given  $B$  as

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

*If  $A$  and  $B$  are independent, then  $\mathbb{P}(A|B) = \mathbb{P}(A)$ .*

A useful result related to conditional probability is the Law of total probability:

**Lemma: Law of total probability**

Let  $B_1, \dots, B_n$  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 also that

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

# Random variables

# Random variables (definitions)

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

The **distribution function** of the random variable  $X$  is

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

- The rv  $X$  is called **discrete** if it only takes values in a countable subset

$$\Delta = \{x_1, x_2, \dots\} \subset \mathbb{R},$$

and its distribution is characterised by the **probability mass function**

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

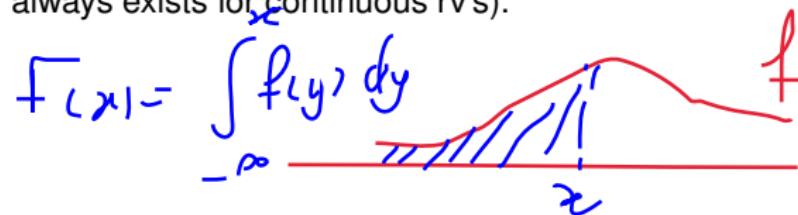
- The rv  $X$  is called **continuous**, if its distribution function is

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

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

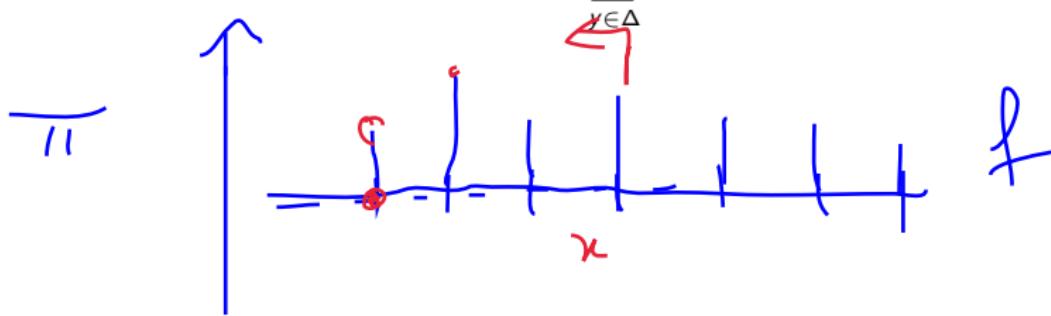
## So... What does this mean?

In general,  $f = F'$  is given by the derivative of the distribution function (which always exists for continuous rv's).



For discrete rv's,  $F$  is a step function with 'PDF'

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



## Important concepts (1/2)

There are many ways to describe random variables or events, and we will focus on the usual ones: mean, variance and covariance.

- The **expected value** of  $X$  is given by

$$\mathbb{E}[X] = \begin{cases} \sum_{x \in \Delta} x \pi(x), & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f(x) dx, & \text{if } X \text{ is continuous} \end{cases}$$

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

## Important concepts (2/2)

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 justifies the definition of **joint distributions**:

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

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

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.

# The simple random walk

## So... What happens when $n \rightarrow \infty$ ?

Turns out we can predict what a simple random walk will do (and we will work on expansions on this) using the Law of Large Numbers:

### Weak law of large numbers (LLN)

Let  $X_1, X_2, \dots \in \mathbb{R}$  be a sequence of iid rv'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 **distribution function** of  $Y_n$  converges to  $\mathbb{1}_{[\mu, \infty)}(x)$  for  $x \neq \mu$ ).

# And...

And in some (most!) situations we can even say more...

## Central limit theorem (CLT)

Let  $X_1, X_2, \dots \in \mathbb{R}$  be a sequence of iid rv'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 distribution, where  $\xi \sim N(0, 1)$  is a **standard Gaussian** with PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

$$Y_n \sim N(n\mu, n\sigma^2)$$

In fact, we can say that, as  $n \rightarrow \infty$ ,

$$(Y_n) = \sum_{k=1}^n X_k = n\mu + \sqrt{n}\sigma\xi + o(\sqrt{n}), \quad \xi \sim N(0, 1).$$

# Lecture 2: Discrete time Markov chains

Susana Gomes

October 8<sup>th</sup>, 2021

# Plan for today

1. Definition of stochastic processes and Markov chains
2. Some examples
3. Some properties of Markov chains
4. (if we have time) Stationary distributions



# Discrete Time Markov Chains

# Discrete-time stochastic processes, and the Markov property

We want to an efficient description for processes such as the simple random walk from yesterday. We will start with some definitions and then will look at some useful examples and properties.

A **discrete-time stochastic process** with **state space**  $S$  is a sequence

$$Y_0, Y_1, \dots = (Y_n : n \in \mathbb{N}_0)$$

of random variables taking values in  $S$ .

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

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

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

## Some notation

We will be working with the generic probability space  $\Omega$ , which is the **path space**

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

Note that  $\Omega$  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.

## Example 1

Consider the simple random walk from yesterday, with a random walker starting at  $X_0 = 0$ . We have

- $Y_0 = X_0 = 0$ .
- The state space  $S = \mathbb{Z}$ .
- Up to time  $N$ ,  $\mathbb{P}$  is a distribution on the finite path space  $\Omega_N$  with

$$\mathbb{P}(\omega) = \begin{cases} p^{\# \text{ of right-steps}} q^{\# \text{ of left-steps}} & , \text{ path } \omega \text{ possible} \\ 0 & , \text{ path } \omega \text{ not possible} \end{cases}$$

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

$$N=3$$

$$p^2 q$$

## More examples

- A **generalised random walk** is a random walk with  $Y_0 = 0$  and increments  $X_{n+1} = Y_{n+1} - Y_n \in \mathbb{R}$ . This is a Markov process with  $S = \mathbb{R}$  and  $\Omega_N = \mathbb{R}^N$ . It has an uncountable number of possible paths.
- A sequence  $Y_0, Y_1, \dots \in S$  of iid rv's is also a Markov process with state space  $S$ .
- Let  $S = \{1, \dots, 52\}$  be a deck of cards, and  $Y_1, \dots, Y_{52}$  be the cards drawn at random without replacement. Is this a Markov process?

# Homogeneous MCs and the Chapman-Kolmogorov equations

Let  $(X_n : n \in \mathbb{N}_0)$  be a homogeneous DTMC with **discrete** state space  $S$ .

Then we can define the **transition function**

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

**Proposition: Chapman-Kolmogorov equations**

If  $(X_n : n \in \mathbb{N}_0)$  is a homogeneous DTMC,  $p_n(x, y)$  is well defined and fulfills the **Chapman Kolmogorov equations**:

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

# Proof.

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

$$P_{n+k}(x, y) = P(X_{n+k} = y \mid X_0 = x)$$

LTP  $\sum_{z \in S} P(X_{n+k} = y \mid X_k = z, X_0 = x) P(X_k = z \mid X_0 = x)$

Markov  $\sum_{z \in S} P(X_{n+k} = y \mid X_k = z) P(X_k = z \mid X_0 = x)$

homog  $\sum_{z \in S} \underbrace{P(X_n = y \mid X_0 = z)}_{P_n(z, y)} \underbrace{P(X_k = z \mid X_0 = x)}_{P_x(x, z)}$

# Transition matrices

It is convenient to write all of this in matrix form. For this, we define the matrix  $P_n = (p_n(x, y) : x, y \in S)$ , and can rewrite the Chapman-Kolmogotov equations as:

$$P_{n+k} = P_n P_k$$

and, in particular,

$$P_{n+1} = P_n P_1.$$

which gives us a recursion relation to the matrices  $P_n$ .

With  $P_0 = \mathbb{I}$ , the obvious solution to this recursion is  $P_n = P^n$ , where we define the **transition matrix**

$$P = P_1 = (p(x, y) : x, y \in S).$$

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  $A_1, \dots, A_k \subset S$

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

# Properties of transition matrices and some notation

**Note that** there is no reason to have a fixed  $X_0$  and instead we can work with an **initial distribution**

$$\pi_0(x) := \mathbb{P}[X_0 = x].$$

The distribution at time  $n$  is then

$$\pi_n(x) = \sum_{y \in S} \sum_{s_1 \in S} \cdots \sum_{s_{n-1} \in S} \pi_0(y)p(y, s_1) \cdots p(s_{n-1}, x)$$

In this case, we can also write

$$\langle \pi_n | = \langle \pi_0 | P^n,$$

where  $\langle \cdot |$  denotes a row vector.

Finally, the transition matrix  $P$  is **stochastic**, i.e.

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

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

## Quick example

Back to the simple random walk...

$$q \quad p$$

$$X_0 = 0 \Rightarrow \pi_0 = \delta_0$$



$$P(X_1 = x \mid X_0 = 0) = p \delta_{1,x} + q \delta_{-1,x}$$

$$P \left( \dots 0 \quad q \quad 0 \quad p \quad 0 \dots \right)$$

$$\pi_1 = \pi_0 P = \begin{pmatrix} 0 \\ q \\ 0 \\ p \\ 0 \\ \vdots \end{pmatrix} \begin{matrix} \swarrow & \downarrow \\ \leftarrow & 0 \\ \leftarrow & 1 \end{matrix}$$

$$\pi_2 = (0 \dots q^2 0 \quad 2pq \quad 0 \quad p^2 0 \dots)$$

## Example: Random walk with boundaries

To make our lives simpler, we can look at MCs with finite state space. A good example of this are random walks with boundaries.

