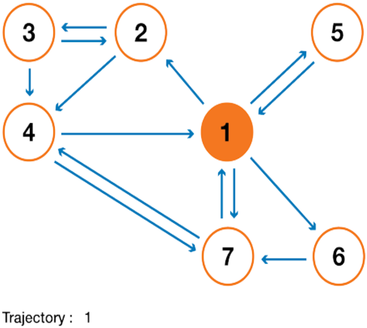
Introduction to Markov Chains

Open Your Eyes



October 9th 2025

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment.

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment. For example,

- Newton's laws determine how a particle moves once initial position and velocity are known, i.e., $F = ma = mx''(t)$ – differential equations (MA 2051, 2431).

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment. For example,

- Newton's laws determine how a particle moves once initial position and velocity are known, i.e., $F = ma = mx''(t)$ – differential equations (MA 2051, 2431).
- The system's future is fixed by its present — uncertainty does not enter the model.

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment. For example,

- Newton's laws determine how a particle moves once initial position and velocity are known, i.e., $F = ma = mx''(t)$ – differential equations (MA 2051, 2431).
- The system's future is fixed by its present — uncertainty does not enter the model.
- If we know $x(t_0)$, we can (in principle) compute $x(t)$ for all later times.

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment. For example,

- Newton's laws determine how a particle moves once initial position and velocity are known, i.e., $F = ma = mx''(t)$ – differential equations (MA 2051, 2431).
- The system's future is fixed by its present — uncertainty does not enter the model.
- If we know $x(t_0)$, we can (in principle) compute $x(t)$ for all later times.

But in many real systems, such exact prediction is impossible:

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment. For example,

- Newton's laws determine how a particle moves once initial position and velocity are known, i.e., $F = ma = mx''(t)$ – differential equations (MA 2051, 2431).
- The system's future is fixed by its present — uncertainty does not enter the model.
- If we know $x(t_0)$, we can (in principle) compute $x(t)$ for all later times.

But in many real systems, such exact prediction is impossible:

- The system is too complex (weather, finance, biology).

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment. For example,

- Newton's laws determine how a particle moves once initial position and velocity are known, i.e., $F = ma = mx''(t)$ – differential equations (MA 2051, 2431).
- The system's future is fixed by its present — uncertainty does not enter the model.
- If we know $x(t_0)$, we can (in principle) compute $x(t)$ for all later times.

But in many real systems, such exact prediction is impossible:

- The system is too complex (weather, finance, biology).
- We do not observe all relevant variables.

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment. For example,

- Newton's laws determine how a particle moves once initial position and velocity are known, i.e., $F = ma = mx''(t)$ – differential equations (MA 2051, 2431).
- The system's future is fixed by its present — uncertainty does not enter the model.
- If we know $x(t_0)$, we can (in principle) compute $x(t)$ for all later times.

But in many real systems, such exact prediction is impossible:

- The system is too complex (weather, finance, biology).
- We do not observe all relevant variables.
- Measurement errors or noise accumulate.

From Deterministic to Probabilistic Thinking (I)

In physics or engineering, we often describe motion or change by a deterministic function, known as a **trajectory**:

$$x = x(t), \quad t > 0.$$

Here, t is time, and $x(t)$ tells us the **exact state** of the system at each moment. For example,

- Newton's laws determine how a particle moves once initial position and velocity are known, i.e., $F = ma = mx''(t)$ – differential equations (MA 2051, 2431).
- The system's future is fixed by its present — uncertainty does not enter the model.
- If we know $x(t_0)$, we can (in principle) compute $x(t)$ for all later times.

But in many real systems, such exact prediction is impossible:

- The system is too complex (weather, finance, biology).
- We do not observe all relevant variables.
- Measurement errors or noise accumulate.

Result: Our best description is not a trajectory, but a *distribution of possible states*.

From Deterministic to Probabilistic Thinking (II)

A **Markov chain** describes dynamics through *probabilities*, not precise values.

- Instead of $x(t)$, we track the probability that the system, as a random variable X_t, is in **each possible state**, using a vector

$$\boldsymbol{\pi}^{(t)} = \left[P\left(X_t = 1\right), P\left(X_t = 2\right), \ldots, P\left(X_t = n\right)\right]$$

From Deterministic to Probabilistic Thinking (II)

A **Markov chain** describes dynamics through *probabilities*, not precise values.

