Introduction to the Markov chains

Marcin Lewandowski

December 19, 2024

Markov chains: intro

Defining *state space* and *state values*

- Let S be a finite set with n elements, $S = \{x_1, x_2, \ldots, x_n\}$.
- The set S is called the *state space* and $x_1, x_2, \ldots, x_n$ are the state values.

Markov chains: intro

Defining *state space* and *state values*

- Let S be a finite set with n elements, $S = \{x_1, x_2, \ldots, x_n\}$.
- The set S is called the *state space* and $x_1, x_2, \ldots, x_n$ are the state values.

Examples:

- State space for an economy: $S = \{\text{Normal Growth}, \text{Mild Recession}, \text{Severe recession}\}$
- State space for an agent: $S = \{\text{Employed}, \text{Unemployed}\}$

Markov chains: intro

A Markov chain $\{X_t\}$ on S is a sequence of random variables that have the **Markov property**:

For any date t and any state $y \in S$

$$P(X_{t+1} = y \mid X_t, X_{t-1}, \ldots, X_0) = P(X_{t+1} = y \mid X_t)$$

Markov chains: intro

A Markov chain $\{X_t\}$ on S is a sequence of random variables that have the **Markov property**:

For any date t and any state $y \in S$

$$P\left(X_{t+1} = y \mid X_t, X_{t-1}, \ldots, X_0\right) = P\left(X_{t+1} = y \mid X_t\right)$$

This implies that:

- Past & future are conditionally independent given the present
- Knowing the current state is enough to know probabilities for future states.

Markov chains: intro

Define a transition matrix $\mathbf{P}$

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

P_{ij} is called a *transition probability*.

Example of a a matrix $\mathbf{P}$ if:

- $S = \{\text{Employed}, \text{Unemployed}\}$

Markov chains: intro

Define a transition matrix $\mathbf{P}$

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

P_{ij} is called a *transition probability*.

Example of a a matrix $\mathbf{P}$ if:

- S = $\{\text{Employed}, \text{Unemployed}\}$
- An unemployed finds a job with probability $\alpha \in (0, 1)$.

Markov chains: intro

Define a transition matrix $\mathbf{P}$

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

P_{ij} is called a *transition probability*.

Example of a a matrix $\mathbf{P}$ if:

- S = $\{\text{Employed}, \text{Unemployed}\}$
- An unemployed finds a job with probability $\alpha \in (0, 1)$.
- An employee loses his job with probability $\beta \in (0, 1)$.

Markov chains: intro

Define a transition matrix P

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

P_{ij} is called a *transition probability*.

Example of a a matrix **P** if:

- S = $\{\text{Employed}, \text{Unemployed}\}$
- An unemployed finds a job with probability $\alpha \in (0, 1)$.
- An employee loses his job with probability $\beta \in (0, 1)$.

Markov chains: intro

Define a transition matrix $\mathbf{P}$

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

P_{ij} is called a *transition probability*.

Example of a a matrix $\mathbf{P}$ if:

- S = $\{\text{Employed}, \text{Unemployed}\}$
- An unemployed finds a job with probability $\alpha \in (0, 1)$.
- An employee loses his job with probability $\beta \in (0, 1)$.

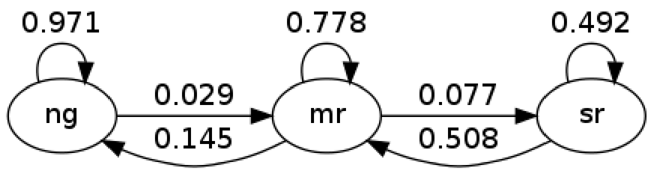
$$\mathbf{P} = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

Example 2: MC through a directed graph lense

Recall the state space for an economy:
$S = \{\text{Normal Growth}, \text{Mild Recession}, \text{Severe recession}\}$

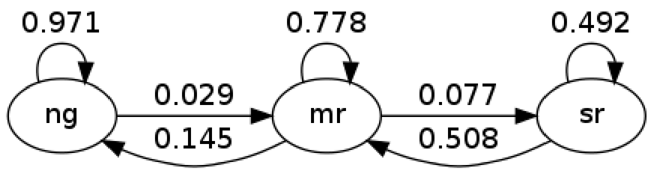
Example 2: MC through a directed graph lense