Let  $(X_n : n \in \mathbb{N}_0)$  be a simple random walk on  $S = \{1, \dots, L\}$  with

$$p(x, y) = p\delta_{y,x+1} + q\delta_{y,x-1}.$$

In this case, we need to tell it what happens once we reach the boundary (1 or  $L$ ). We can have the following boundary conditions:

- **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  $p(1, 2) = 1, p(L, L - 1) = 1.$

Periodic,  $L = 4$

$$\begin{pmatrix} 0 & p & 0 & q \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ p & 0 & q & 0 \end{pmatrix}$$

$$\pi_0 = (1, 0, 0, 0)$$

$$\Rightarrow \pi_1 = \pi_0 \cdot T$$

$$= (0, p, 0, q)$$

# Lecture 3: Discrete time Markov chains (continued) and intro to continuous time Markov chains

Susana Gomes

October 14<sup>th</sup>, 2021

# Plan for today

1. Stationary and reversible distributions
2. Absorbing states
3. Basics of Continuous time Markov chains

# Stationary and reversible distributions

Let  $(X_n : n \in \mathbb{N}_0)$  be a homogeneous DTMC with state space  $S$ .

We say that the distribution  $\pi(x)$ ,  $x \in S$  is **stationary** if for all  $y \in S$

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

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

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

**Note that** if a distribution  $\pi$  is reversible then it is stationary:

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

# Existence of stationary distributions

If we see a stationary distribution as a row vector  $\langle \pi | = (\pi(x) : x \in S)$ , then we can see that stationary distributions are **left eigenvectors** of  $P$  with **eigenvalue** 1:

$$\langle \pi | = \langle \pi | P.$$

This immediately implies the following:

## Existence

Every DTMC **with finite state space** has at least one stationary distribution.

**Proof.** Since  $P$  is stochastic, we have  $P|\mathbf{1}\rangle = |\mathbf{1}\rangle$ .

So 1 is an eigenvalue of  $P$ , and its corresponding left eigenvector(s) can be shown to have non-negative entries and thus can be normalized to be stationary distribution(s)  $\langle \pi |$ .

This last part relies on the Perron-Frobenius theorem, which we will look at in more detail later.

## Example

However, there is no reason for a stationary distribution to be unique...

$$p_r(x,y) = P(X_n = y \mid X_0 = x) > 0$$

!

### Irreducible DTMCs

A DTMC is called **irreducible**, if for all  $x, y \in S$  we have

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

Using the Perron-Frobenius theorem again, we will be able to prove (in a week or so) that irreducible DTMCs have a **unique** stationary distribution.

# Absorbing states

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

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

In other words, once the chain reached the state  $s$ , it won't leave it.

A good example of a DTMC with an absorbing state is a **simple Random Walk with absorbing boundary conditions**.

Recall:

A simple random walk in  $S = \{1, \dots, L\}$  with absorbing BCs verifies:

- $p(x, y) = p\delta_{y,x+1} + q\delta_{y,x-1}$
- $p(L, L) = 1, p(1, 1) = 1.$

We can use this to compute interesting things, like the **absorption probability**. In practice, we can also have absorbing "cycles", or other interesting behaviours without absorbing states. We will say more about this later.

# Absorption probability

Let  $h_k$  be the **absorption probability** for  $X_0 = k \in S = \{1, \dots, L\}$ , i.e.

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

First of all, we can clearly see that  $h_1 = h_L = 1$ .

Now, for the other starting points, we condition on the first jump and use the Markov property:

$$h_k = ph_{k+1} + qh_{k-1} \quad \text{for } k = 2, \dots, L-1,$$

and this gives us a recursive relation.

Assume that the solution is of the form  $h_k = \lambda^k$ ,  $\lambda \in \mathbb{C}$ .

$$\lambda^k = p \lambda^{k+1} + q \lambda^{k-1} \Rightarrow \lambda = p \lambda^2 + q$$

$$\lambda_1 = 1 \quad \text{or} \quad \lambda_2 = \frac{q}{p}$$

$$\Rightarrow h_k = a \lambda_1^k + b \lambda_2^k = a + b \left(\frac{q}{p}\right)^k$$

$$\begin{aligned} h_1 &= 1 \Rightarrow a + b \frac{q}{p} = 1 & a &= 1 & \Rightarrow h_k &= 1 \\ h_L &= 1 \Rightarrow a + b \left(\frac{q}{p}\right)^L = 1 & b &= 0 & \forall k \end{aligned}$$

## Distribution at time $n$

The last topic we will consider for DTMCs is their distribution at time  $n$ .

Consider a DTMC on a finite state space with  $|S| = L$ , and let  $\lambda_1, \dots, \lambda_L \in \mathbb{C}$  be the **eigenvalues** of the transition matrix  $P$  with corresponding

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

If we assume that **all eigenvalues are distinct** we can always write

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

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

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

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

## Distribution at time $n$ - some observations

The **Gershgorin theorem** (see handout 1) implies that  $|\lambda_i| \leq 1$ .

This also means that contributions with  $|\lambda_i| < 1$  decay exponentially.

$\lambda_1 = 1$  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**.



# Continuous time Markov chains

# Continuous-time Markov chains

We'll spend the rest of today (and probably tomorrow) generalising the previous results for the case of **continuous time**. So... as before:

- 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  $A \subset S$ ,  $n \in \mathbb{N}$ ,  $t_1 < \dots < t_{n+1} \in [0, \infty)$  and  $s_1, \dots, s_n \in S$ , we have

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

- A Markov process (MP) is called **homogeneous** if for all  $A \subset S$ ,  $t, u > 0$  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)**.

## A couple of technical details

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

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

$\mathbb{P}$  is a probability distribution on  $\Omega$ .

By **Kolmogorov's extension theorem**  $\mathbb{P}$  is fully specified by its **finite dimensional distributions (FDDs)**. These are the distributions of the form

$$\mathbb{P}[X_{t_1} \in A_1, \dots, X_{t_n} \in A_n], \quad n \in \mathbb{N}, \quad t_i \in [0, \infty), \quad A_i \subset S.$$

# Chapman-Kolmogorov equations for CTMCs

We can still define a transition function as before and the Chapman-Kolmogorov equations are valid.

Let  $(X_t : t \geq 0)$  by a homogeneous CTMC with state space  $S$ . Then for all  $t \geq 0$  the **transition function** is given by

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

**Proposition:** Chapman-Kolmogorov equations

The transition function is well defined and fulfills the **Chapman Kolmogorov equations**

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

# Generator of a CTMC

As before, we can write this in matrix notation:

We define  $P_t = (p_t(x, y) : x, y \in S)$  and we can write

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

In particular,

$$\frac{P_{t+\Delta t} - P_t}{\Delta t} = P_t \frac{P_{\Delta t} - \mathbb{I}}{\Delta t} = \frac{P_{\Delta t} - \mathbb{I}}{\Delta t} P_t.$$

We now take  $\Delta t \searrow 0$  and get the so-called **forward and backward equations**

$$\frac{d}{dt} P_t = P_t G = G P_t, \quad \text{where} \quad G = \left. \frac{dP_t}{dt} \right|_{t=0}$$

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

# Forward and backward equations for CTMCs

The solution for these equations is given by the matrix exponential

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

And from this we can obtain the **distribution**  $\pi_t$  at time  $t > 0$ :

$$\langle \pi_t | = \langle \pi_0 | \exp(tG) \quad \text{which solves} \quad \frac{d}{dt} \langle \pi_t | = \langle \pi_t | G.$$

Note that, just like before, if  $S$  is finite, we can compute the eigenvalues of  $G$ ,  $\lambda_1, \dots, \lambda_L \in \mathbb{C}$ . Then,  $P_t$  has eigenvalues  $\exp(t\lambda_i)$  with the same eigenvectors  $\langle v_i |$ ,  $| u_i \rangle$ .

If the  $\lambda_i$  are distinct, we can still expand the initial condition in the eigenvector basis

$$\langle \pi_0 | = \alpha_1 \langle v_1 | + \dots + \alpha_L \langle v_L |,$$

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

$$\langle \pi_t | = \alpha_1 \langle v_1 | e^{\lambda_1 t} + \dots + \alpha_L \langle v_L | e^{\lambda_L t} .$$

## Transition rates

Using the expression for  $P_t$  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 the  $g(x, y) \geq 0$  can be interpreted as **transition rates**.

We also have

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

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

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

# The Master equation

Using the results from the previous slide, we can rewrite the equation for the distribution at time  $t$ :

$$\langle \pi_t | = \langle \pi_0 | \exp(tG) \quad \text{which solves} \quad \frac{d}{dt} \langle \pi_t | = \langle \pi_t | G,$$

as the **Master equation**

$$\frac{d}{dt} \pi_t(x) = \underbrace{\sum_{y \neq x} \pi_t(y) g(y, x)}_{\text{gain term}} - \underbrace{\sum_{y \neq x} \pi_t(x) g(x, y)}_{\text{loss term}} \quad \text{for all } x \in S.$$

**Note that:** the Gershgorin theorem now implies that either  $\lambda_i = 0$  or  $\operatorname{Re}(\lambda_i) < 0$  for the eigenvalues of  $G$ , so there are **no persistent oscillations for CTMCs**.

# Lecture 4: Stationary distributions for CTMCs

Susana Gomes

October 15<sup>th</sup>, 2021

## Plan for today

1. Chapman-Kolmogorov equations, generators and the Master equation.
2. Stationary and irreducible distributions for CTMCs
3. Discussion about Assignment 1

# Chapman-Kolmogorov equations for CTMCs

We can still define a transition function as before and the Chapman-Kolmogorov equations are valid.