- Instead of $x(t)$, we track the probability that the system, as a random variable X_t, is in **each possible state**, using a vector

$$\boldsymbol{\pi}^{(t)} = [P(X_t = 1), P(X_t = 2), \ldots, P(X_t = n)]$$

- Transitions between states are governed by a **transition matrix** (defined later).

From Deterministic to Probabilistic Thinking (II)

A **Markov chain** describes dynamics through *probabilities*, not precise values.

- Instead of $x(t)$, we track the probability that the system, as a random variable X_t, is in **each possible state**, using a vector

$$\boldsymbol{\pi}^{(t)} = [P(X_t = 1), P(X_t = 2), \ldots, P(X_t = n)]$$

- Transitions between states are governed by a **transition matrix** (defined later).
- This matrix plays the role of Newton's equations — it determines how uncertainty **evolves** in time.

From Deterministic to Probabilistic Thinking (II)

A **Markov chain** describes dynamics through *probabilities*, not precise values.

- Instead of $x(t)$, we track the probability that the system, as a random variable X_t, is in **each possible state**, using a vector

$$\boldsymbol{\pi}^{(t)} = \left[P\left(X_t = 1\right), P\left(X_t = 2\right), \ldots, P\left(X_t = n\right)\right]$$

- Transitions between states are governed by a **transition matrix** (defined later).
- This matrix plays the role of Newton's equations — it determines how uncertainty **evolves** in time.

Why is this perspective powerful?

- It models systems with incomplete information or inherent randomness.

From Deterministic to Probabilistic Thinking (II)

A **Markov chain** describes dynamics through *probabilities*, not precise values.

- Instead of $x(t)$, we track the probability that the system, as a random variable X_t, is in **each possible state**, using a vector

$$\boldsymbol{\pi}^{(t)} = [P(X_t = 1), P(X_t = 2), \ldots, P(X_t = n)]$$

- Transitions between states are governed by a **transition matrix** (defined later).
- This matrix plays the role of Newton's equations — it determines how uncertainty **evolves** in time.

Why is this perspective powerful?

- It models systems with incomplete information or inherent randomness.
- It can describe populations, networks, queues, and learning processes where exact laws are unknown.

From Deterministic to Probabilistic Thinking (II)

A **Markov chain** describes dynamics through *probabilities*, not precise values.

- Instead of $x(t)$, we track the probability that the system, as a random variable X_t, is in **each possible state**, using a vector

$$\boldsymbol{\pi}^{(t)} = [P(X_t = 1), P(X_t = 2), \ldots, P(X_t = n)]$$

- Transitions between states are governed by a **transition matrix** (defined later).
- This matrix plays the role of Newton's equations — it determines how uncertainty **evolves** in time.

Why is this perspective powerful?

- It models systems with incomplete information or inherent randomness.
- It can describe populations, networks, queues, and learning processes where exact laws are unknown.
- It shifts focus from "What *will* happen?" to "What is *likely* to happen?"

From Deterministic to Probabilistic Thinking (II)

A **Markov chain** describes dynamics through *probabilities*, not precise values.

- Instead of $x(t)$, we track the probability that the system, as a random variable X_t, is in **each possible state**, using a vector

$$\boldsymbol{\pi}^{(t)} = [P(X_t = 1), P(X_t = 2), \ldots, P(X_t = n)]$$

- Transitions between states are governed by a **transition matrix** (defined later).
- This matrix plays the role of Newton's equations — it determines how uncertainty **evolves** in time.

Why is this perspective powerful?

- It models systems with incomplete information or inherent randomness.
- It can describe populations, networks, queues, and learning processes where exact laws are unknown.
- It shifts focus from "What *will* happen?" to "What is *likely* to happen?"

In short: deterministic dynamics describe *what must happen*; Markov dynamics describe *what tends to happen*.

What is a Markov Chain?

A **Discrete State Discrete Time Markov chain** is a sequence of random variables

$$X_0, X_1, X_2, \ldots, \quad X_i \in \text{Countable or Finite Set}$$

that satisfies the **Markov property**:

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\ = P(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

The future depends only on the present, not on the past. In other words, the future state forgets about history – **memoryless**.

What is a Markov Chain?

