Markov chains I

Stochastic Processes 2024/25

Prof. José Niño Mora

Master in Statistics for Data Science

Universidad
Carlos III de Madrid
www.uc3m.es

Outline

- Introduction and motivation
- Discrete-time Markov chains
- Chapman–Kolmogorov equations
- Classification of states
- Limit theorems
- Extensions and applications

What is a Markov chain?

A Markov chain on the state space S (a discrete set) is a sequence of r.v. $X_0, X_1, \ldots$ taking values in S, which satisfies the Markov property:

$$\mathbb{P}\{X_{n+1} = j \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\} = \mathbb{P}\{X_{n+1} = j \mid X_n = i\}$$

- In words: the future depends on the past only through the present
- Common terminology in Markov chains: if $X_n = i$, we say that the chain visits or hits state i at time n
- A Markov chain is time-homogeneous if, for all n,

$$\mathbb{P}\{X_{n+1} = j \mid X_n = i\} = \mathbb{P}\{X_1 = j \mid X_0 = i\} = p_{ij}$$

We focus on such chains, defined by their state transition probabilities p_{ij}, through the transition (probability) matrix $\mathbf{P} = (p_{ij})_{i,j \in S}$

Transition matrices and stochastic matrices

- A stochastic matrix is a square matrix $\mathbf{P} = (p_{ij})_{i,j \in S}$ satisfying (i) $p_{ij} \geqslant 0$ for all i, j; and (ii) for each row i, $\sum_j p_{ij} = 1$

- The transition matrix $\mathbf{P}$ of a Markov chain is a stochastic matrix

Example: flipping of a fair coin (in the real world)

• Flipping a fair coin generates an i.i.d. sequence of Bernoulli r.v. (1: heads, 0: tails) with probab. $p = 1/2$, right?

• That's what we might think. But Diaconis (2007) carried out an experimental study of how people actually flip coins. He found otherwise

• The generated sequence $X_0, X_1, \ldots$ is NOT i.i.d. It is more accurately modeled by a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.51 & 0.49 \\ 0.49 & 0.51 \end{pmatrix}$$

Example: Vowels and consonants in poetry

• Andrei A. Markov (1856–1922) himself investigated the sequence of vowels and consonants in a very long poem by Alexander Pushkin

• He found that the generated sequence $X_0, X_1, \dots$ (with $X_n = 0$ if the nth letter is a vowel and $X_n = 1$ if it is a consonant) is NOT i.i.d. It is accurately modeled by a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.175 & 0.825 \\ 0.526 & 0.474 \end{pmatrix}$$

Example: Gambler's ruin

• In each round of a game a player either wins 1 €, w/ prob. p, or loses 1 €, w. prob. $1 - p$

• The gambler starts with k €

• The game stops when the player either loses all their money, or gains a total of n € (w/ $n > k$)

• The gambler's successive fortunes form a Markov chain $X_0, X_1, \ldots$ on $S = \{0, 1, \ldots, n\}$ with $X_0 = k$ and transition matrix $\mathbf{P} = (p_{ij})_{i,j \in S}$ given by

$$p_{ij} = \begin{cases} p, & \text{if } j = i + 1, 0 < i < n \\ 1 - p, & \text{if } j = i - 1, 0 < i < n \\ 1, & \text{if } i = j = 0 \text{ or } i = j = n \\ 0, & \text{otherwise} \end{cases}$$

Example: Birth–death chain

• A birth–death chain is a model for the evolving size of a population given by a Markov chain w/ state space $S = \{0, 1, \ldots\}$ and two types of transitions: births from i to $i + 1$ and deaths from i to $i - 1$

• The successive population sizes form a Markov chain $X_0, X_1, \ldots$ on S with transition matrix $\mathbf{P} = (p_{ij})_{i,j \in S}$ given by

$$p_{ij} = \begin{cases} q_i, & \text{if } j = i - 1, i > 0 \\ p_i, & \text{if } j = i + 1 \\ 1 - p_i - q_i, & \text{if } j = i \\ 0, & \text{otherwise} \end{cases}$$

where $p_i, q_i \geqslant 0$, $p_i + q_i \leqslant 1$, and $q_0 = 0$

Computing n-step transition probabilities

- How to compute the n-step transition probab. $\mathbb{P}\{X_n = j \mid X_0 = i\}$?
- We have

$$\begin{aligned}
\mathbb{P}\{X_n = j \mid X_0 = i\} &= \sum_k \mathbb{P}\{X_n = j \mid X_{n-1} = k, X_0 = i\}\mathbb{P}\{X_{n-1} = k \mid X_0 = i\} \\
&= \sum_k \mathbb{P}\{X_n = j \mid X_{n-1} = k\}\mathbb{P}\{X_{n-1} = k \mid X_0 = i\} \\
&= \sum_k p_{kj}\mathbb{P}\{X_{n-1} = k \mid X_0 = i\}
\end{aligned}$$

- That is, writing $p_{ij}^{(n)} \triangleq \mathbb{P}\{X_n = j \mid X_0 = i\}$, we have

$$p_{ij}^{(n)} = \sum_k p_{ik}^{(n-1)} p_{kj}, \quad n = 2, 3, \ldots$$

or, in matrix notation, $\mathbf{P}^{(n)} = \mathbf{P}\mathbf{P}^{(n-1)}$

- It follows that $\mathbf{P}^{(n)} = \mathbf{P}^n$ (matrix power)