Let  $(X_t : t \geq 0)$  by a homogeneous CTMC with state space  $S$ . Then for all  $t \geq 0$  the **transition function** is given by

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

**Proposition:** Chapman-Kolmogorov equations

The transition function is well defined and fulfills the **Chapman Kolmogorov equations**

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

# Generator of a CTMC

As before, we can write this in matrix notation:

We define  $P_t = (p_t(x, y) : x, y \in S)$  and we can write

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

In particular,

$$\frac{P_{t+\Delta t} - P_t}{\Delta t} = P_t \frac{P_{\Delta t} - \mathbb{I}}{\Delta t} = \frac{P_{\Delta t} - \mathbb{I}}{\Delta t} P_t.$$

We now take  $\Delta t \searrow 0$  and get the so-called **forward and backward equations**

$$\frac{d}{dt} P_t = P_t G = GP_t, \quad \text{where}$$

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

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

# Forward and backward equations for CTMCs

The solution for these equations is given by the matrix exponential

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

And from this we can obtain the **distribution**  $\pi_t$  at time  $t > 0$ :

$$\langle \pi_t | = \langle \pi_0 | \exp(tG) \quad \text{which solves} \quad \frac{d}{dt} \langle \pi_t | = \langle \pi_t | G.$$

Note that, just like before, if  $S$  is finite, we can compute the eigenvalues of  $G$ ,  $\lambda_1, \dots, \lambda_L \in \mathbb{C}$ . Then,  $P_t$  has eigenvalues  $\exp(t\lambda_i)$  with the same eigenvectors  $\langle v_i |$ ,  $| u_i \rangle$ .

If the  $\lambda_i$  are distinct, we can still expand the initial condition in the eigenvector basis

$$\langle \pi_0 | = \alpha_1 \langle v_1 | + \dots + \alpha_L \langle v_L |,$$

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

$$\langle \pi_t | = \alpha_1 \langle v_1 | e^{\lambda_1 t} + \dots + \alpha_L \langle v_L | e^{\lambda_L t}.$$

# Transition rates

Using the expression for  $P_t$  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 the  $g(x, y) \geq 0$  can be interpreted as **transition rates**.

We also have

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

$P_t$  is stochastic

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

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

$$\begin{aligned} &= 1 + g(x, x)\Delta t + \sum_{y \neq x} g(x, y)\Delta t = 1 - g(x, x)\Delta t \\ &\quad + \sum_{y \neq x} g(x, y)\Delta t = 0 \end{aligned}$$

# The Master equation

Using the results from the previous slide, we can rewrite the equation for the distribution at time  $t$ :

$$\langle \pi_t | = \langle \pi_0 | \exp(tG) \quad \text{which solves} \quad \frac{d}{dt} \langle \pi_t | = \langle \pi_t | G,$$

as the **Master equation**

$$\frac{d}{dt} \pi_t(x) = \underbrace{\sum_{y \neq x} \pi_t(y) g(y, x)}_{\text{gain term}} - \underbrace{\sum_{y \neq x} \pi_t(x) g(x, y)}_{\text{loss term}} \quad \text{for all } x \in S.$$

**Note that:** the Gershgorin theorem now implies that either  $\lambda_i = 0$  or  $\operatorname{Re}(\lambda_i) < 0$  for the eigenvalues of  $G$ , so there are **no persistent oscillations for CTMCs**.

# Stationary and reversible distributions

These definitions are similar to the discrete-time case...

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  $\langle \pi | G = \langle 0 |$ , or for all  $y \in S$

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

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

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

- As before, **reversibility implies stationarity**.
- Stationary distributions are left **eigenvectors** of  $G$  with **eigenvalue** 0.
- $\langle \pi | G = \langle 0 |$  implies  $\langle \pi | P_t = \langle \pi | (\mathbb{I} + \sum_{k \geq 1} t^k G^k / k! ) = \langle \pi |$  for all  $t \geq 0$

# Stationary distributions (existence)

Proposition (existence):

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

**Proof.** The proof is similar to the discrete-time case. Since  $G$  has row sum 0, we have  $G|\mathbf{1}\rangle = |\mathbf{0}\rangle$ .

So 0 is an eigenvalue of  $G$ , and its corresponding left eigenvector(s) can be shown to have non-negative entries and thus can be normalized to be stationary distributions  $\langle\pi|$ .

Remark:

Note that if  $S$  is countably infinite, stationary distributions may not exist.

This is the case, for example, for the SRW on  $\mathbb{Z}$  or the Poisson process on  $\mathbb{N}$  (which we will see later).

# Stationary distributions (uniqueness)

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

$$p_t(x, y) > 0 \text{ for some } t > 0.$$

**Note that** for continuous time irreducibility implies  $p_t(x, y) > 0$  for **all**  $t > 0$ .

**Proposition (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:

- We know that  $\lambda_1 = 1$  is an eigenvalue of  $P$ .
- From PF, we can conclude that this eigenvalue is singular if and only if the chain is irreducible.
- As before, we can show that its corresponding left and right eigenvectors have non-negative entries and so we obtained the distribution.

# More on stationary distributions

of G

The Perron-Frobenius theorem also implies the following:

- If the chain is continuous-time, all remaining eigenvalues  $\lambda_i \in \mathbb{C}$ ,  $i \neq 1$  satisfy  $\text{Re}(\lambda_i) < 0$ .
- If the chain is discrete-time aperiodic (no persistent oscillations), all remaining eigenvalues  $\lambda_i \in \mathbb{C}$ ,  $i \neq 1$  satisfy  $|\lambda_i| < 1$ . → of P
- The second part of the Perron Frobenius theorem also implies convergence of the transition functions to the stationary distribution, since

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

This is usually called ergodicity (and we will see more of it next week).

$e^{-t} \rightarrow 0$  as  $t \rightarrow \infty$

# Lecture 5: Sample paths, ergodicity, and MCMC

Susana Gomes

October 21<sup>st</sup>, 2021

# Plan for today

1. Sample paths
2. Some examples
3. Ergodicity
4. An application: MCMC

## Sample paths (holding times)

Another useful way to characterise CTMCs is by looking at sample paths and their properties.

A **sample path**  $t \mapsto X_t(\omega)$  is a function that describes the state of our CTMC at time  $t$ . It is a piecewise constant and right-continuous function by convention.

Before we actually see how to write down sample paths, we will look at another concept:

For  $X_0 = x$ , we define the **holding time**  $W_x := \inf\{t > 0 : X_t \neq x\}$ .

### Proposition:

The holding time  $W_x$  is **exponentially distributed** with mean  $\frac{1}{|g(x,x)|}$ , i.e.,  $W_x \sim \text{Exp}(|g(x,x)|)$ .

If  $|g(x,x)| > 0$ , the chain jumps to  $y \neq x$  after time  $W_x$  with probability  $\frac{g(x,y)}{|g(x,x)|}$ .

## Proof

For any  $t, u > 0$ , we have

$$P(W_x > t+u | W_x > t) = P(W_x > t+u | X_t = \infty)$$

$\xrightarrow{W_x > t \text{ means } X_t = \infty}$

$$= P(W_x > u) \xrightarrow{\text{memoryless property}}$$

$W_x > t$  means  $X_s = \infty$  if  $0 < s \leq t$   
 Since  $X_t$  is indep. of  $X_s$  for  $0 < s < t$  we can say this.

$$\Rightarrow P(W_x > t+u) = P(W_x > t+u | W_x > t) P(W_x > t)$$

$\xrightarrow{\text{law of total prob.}}$

$$= P(W_x > u) P(W_u > t)$$

$\checkmark$  same arg. as Friday

$$\Rightarrow P(W_x > t) = e^{\gamma t} \quad \text{with} \quad \gamma = \frac{d}{dt} P(W_x > t) \Big|_{t=0}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{P(x, x_0 + \Delta t) - 1}{\Delta t} = g(u, x) \leq 0.$$

$\uparrow$  same as Friday

## Proof (continued)

Condition on chain leaving  $x$  shortly:

$P(X_{t+\Delta t} = y | X_t = x, W_{t+\Delta t} < \Delta t) = \text{will explain this separately}$

$$= \frac{P(X_{t+\Delta t} = y | X_t = x)}{P(W_{t+\Delta t} < \Delta t | X_t = x)}$$

$$\underset{\substack{\downarrow \\ \text{Friday}}}{\lim_{\Delta t \rightarrow 0}} \frac{P_{0t}(x, y)}{1 - P_{0t}(x, x)} = \underset{\substack{\uparrow \\ \text{Friday}}}{\lim_{\Delta t \rightarrow 0}} \frac{\Delta t g(x, y)}{1 - (1 + \Delta t g(x, x))}$$

$$= \frac{g(x, y)}{1 - g(x, x)}$$

## Sample paths (jump times)

Once we have the holding times defined, we can define **jump times**:  $J_0, J_1, \dots$ . These are defined recursively as

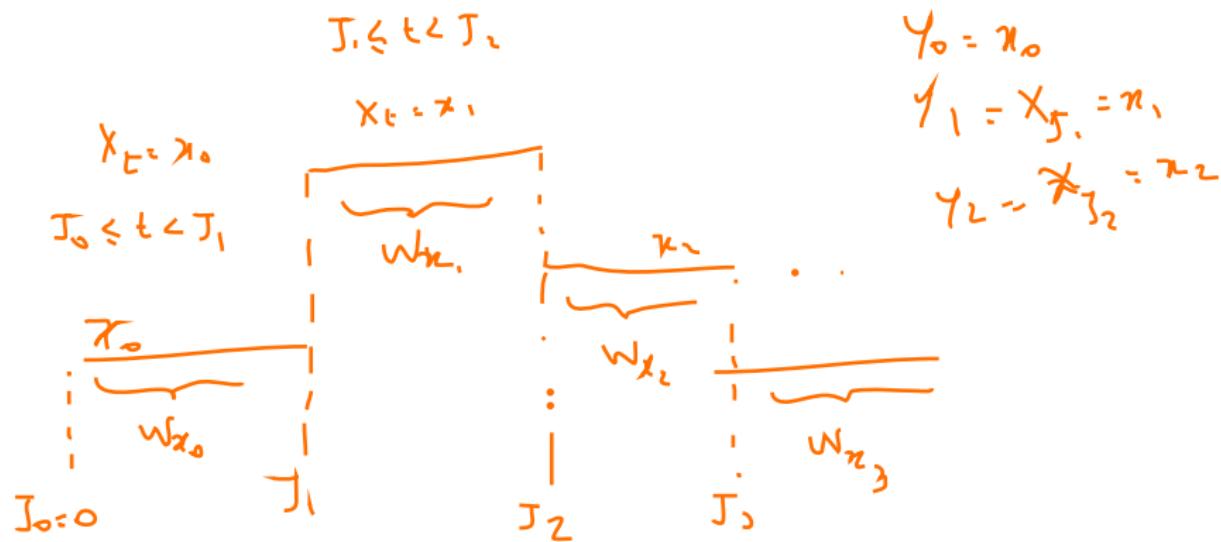
$$J_0 = 0 \quad \text{and} \quad J_{n+1} = \inf\{t > J_n : X_t \neq X_{J_n}\}.$$

- Jump times are an example of "**stopping times**" because we are working with right-continuous paths.
- This means that for all  $t \geq 0$ , the event  $\{J_n \leq t\}$  depends **only** on  $(X_s : 0 \leq s \leq t)$ .
- This is justified by the **strong Markov property**: If we condition on a stopping time  $\tau$  with  $X_\tau = i$  then  $X_{\tau+s}$  is independent of  $X_s$  for all  $s \leq \tau$ .
- i.e., subsequent holding times and jump probabilities are all independent.