A **Discrete State Discrete Time Markov chain** is a sequence of random variables

$$X_0, X_1, X_2, \ldots, \quad X_i \in \text{Countable or Finite Set}$$

that satisfies the **Markov property**:

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0)$$
$$= P(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

The future depends only on the present, not on the past. In other words, the future state forgets about history – **memoryless**.

Typical examples include:

- Weather transitions: sunny $\leftrightarrow$ cloudy $\leftrightarrow$ rainy.
- Number of customers in the queue and at the server.

What is a Markov Chain?

A **Discrete State Discrete Time Markov chain** is a sequence of random variables

$$X_0, X_1, X_2, \ldots, \quad X_i \in \text{Countable or Finite Set}$$

that satisfies the **Markov property**:

$$\begin{aligned} &P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\ &= P(X_{n+1} = x_{n+1} \mid X_n = x_n). \end{aligned}$$

The future depends only on the present, not on the past. In other words, the future state forgets about history – **memoryless**.

Typical examples include:

- Weather transitions: sunny $\leftrightarrow$ cloudy $\leftrightarrow$ rainy.
- Number of customers in the queue and at the server.
- Machine states: operational $\leftrightarrow$ failed.

Transition Probability Matrix

The transitions between states are summarized in a matrix:

$$P = [p_{ij}], \qquad p_{ij} = P(X_{n+1} = j \mid X_n = i),$$

where each row represents the probabilities of leaving state i.

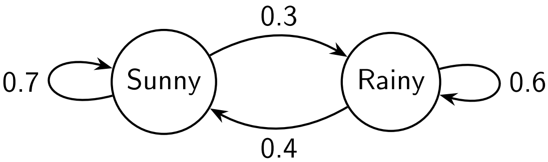
Properties

- $p_{ij} \geq 0$ for all i, j, (Axiom 1)
- $\sum_j p_{ij} = 1$ for each i. (Axiom 3)

Example: A Simple Weather Model

Example: Two-state weather model

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}, \quad \text{States: } 1 = \text{Sunny},\ 2 = \text{Rainy}.$$

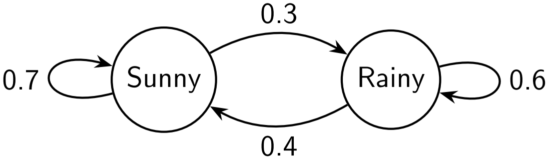


If today is Sunny, the probability it stays Sunny tomorrow is 0.7; if today is Rainy, it stays Rainy with probability 0.6.

Example: A Simple Weather Model

Example: Two-state weather model

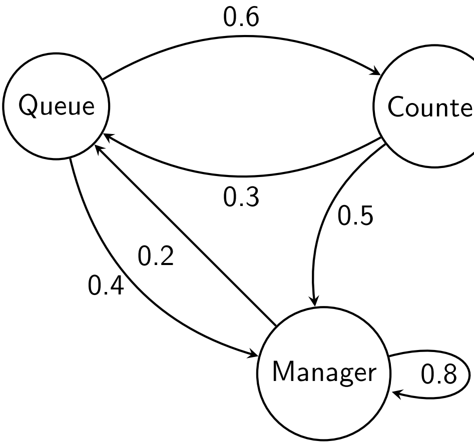
$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}, \quad \text{States: } 1 = \text{Sunny},\ 2 = \text{Rainy}.$$



If today is Sunny, the probability it stays Sunny tomorrow is 0.7; if today is Rainy, it stays Rainy with probability 0.6.

Real weather models are far more complex, but they do involve basic steps that use (continuous-time) Markov chains.

Example: Three-State System



$$P = \begin{bmatrix} 0 & 0.6 & 0.4 \\ 0.3 & 0 & 0.7 \\ 0.2 & 0 & 0.8 \end{bmatrix}$$

Interpretation:

- Rows sum to 1 (law of total probability).
- $P_{ij} = \text{Prob}(\text{next state } j \mid \text{current state } i)$.
- The matrix drives how probabilities evolve in time.

n-step Transitions

The n-step transition probabilities are defined by

$$p_{ij}^{(n)} = P(X_{n+k} = j \mid X_k = i).$$

They are obtained from powers of the transition matrix:

$$P^{(n)} = P^n. \text{ Why? Next slide.}$$

In words, the n-step transition matrix is the one-step matrix raised to the pth power.

n-step Transitions

The n-step transition probabilities are defined by

$$p_{ij}^{(n)} = P(X_{n+k} = j \mid X_k = i).$$

They are obtained from powers of the transition matrix:

$$P^{(n)} = P^n. \text{ Why? Next slide.}$$

In words, the n-step transition matrix is the one-step matrix raised to the pth power. **Example:** If

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}, \quad \text{then} \quad P^2 = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}.$$

After two days, the chain “forgets” a bit of its starting point.

