Why move beyond direct sampling?

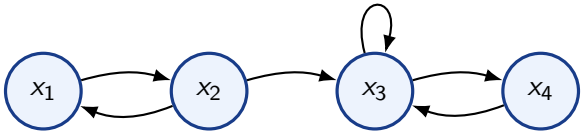
- Inverse transform sampling needs a tractable quantile function.
- Rejection sampling needs an easy proposal q and a practical envelope cq.
- In multiple dimensions, these requirements can become difficult.

MCMC idea

Construct $X_0, X_1, \ldots$ so that, under suitable conditions, the law of X_t approaches a target distribution π.

The samples are generally dependent; that is not a mistake, but part of the method.

A motivating picture: moving between states



- A **state** describes where the process is currently located.
- A **transition rule** describes how the next state is chosen.
- We want the chain to spend time in states according to the target π.

Following the notes, first consider a **finite state space**:

$$\mathcal{S} = \{x_1, x_2, \ldots, x_s\}.$$

A discrete-time stochastic process is a sequence of random variables

$$X_0, X_1, \ldots, X_t, \ldots, \qquad X_t \in \mathcal{S}.$$

Examples

- Position of a traveller on one of 10 islands.
- Weather state: sunny, cloudy, or rainy.
- A parameter location on a finite grid.

The same principles extend to continuous-state MCMC, but finite states make the matrix calculations transparent.

Definition: the Markov property

Markov chain

A process is Markov if, for every admissible history,

$$\mathbb{P}(X_t = x_j \mid X_{t-1}, \ldots, X_0) = \mathbb{P}(X_t = x_j \mid X_{t-1}).$$

- Once we know the **present state**, the earlier trajectory adds no information about the next state.
- The next state can still be random.
- The chain is generally not a sequence of independent observations.

Markov dependence is not independence

Independent sequence

$\mathbb{P}(X_t \in A \mid X_{t-1}, \ldots) = \mathbb{P}(X_t \in A)$. The next draw ignores the current state.

Markov sequence

$\mathbb{P}(X_t \in A \mid X_{t-1}, \ldots) = \mathbb{P}(X_t \in A \mid X_{t-1})$. The next draw may depend strongly on the current state.

Teaching point

Independence is a special case of the Markov property, not the definition of a Markov chain.

The Markov property does not require the transition rule to stay fixed over time.

Homogeneity

The chain is **homogeneous** when

$$\boxed{\mathbb{P}(X_t = x_j \mid X_{t-1} = x_i) = T_{ij}}$$

does *not* depend on t.

- One fixed transition matrix T describes every step.
- This is the setting used for the rest of Section 7.1.

One-step transition probabilities

For $\mathcal{S} = \{x_1, \ldots, x_s\}$, define

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

- i indexes the **current** state (matrix row).
- j indexes the **next** state (matrix column).
- $T_{ij} \geq 0$ for all i, j.

Row-sum condition

$$\sum_{j=1}^{s} T_{ij} = 1 \qquad \text{for each } i.$$

The probabilities are organised into the matrix

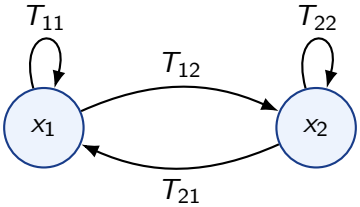
$$T = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1s} \\ T_{21} & T_{22} & \cdots & T_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ T_{s1} & T_{s2} & \cdots & T_{ss} \end{pmatrix}.$$

Every entry is nonnegative and every row sums to one.

Interpretation

Row i contains the full conditional probability distribution of X_t given $X_{t-1} = x_i$.

How to read a transition diagram



- An arrow $x_i \to x_j$ indicates $T_{ij} > 0$.
- A self-loop $x_i \to x_i$ means the chain may remain in the same state.
- The outgoing arrow probabilities must sum to one.

Let $\pi_{t-1}(x_i) = \mathbb{P}(X_{t-1} = x_i)$. Then

$$\boxed{\pi_t(x_j) = \sum_{i=1}^{s} \pi_{t-1}(x_i)\, T_{ij}.} \tag{7.1}$$

Meaning

To find the probability of arriving at x_j, add the contributions from *every possible previous state* x_i.

Each contribution is "probability of being at x_i" $\times$ "probability of moving from x_i to x_j".

Partition according to the previous state:

$$\begin{aligned}\pi_t(x_j) &= \mathbb{P}(X_t = x_j) \\ &= \sum_{i=1}^{s} \mathbb{P}(X_{t-1} = x_i,\, X_t = x_j) \\ &= \sum_{i=1}^{s} \mathbb{P}(X_{t-1} = x_i)\mathbb{P}(X_t = x_j \mid X_{t-1} = x_i) \\ &= \sum_{i=1}^{s} \pi_{t-1}(x_i)\, T_{ij}.\end{aligned}$$

Key observation

No assumption of independent consecutive states is made.