# Sample paths

Using the previous results, we can now define the **jump chain**:

$$(Y_n : n \in N_0) \quad \text{with} \quad Y_n := X_{J_n}$$



# Sample paths

Using the previous results, we can now define the **jump chain**:

$$(Y_n : n \in \mathbb{N}_0) \quad \text{with} \quad Y_n := X_{J_n}$$

This is a discrete-time Markov chain with transition matrix

$$p^Y(x, y) = \begin{cases} 0 & , x = y \\ g(x, y)/|g(x, x)| & , x \neq y \end{cases}$$

if  $g(x, x) < 0$ .

If  $g(x, x) = 0$ , we say (by convention) that

$$p^Y(x, y) = \delta_{x,y}.$$

## Sample paths

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

$$J_n = \sum_{k=0}^{n-1} W_{Y_k}.$$

A slide to sketch this if needed

Simulate CTRC  $X_t$ :

Start with  $x_0$ , define  $y_0 = x_0$

Compute  $w_{y_0} \rightarrow J_0 = w_{y_0}$

use  $g(y_0, y_1)$   
and sample from  
 $\exp(1 g(y_0, y_1))$

$y_1$  = next step of DTMC

(advance using  $p^T(y_0, y)$  from last slide)

$\rightarrow x_t = y_1$  for  $J_0 \leq t < J_1$  ( $J_0$  computed)

Compute  $w_{y_1}$  (sample from  $\exp(1 g(y_1, y_2))$ )

and  $J_1 - w_{y_0} + w_{y_1}$

Compute  $y_2$ , repeat.

## Example 1 - Poisson Process

Suppose you want to model the arrival of customers to a waiting line. If we assume the following:

1. People arrive alone (never in groups).
2. The probability  $p$  that an arrival occurs during a time interval of (small) length  $\Delta t$  is proportional to  $\Delta t$ :  $p = \lambda\Delta t$ .
3. The number of arrivals on disjoint intervals is independent.

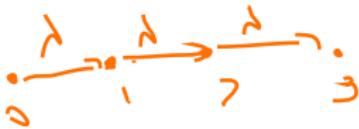
In this context, we would like to know, e.g., the law of the number of arrivals  $N_t$  in the interval  $[0, t]$ , or the number of arrivals per unit time.

The **Poisson process** is a good way to model this. Before we define it, we need to point out a couple of assumptions.

We can assume that people arrive in the waiting line completely at random. also, for 2. to be valid, we need to think in an "infinitesimal" sense, i.e., we should make sure that

$$\lim_{\Delta t \rightarrow 0} \frac{p}{\Delta t} = \lambda.$$

# Poisson Process



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

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

We can show that the PP( $\lambda$ ) has **stationary and independent increments** with

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

This can be shown by dividing the time interval  $[0, t]$  into intervals of length  $\Delta t = t/n$  for  $n$  big enough and using properties of the Binomial law.

This is also related to the fact that  $\pi_t(k) = p_t(0, k)$  solves the Master equation

$$\frac{d}{dt} \pi_t(k) = (\pi_t G)(k).$$

## Example 2 - Birth-Death chains

A **birth-death chain** with **birth rates**  $\alpha_x$  and **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  $\beta_0 = 0$ .

These chains are used to model all sorts of things, from server queues to population sizes, to the evolution of an epidemic.

Special cases include

- **M/M/1 server queues:**  $\alpha_x \equiv \alpha > 0$ ,  $\beta_x \equiv \beta > 0$  for  $x > 1$ .  
e.g. a queue with Poisson arrivals but where one customer is served at a time, with random service time (following an exponential law).
- **M/M/ $\infty$  server queues:**  $\alpha_x \equiv \alpha > 0$ ,  $\beta_x = x\beta$ .  
same as before but with immediate service.
- **population growth model:**  $\alpha_x = x\alpha$ ,  $\beta_x = x\beta$ .

# Ergodicity

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

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

Theorem:

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

The proof of this follows from the Perron-Frobenius theorem:  
We finished our lecture last Friday by saying that

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

and this implies the theorem.  
mention countably infinite state space.

# Ergodicity

A very important result for ergodic Markov Chains is the Ergodic Theorem.

Theorem (Ergodic Theorem):

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

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

For a proof of this theorem, you can check the book by Grimmett and Stirzaker (2001), chapter 9.5.

What this means is that stationary expectations can be approximated by time averages, which is the basis for **Markov chain Monte Carlo** (which we will see next).

An immediate example is that if we choose the indicator function  $f = \mathbb{1}_x$  we get  $\mathbb{E}_\pi[f] = \pi(x)$ .

# Markov Chain Monte Carlo (MCMC)

MCMC is used in several applications, when one needs to sample from some distribution  $\pi$  on a very large state space  $S$ .  
*(examples)*

Often, in these problems, one needs to compute complicated integrals which are not straightforward. Examples include:

- In general, compute **expectations**:

$$\mathbb{E}_\pi[f] = \sum_{x \in S} f(x)\pi(x) \quad \text{or} \quad \mathbb{E}_\pi[f] = \int f(x)\pi(x) dx$$

- In statistical mechanics, compute **Gibbs measures**:

If  $\pi(x) = \frac{1}{Z(\beta)} e^{-\beta H(x)}$ , compute the **partition function**

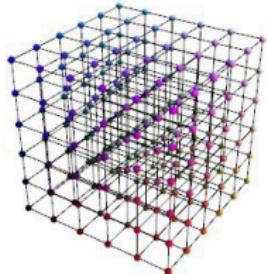
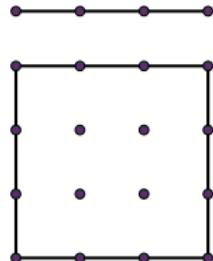
$$Z(\beta) = \sum_{x \in S} e^{-\beta H(x)}.$$

# Why do we need Monte Carlo Methods?

Suppose you want to compute an integral in a domain  $D = [0, 1]^d$  and want to use numerical quadrature to evaluate the integral.

1. Choose mesh of grid points within state space, with mesh-size  $h$ .
2. Evaluate  $f(x_i)\pi(x_i)$  for every grid point  $x_i$ .
3. Use quadrature scheme to approximate integral.

- With the standard quadrature approach, error is typically  $O(h^k)$ , for some  $k \geq 2$ .
- The number of points evaluated is  $M \sim O(h^{-d})$
- $\Rightarrow$  error  $\sim O(M^{-k/d})$ .



**The computational cost grows exponentially with dimension** (if we want to maintain the same error) This is usually known as curse of dimensionality.

# MCMC

The reason that MCMC works is that we can use the **ergodic theorem** to estimate expectations by time averages! (together with reversibility)

For an MCMC algorithm, we need to:

- assume that  $\pi(x) > 0$  for all  $x \in S$  (otherwise restrict  $S$ ).
- come up with a DTMC with transition function  $p(x, y)$  (CTMC with generator  $g(x, y)$ ) such that  $\pi$  is its stationary distribution.
- This can be done, e.g., via **detailed balance**:

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

$$\pi(x)p(x, y) = \pi(y)p(y, x) \quad (\text{discrete}).$$

- For example, for Gibbs measures, we know that

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

## How does this actually work?

The main point of MCMC is that then we will *sample from the stationary distribution*  $\pi$ . To do this, we need to find the right  $p(x, y)$ . Suppose we have an associated distribution  $q(\cdot|x)$  which is easy to sample.

1. Write  $p(x, y) = q(x, y) a(x, y)$
2. Propose a move from  $x$  to  $y$  with probability  $q(x, y)$
3. Accept this move from with probability  $a(x, y)$ .

To make sure this works, we need to check that the Markov chain we generated with  $p(x, y)$  has  $\pi$  as its stationary distribution.

I will do this for the discrete case (and for one particular example), but the continuous case is analogous.

# Metropolis-Hastings algorithm

This is one of the simplest MCMC algorithms you can think of.

Suppose that the chain is at the state  $X_n$  at time  $n$ . Then

1. Generate  $Y \sim q(X_n, y)$ .
2. Set  $X_{n+1} = Y$  with probability  $a(X_n, Y)$ , where

$$a(x, y) = \min \left\{ 1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right\}$$

3. Otherwise, reject  $Y$  and set  $X_{n+1} = X_n$ .

We can easily prove that  $\pi$  is reversible with respect to the transition density of this DTMC, and therefore it is a stationary distribution!

We would then need to show it is ergodic for it all to work, but we won't do that here.

