WARWICK
THE UNIVERSITY OF WARWICK

Lecture 4: Introduction to CTMCs

Susana Gomes

Last updated October 1, 2023

Plan for today

1. Basics of Continuous time Markov chains
2. Chapman-Kolmogorov equations, generators and the Master equation.
3. Stationary and irreducible distributions for CTMCs
4. Sample paths
5. Some examples
6. Ergodicity
7. An application: MCMC

Continuous time Markov chains

Continuous-time Markov chains

In this part of the module, we'll mostly spend our time 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 $A \subset S$, $n \in \mathbb{N}$, $t_1 < \ldots < t_{n+1} \in [0, \infty)$ and $s_1, \ldots, s_n \in S$, we have

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

- A Markov process (MP) is called **homogeneous** if for all $A \subset S$, $t, u > 0$ 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 Ω of a CTMC is the space of **right-continuous paths**

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

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

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

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 $P_t = \big(p_t(x,y) : x, y \in S\big)$ 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**

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

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 π_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 λ_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 $G = \big(g(x,y) : x, y \in S\big)$,

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

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 $\mathrm{Re}(\lambda_i) < 0$ for the eigenvalues of G, so there are **no persistent oscillations for CTMCs**.

Chapman-Kolmogorov equations for CTMCs

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

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

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

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:

Generator of a CTMC

As before, we can write this in matrix notation:

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

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

Generator of a CTMC

As before, we can write this in matrix notation:

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

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

In particular,

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

Generator of a CTMC

As before, we can write this in matrix notation:

We define $P_t = \big(p_t(x,y) : x, y \in S\big)$ 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**

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

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

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

Forward and backward equations for CTMCs

The solution for these equations is given by the matrix exponential

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

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

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

Forward and backward equations for CTMCs

The solution for these equations is given by the matrix exponential

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

And from this we can obtain the **distribution π_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, $\lambda_1, \ldots, \lambda_L \in \mathbb{C}$. Then, P_t **has eigenvalues** $\exp(t\lambda_i)$ with the same eigenvectors $\langle v_i |$, $| u_i \rangle$.

Forward and backward equations for CTMCs

The solution for these equations is given by the matrix exponential

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

And from this we can obtain the **distribution π_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, $\lambda_1, \ldots, \lambda_L \in \mathbb{C}$. Then, P_t **has eigenvalues** $\exp(t\lambda_i)$ with the same eigenvectors $\langle v_i |$, $| u_i \rangle$.

If the λ_i are distinct, we can still expand the initial condition in the eigenvector basis

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

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

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

Transition rates

Using the expression for P_t we have, for $G = \big(g(x,y) : x, y \in S\big)$,

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

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

Transition rates

Using the expression for P_t we have, for $G = \big(g(x,y) : x, y \in S\big)$,

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

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

We also have

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

Transition rates

Using the expression for P_t we have, for $G = \big(g(x,y) : x, y \in S\big)$,

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

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**

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

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 $\mathrm{Re}(\lambda_i) < 0$ 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 $y \in S$

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

π 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.
- $\langle \pi | G = \langle 0 | \quad \text{implies} \quad \langle \pi | P_t = \langle \pi | \big(\mathbb{I} + \sum_{k \geq 1} t^k G^k / k!\big) = \langle \pi | \quad \text{for all } t \geq 0$

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

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

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

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

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., $W_x \sim \mathrm{Exp}(|g(x,x)|)$.
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

Sample paths (jump times)

Once we have the holding times defined, we can define **jump times**: $J_0, J_1, \ldots$.
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 τ with $X_\tau = i$ then $X_{\tau+t}$ 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**:

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

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

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 Δt is proportional to Δt: $p = \lambda \Delta t$.
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 ve valid, we need to think in an “infinitesimal" sense, i.e., we should make sure that

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

Poisson Process

A **Poisson process** with **rate** λ (short $\mathrm{PP}(\lambda)$) 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 $\mathrm{PP}(\lambda)$ 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** α_x and **death rates** β_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/∞ 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 π and

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

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

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

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 π. Then for every bounded function $f : S \to \mathbb{R}$ we have with probability 1

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

For a proof 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 π 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**:

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

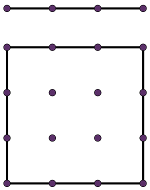
- In statistical mechanics, compute **Gibbs measures**:

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

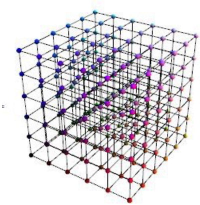
$$Z(\beta) = \sum_{x \in S} e^{-\beta H(x)}.$$

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 π is its stationary ditribution.
- 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* π. 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 π 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 π 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

Other examples

- **Independence sampler:** If $q(x,y) = q(y)$ (independent of the current state)
$$a(x,y) = \min\left\{1, \frac{\pi(y)q(x)}{\pi(x)q(y)}\right\}.$$

- **Random walk MH:** when $q(x,y) = q(y,x)$,
$$a(x,y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}.$$

- **Langevin proposals:** come from solving an SDE.

- **and others...** Gibbs sampling, simulated annealing, ...

and many others. Ask me if you are interested!