Matrix form: $\pi_t = \pi_{t-1} T$

Represent the distribution as a **row vector**:

$$\boldsymbol{\pi}_t = \big(\pi_t(x_1), \ldots, \pi_t(x_s)\big).$$

Then Eq. (7.1) is simply

$$\boxed{\boldsymbol{\pi}_t = \boldsymbol{\pi}_{t-1} T.}$$

- The row-vector convention matches the lecture notes.
- Each component of $\boldsymbol{\pi}_t$ is nonnegative.
- Its entries sum to 1 because T is row-stochastic.

Repeated application yields

$$\pi_1 = \pi_0 T,$$
$$\pi_2 = (\pi_0 T) T = \pi_0 T^2,$$
$$\pi_t = \boxed{\pi_0 T^t}.$$

Likewise, the (i, j)-entry of T^t is the ***t*-step transition probability**:

$$(T^t)_{ij} = \mathbb{P}(X_t = x_j \mid X_0 = x_i).$$

This is the finite-state version of propagating probability distributions forward in time.

Worked example from the notes: three states

The notes give the transition matrix

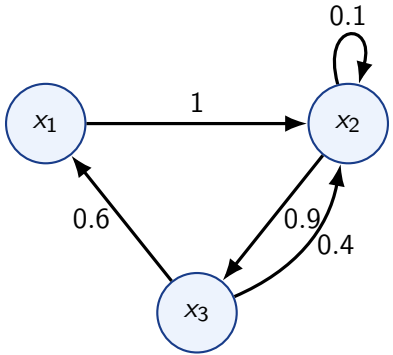
$$T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{pmatrix}.$$

- From x_1, move to x_2 with probability 1.
- From x_2, stay with probability 0.1, or move to x_3 with 0.9.
- From x_3, move to x_1 with 0.6 or x_2 with 0.4.

Quick check

Each row sums to one: 1, $0.1 + 0.9 = 1$, $0.6 + 0.4 = 1$.

The three-state example as a graph



All arrows and labels are taken from the transition matrix in Section 7.1.

Start at state x_1 with certainty:

$$\boldsymbol{\pi}_0 = (1, 0, 0).$$

Then

$$\boldsymbol{\pi}_1 = \boldsymbol{\pi}_0 T = (0, 1, 0),$$

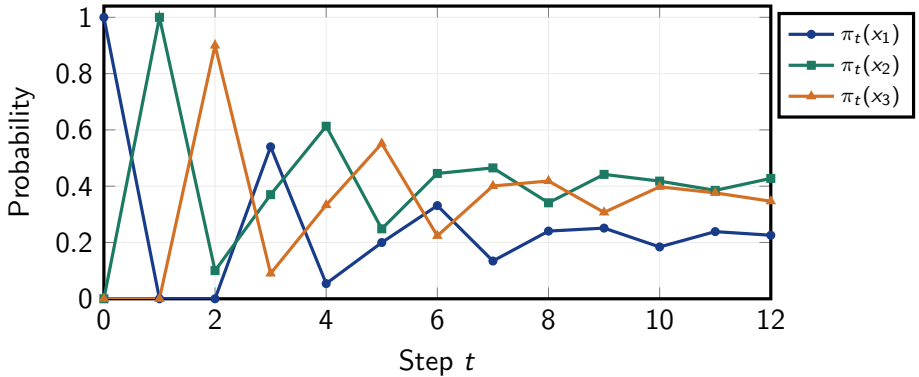
$$\boldsymbol{\pi}_2 = \boldsymbol{\pi}_1 T = (0, 0.1, 0.9),$$

$$\boldsymbol{\pi}_3 = \boldsymbol{\pi}_2 T = (0.54, 0.37, 0.09).$$

Check $\boldsymbol{\pi}_3$

For example, $\pi_3(x_1) = 0 \times 0 + 0.1 \times 0 + 0.9 \times 0.6 = 0.54$.

Visualising distribution evolution



The distribution changes from step to step, and ultimately settles towards a fixed distribution.

Definition from Section 7.1

A probability row vector $\boldsymbol{\pi}$ is **invariant** (stationary) for T if

$$\boldsymbol{\pi} T = \boldsymbol{\pi}, \qquad \pi_i \geq 0, \quad \sum_i \pi_i = 1.$$

- If $X_t \sim \boldsymbol{\pi}$, then $X_{t+1} \sim \boldsymbol{\pi}$.
- The *distribution* remains unchanged even though individual states can move.
- Finding $\boldsymbol{\pi}$ means solving a linear system with a normalisation constraint.