Recall the state space for an economy:
$S = \{\text{Normal Growth}, \text{Mild Recession}, \text{Severe recession}\}$



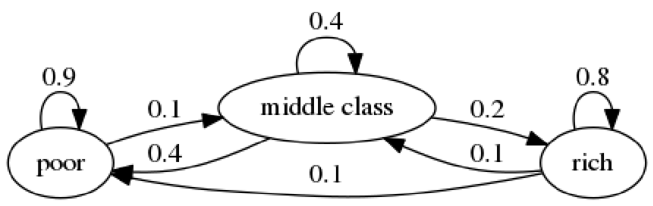
Example 2: MC through a directed graph lense

Recall the state space for an economy:
$S = \{\text{Normal Growth}, \text{Mild Recession}, \text{Severe recession}\}$



$$\mathsf{P} = \begin{bmatrix} 0.971 & 0.029 & 0 \\ 0.145 & 0.779 & 0.077 \\ ? & ? & ? \end{bmatrix}$$

Concept check: write a P matrix



$$P = \begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix}$$

Marginal Distribution

- Let's consider a Markov chain $\{X_t\}$ with a transition matrix P.
- Suppose we know the distribution of X_t it is ψ_t.
- How would we get the distribution of X_{t+1}?

Marginal Distribution

- Let's consider a Markov chain $\{X_t\}$ with a transition matrix P.
- Suppose we know the distribution of X_t it is ψ_t.
- How would we get the distribution of X_{t+1}?

For any $x_j \in S$

$$P(X_{t+1} = x_j) = \sum_{x_i \in S} P(X_{t+1} = x_j \mid X_t = x_i) \cdot P(X_t = x_i).$$

Therefore,

$$\psi_{t+1}(x_j) = \sum_{x_i \in S} \psi_t(x_i) \cdot P_{ij}$$

We can perform this calculation for all $x_j \in S$ to obtain ψ_{t+1}.

Marginal Distribution

Alternatively:

If ψ_t is a row vector, i.e.,

$$\psi_t = [\psi_t(x_1), \ldots, \psi_t(x_n)],$$

then we have

$$\psi_{t+1} = \psi_t P$$

and, for any nonnegative integer s,

$$\psi_{t+s} = \psi_t P^s.$$

Example:

Recall a transition matrix P for unemployed/employed:

$$\mathsf{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

Assume: $\psi_t = [0.1, 0.9]$

Example:

Recall a transition matrix P for unemployed/employed:

$$\mathsf{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

Assume: $\psi_t = [0.1, 0.9]$

$$\begin{aligned} \psi_{t+1} &= [0.1, 0.9] \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix} \\ &= \begin{bmatrix} 0.1 \cdot 0.9 + 0.9 \cdot 0.05 & 0.1 \cdot 0.1 + 0.9 \cdot 0.95 \end{bmatrix} \\ &= \\ &= \qquad\qquad . \end{aligned}$$

Example:

Recall a transition matrix **P** for unemployed/employed:

$$\mathbf{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

Assume: $\psi_t = [0.1, 0.9]$

$$\begin{aligned}\psi_{t+1} &= [0.1, 0.9] \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix} \\ &= \begin{bmatrix} 0.1 \cdot 0.9 + 0.9 \cdot 0.05 & 0.1 \cdot 0.1 + 0.9 \cdot 0.95 \end{bmatrix} \\ &= \begin{bmatrix} 0.09 + 0.045 & 0.01 + 0.855 \end{bmatrix} \\ &= \qquad\qquad\qquad .\end{aligned}$$

Example:

Recall a transition matrix **P** for unemployed/employed:

$$\mathsf{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

Assume: $\psi_t = [0.1, 0.9]$

$$\begin{aligned}
\psi_{t+1} &= [0.1, 0.9] \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix} \\
&= \begin{bmatrix} 0.1 \cdot 0.9 + 0.9 \cdot 0.05 & 0.1 \cdot 0.1 + 0.9 \cdot 0.95 \end{bmatrix} \\
&= \begin{bmatrix} 0.09 + 0.045 & 0.01 + 0.855 \end{bmatrix} \\
&= \begin{bmatrix} 0.135 & 0.865 \end{bmatrix}.
\end{aligned}$$

Example continued:

What if we would like to get ψ_{t+2}?

$$\psi_{t+2} = \psi_{t+1} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

.