Space for notes if needed

$$p(x, \pi) = ? , p(\pi, y) = ?$$

1)  $p(\pi, y) = q(\pi, y) \alpha(\pi, y)$  ✓ (propose move from  $\pi$  to  $y$  and accept it)

2)  $p(\pi, \pi) = q(\pi, \pi) \alpha(\pi, \pi) + \sum_{z \in S} q(\pi, z) (1 - \alpha(\pi, z))$   
propose  $\pi \neq \pi$  & accept or propose  $z \neq \pi \in S$  and reject it

$$\Rightarrow p(\pi, y) = q(\pi, y) \alpha(\pi, y) + \delta_{\pi, y} \sum_{z \in S} q(\pi, z) (1 - \alpha(z, z)).$$

→ Need to show it is reversible (to show stationary)

$$\pi(x) p(x, y) = \pi(x) q(y, x) \alpha(y, x) = \pi(x) q(y, x) \min\left\{1, \frac{\pi(y) q(y, x)}{\pi(x) q(x, y)}\right\}$$

$$\begin{aligned} \uparrow \\ z=y:: &= \min\{\pi(x) q(x, y), \pi(y) q(y, x)\} = \\ \text{out:} &= \min\left\{\frac{\pi(x) q(x, y)}{\pi(y) q(y, x)}, \min\{\pi(y) q(y, x), \pi(x) q(x, y)\}\right\} = \end{aligned}$$

$$\min\left\{\frac{\pi(x) q(x, y)}{\pi(y) q(y, x)}, \min\{\pi(y) q(y, x), \pi(x) q(x, y)\}\right\} = \pi(y) q(y, x) \alpha(y, x) - \pi(y) p(y, x)$$

so reversible!

## Other examples

- **Independence sampler:** If  $q(x, y) = q(y)$  (independent of the current state)

$$a(x, y) = \min \left\{ 1, \frac{\pi(y)q(x)}{\pi(x)q(y)} \right\}.$$



WARWICK  
THE UNIVERSITY OF WARWICK

# Lecture 6: Reversibility, and countably infinite state spaces

Susana Gomes

October 22<sup>nd</sup>, 2021  
Reminder about class swap in a couple of weeks!

## Plan for today

1. Reversibility
2. Communicating classes
3. Explosions
4. Some time to catch up (next week we'll start the continuous state space case)

# Reversibility

I mentioned before that we would look again at reversibility and the detailed balance condition. The main nice thing that comes out of reversibility is that we can “reverse time”. Before we see how, let us first look at the following:

## Proposition (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

$$g^Y(x, y) = \frac{\pi(y)}{\pi(x)} g^X(y, x)$$

and stat. prob.  $\pi$ .

**Note that:** This means that the definition of stationary chains can be extended to negative times:  $(X_t : t \in \mathbb{R})$ , with the time reversed chain given by  $Y_t := X_{-t}$ .

## Reversibility

Proof.

$$\begin{aligned} & \text{P}(Y_t = y | Y_s = x) = \text{P}(X_{-t} = y | X_{-s} = x) \\ &= \frac{\text{P}(X_{-t} = y \cap X_{-s} = x)}{\text{P}(X_{-s} = x)} \pi(y) \\ &= \frac{\cancel{\text{P}(X_{-s} = x)}}{\text{P}(X_{-s} = x | X_{-t} = y) \cdot \text{P}(X_{-t} = y)} \cdot \underbrace{\text{P}(X_{-s} = x)}_{\pi(x)} \quad \square \\ &= \pi(x) \text{P}(X_{-t} = y | X_{-s} = x) \end{aligned}$$

# Reversibility

## Proof.

- An analogous statement (with similar proof!) 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**, i.e.,

$$g^Y(x, y) = g^X(x, y), \quad (\text{or } p^Y(x, y) = p^X(x, y)).$$

- The time reversal of non-stationary MCs is in general **not** a homogeneous MC, and using Bayes' Theorem (for DTMCs) we get

$$p^Y(x, y; n) = \frac{\pi_{N-n-1}(y)}{\pi_{N-n}(x)} p^X(y, x).$$

# Countably infinite state space

The last topic we will look at for CTMCs is what happens when we are in a **countably infinite state space**. This is the case, e.g., of a simple random walk in  $\mathbb{Z}$ .

**Note that:** The SRW is a DTMC!, and the results in this section are valid for both CTMCs and DTMCs.

The reason we care about this, is that for infinite state space, Markov chains can get ‘lost at infinity’ and have no stationary distribution.

Again, think of the SRW. It is irreducible, but it is not ergodic and does **not** have a stationary distribution!

So... What can we do? We need to do a bit more to be able to characterise these Markov chains.

Yesterday we defined holding times. Today we will work with a “similar” time:

## Return times

We define the first **return time** to a state  $x$  by  $T_x := \inf\{t > J_1 : X_t = x\}$

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

# Classification of states and communication classes

We will use the return times to say something about the states of our MCs:

## Definition (classification of states)

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

$$\pi(x) = \frac{1}{\mathbb{E}[T_x | X_0 = x]} \mathbb{E} \left[ \int_0^{T_x} \mathbb{1}_x(X_s) ds | X_0 = x \right].$$

# Explosions

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

## Definition

For a CTMC with jumping times  $J_n$ , we define the **explosion time** by

$$J_\infty := \lim_{n \rightarrow \infty} J_n \in (0, \infty].$$

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

**Note that:** if the chain is explosive, what this means is that the chain does *infinite jumps in a finite amount of time*.

# Explosions

It is easy to see that if the exit rates are uniformly bounded, i.e.

$$\sup_{x \in S} |g(x, x)| < \infty,$$

then the chain is non-explosive, and this is always the case if  $S$  is finite.

However, an example of an explosive CTMC is a **pure birth chain**. Consider  $X_0 = 1$  and the generator of this CTMC is

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

If  $\alpha_x \rightarrow \infty$  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  $\mathbb{P}[J_\infty = \infty] = 0 < 1$ .

# Lecture 7: Processes with continuous state space

Susana Gomes

October 28<sup>th</sup>, 2021

## Plan for today

1. Summary of what we have done so far
2. Generalisation to continuous state space and a couple of examples
3. Stationary independent increments and Brownian motion

# Processes with continuous state space

# Markov processes with $S = \mathbb{R}$

Recall the definition of continuous time Markov process from week 3...

- 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  $A \subset S$ ,  $n \in \mathbb{N}$ ,  $t_1 < \dots < t_{n+1} \in [0, \infty)$  and  $s_1, \dots, s_n \in S$ , we have

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

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

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

We spent the last 3 weeks talking about what happens when the state space  $S$  is finite. The next couple of weeks will be focused on when  $S = \mathbb{R}$ .

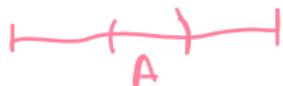
# Kernels, densities, and the Chapman-Kolmogorov equations

Let  $(X_t : t \geq 0)$  be a homogeneous MP as in previous slide, with state space  $S = \mathbb{R}$ .

For all  $t \geq 0$  and (measurable)  $A \subset \mathbb{R}$ , the **transition kernel** for all  $x \in \mathbb{R}$

$$P_t(x, A) := \mathbb{P}(X_t \in A | X_0 = x) = \mathbb{P}(X_{t+u} \in A | X_u = x) \quad \forall u \geq 0$$

is well defined.



## Proposition (Chapman-Kolmogorov equations)

If  $P_t(x, A)$  is absolutely continuous, we can define the **transition density**  $p_t$

$$P_t(x, A) = \int_A p_t(x, y) dy$$

replace  $\sum_{y \in A}$  by  $\int_A$

and it fulfills the **Chapman Kolmogorov equations**

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

## A note on finite dimensional distributions

Similarly to what happened with CTMCs, we need to say something about time and finite dimensional distributions.

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

This means that for all  $n \in \mathbb{N}$ ,  $0 < t_1 < \dots < t_n$  and  $x_1, \dots, x_n \in \mathbb{R}$ , we can write.

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

However, there is no general solution formula for the CK equations and we have to consider several types of processes separately.

## Example 1: Jump processes

We can think of two extremes for continuous state space MPs. One of them is Gaussian processes, which have **continuous movement**. The opposite extreme is when our MP has **discrete movements**.

Our first example considers the discrete movements case: jump processes.

Jump processes are similar to CTMCs, except that now the state space is continuous.

### Jump processes

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

- a **jump rate density**  $r(x, y) \geq 0$ , and
- a uniformly bounded **total exit rate**  $R(x) = \int_{\mathbb{R}} r(x, y) dy < \bar{R} < \infty$  for all  $x \in \mathbb{R}$ .

In this case, we can simplify the Chapman-Kolmogorov equations...

# Jump processes - Kolmogorov-Feller equation

To try and obtain a better expression for the transition rates, we can try to solve the Chapman-Kolmogorov equations.

- Recall that for DTMCs we obtained a recurrence relation that told us what  $P$  was.
- Similarly, for CTMCs we used it to obtain the forward-backward equations, which gave us  $G$ .

To do this, we make an **ansatz** for the transition function as  $\Delta t \rightarrow 0$ :

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

and plug this into the Chapman Kolmogorov equations.

# Jump processes - Kolmogorov-Feller equation

The Chapman-Kolmogorov equations say:

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

Plugging the previous ansatz, we obtain

$$\begin{aligned} p_{t+\Delta t}(x, y) - p_t(x, y) &= \int_{\mathbb{R}} p_t(x, z) p_{\Delta t}(z, y) dz - p_t(x, y) = \\ &= \int_{\mathbb{R}} p_t(x, z) r(z, y) \Delta t dz + \int_{\mathbb{R}} (1 - R(z) \Delta t - 1) p_t(x, z) \delta(y - z) dz. \end{aligned}$$

This allows us to get the **Kolmogorov-Feller equation** ( $x$  is a fixed initial condition)

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

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

## Example 2: Gaussian processes

The other extreme example (with continuous movement) is that of Gaussian processes, which we define now.

We say that the random variable  $\mathbf{X} = (X_1, \dots, X_n) \sim \mathcal{N}(\mu, \Sigma)$  is a **multivariate Gaussian** in  $\mathbb{R}^n$  if it has the following Probability Density Function (PDF):

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

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

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

$\Sigma$  is symmetric and invertible (unless in degenerate cases with vanishing variance, which we won't look at that much).