Solve the invariant distribution for the notes' matrix

Let $\boldsymbol{\pi} = (a, b, d)$ for the three-state matrix:

$$\boldsymbol{\pi} T = \boldsymbol{\pi} \quad \Longleftrightarrow \quad \begin{cases} a = 0.6d, \\ b = a + 0.1b + 0.4d, \\ d = 0.9b, \\ a + b + d = 1. \end{cases}$$

Substitute $d = 0.9b$ and $a = 0.6d = 0.54b$:

$$(0.54 + 1 + 0.9)b = 1 \quad \Longrightarrow \quad b = \frac{25}{61}.$$

Hence

$$\boxed{\boldsymbol{\pi} = \left(\frac{27}{122}, \frac{50}{122}, \frac{45}{122} \right).}$$

Verify the stationary distribution

Multiply the candidate stationary vector by T:

$$\left(\frac{27}{122}, \frac{50}{122}, \frac{45}{122}\right) T = \frac{1}{122}(45(0.6),\ 27 + 50(0.1) + 45(0.4),\ 50(0.9))$$
$$= \frac{1}{122}(27, 50, 45) = \boldsymbol{\pi}.$$

Conclusion

$$\boxed{\boldsymbol{\pi} T = \boldsymbol{\pi}, \quad \boldsymbol{\pi} \approx (0.2213, 0.4098, 0.3689).}$$

The stationary distribution is neither a row of T nor simply a uniform distribution.

Stationary does not automatically mean convergence

There are two different statements:

1. **Invariance**: if the chain starts with $X_0 \sim \pi$, then $X_t \sim \pi$ for every t.
2. **Convergence**: if the chain starts from an arbitrary distribution, its law approaches π as $t \to \infty$.

Important distinction

The equation $\pi T = \pi$ guarantees the first statement, not the second without further assumptions.

Detailed balance with respect to π

For all pairs of states x_i, x_j,

$$\boxed{\pi_i T_{ij} = \pi_j T_{ji}.} \tag{7.2}$$

- $\pi_i T_{ij}$: probability flow from state i to state j when the chain has distribution π.
- $\pi_j T_{ji}$: flow in the opposite direction.
- Detailed balance means the two flows match **pair by pair**.

A chain satisfying detailed balance is called *reversible* with respect to π.

Proof: detailed balance implies invariance

For each j, sum Eq. (7.2) over i:

$$(\boldsymbol{\pi} T)_j = \sum_i \pi_i T_{ij} \overset{\text{DB}}{=} \sum_i \pi_j T_{ji}$$

$$= \pi_j \underbrace{\sum_i T_{ji}}_{\text{row } j \text{ sums to } 1} = \pi_j.$$

Therefore

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

What this proves

Detailed balance is sufficient for stationarity. This calculation is the one given in the notes after Eq. (7.2).

Is detailed balance necessary? No!

Take the *same* three-state example from Section 7.1:

$$T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{pmatrix}, \qquad \boldsymbol{\pi} = \frac{1}{122}(27, 50, 45).$$

- It has a stationary distribution: $\boldsymbol{\pi} T = \boldsymbol{\pi}$.
- However, for states 1 and 2,

$$\pi_1 T_{12} = \frac{27}{122}, \qquad \pi_2 T_{21} = 0.$$

Conclusion

Detailed balance *fails*, yet the distribution is stationary. Stationarity is weaker than reversibility.

Detailed balance alone does not ensure convergence

Consider the two-state matrix

$$T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \boldsymbol{\pi} = \left(\frac{1}{2}, \frac{1}{2}\right).$$

- $\boldsymbol{\pi} T = \boldsymbol{\pi}$, and detailed balance is satisfied.
- But starting at state 1 yields

$$(1,0) \xrightarrow{T} (0,1) \xrightarrow{T} (1,0) \xrightarrow{T} \cdots$$

- The law oscillates rather than converging to $\boldsymbol{\pi}$.

Lesson

Detailed balance proves *invariance*, but a periodic chain may fail to converge.

When does a finite-state chain converge?

For a **finite, time-homogeneous** Markov chain, a standard sufficient set of conditions is:

1. **Irreducible**: every state can eventually be reached from every other state.
2. **Aperiodic**: the chain is not forced into a rigid repeating cycle.

Then there is a unique stationary distribution and

$$\boxed{\boldsymbol{\pi}_t = \boldsymbol{\pi}_0 T^t \longrightarrow \boldsymbol{\pi} \quad (t \to \infty).}$$

The notes' three-state example

It is irreducible, and state 2 has a self-loop $T_{22} = 0.1 > 0$; hence the chain is aperiodic and converges.