Example continued:

What if we would like to get ψ_{t+2}?

$$\psi_{t+2} = \psi_{t+1} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

$$= \begin{bmatrix} 0.135 & 0.865 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

.

Example continued:

What if we would like to get ψ_{t+2}?

$$\psi_{t+2} = \psi_{t+1} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

$$= \begin{bmatrix} 0.135 & 0.865 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

Or:

.

Example continued:

What if we would like to get ψ_{t+2}?

$$\begin{aligned}\psi_{t+2} &= \psi_{t+1}\begin{bmatrix}0.9 & 0.1\\ 0.05 & 0.95\end{bmatrix}\\ &= \begin{bmatrix}0.135 & 0.865\end{bmatrix}\begin{bmatrix}0.9 & 0.1\\ 0.05 & 0.95\end{bmatrix}\end{aligned}$$

Or:

$$\psi_{t+2} = \psi_t\begin{bmatrix}0.9 & 0.1\\ 0.05 & 0.95\end{bmatrix}\begin{bmatrix}0.9 & 0.1\\ 0.05 & 0.95\end{bmatrix}.$$

Example continued:

What if we would like to get ψ_{t+2}?

$$\psi_{t+2} = \psi_{t+1} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

$$= \begin{bmatrix} 0.135 & 0.865 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}$$

Or:

$$\psi_{t+2} = \psi_t \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.05 & 0.95 \end{bmatrix}.$$

You see why:

$$\psi_{t+s} = \psi_t P^s.$$

Cross-sectional distributions perspective:

- The ψ_t can be viewed as PMF or cross-sectional frequencies in large samples.

Cross-sectional distributions perspective:

- The ψ_t can be viewed as PMF or cross-sectional frequencies in large samples.
- In unemployement example assume there are many identical workers in the economy:

Cross-sectional distributions perspective:

- The ψ_t can be viewed as PMF or cross-sectional frequencies in large samples.
- In unemployement example assume there are many identical workers in the economy:
    - They are independent of each other and all have the same stochastic matrix P.

Cross-sectional distributions perspective:

- The ψ_t can be viewed as PMF or cross-sectional frequencies in large samples.
- In unemployement example assume there are many identical workers in the economy:
    - They are independent of each other and all have the same stochastic matrix P.
    - ψ_t is the current cross-sectional distribution over unemployed/employed.

Cross-sectional distributions perspective:

- The ψ_t can be viewed as PMF or cross-sectional frequencies in large samples.
- In unemployement example assume there are many identical workers in the economy:
    - They are independent of each other and all have the same stochastic matrix P.
    - ψ_t is the current cross-sectional distribution over unemployed/employed.
    - i.e., we can interpret $\psi_t(1)$ as the unemployment rate at t.

Cross-sectional distributions perspective:

- The ψ_t can be viewed as PMF or cross-sectional frequencies in large samples.
- In unemployement example assume there are many identical workers in the economy:
    - They are independent of each other and all have the same stochastic matrix P.
    - ψ_t is the current cross-sectional distribution over unemployed/employed.
    - i.e., we can interpret $\psi_t(1)$ as the unemployment rate at t.
- The cross-sectional distribution tomorrow is

$$\psi_{t+1} = \psi_t P.$$

The stationary distribution

We are often interested in ψ^* such that:

$$\psi^* = \psi^* P. \tag{1}$$

We call distribution ψ^* *stationary*.

- Note: $\psi^* P$ is a distribution in the next period

The stationary distribution

We are often interested in ψ^* such that:

$$\psi^* = \psi^* P. \tag{1}$$

We call distribution ψ^* *stationary*.

- Note: $\psi^* P$ is a distribution in the next period
- The Equation (1) requires that ψ^* *doesn't* change over time

The stationary distribution

We are often interested in ψ^* such that:

$$\psi^* = \psi^* P. \tag{1}$$

We call distribution ψ^* *stationary*.