Definition: Gaussian process

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

# Some quick notes on Gaussian processes

We will spend some time over the next couple of weeks discussing Gaussian processes (Brownian motion, other diffusion processes), but for now we will just see one of its key features:

## Proposition

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

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

In fact, it is also possible to prove the following:

## Proposition

For any function  $\mu : [0, \infty) \rightarrow \mathbb{R}$  and any non-negative definite function  $C : [0, \infty) \times [0, \infty) \rightarrow \mathbb{R}$ , there exists a Gaussian process  $X_t$  such that

$$\mathbb{E}[X_t] = \mu(t) \quad \text{and} \quad \text{Cov}[X_s, X_t] = C(s, t).$$

# Stationary independent increments and Brownian motion

# Stationary independent increments

The best way to look at continuous-time, continuous state space processes is by considering their increments.

## Definition

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  $0 \leq t_1 < \dots < t_n$

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

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

# Stationary indep. increments and Gaussian Processes

The following is a very useful property:

Proposition:

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  $\sigma(s, t) = \min\{s, t\}$ .

**Note that** stationary independent increments have **stable distributions** such as Gaussian or Poisson.

# Brownian motion

One of the most famous (Gaussian) stochastic processes which you will probably have heard of before is the Brownian motion.

There are many ways that people choose to define it (for an example of an alternative, check 2019 lecture notes in the module resources), and we will use this one:

## Definition

We define a **Standard Brownian motion** (SBM)  $(B_t : t \geq 0)$  to be a real-valued stochastic process such that

- (i)  $B_0 = 0$
- (ii)  $B_t$  is continuous almost surely, i.e.,

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

- (iii)  $B_t - B_s \sim \mathcal{N}(0, t - s)$  for all  $0 \leq s \leq t$
- (iv)  $B_t$  has independent increments, i.e.,  $\forall n \in \mathbb{N}, \forall 0 \leq t_1 < t_2 < \dots < t_n$  we have that  $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$  are independent random variables.

$$B_{t_2} - B_{t_1} \quad B_{t_n} - B_{t_{n-1}}$$

# Brownian motion and Wiener

Wiener proved in 1923 that “the Brownian motion exists”:

Theorem (Wiener, 1923):

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

For this reason, the SBM is often also called a Wiener process, and a lot of books (or me, if I am distracted!) will write  $W_t$  instead of  $B_t$ !

His proof is beyond the scope of this module, but the idea is to construct the process on  $\Omega = \mathbb{R}^{[0, \infty)}$  using **Kolmogorov's extension theorem**.

This shows that for every ‘consistent’ description of finite dimensional distributions (fdds) there exists a ‘canonical’ process  $X_t[\omega] = \omega(t)$  characterised by a law  $\mathbb{P}$  on  $\Omega$ .

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  $\omega$ .

# Properties of Brownian motion (1)

The Brownian motion has several useful properties (which is why it is so widely used!)

First of all, from the definition (and previous proposition), it follows:

- The SBM is a time-homogeneous Gaussian process.
- We have  $m(t) = \mathbb{E}(B_t) = 0$  and

$$\text{Cov}(B_t, B_s) = \mathbb{E}(B_t B_s) = \min(t, s)$$

- For all  $a \leq b$ , we have

$$\mathbb{P}(B_t \in (a, b)) = \frac{1}{2\pi t} \int_a^b \exp\left(-\frac{x^2}{2t}\right) dx$$

**Note that:** the SBM can be seen as the limit of a random walk, and this can be seen from the “functional central limit theorem” (Donsker’s Theorem), which we will not cover.

## Properties of Brownian motion (2)

Some more “advanced” properties are the following:

- For  $\sigma > 0$  and a given number  $\mu$ ,  $\sigma B_t + \mu$  is a (general) BM with  $B_t \sim \mathcal{N}(\mu, \sigma^2 t)$ .

Its 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 transition density is also called the **heat kernel**, since it solves the **heat/diffusion equation**

$$\frac{\partial}{\partial t} p_t(x, y) = \frac{\sigma^2}{2} \frac{\partial^2}{\partial y^2} p_t(x, y) \quad \text{with} \quad p_0(x, y) = \delta(y - x).$$

We will see more about this later!

- The BM has the following scaling properties: If  $B_t$  is a SBM, so is

$$\begin{aligned} X_t &:= B_{t+s} - B_s && \text{with fixed } s > 0, \text{ and} \\ Y_t &:= B_{ct}/\sqrt{c} && \text{with fixed } c > 0. \end{aligned}$$

## Properties of Brownian motion (3)

Finally, some very useful properties:

- The SBM is **self-similar** with **Hurst exponent**  $H = 1/2$ , i.e.

$$(B_{\lambda t} : t \geq 0) \sim \lambda^H (B_t : t \geq 0) \quad \text{for all } \lambda > 0.$$

- It is also **Hölder continuous**: For all  $T > 0$  and  $0 < \alpha < \frac{1}{2}$  there exists a random variable  $C$  such that

$$|B_t - B_s| \leq C|t - s|^\alpha, \quad \forall 0 \leq s, t \leq T.$$

- Most importantly, if we see the SBM as a function,  $t \mapsto B_t$ , it is  $\mathbb{P} - \text{a.s.}$  **not differentiable** at  $t$  for all  $t \geq 0$ !

If it was, then for fixed  $h > 0$  define  $\xi_t^h := (B_{t+h} - B_t)/h \sim \mathcal{N}(0, 1/h)$ , 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  $\xi_t := \lim_{h \rightarrow 0} \xi_t^h$  is called **white noise** and is formally a mean-0 Gaussian process with covariance  $\sigma(s, t) = \delta(t - s)$ .

# Lecture 8: Generators as operators and diffusion processes

Susana Gomes

October 29<sup>th</sup>, 2021

## Plan for today

1. Generators as operators
2. Scaling limits and diffusion processes



# Generators as operators

## Generators as operators

Recall from week 2 that, for a CTMC  $(X_t : t \geq 0)$  with discrete state space  $S$ , we could write an ODE for the distribution at time  $t$ :

$$\frac{d}{dt} \langle \pi_t | = \langle \pi_t | G.$$

Furthermore, we also know that, given a function  $f : S \rightarrow \mathbb{R}$ , we can compute its expectation:

$$\mathbb{E}(f(X_t)) = \sum_{x \in S} \pi_t(x) f(x) = \langle \pi_t | f \rangle.$$

Therefore, we can use this to write an ODE to  $\mathbb{E}(f(X_t))$ :

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

# Generators as operators

We can do the same when  $S = \mathbb{R}$ , and this motivates the definition of the generator as an (differential) operator acting on functions  $f : S \rightarrow \mathbb{R}$ :

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

**Note that:** When we do this, we usually write  $\mathcal{L}$  instead of  $G$  but I will try to be clear when doing that :)

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

$$(\mathcal{L}f)(x) = \int_{\mathbb{R}} r(x, y)[f(y) - f(x)] dy.$$

## Example: Brownian motion

Let us see what the generator is for the **Brownian motion**.

Recall that we mentioned yesterday that the transition density solves the **heat equation**:

$$p_t(x, y) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(y-x)^2}{2\sigma^2 t}\right), \quad \text{solves} \quad \frac{\partial}{\partial t} p_t(x, y) = \frac{\sigma^2}{2} \frac{\partial^2}{\partial y^2} p_t(x, y)$$

Using this, we obtain, for  $f \in C^2(\mathbb{R})$ ,

$$\frac{d}{dt} \mathbb{E}_x(f(X_t)) = \int_{\mathbb{R}} \partial_t p_t(x, y) f(y) dy = \frac{\sigma^2}{2} \int_{\mathbb{R}} \partial_y^2 p_t(x, y) f(y) dy.$$

Now we integrate by parts:

$$\frac{\sigma^2}{2} \int_{\mathbb{R}} \partial_y^2 p_t(x, y) f(y) dy = \frac{\sigma^2}{2} \int_{\mathbb{R}} p_t(x, y) \partial_y^2 f(y) dy = \mathbb{E}_x((\mathcal{L}f)(X_t)).$$

This means that the **generator of BM** is

$$(\mathcal{L}f)(x) = \frac{\sigma^2}{2} \Delta f(x) \quad \left( \text{or } \underbrace{\frac{\sigma^2}{2} f''(x)} \right).$$

# Brownian motion as scaling limit

An interesting consequence of this is that we can see the Brownian motion as the scaling limit of a jump process:

**Proposition:**

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

Then, for all  $T > 0$  the rescaled process  $(\epsilon X_{t/\epsilon^2} : t \in [0, T])$  converges in distribution to a BM with generator  $\mathcal{L} = \frac{1}{2}\sigma^2 \Delta$  for all  $T > 0$  as  $\epsilon \rightarrow 0$ , i.e.

$$(\epsilon X_{t/\epsilon^2} : t \in [0, T]) \longrightarrow (B_t : t \in [0, T]) \quad \text{as } \epsilon \rightarrow 0.$$

**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  $[0, T]$ ).

# Diffusion processes

# Diffusion processes

We can now define a general class of Markov processes.

## Definition

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

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

## Examples.

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

$$(\mathcal{L}f)(x) = -\alpha x f'(x) + \frac{1}{2} \sigma^2 f''(x), \quad \alpha, \sigma^2 > 0.$$

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**.

- **Brownian bridge** is a Gaussian diffusion with  $X_0 = 0$  and generator

$$(\mathcal{L}f)(x) = -\frac{x}{1-t} f'(x) + \frac{1}{2} f''(x).$$

# Time evolution of diffusion processes

Generators are defined on functions  $f$  of the state space. However, they are very useful, as they tell us a lot about the evolution of the underlying probability distributions.