The Multiplication Rule Revisited

Recall: For any three events A, B, C,

$$P(A) = \sum_b P(A \mid B = b)\, P(B = b)$$

The Multiplication Rule Revisited

Recall: For any three events A, B, C,

$$P(A) = \sum_b P(A \mid B = b)\, P(B = b)$$

If we know C, we can condition everything on it. By the **multiplication rule for conditional probability**,

$$P(A \mid C) = \sum_b P(A \mid B = b, C)\, P(B = b \mid C)$$

The Multiplication Rule Revisited

Recall: For any three events A, B, C,

$$P(A) = \sum_b P(A \mid B = b)\, P(B = b)$$

If we know C, we can condition everything on it. By the **multiplication rule for conditional probability**,

$$P(A \mid C) = \sum_b P(A \mid B = b, C)\, P(B = b \mid C)$$

Interpretation:

- We sum over all possible intermediate events $B = b$.
- The second law expresses how we “go through” the middle layer B.

Why Does $\mathbf{P}^{(2)} = \mathbf{P}^2$?

Idea: A two-step transition probability is just a *sum over all possible intermediate states*, by the **law of total (conditional) probabilities.**

$$A = \{X_{t+2} = j\}, \quad B = \{X_{t+1} = k\}, \quad C = \{X_t = i\}.$$

Why Does $\mathbf{P}^{(2)} = \mathbf{P}^2$?

Idea: A two-step transition probability is just a *sum over all possible intermediate states*, by the **law of total (conditional) probabilities.**

$$A = \{X_{t+2} = j\}, \quad B = \{X_{t+1} = k\}, \quad C = \{X_t = i\}.$$

Then the formula becomes:

$$P(X_{t+2} = j \mid X_t = i) = \sum_k \boxed{P(X_{t+2} = j \mid X_{t+1} = k, X_t = i)} P(X_{t+1} = k \mid X_t = i).$$

Why Does $\mathbf{P}^{(2)} = \mathbf{P}^2$?

Idea: A two-step transition probability is just a *sum over all possible intermediate states*, by the **law of total (conditional) probabilities.**

$$A = \{X_{t+2} = j\}, \quad B = \{X_{t+1} = k\}, \quad C = \{X_t = i\}.$$

Then the formula becomes:

$$P(X_{t+2} = j \mid X_t = i) = \sum_k \boxed{P(X_{t+2} = j \mid X_{t+1} = k, X_t = i)}\, P(X_{t+1} = k \mid X_t = i).$$

By the Markov property:

$$\boxed{P(X_{t+2} = j \mid X_{t+1} = k, X_t = i)} = P(X_{t+2} = j \mid X_{t+1} = k).$$

Why Does $\mathbf{P}^{(2)} = \mathbf{P}^2$?

Idea: A two-step transition probability is just a *sum over all possible intermediate states*, by the **law of total (conditional) probabilities.**

$$A = \{X_{t+2} = j\}, \quad B = \{X_{t+1} = k\}, \quad C = \{X_t = i\}.$$

Then the formula becomes:

$$P(X_{t+2} = j \mid X_t = i) = \sum_k \boxed{P(X_{t+2} = j \mid X_{t+1} = k, X_t = i)}\, P(X_{t+1} = k \mid X_t = i).$$

By the Markov property:

$\boxed{P(X_{t+2} = j \mid X_{t+1} = k, X_t = i)} = P(X_{t+2} = j \mid X_{t+1} = k).$

So,

$$P(X_{t+2} = j \mid X_t = i) = \sum_k P(X_{t+2} = j \mid X_{t+1} = k)\, P(X_{t+1} = k \mid X_t = i),$$

which in matrix form is simply

$$\boxed{\mathbf{P}^{(2)} = \mathbf{P}^2.}$$

Why Does $\mathbf{P}^{(2)} = \mathbf{P}^2$?

