Why introduce Markov chains?

Our original motivation was **sampling**: given a target distribution, we want a method for generating samples from it.

A stochastic process is a collection of random variables indexed by time:

$$X_1, X_2, X_3, \ldots$$

The Markov property

Given the current state, the future does not depend on the earlier history.

How can this property help us construct a sampling method?

Recap: convergence to a stationary distribution

Under suitable conditions, the marginal distribution of a Markov chain converges to a stationary distribution.

In the earlier classroom example, two different initial conditions gave distributions that were already close after about five iterations.

The terms *stationary*, *invariant*, and *equilibrium* distribution refer to the same invariant probability distribution here.

Mathematical clarification

The five iterations refer to that particular example. They are not a general convergence guarantee.

The target distribution

The sampling task

Given a target distribution π, generate samples from it.

The target can be discrete or continuous. A Markov chain can also have a discrete or continuous state space.

The earlier discrete examples were chosen for simplicity.

Mathematical clarification

Continuous state space does not mean continuous time. Here the iterations still have indices $t = 0, 1, 2, \ldots$. Convergence conditions depend on the setting.

Markov chain Monte Carlo (MCMC) uses the stationary distribution of a chain.

1. Construct a Markov chain whose stationary distribution is the target π.
2. Ensure that the chain has the required convergence properties.
3. Simulate the chain for sufficiently long.

The distribution of a late state is then close to the target distribution.

How long is sufficiently long?

The chain may need many steps to approach its stationary distribution.

The lecture uses **100 steps** as an illustration: if this is sufficiently long for the particular chain, the state then has a distribution close to π.

Mathematical clarification

There is no universal number of steps that guarantees convergence. At a finite time, the distribution is generally only approximately the target, unless the chain is already in stationarity.

Having a stationary distribution alone does not guarantee convergence to it.

Simulating the chain on a computer

Starting from an initial state, simulate the next state, then the next:

$$X_0, X_1, X_2, X_3, \ldots$$

Under suitable convergence conditions, later states provide approximate samples from the target distribution.

Determining how long to run the simulation is a difficult question, left aside for now.

Mathematical clarification

Successive MCMC states are generally dependent, even in stationarity. They are not generally independent target draws.

The main construction problem

How can we construct a chain with the desired target distribution?

We need both:

- the required long-run convergence behaviour,
- the particular target π as a stationary distribution.

Detailed balance is a condition that helps establish the second requirement.

Detailed balance: the discrete setup

Let the target distribution be

$$\pi = (\pi_1, \pi_2, \ldots),$$

where π_i is the probability assigned to state x_i. Let $T = (a_{ij})$ be a transition matrix, with

$$a_{ij} = \mathbb{P}(X_{t+1} = x_j \mid X_t = x_i).$$

Thus a_{ij} describes a transition from state x_i to state x_j.

An analogous balance condition exists for continuous state spaces.

Detailed balance with respect to π

For every pair of states i, j,

$$\boxed{\pi_i a_{ij} = \pi_j a_{ji}.}$$

This condition relates the transition probabilities to the specified distribution π.

It must hold for **every pair**, using the same target distribution.

If detailed balance holds, then π is stationary:

$$\boxed{\pi T = \pi.}$$

Under additional convergence conditions, the marginal distribution approaches π.

Mathematical clarification

Detailed balance alone does not guarantee convergence. For a finite-state chain, irreducibility and aperiodicity are standard sufficient conditions for convergence to a unique stationary distribution from every initial distribution.

Suppose the current distribution is π. Then

$$\mathbb{P}(X_t = x_i, X_{t+1} = x_j) = \pi_i a_{ij},$$
$$\mathbb{P}(X_t = x_j, X_{t+1} = x_i) = \pi_j a_{ji}.$$

Detailed balance makes these two joint probabilities equal.

In stationarity, the forward and reverse two-state transitions have the same probability.

Mathematical clarification

The weighted flows are equal. The condition does not generally require $a_{ij} = a_{ji}$.

Detailed balance as a mathematical tool

Detailed balance has an important interpretation in physical models and in time reversal.

Here, the focus is its use in constructing a sampling algorithm. The condition to remember is

$$\pi_i a_{ij} = \pi_j a_{ji}.$$

We use a Markov chain as a mathematical tool for generating samples.

Designing a sampler for the target

Choose the target π first. Construct a transition rule that satisfies detailed balance with respect to **that target**:

$$\pi_i a_{ij} = \pi_j a_{ji}.$$

Then π is stationary for the constructed chain.

If the appropriate convergence conditions also hold, the chain approaches the desired target distribution.

The mathematical foundation of MCMC

The purpose of the mathematics is to explain why the algorithm works.

1. Construct a suitable chain with target stationary distribution π.
2. Simulate that chain on a computer.
3. Use its later states as approximate target samples.

We now move from this construction principle to a specific algorithm.

Metropolis–Hastings is a widely used procedure for constructing a chain with the required stationary distribution.