Recall that the generator is given by

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

## Time evolution of the mean:

Use  $\frac{d}{dt} \mathbb{E}[f(X_t)] = \mathbb{E}[(\mathcal{L}f)(X_t)]$  with  $f(x) = x$  to obtain

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

$$f'(x) = 1, \quad f''(x) = 0$$

# Time evolution of diffusion processes

Time evolution of the transition density:

With  $X_0 = x$  we have for  $p_t(x, y)$

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

As before, we can use integration by parts to get the **Fokker-Planck equation**:

$$\boxed{\frac{\partial}{\partial t} p_t(x, y) = -\frac{\partial}{\partial y} (a(y, t)p_t(x, y)) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (\sigma^2(y, t)p_t(x, y)).}$$

$$\begin{aligned} & \int p_t(x, y) \left( a(y, t) f'(y) + \frac{1}{2} \sigma^2(y, t) f''(y) \right) dy \\ &= \int \underbrace{p_t(x, y) a(y, t)}_{= -\partial_y (p_t(x, y) a(y, t))} f'(y) dy + \frac{1}{2} \int p_t(x, y) \sigma^2(y, t) f''(y) dy \\ &= - \int \partial_y (p_t(x, y) a(y, t)) f dy + \frac{1}{2} \int \partial_y^2 (p_t(x, y)) \sigma^2(y, t) f'(y) dy \end{aligned}$$

# Time evolution of diffusion processes

Finally, we can also look at **stationary distributions** for time-independent  $a(y) \in \mathbb{R}$  and  $\sigma^2(y) > 0$ .

A stationary distribution  $p^*$  satisfies  $\frac{\partial p^*}{\partial t} = 0$  and so we have

$$\frac{d}{dy}(a(y)p^*(y)) = \frac{1}{2} \frac{d^2}{dy^2}(\sigma^2(y)p^*(y)).$$

With this, we can solve for a stationary density (modulo normalisation fixing  $p^*(0)$ )

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

Note the need for computing a normalisation constant here - connection to MCMC

# Lectures 9 and 10: Beyond diffusion and intro to SDEs

Susana Gomes

November 4<sup>th</sup> and November 8<sup>th</sup>, 2021  
(remember our swapped lecture with MA930)

# Plan for this week

1. Beyond diffusion
2. Intro to stochastic differential equations

# Beyond diffusion - Lévy processes

We can define processes other than diffusion processes using generators, e.g., processes that combine jumps and diffusion...

## Definition (Lévy process)

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

These processes have generators which have

- a part with **constant drift**  $a \in \mathbb{R}$ ,
- constant **diffusion**  $\sigma^2 \geq 0$ ,
- and a translation invariant **jump part** with density  $q(z)$  that fulfills

$$\int_{|z|>1} q(z)dz < \infty \quad \text{and} \quad \int_{0<|z|<1} z^2 q(z)dz < \infty.$$

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

# Examples of Lévy processes

1. Diffusion processes are Lévy processes. In particular the **Brownian Motion** with  $a = 0$ ,  $\sigma^2 > 0$  and  $q(z) \equiv 0$ .
2. Jump processes are also Lévy processes. An example is the **Poisson process** with  $a = \sigma = 0$  and  $q(z) = \lambda\delta(z - 1)$ .
3. A new example: the process with  $a = \sigma = 0$  and heavy-tailed jump distribution

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

is called  **$\alpha$ -stable symmetric Lévy process** or **Lévy flight**.

The Lévy flight is **self-similar**:

$$(X_{\lambda t} : t \geq 0) \sim \lambda^H (X_t : t \geq 0), \quad \lambda > 0 \quad \text{with } H = 1/\alpha$$

and exhibits something we call **super-diffusive behaviour** with  $\mathbb{E}[X_t^2] \propto t^{2/\alpha}$ .

This is an example of a **Markov process which is not Gaussian**.

# Beyond diffusion - anomalous diffusion

In general, we say that a process  $(X_t : t \geq 0)$  exhibits **anomalous diffusion** if

$$\frac{\text{Var}[X_t]}{t} \rightarrow \begin{cases} 0, & \text{(sub-diffusive)} \\ \infty, & \text{(super-diffusive)} \end{cases} \quad \text{as } t \rightarrow \infty.$$

This leads us to introduce a process in another extreme: one that is Gaussian but **not Markov**.

## Definition (fractional Brownian motion)

A **fractional Brownian motion** (fBM)  $(B_t^H : t \geq 0)$  with **Hurst index**  $H \in (0, 1)$  is a **mean-zero Gaussian process** with continuous paths,  $B_0^H = 0$  and covariances given by

$$\mathbb{E}(B_t^H B_s^H) = \min(t, s) \quad (H = \frac{1}{2})$$

$$\mathbb{E}(B_t^H B_s^H) = \frac{1}{2} (t^{2H} + s^{2H} - |t-s|^{2H}) \quad \text{for all } s, t \geq 0.$$

$$H = \frac{1}{2}$$

$$\begin{aligned} \frac{1}{2} (t+s - |t-s|) &= \frac{1}{2} (t+s - (t-s)) \\ &= \frac{1}{2} (2s) = s \end{aligned}$$

# Fractional Brownian Motion

Some properties of fBM:

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

$$B_t^H - B_s^H \sim B_{t-s}^H \sim \mathcal{N}\left(0, (t-s)^{2H}\right).$$

For  $H \neq 1/2$ , these increments are **not** independent and the process is **non-Markov**.

- The fBM is **self-similar**, i.e.

$$(B_{\lambda t}^H : t \geq 0) \sim \lambda^H (B_t^H : t \geq 0) \quad \text{for all } \lambda > 0.$$

- The fBM exhibits **anomalous diffusion** with  $\text{Var}[B_t^H] = t^{2H}$ . If
  - ★  $H > 1/2$ , it is super-diffusive with positively correlated increments.
  - ★  $H < 1/2$  it is sub-diffusive with negatively correlated increments.

$$\mathbb{E}_{\color{red}t} [B_t^H (B_{t+1}^H - B_t^H)] = \frac{(t+1)^{2H} - 2t^{2H} + (t-1)^{2H}}{2} \underset{t \rightarrow \infty}{\simeq} H(2H-1)t^{2(H-1)}$$

# Spectral densities and noise

For a **stationary process** ( $X_t : t \geq 0$ ) we define **autocorrelation function**

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

The Fourier transform of this function is called the **spectral density**

$$S(\omega) := \int_{\mathbb{R}} c(t) e^{-i\omega t} dt.$$

We can use this to describe noise:

- **White noise** ( $\xi_t : t \geq 0$ ), is a stationary GP with mean zero and

$$c(t) = \delta(t) \quad \Rightarrow \quad S(\omega) \equiv 1.$$

- **fractional noise** ( $\xi_t^H : t \geq 0$ ), is a stationary GP formally defined as the “derivative” of the fractional BM. It has mean zero and

$$c(t) = \frac{H(2H-1)}{|t|^{2(1-H)}} \quad \Rightarrow \quad S(\omega) \propto \frac{1}{|\omega|^{2H-1}}$$

- If  $H \rightarrow 1$ ,  $S(\omega) \propto \frac{1}{\omega}$  and we call this **1/f-noise** or “**pink noise**”. Similarly, if  $H \rightarrow 0$  we have  $S(\omega) \propto \omega$  and we have “**blue noise**”.



# Introduction to Stochastic Differential Equations

# SDEs and some definitions

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 as described before, and we interpret it in its integrated form.

To understand why, use our intuition from ODEs, and “conclude” that the solution is given by

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

where  $X_0$  is the initial condition (which can be deterministic or random).

The **problem** here is that we have an integral that we don't know how to compute:

$$\mathcal{I} = \int_0^t \sigma(X_s, s)dB_s,$$

which is a **stochastic integral!**

# The stochastic integral

The problem with the stochastic integral  $\mathcal{I} = \int_0^t f(X_s, s) dB_s$  is that we are trying to integrate a stochastic process  $X_t$  or a function of a stochastic process  $f(X_t)$  **with respect to another stochastic process**.

This means that the stochastic integral  $\mathcal{I}$  **is a random variable**! So... How do we compute it?

Let's think about the definition of Riemann integral.

We discretise the interval:

and we define

$$\mathcal{I}(t) := \lim_{K \rightarrow \infty} \sum_{k=1}^K f(\tau_k) (B_{t_k} - B_{t_{k-1}}).$$

The important part now is that the definition of the stochastic integral depends on our choice of  $\tau_k$ !

# Why is this a problem?

$$\frac{t_K - t_{K-1}}{\Delta t} = \frac{\tau_K}{\Delta t} = T$$

Recall that  $B_t$  is a sBM, so we know that  $B_{t_k} - B_{t_{k-1}}$  are increments of a BM and therefore they are independent and  $B_{t_k} - B_{t_{k-1}} \sim \mathcal{N}(0, \Delta t)$ .

This sort of makes sense to compute the integral. However, we would expect the limit to be **independent of the chosen  $\tau_k$** . This will **not** be the case for us.

**Example:** Let's try to compute the integral  $I = \int_0^T B_t dB_t$ .

$$I = \sum B_{t_k} (B_{t_k} - B_{t_{k-1}}) \quad t_k \in [t_{k-1}, t_k]$$

Case 1: choose  $t_k = t_{k-1}$  (left)

$$I^{(l)} = \sum B_{t_{k-1}} (B_{t_k} - B_{t_{k-1}}) \Rightarrow \mathbb{E}(I^{(l)}) = \sum \mathbb{E}(B_{t_{k-1}} (B_{t_k} - B_{t_{k-1}})) = 0$$

Case 2: choose  $t_k = t_k$  (right)

$$I^{(r)} = \sum B_{t_k} (B_{t_k} - B_{t_{k-1}}) \Rightarrow \mathbb{E}(I^{(r)}) = \sum \mathbb{E}(B_{t_k} (B_{t_k} - B_{t_{k-1}}))$$

$$= \sum \mathbb{E}((B_{t_k} - B_{t_{k-1}} + B_{t_{k-1}})(B_{t_k} - B_{t_{k-1}}))$$

$$= \sum \mathbb{E}((B_{t_k} - B_{t_{k-1}})^2 + \underbrace{B_{t_{k-1}}(B_{t_k} - B_{t_{k-1}})}_{=0}) = \sum \Delta t = T$$

## Various definitions of stochastic integral

$$X_t = f(x(t)) \quad \frac{df}{dt} = f'(x(t)) \frac{dx}{dt}$$

This happens because the BM is a.s. non-differentiable; and this means it "varies too much" in the interval  $[t_{k-1}, t_k]$ .

