Markov Chains

Transition Matrices and Stationary Distributions

Computational Statistics

The transition matrix

The possible states are $x_1, \ldots, x_s$.

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

The entry a_{ij} is the probability of moving from x_i to x_j, given that the current state is x_i:

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

The example $a_{12} = 0.2$

If the chain is currently at x_1, the probability that the next state is x_2 is 20%.

Each row describes a **conditional distribution**:

- Row 1 describes the next state, given that the current state is x_1.
- Row 2 describes the next state, given that the current state is x_2.
- The same interpretation applies to every row.

Mathematical clarification

Entries are nonnegative and each row sums to one. Strict positivity of every entry is not required.

The next marginal distribution

Write the current marginal distribution as the **row vector**

$$\pi_{t-1} = (p_1, p_2, \ldots, p_s), \qquad p_i = \mathbb{P}(X_{t-1} = x_i).$$

Given π_{t-1} and the current transition matrix T_{t-1}, how do we compute π_t?

The update formula

$$\boxed{\pi_t = \pi_{t-1} T_{t-1}.}$$

Notation: T_{t-1} describes the transition from $t-1$ to t. Below, its entries are written as a_{ij}, with the time index suppressed.

The first entry of the product

Multiply the row vector of marginal probabilities by the first column of the transition matrix:

$$(p_1, p_2, \ldots, p_s) \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{s1} \end{pmatrix} = p_1 a_{11} + p_2 a_{21} + \cdots + p_s a_{s1}.$$

This is the first entry of $\pi_{t-1} T_{t-1}$.

What does each term mean?

The term $p_1 a_{11}$

p_1 is the probability that the current state is x_1.

a_{11} is the probability of staying at x_1 at the next time, given that the current state is x_1.

The term $p_2 a_{21}$

p_2 is the probability that the current state is x_2.

a_{21} is the probability of moving to x_1, given that the current state is x_2.

The products describe being at the starting state *and* then at x_1.

The conditional probability is

$$a_{11} = \mathbb{P}(X_t = x_1 \mid X_{t-1} = x_1).$$

The product is a **joint probability**:

$$\begin{aligned} p_1 a_{11} &= \mathbb{P}(X_{t-1} = x_1)\mathbb{P}(X_t = x_1 \mid X_{t-1} = x_1) \\ &= \mathbb{P}(X_{t-1} = x_1,\ X_t = x_1). \end{aligned}$$

There are two random variables: the state at time $t-1$ and the state at time t.

All transitions into state x_1

Each term accounts for one possible starting state.

State at time $t-1$	State at time t	Joint probability
x_1	x_1	$p_1 a_{11}$
x_2	x_1	$p_2 a_{21}$
x_3	x_1	$p_3 a_{31}$
$\vdots$	$\vdots$	$\vdots$
x_s	x_1	$p_s a_{s1}$

Adding all possibilities

To be at x_1 at time t, the chain could have come from any of the states $x_1, \ldots, x_s$.

$$\begin{aligned}\mathbb{P}(X_t = x_1) &= \sum_{i=1}^{s} \mathbb{P}(X_{t-1} = x_i,\ X_t = x_1) \\ &= p_1 a_{11} + p_2 a_{21} + \cdots + p_s a_{s1}.\end{aligned}$$

Thus the first entry of the matrix product gives the probability that the next state is x_1.

The same argument for every state

The second entry gives the probability of being at x_2 at the next time point. The same argument applies to every x_j:

$$\mathbb{P}(X_t = x_j) = \sum_{i=1}^{s} p_i a_{ij}, \qquad j = 1, \ldots, s.$$

Collecting these probabilities into a row vector gives

$$\boxed{\pi_t = \pi_{t-1} T_{t-1}.}$$

This is how we compute the next marginal distribution.

Repeated updates for prediction

Suppose the initial marginal distribution is π_1.

$$\pi_2 = \pi_1 T_1,$$

$$\pi_3 = \pi_2 T_2,$$

$$\pi_4 = \pi_3 T_3, \quad \ldots$$

We can compute the marginal distribution at any later time, provided that we know the transition matrix at each time point.

The initial time can be labelled $t = 0$ or $t = 1$, as long as the convention is consistent.

The marginal distribution as a prediction

Weather prediction

A prediction gives probabilities for the possible future states.

These probabilities form the marginal distribution at the future time.

Given the initial marginal distribution and the transition matrices at each time, we can compute these future probabilities.

In general, the transition matrix may change with time.

An important special case uses the **same transition matrix** at every time:

$$T_{t-1} = T.$$

Homogeneous Markov chain

A Markov chain is homogeneous if its transition matrix is independent of time.

The same transition rule applies at $t = 1$, $t = 2$, $t = 3$, and so on.

What homogeneity means

- The transition rule remains the same over time.
- The marginal distribution can still change over time.
- Under appropriate conditions, the initial distribution has less influence on the distribution at large times.

Mathematical clarification

The initial distribution and transition matrix together specify the law of a homogeneous Markov chain. Independence of the long-run limit from the initial distribution requires additional conditions.

Why focus on the homogeneous case?

Some stochastic processes are not homogeneous.

For many physical or biological models, however, the governing laws remain stable over time. Homogeneous Markov chains can model this situation.

Connection with the course

The lecture introduces Markov chains in preparation for the sampling problem. For that purpose, the focus here is on homogeneous Markov chains.