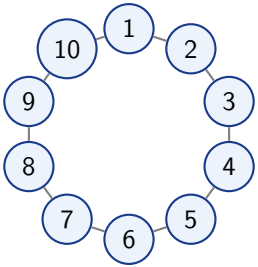
The irreducibility/aperiodicity clarification is a mathematical supplement to Section 7.1.

The King and ten islands: target distribution

The attached slides introduce **10 islands** arranged in a ring.

- Island i has size i.
- The King can only propose moving to a neighbouring island.
- We want long-run visits proportional to island size:

$$\boxed{\pi_i = \frac{i}{1+2+\cdots+10} = \frac{i}{55}.}$$



The King: a transition rule

From island i:

1. Propose a neighbour j (clockwise or anticlockwise), each with probability $1/2$.
2. Accept the move with probability

$$\alpha(i,j) = \min\left\{1, \frac{j}{i}\right\}.$$

3. Otherwise **remain at island i**.

Examples

$4 \to 5$: accept with probability 1. $5 \to 4$: accept with probability $4/5$.

This rule is an example of the Metropolis idea; the full construction is developed in Section 7.2.

Why the island chain has the right stationary law

For neighbouring $i \neq j$, its transition probability is

$$T_{ij} = \frac{1}{2} \min\left\{1, \frac{j}{i}\right\}.$$

Using $\pi_i = i/55$:

$$\begin{aligned} \pi_i T_{ij} &= \frac{i}{55} \cdot \frac{1}{2} \min\{1, j/i\} \\ &= \frac{\min\{i, j\}}{110} \\ &= \pi_j T_{ji}. \end{aligned}$$

Detailed balance $\Rightarrow$ stationarity

$\pi_i = i/55$ is stationary. The chain can reach any island, and rejections create self-loops, so it also converges.

Numerical check: theory versus simulation

- The stationary distribution derived from the notes' matrix is

$$\boldsymbol{\pi} = \frac{1}{122}(27, 50, 45) \approx (0.2213, 0.4098, 0.3689).$$

- A long simulation should have empirical visitation frequencies close to these values.
- The draws $X_1, X_2, \ldots$ are **dependent**, so usual iid error formulas cannot be applied without adjustment.

Related Monte Carlo idea

Under suitable ergodicity assumptions,

$$\frac{1}{N}\sum_{t=1}^{N} f(X_t) \longrightarrow \sum_{i=1}^{s} \pi_i f(x_i).$$

Quick exercise: a two-state Markov chain

Let

$$T = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}.$$

1. Verify that T is a valid transition matrix.
2. Starting from $\boldsymbol{\pi}_0 = (1, 0)$, find $\boldsymbol{\pi}_1$.
3. Find a stationary distribution $\boldsymbol{\pi} = (a, b)$.
4. Does T satisfy detailed balance with respect to $\boldsymbol{\pi}$?

Try independently before viewing the solution.

1. The row sums are $0.8 + 0.2 = 1$ and $0.3 + 0.7 = 1$.
2. $\boldsymbol{\pi}_1 = (1, 0)T = \boxed{(0.8, 0.2)}$.
3. Stationarity requires

$$a = 0.8a + 0.3b, \qquad a + b = 1 \quad \Longrightarrow \quad \boxed{\boldsymbol{\pi} = (0.6, 0.4)}.$$

4. Check detailed balance:

$$\pi_1 T_{12} = 0.6(0.2) = 0.12, \qquad \pi_2 T_{21} = 0.4(0.3) = 0.12.$$

Thus detailed balance holds, and the stationary distribution is $(0.6, 0.4)$.

Summary: what Section 7.1 establishes

1. **Markov property:** the next-state law depends only on the present state.
2. **Homogeneous chain:** one fixed transition matrix T.
3. **Evolution:** $\boxed{\pi_t = \pi_{t-1} T}$ (7.1)
4. **Stationarity:** $\boxed{\pi T = \pi}$.
5. **Detailed balance:** $\boxed{\pi_i T_{ij} = \pi_j T_{ji}}$ (7.2)
6. **Convergence:** irreducibility and aperiodicity supply the missing conditions for a finite chain.

The MCMC design goal

Choose transitions so that the *desired* distribution is invariant, and the chain can explore the full state space.

We now know *what* to aim for:

$$\pi T = \pi.$$

The next question is *how to construct* such a transition rule for a target density π that is difficult to sample directly.

Preview

- Propose a new state using a convenient proposal distribution.
- Accept or reject using a carefully chosen probability.
- Use detailed balance to prove the target is invariant.

From Markov-chain theory to MCMC algorithms.