**Note that:** in "normal" integrals  $\int f(x) dg(x)$  it was required that  $g(x)$  had bounded total variation  $\rightarrow$  this is what fails here.

There is no way around this problem. So we always need to specify our choice of  $\tau_k$  when computing stochastic integrals. The most popular choices are:

- $\tau_k = t_{k-1}$   $\rightarrow$  **Itô stochastic integral**  
commonly used in finance and biology
- $\tau_k = \frac{t_k + t_{k-1}}{2}$   $\rightarrow$  **Stratonovich stochastic integral**  
mostly used in physics and engineering
- $\tau_k = t_k$   $\rightarrow$  **Klimontovich stochastic integral**  
commonly used in statistical physics

$$\text{Itô's formula } \frac{df}{dt}(x(t)) = f'(x(t)) \frac{dx}{dt} + df(x_t) =$$

$$dx_t = \underbrace{a(x_t, t) dt}_{\text{drift}} + \underbrace{\sigma(x_t, t) dB_t}_{\text{diffusion}} \quad df(x_t) = f'(x_t) \cdot dx_t$$

In this module, we will only use the **Itô interpretation**. This is because it has a lot of nice properties that you would expect of an integral.

However, it doesn't have a very important property: **the chain rule does not hold**. To overcome this, we use one of the most important results in stochastic calculus...

### Theorem (Itô's formula for diffusions)

Let  $(X_t : t \geq 0)$  be a diffusion with generator  $\mathcal{L}$  and  $f : \mathbb{R} \rightarrow \mathbb{R}$  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

$$df(X_t) = \underbrace{a(X_t, t)f'(X_t)dt}_{\text{drift}} + \underbrace{\frac{1}{2}\sigma^2(X_t, t)f''(X_t)dt}_{\text{diffusion}} + \underbrace{\sigma(X_t, t)f'(X_t)dB_t}_{\text{noise}}$$

*correction term!*

# Back to SDEs

$$Y_t = f(X_t)$$

$$\hookrightarrow dY_t$$

$$X_t = f^{-1}(Y_t)$$

Recall we are looking into SDEs of the form

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

Suppose we want to change variables to some r.v.  $Y_t = f(X_t)$  for some nice invertible function  $f \in C^2$ . Itô's formula for diffusions implies the following.

**Proposition:**

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

**drift**  $a(x, t)f'(x) + \frac{1}{2}\sigma^2(x, t)f''(x)$  and **diffusion**  $\sigma(x, t)f'(x)$ ,

i.e., it solves the SDE

$$\begin{aligned} dY_t &= (a(X_t, t)f'(X_t) + \frac{1}{2}\sigma^2(X_t, t)f''(X_t)) dt + \sigma(X_t, t)f'(X_t) dB_t \\ &= f'(X_t)(a(X_t, t) dt + \sigma(X_t, t) dB_t) + f''(X_t)\frac{1}{2}\sigma^2(X_t, t) dt \end{aligned}$$

Correction

## Example - geometric Brownian motion

We saw the geometric random walk in one of our problem sheets, and a similar process is the gBM: Let  $Y_t := e^{\theta B_t}$ , so that  $Y_t = f(X_t)$  with  $f(x) = e^{\theta x}$ .

Since  $f'(x) = \theta f(x)$  and  $f''(x) = \theta^2 f(x)$  and the sBM is a diffusion with  $a \equiv 0$ ,  $\sigma^2 \equiv 1$ , for  $\theta \in \mathbb{R}$  we have that  $(Y_t : t \geq 0)$  is a diffusion process with SDE

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

Alternatively, suppose  $X_t$  is a gBM and we start with its SDE

$$dX_t = \theta X_t dt + \sigma X_t dB_t \quad \longrightarrow \quad \frac{dX_t}{X_t} = \theta dt + \sigma dB_t.$$

So... Define  $Y_t = f(X_t)$  with  $f(x) = \ln(x)$ . We have

$$f'(x) = \frac{1}{x}, \quad \text{and} \quad f''(x) = -\frac{1}{2x^2},$$

and we can use Itô's formula.

# gBM continued

Applying Itô's formula, we have

$$\begin{aligned} dY_t &= d(\ln X_t) = \left( \frac{1}{X_t} \theta X_t - \frac{1}{2X_t^2} \sigma^2 X_t^2 \right) dt + \frac{1}{X_t} \sigma X_t dB_t \\ &= \left( \theta - \frac{\sigma^2}{2} \right) dt + \sigma dB_t. \end{aligned}$$

This is something we can integrate!

$$\ln \left( \frac{X_t}{X_0} \right) = \left( \theta - \frac{\sigma^2}{2} \right) t + \sigma B_t,$$

... and this gives

$$X_t = X_0 \exp \left( \left( \theta - \frac{\sigma^2}{2} \right) t + \sigma B_t \right).$$

## A note on gBM

We can show that the law of the gBM is a log-normal (like with the gRW) with mean  $\mathbb{E}(X_t) = X_0 e^{\theta t}$  and variance  $\text{Var}(X_t) = X_0^2 e^{2\theta t} (e^{\sigma^2 t} - 1)$ .

The gBM is the most widely used model for stock price behaviour in mathematical finance.

## Example - Ornstein-Uhlenbeck process

The OU process is a diffusion process which solves the SDE

$$dX_t = -\alpha X_t dt + \sigma dB_t, \quad X_0 = x, \quad \alpha, \sigma > 0.$$

We can solve this SDE analytically by considering the ODE analogue:

$$\frac{dx}{dt} = -\alpha x + f(t) \quad \Rightarrow x(t) = e^{-\alpha t} x(0) + \int_0^t e^{-\alpha(t-s)} f(s) ds.$$

So, for the OU process, we obtain

$$X_t = e^{-\alpha t} X_0 + \int_0^t e^{-\alpha(t-s)} dW_s,$$

which we can also check using Itô's formula.

## More on the OU process

The OU process also has some nice properties. For example, if  $X_0 = x$  is deterministic, then  $X_t \sim \mathcal{N}\left(e^{-\alpha t}x, \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})\right)$ , i.e.

$$\mathbb{E}(X_t) = e^{-\alpha t}x \quad \text{and} \quad \text{Var}(X_t) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})$$

$$\mathbb{E}\left(\int_0^T f dW_t\right)^2 = \mathbb{E}\left(\int_0^T f^2(t) dt\right). \quad \text{Use Itô's isometry}$$

- The OU process describes the movement of a Brownian particle moving within a fluid with random "kicks" due to friction with other particles.
- It is a mean-reverting process (the mean acts as an equilibrium state)
- In mathematical finance, it is used to model interest rates and currency exchange rates.

## Other examples

Other examples of SDEs that appear in applications include

- Cox-Ingersoll-Ross model for interest rates

$$dX_t = \alpha(\beta - X_t) dt + \sigma \sqrt{X_t} dB_t,$$

- Stochastic Verhulst equation for population dynamics

$$dX_t = (\lambda X_t - X_t^2) dt + \sigma X_t dB_t$$

- Langevin equation (similar to OU with a particle also having potential energy)

$$\begin{cases} dQ_t &= P_t dt \\ dP_t &= (-\lambda P_t - V'(Q_t)) dt + \sigma dB_t \end{cases}$$

... and many others.

# A note on solving SDEs numerically

We can't always solve SDEs analytically, so we must sometimes revert to numerical techniques.

The most commonly used numerical integration technique for SDEs is the **Euler-Maruyama scheme**.

Using the Markov property of diffusions, we can assume that  $a(X_t)$  and  $\sigma(X_t)$  don't change too much in a small time interval, so we can write, for  $[t_n, t_{n+1}]$  with  $t_{n+1} - t_n = \Delta t$

$$\begin{aligned} X_{t_{n+1}} &= X_{t_n} + \int_{t_n}^{t_{n+1}} a(X_s) \, ds + \int_{t_n}^{t_{n+1}} \sigma(X_s) \, dB_s \\ &\approx X_{t_n} + a(X_{t_n}) \int_{t_n}^{t_{n+1}} \, ds + \sigma(X_{t_n}) \int_{t_n}^{t_{n+1}} \, dB_s \end{aligned}$$

If we let  $X_n = X_{t_n}$ , this gives

$$X_{n+1} = X_n + a(X_n) \Delta t + \sigma(X_n) \Delta B_n,$$

with  $\Delta B_n = B_{t_{n+1}} - B_{t_n} \sim \mathcal{N}(0, \Delta t)$ .

## A little more on numerical solution of SDEs

The Euler-Maruyama scheme is often the best we can do, especially with constant  $\sigma$ . For example, it is possible to show it converges (in some sense) to the right process with an optimal rate.

However, an alternative (for non-constant  $\sigma$ ) is the **Milstein scheme**, which improves on the approximation

$$\int_{t_n}^{t_{n+1}} \sigma(X_s) dB_s \approx \sigma(X_n) \int_{t_n}^{t_{n+1}} dB_s.$$

This scheme gives

$$X_{n+1} = X_n + a(X_n) \Delta t + \sigma(X_n) \Delta B_n + (\sigma' \sigma)(X_n) ((\Delta B_n)^2 - \Delta t).$$

We will not discuss numerical solution of SDEs in this module, but if you are interested, see this **paper by Des Higham** or ask me to borrow his book :)