- Note: $\psi^* P$ is a distribution in the next period
- The Equation (1) requires that ψ^* *doesn't* change over time
- If the chain starts in $\psi^* P$ if will have such distribtion forever!

The stationary distribution

We are often interested in ψ^* such that:

$$\psi^* = \psi^* P. \tag{1}$$

We call distribution ψ^* *stationary*.

- Note: $\psi^* P$ is a distribution in the next period
- The Equation (1) requires that ψ^* *doesn't* change over time
- If the chain starts in $\psi^* P$ if will have such distribtion forever!
- We can treat $\psi^* P$ as some long-run "equilibrium"

How to find the stationary distribution

Under irreducibility and aperiodicity:

$$\psi_0 P^t \to \psi^*$$

as $t \to \infty$

How to find the stationary distribution

Under irreducibility and aperiodicity:

$$\psi_0 P^t \to \psi^*$$

as $t \to \infty$

- One way: start with ψ_0, calculate $\psi_t = \psi_0 P^t$, t is (very?) large.
- Other way: solve the Equation (1) with some linear algebra

Approximation

Approximation

We often assume a continuous stochastic process
For example, consider a linear Gaussian AR(1) model:

$$X_{t+1} = \rho X_t + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma_\epsilon^2), \quad |\rho| < 1.$$

- The model has a unique stationary distribution:

$$\psi^* = N\left(\mu_X, \sigma_X^2\right), \quad \mu_X := 0, \quad \sigma_X^2 := \frac{\sigma_\epsilon^2}{1 - \rho^2}.$$

Approximation

We often assume a continuous stochastic process
For example, consider a linear Gaussian AR(1) model:

$$X_{t+1} = \rho X_t + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma_\epsilon^2), \quad |\rho| < 1.$$

- The model has a unique stationary distribution:

$$\psi^* = N\left(\mu_X, \sigma_X^2\right), \quad \mu_X := 0, \quad \sigma_X^2 := \frac{\sigma_\epsilon^2}{1 - \rho^2}.$$

Goal: Find a discrete approximation.

- This means finding a grid X and a transition matrix P such that ψ^* is close to the stationary distribution of the Markov chain with transition matrix P.

Approximation (continued)

Once we do it, we can work with the discrete approximation instead of the continuous model.

There are several ways to do it. For example:

1. Tauchen method (this one we will discuss in more detail)
2. Rouwenhurst method

Tauchen Method

1. We start by creating the grid X.
2. Pick a number $m > 0$ and an integer N.
3. Set: $x_1 = -m\sigma_x$ and $x_N = m\sigma_x$.
4. Set:
$$x_{i+1} = x_1 + \frac{i-1}{N-1}(x_N - x_1) \quad \text{for } i = 2, \ldots, N-1.$$
5. This is our state space x.

Tauchen Method

1. We start by creating the grid X.
2. Pick a number $m > 0$ and an integer N.
3. Set: $x_1 = -m\sigma_x$ and $x_N = m\sigma_x$.
4. Set:
$$x_{i+1} = x_1 + \frac{i-1}{N-1}(x_N - x_1) \quad \text{for } i = 2, \ldots, N-1.$$
5. This is our state space x.

Note:

$$\text{for } i = 1, \ldots, N-1: \quad x_{i+1} - x_i \equiv d$$

d is just a constant.

Tauchen Method (continued)

The space between states is allocated equally:

$$\begin{aligned}
X &\in (-\infty, x_1 + 0.5d) \rightarrow x_1 \\
X &\in (x_2 - 0.5d, x_2 + 0.5d) \rightarrow x_2 \\
&\vdots \\
X &\in (x_N - 0.5d, \infty) \rightarrow x_N
\end{aligned}$$

Tauchen Method (continued)

Now we need to find the transition matrix P.

- The idea is to discretize the conditional distribution for each point on the grid.

Tauchen Method (continued)

Now we need to find the transition matrix P.

- The idea is to discretize the conditional distribution for each point on the grid.
- Let F be the CDF of $N(0, \sigma_\epsilon^2)$. For $j \in \{2, \ldots, N-1\}$:

Tauchen Method (continued)

Now we need to find the transition matrix P.