Idea: A two-step transition probability is just a *sum over all possible intermediate states*, by the **law of total (conditional) probabilities.**

$$A = \{X_{t+2} = j\}, \quad B = \{X_{t+1} = k\}, \quad C = \{X_t = i\}.$$

Then the formula becomes:

$$P(X_{t+2} = j \mid X_t = i) = \sum_k \boxed{P(X_{t+2} = j \mid X_{t+1} = k, X_t = i)}\, P(X_{t+1} = k \mid X_t = i).$$

By the Markov property:

$$\boxed{P(X_{t+2} = j \mid X_{t+1} = k, X_t = i)} = P(X_{t+2} = j \mid X_{t+1} = k).$$

So,

$$P(X_{t+2} = j \mid X_t = i) = \sum_k P(X_{t+2} = j \mid X_{t+1} = k)\, P(X_{t+1} = k \mid X_t = i),$$

which in matrix form is simply

$$\boxed{\mathbf{P}^{(2)} = \mathbf{P}^2.}$$

Interpretation:

- P_{ik} = probability of going from i to k in one step.
- P_{kj} = probability of going from k to j in one step.
- Summing over k accounts for *all possible middle stops.*

Why (and how) Does the Transition Matrix Evolve the Distribution?

Let the system have states $1, 2, \ldots, m$. At time t, the distribution is

$$\boldsymbol{\pi}^{(t)} = \begin{bmatrix} P(X_t = 1) & P(X_t = 2) & \cdots & P(X_t = m) \end{bmatrix}.$$

Why (and how) Does the Transition Matrix Evolve the Distribution?

Let the system have states $1, 2, \ldots, m$. At time t, the distribution is

$$\boldsymbol{\pi}^{(t)} = \begin{bmatrix} P(X_t = 1) & P(X_t = 2) & \cdots & P(X_t = m) \end{bmatrix}.$$

We want the probabilities $\pi_j^{(t+1)} = P(X_{t+1} = j)$.

Why (and how) Does the Transition Matrix Evolve the Distribution?

Let the system have states $1, 2, \ldots, m$. At time t, the distribution is

$$\boldsymbol{\pi}^{(t)} = \begin{bmatrix} P(X_t = 1) & P(X_t = 2) & \cdots & P(X_t = m) \end{bmatrix}.$$

We want the probabilities $\pi_j^{(t+1)} = P(X_{t+1} = j)$.
By the **law of total probability**:

$$\pi_j^{(t+1)} = P(X_{t+1} = j) = \sum_{i=1}^{m} P(X_{t+1} = j \mid X_t = i)\, P(X_t = i).$$

Why (and how) Does the Transition Matrix Evolve the Distribution?

Let the system have states $1, 2, \ldots, m$. At time t, the distribution is

$$\boldsymbol{\pi}^{(t)} = \begin{bmatrix} P(X_t = 1) & P(X_t = 2) & \cdots & P(X_t = m) \end{bmatrix}.$$

We want the probabilities $\pi_j^{(t+1)} = P(X_{t+1} = j)$.
By the **law of total probability**:

$$\pi_j^{(t+1)} = P(X_{t+1} = j) = \sum_{i=1}^{m} P(X_{t+1} = j \mid X_t = i)\, P(X_t = i).$$

Writing this in vector form:

$$\boldsymbol{\pi}^{(t+1)} = \boldsymbol{\pi}^{(t)}\, \mathbf{P},$$

where $\mathbf{P} = [p_{ij}]$ is the transition matrix.

Why (and how) Does the Transition Matrix Evolve the Distribution?

Let the system have states $1, 2, \ldots, m$. At time t, the distribution is

$$\boldsymbol{\pi}^{(t)} = \begin{bmatrix} P(X_t = 1) & P(X_t = 2) & \cdots & P(X_t = m) \end{bmatrix}.$$

We want the probabilities $\pi_j^{(t+1)} = P(X_{t+1} = j)$.
By the **law of total probability**:

$$\pi_j^{(t+1)} = P(X_{t+1} = j) = \sum_{i=1}^{m} P(X_{t+1} = j \mid X_t = i)\, P(X_t = i).$$

Writing this in vector form:

$$\boldsymbol{\pi}^{(t+1)} = \boldsymbol{\pi}^{(t)}\, \mathbf{P},$$

where $\mathbf{P} = [p_{ij}]$ is the transition matrix.
Interpretation:

- The transition matrix does not change individual outcomes.
- It *moves probability mass* across states.
- This is how the distribution *evolves in time*.