Design principle

Its acceptance rule ensures detailed balance with respect to the target π.

The lecture states this property without proving it, and focuses on the steps of the algorithm.

Choose an initial point x_0. Then iterate for

$$t = 0, 1, 2, \ldots$$

Here x_t denotes the realised state at iteration t, and $x^\star$ denotes a proposed state.

Write $q(y \mid x)$ for the proposal density or probability mass function.

Mathematical clarification

A straightforward safe initial choice satisfies $\pi(x_0) > 0$. Initialisation can affect convergence speed, and the proposal must allow adequate exploration of the target.

A proposal distribution

As in rejection sampling, we need a proposal distribution that we can sample from.

For Metropolis–Hastings, it is written conditionally on the current state:

$$q(\cdot \mid x_t).$$

It can suggest different candidate points depending on where the chain currently is.

Mathematical clarification

A proposal can also be independent of the current state: $q(y \mid x) = q(y)$. Actual state dependence is not required.

Step 1: draw a proposed state

At iteration t, sample

$$\boxed{x^\star \sim q(\cdot \mid x_t).}$$

Condition on the **current state** when generating the proposal.

At the first iteration, $t = 0$, this becomes

$$x^\star \sim q(\cdot \mid x_0).$$

The proposed point is a candidate for the next state.

Reconstructed equation

The captions refer to a displayed formula but do not record it. The standard Metropolis–Hastings acceptance probability is

$$A = \min\left\{1, \frac{\pi(x^\star)q(x_t \mid x^\star)}{\pi(x_t)q(x^\star \mid x_t)}\right\}.$$

π is the target density or probability mass function, and q is the proposal density or probability mass function.

The ratio is used where the denominator is positive. A zero reverse-proposal term gives $A = 0$.

Why A is a valid probability

All density or probability terms in the ratio are nonnegative.

Taking the minimum with 1 prevents the result from exceeding 1.

$$\boxed{0 \leq A \leq 1.}$$

Thus A can be used as the probability of accepting the proposal.

Step 3: draw a uniform random number

Draw

$$\boxed{U \sim \text{Uniform}(0, 1).}$$

The uniform draw will implement the acceptance decision by comparison with A.

Mathematical clarification

Use a fresh uniform variable, independent of the proposal and the previous random draws.

Step 4: accept or reject

Compare U with the acceptance probability A:

$$x_{t+1} = \begin{cases} x^\star, & U < A, \\ x_t, & U \geq A. \end{cases}$$

If the proposal is accepted, use the proposed state. If it is rejected, keep the current state.

Mathematical clarification

Rejection still counts as an iteration. Record the current state again as x_{t+1}.

Why this particular acceptance rule?

The proposal and acceptance steps together define the transition rule of the Markov chain.

The Metropolis–Hastings rule ensures detailed balance with respect to the target distribution.

The consequence

The target π is stationary for the resulting chain. Convergence still requires the appropriate additional conditions.

The proof is not included in this lecture.

The four steps of each iteration

Choose a proposal distribution and an initial state x_0.
For $t = 0, 1, 2, \ldots$:

1. Draw $x^\star \sim q(\,\cdot \mid x_t)$.
2. Compute
$$A = \min\left\{1, \frac{\pi(x^\star) q(x_t \mid x^\star)}{\pi(x_t) q(x^\star \mid x_t)}\right\}.$$
3. Draw $U \sim \text{Uniform}(0, 1)$ independently.
4. If $U < A$, set $x_{t+1} = x^\star$. Otherwise, set $x_{t+1} = x_t$.

Generating the chain

Starting from x_0:

- the iteration at $t = 0$ produces x_1,
- the iteration at $t = 1$ produces x_2,
- the next iteration produces x_3, and so on.

The resulting sequence is a Markov chain with stationary distribution π.

Under suitable convergence conditions, the marginal distribution approaches the target.

Intuition: propose a move, then decide

Suppose the chain is currently at x_t.

1. Propose a candidate $x^\star$: "Why not move to this point?"
2. Compute the probability A of accepting that proposal.
3. Carry out the random decision using U.

Given the current state and proposal, accept with probability A and reject with probability $1 - A$.

Why the uniform comparison works

For a fixed $A \in [0,1]$, a uniform draw satisfies

$$\mathbb{P}(U < A) = A, \qquad \mathbb{P}(U \geq A) = 1 - A.$$

Therefore the comparison implements exactly the required acceptance probability.

On rejection, the chain stays at x_t.

Mathematical clarification

These probabilities are conditional on the current state and proposal. If the proposal equals the current state, accepting it still leaves the location unchanged.

There is randomness in the proposal and in the acceptance decision.

Acceptance probability $A = 0.99$

The proposal is accepted with probability 99%. There is still a 1% chance of rejection, in which case the chain stays at the current state.

The lecture describes this movement informally as a "random walk."

Mathematical clarification

General Metropolis–Hastings need not be random-walk Metropolis. That term usually refers to a proposal formed by adding a random increment to the current state.