Prediction with a single transition matrix

We now need only:

- an initial marginal distribution π_1,
- one transition matrix T.

$$\pi_2 = \pi_1 T,$$

$$\pi_3 = \pi_2 T,$$

$$\pi_4 = \pi_3 T, \quad \ldots$$

Repeated vector–matrix multiplication gives the distribution at later times.

Keep computing $\pi_2, \pi_3, \pi_4, \ldots$ using the same matrix T.

What happens as $t \to \infty$?

Convergence under suitable conditions

$$\pi_t \longrightarrow \pi_\star \qquad \text{as } t \longrightarrow \infty.$$

The marginal distribution approaches a particular distribution $\pi_\star$.

This is not always true. The lecture does not develop the theoretical guarantees.

The intuition is that the distribution eventually settles down.

Stationary, invariant, and equilibrium

The distribution $\pi_\star$ has several names:

Stationary distribution The distribution does not change under another transition.

Invariant distribution The next part explains the precise meaning of invariance.

Equilibrium distribution A physical interpretation is a state of balance.

Mathematical clarification in the corrected transcript

- The *distribution* is stationary. Individual sample paths can continue moving between states.
- A stationary distribution need not be a limiting distribution from every starting point.
- For a finite-state chain, irreducibility and aperiodicity are standard sufficient conditions for convergence to a unique stationary distribution from every initial distribution.

What does *invariant* mean?

The distribution at two late times

Assume that $\pi_t \to \pi_\star$.

For a sufficiently large time t', both distributions are close to the limit:

$$\pi_{t'} \approx \pi_\star, \qquad \pi_{t'+1} \approx \pi_\star.$$

At the same time, the update equation holds exactly:

$$\pi_{t'+1} = \pi_{t'} T.$$

Taking limits in the update equation gives

$$\boxed{\pi_\star = \pi_\star T.}$$

Applying the transition matrix to $\pi_\star$ leaves the distribution unchanged. This is why it is called **invariant**.

Stationary, invariant, and equilibrium distributions satisfy this equation.

Mathematical clarification

At a finite, large time, proximity to the limit is generally approximate. The exact identity follows by taking limits in $\pi_{t+1} = \pi_t T$.

The influence of the initial distribution

Under the appropriate convergence conditions, the marginal distribution approaches the same stationary distribution, regardless of the initial distribution.

As the chain runs for longer, the influence of the initial state or initial marginal distribution becomes small.

Long-run behaviour

The initial condition no longer determines the limiting distribution.

Computing the stationary distribution by iteration

Suppose we know the transition matrix and want to find its stationary distribution.

Under the convergence conditions:

1. Start from an initial distribution π_1.
2. Compute $\pi_2, \pi_3, \pi_4, \ldots$ by repeated multiplication by T.
3. Continue until the distribution changes very little.

This gives an approximate stationary distribution. The difficulty is that we do not know in advance how many steps will be needed.

Computing the stationary distribution from equations

Another approach uses the identity

$$\boxed{\pi_\star = \pi_\star T.}$$

For a small example, this gives a direct way to compute the stationary distribution.

We will write out this equation for three states.

The three-state setup

For simplicity, let $s = 3$. The state space is

$$\{x_1, x_2, x_3\}.$$

Write the stationary distribution as

$$\pi_\star = (p_1, p_2, p_3),$$

and the transition matrix as

$$T = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The formula $\pi_\star = \pi_\star T$ becomes

$$(p_1, p_2, p_3) = (p_1, p_2, p_3) \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

We now compute the three entries on the right-hand side.

Expanding the matrix product

The entries of $(p_1, p_2, p_3)T$ are:

$$\text{First entry:} \quad p_1 a_{11} + p_2 a_{21} + p_3 a_{31},$$

$$\text{Second entry:} \quad p_1 a_{12} + p_2 a_{22} + p_3 a_{32},$$

$$\text{Third entry:} \quad p_1 a_{13} + p_2 a_{23} + p_3 a_{33}.$$

Using three states makes each term easy to write out.

Three linear equations

Equality of the two vectors gives

$$p_1 = p_1 a_{11} + p_2 a_{21} + p_3 a_{31},$$
$$p_2 = p_1 a_{12} + p_2 a_{22} + p_3 a_{32},$$
$$p_3 = p_1 a_{13} + p_2 a_{23} + p_3 a_{33}.$$

- All entries a_{ij} are known constants.
- The unknowns are p_1, p_2, p_3.
- The goal is to determine $\pi_\star = (p_1, p_2, p_3)$.

Three equations determine three unknowns uniquely only when the relevant system has sufficient independent equations.

The three stationarity equations are linearly dependent.

In the rank-two case, only two are independent. The third depends on the other two.

Mathematical clarification

For a three-state chain with a unique stationary distribution, including an irreducible chain, $\operatorname{rank}(T - I) = 2$. In general, the rank is at most two and can be smaller.

Normalisation supplies the missing equation

The probabilities must sum to one:

$$\boxed{p_1 + p_2 + p_3 = 1.}$$

In the rank-two case, combine two **linearly independent** stationarity equations with this normalisation equation. Solve for p_1, p_2, p_3.

Mathematical clarification

Any two stationarity equations can be used in the irreducible three-state case. Rank two alone guarantees only that some pair is independent. The probabilities must also satisfy $p_i \geq 0$.