Setup (states)

We consider a discrete state space $\Omega = \{0, 1, 2, 3\}$ (positions along a short line/cycle). Start the system at state 0 (initial distribution $\pi^{(0)} = (1, 0, 0, 0)$).

We compare:

- A deterministic *permutation* dynamics (constant-speed motion around the 4 states).

Setup (states)

We consider a discrete state space $\Omega = \{0, 1, 2, 3\}$ (positions along a short line/cycle). Start the system at state 0 (initial distribution $\pi^{(0)} = (1, 0, 0, 0)$).

We compare:

- A deterministic *permutation* dynamics (constant-speed motion around the 4 states).
- A realistic-looking *Markov* dynamics with probabilistic transitions (local moves, some inertia).

Transition matrices

Deterministic (permutation) matrix:

$$P_{\mathrm{det}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

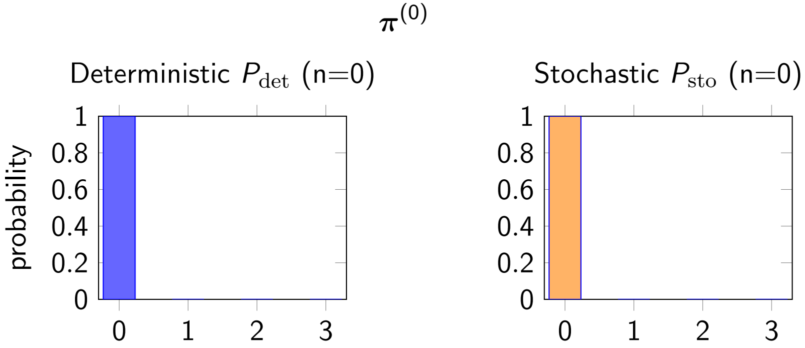
This maps $0 \mapsto 1 \mapsto 2 \mapsto 3 \mapsto 0$: a perfect clockwise step each tick.

Stochastic Markov matrix (richer behaviour):

$$P_{\mathrm{sto}} = \begin{pmatrix} 0.10 & 0.90 & 0 & 0 \\ 0.20 & 0.10 & 0.70 & 0 \\ 0 & 0.30 & 0.10 & 0.60 \\ 0 & 0 & 0.40 & 0.60 \end{pmatrix}$$

Rows sum to 1; entries $p_{ij} = P(\text{next} = j \mid \text{current} = i)$.

Time evolution: $n = 0$

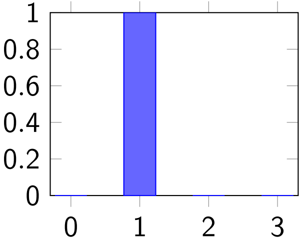


Initial distribution $\pi^{(0)} = (1, 0, 0, 0)$ for both models.

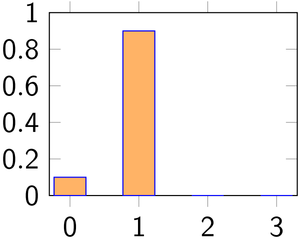
Time evolution: $n = 1$

$$\pi^{(1)} = \pi^{(0)}\,\mathbf{P},$$

Deterministic P_{det} (n=1)

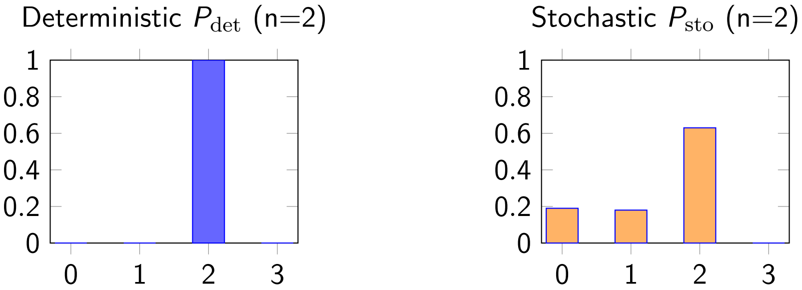


Stochastic P_{sto} (n=1)