- The idea is to discretize the conditional distribution for each point on the grid.
- Let F be the CDF of $N(0, \sigma_\epsilon^2)$. For $j \in \{2, \ldots, N-1\}$:

Tauchen Method (continued)

Now we need to find the transition matrix P.

- The idea is to discretize the conditional distribution for each point on the grid.
- Let F be the CDF of $N(0, \sigma_\epsilon^2)$. For $j \in \{2, \dots, N-1\}$:

$$
\begin{aligned}
P_{ij} := P(X_{t+1} = x_j \mid X_t = x_i) &= P\left(x_j - 0.5d < \rho x_i + \epsilon < x_j + 0.5d\right) \\
&= P\left(x_j - 0.5d - \rho x_i < \epsilon < x_j + 0.5d - \rho x_i\right) \\
&= F\left(x_j + 0.5d - \rho x_i\right) - F\left(x_j - 0.5d - \rho x_i\right)
\end{aligned}
$$

Tauchen Method (continued)

Now we need to find the transition matrix P.

- The idea is to discretize the conditional distribution for each point on the grid.
- Let F be the CDF of $N(0, \sigma_\epsilon^2)$. For $j \in \{2, \ldots, N-1\}$:

$$\begin{aligned} P_{ij} := P(X_{t+1} = x_j \mid X_t = x_i) &= P\left(x_j - 0.5d < \rho x_i + \epsilon < x_j + 0.5d\right) \\ &= P\left(x_j - 0.5d - \rho x_i < \epsilon < x_j + 0.5d - \rho x_i\right) \\ &= F\left(x_j + 0.5d - \rho x_i\right) - F\left(x_j - 0.5d - \rho x_i\right) \end{aligned}$$

For $j = 1$ and $j = N$:

$$\begin{aligned} P_{i1} &= F\left(x_1 + 0.5d - \rho x_i\right) \\ P_{iN} &= 1 - F\left(x_N + 0.5d - \rho x_i\right) \end{aligned}$$

Tauchen Method (final)

- Usually, we pick m equal to three. Our discrete state space is three times the standard deviation of the stationary distribution in each direction.

Tauchen Method (final)

- Usually, we pick m equal to three. Our discrete state space is three times the standard deviation of the stationary distribution in each direction.
- If $\rho \approx 1$ in the AR(1) process, it is difficult to obtain a discrete approximation that matches both the unconditional variance and the first-order auto-correlation, unless the number of grid points is large.

Tauchen Method (final)

- Usually, we pick m equal to three. Our discrete state space is three times the standard deviation of the stationary distribution in each direction.
- If $\rho \approx 1$ in the AR(1) process, it is difficult to obtain a discrete approximation that matches both the unconditional variance and the first-order auto-correlation, unless the number of grid points is large.
- A wide domain is needed because realizations can be far from the unconditional mean.

Tauchen Method (final)

- Usually, we pick m equal to three. Our discrete state space is three times the standard deviation of the stationary distribution in each direction.
- If $\rho \approx 1$ in the AR(1) process, it is difficult to obtain a discrete approximation that matches both the unconditional variance and the first-order auto-correlation, unless the number of grid points is large.
- A wide domain is needed because realizations can be far from the unconditional mean.
- Many points are needed because the conditional distributions are centered around the current realizations.

Tauchen Method (final)

- Usually, we pick m equal to three. Our discrete state space is three times the standard deviation of the stationary distribution in each direction.
- If $\rho \approx 1$ in the AR(1) process, it is difficult to obtain a discrete approximation that matches both the unconditional variance and the first-order auto-correlation, unless the number of grid points is large.
- A wide domain is needed because realizations can be far from the unconditional mean.
- Many points are needed because the conditional distributions are centered around the current realizations.

Tauchen Method (final)

- Usually, we pick m equal to three. Our discrete state space is three times the standard deviation of the stationary distribution in each direction.
- If $\rho \approx 1$ in the AR(1) process, it is difficult to obtain a discrete approximation that matches both the unconditional variance and the first-order auto-correlation, unless the number of grid points is large.
- A wide domain is needed because realizations can be far from the unconditional mean.
- Many points are needed because the conditional distributions are centered around the current realizations.

There exist a better alternative: the Rouwenhorst method guarantees to match these 2 moments even with a small number of points.