Time evolution: $n = 2$

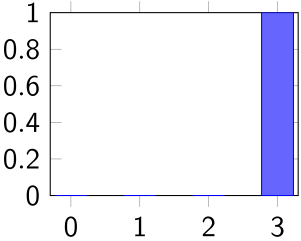
$$\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)}\,\mathbf{P},$$



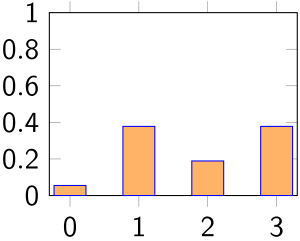
Time evolution: $n = 3$

$$\boldsymbol{\pi}^{(3)} = \boldsymbol{\pi}^{(2)}\,\mathbf{P},$$

Deterministic P_{det} (n=3)

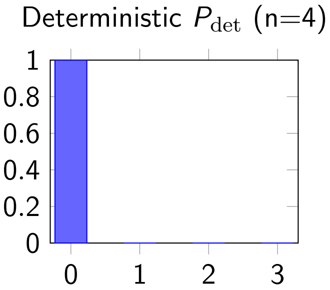


Stochastic P_{sto} (n=3)



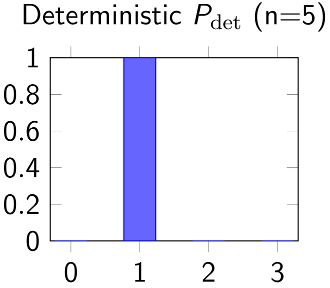
Time evolution: $n = 4$

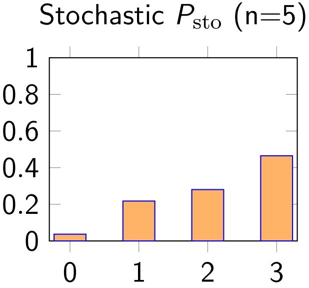
$$\boldsymbol{\pi}^{(4)} = \boldsymbol{\pi}^{(3)}\,\mathbf{P},$$



Time evolution: $n = 5$

$$\boldsymbol{\pi}^{(5)} = \boldsymbol{\pi}^{(4)}\,\mathbf{P},$$





Remarks and interpretation

- **Deterministic chain:** the distribution is always a point mass that cycles through the 4 states — perfect predictability.

Remarks and interpretation

- **Deterministic chain:** the distribution is always a point mass that cycles through the 4 states — perfect predictability.
- **Stochastic chain:** the mass spreads, reflecting uncertainty and back-and-forth moves.

Remarks and interpretation

- **Deterministic chain:** the distribution is always a point mass that cycles through the 4 states — perfect predictability.
- **Stochastic chain:** the mass spreads, reflecting uncertainty and back-and-forth moves.
- Even when the stochastic transitions favor forward motion (like our matrix), randomness creates dispersion and memory loss of the exact initial position.

Remarks and interpretation

- **Deterministic chain:** the distribution is always a point mass that cycles through the 4 states — perfect predictability.
- **Stochastic chain:** the mass spreads, reflecting uncertainty and back-and-forth moves.
- Even when the stochastic transitions favor forward motion (like our matrix), randomness creates dispersion and memory loss of the exact initial position.
- To model a true discrete Newton's second law (position & velocity), one would enlarge the state space to include velocity and obtain a deterministic permutation on the extended state space; marginalizing out velocity then produces a stochastic-looking chain.

From Determinism to Probability: The Journey of Randomness

In deterministic models, the future is certain, given the present.

In probabilistic models, the future is a distribution —
a landscape of possibilities shaped by uncertainty.

From Determinism to Probability: The Journey of Randomness

In deterministic models, the future is certain, given the present.

In probabilistic models, the future is a distribution —
a landscape of possibilities shaped by uncertainty.

Markov Chains remind us that knowing *everything* is rare,
but understanding *how uncertainty evolves* is power.

"Chance is only the measure of our ignorance."

— Henri Poincaré "Chance", Part I, Chapter 4 of Science and Method.

End of Lecture Series on Probability Theory

MA2631, Fall 2025

Ready to experiment?

Time to take your basics and run, and fly high.
Thanks for sticking around!