Example: Flipping a fair coin (in the real world)

$$\mathbf{P} = \begin{pmatrix} 0.51 & 0.49 \\ 0.49 & 0.51 \end{pmatrix}, \ \mathbf{P}^2 = \begin{pmatrix} 0.5002 & 0.4998 \\ 0.4998 & 0.5002 \end{pmatrix}, \ \mathbf{P}^3 = \begin{pmatrix} 0.500004 & 0.499996 \\ 0.499996 & 0.500004 \end{pmatrix}$$

$$\mathbf{P}^4 \approx \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

Working with matrix powers in R

- Will use the script `utilitities.R` (in Aula Global), by R.P. Dobrow

```
P <- matrix(c(0.51, 0.49, 0.49, 0.51), nrow=2, byrow=T)
```

```
P %*% P
```

```
matrixpower(P, 3)
```

Example: Vowels and consonants in poetry

$$\mathbf{P} = \begin{pmatrix} 0.175 & 0.825 \\ 0.526 & 0.474 \end{pmatrix}, \quad \mathbf{P}^2 = \begin{pmatrix} 0.464575 & 0.535425 \\ 0.341374 & 0.658626 \end{pmatrix},$$

$$\mathbf{P}^3 = \begin{pmatrix} 0.362934 & 0.637066 \\ 0.406178 & 0.593822 \end{pmatrix}, \quad \mathbf{P}^4 = \begin{pmatrix} 0.39861 & 0.60139 \\ 0.383432 & 0.616568 \end{pmatrix}$$

$$\mathbf{P}^5 = \begin{pmatrix} 0.386088 & 0.613912 \\ 0.391416 & 0.608584 \end{pmatrix}, \quad \mathbf{P}^6 = \begin{pmatrix} 0.390483 & 0.609517 \\ 0.388613 & 0.611387 \end{pmatrix}$$

$\ldots$

$$\mathbf{P}^{14} = \begin{pmatrix} 0.389341 & 0.610659 \\ 0.389341 & 0.610659 \end{pmatrix}, \quad \mathbf{P}^{15} = \begin{pmatrix} 0.389341 & 0.610659 \\ 0.389341 & 0.610659 \end{pmatrix}$$

Chapman–Kolmogorov equations

• From the matrix identity $\mathbf{P}^{m+n} = \mathbf{P}^m \mathbf{P}^n$ we get

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)},$$

i.e.,

$$\mathbb{P}\{X_{m+n} = j \mid X_0 = i\} = \sum_k \mathbb{P}\{X_m = k \mid X_0 = i\} \mathbb{P}\{X_{m+n} = j \mid X_m = k\}$$

• Transitioning from i to j in $m+n$ steps is equivalent to transitioning from i to some state, say k, in m steps and then moving from that state to j in the remaining n steps

Computing the distribution of X_n

- How to compute $p_j^{(n)} \triangleq \mathbb{P}\{X_n = j\}$ when $\mathbb{P}\{X_0 = i\} = \alpha_i$ for all i (initial state distribution)
- We have

$$\begin{aligned}\mathbb{P}\{X_n = j\} &= \sum_i \mathbb{P}\{X_n = j \mid X_0 = i\}\mathbb{P}\{X_0 = i\} \\ &= \sum_i \alpha_i \mathbb{P}\{X_n = j \mid X_0 = i\}\end{aligned}$$

- Hence, writing $\mathbf{p}^{(n)} = (p_j^{(n)})_{j \in S}$ and $\boldsymbol{\alpha} = (\alpha_i)_{i \in S}$ as row vectors, have

$$\boxed{\mathbf{p}^{(n)} = \boldsymbol{\alpha}\mathbf{P}^n}$$

Example

Consider a Markov chain model with states $S = \{1, 2, 3\}$ for daily weather evolution in a city, w/ $1 =$ rain, $2 =$ snow and $3 =$ clear, and with trans. probab.

$$\mathbf{P} = \begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.6 & 0.3 \end{pmatrix}$$

Suppose that, for tomorrow, a 50 %-50 % chance of rain and snow is predicted. Find the probability that it will snow 2 days later

- We are asked for $p_2^{(2)}$ starting from $X_0 \sim \boldsymbol{\alpha} = (0.5, 0.5, 0)$
- We have

$$\begin{aligned}\mathbf{p}^{(2)} = \boldsymbol{\alpha}\mathbf{P}^2 &= (0.5, 0.5, 0) \begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.6 & 0.3 \end{pmatrix}^2 \\ &= (0.5, 0.5, 0) \begin{pmatrix} 0.12 & 0.72 & 0.16 \\ 0.11 & 0.76 & 0.13 \\ 0.11 & 0.72 & 0.17 \end{pmatrix} = (0.115, 0.74, 0.145)\end{aligned}$$

More on the Markov property

- For a Markov chain $\{X_n\}$ we have, for $m < n$,

$$\begin{aligned}\mathbb{P}\{X_{n+1} = j \mid X_0 = i_0, \ldots, X_{n-m-1} = i_{n-m-1}, X_{n-m} = i\} \\ = \mathbb{P}\{X_{n+1} = j \mid X_{n-m} = i\}\end{aligned}$$

Computing the joint distribution

• Consider a Markov chain $\{X_n\}$ with transition matrix $\mathbf{P}$ and initial distribution $X_0 \sim \boldsymbol{\alpha}$

• For $0 \leqslant n_1 < \cdots < n_k$, how to compute the joint distribution of $(X_{n_1}, \ldots, X_{n_k})$?

$$\mathbb{P}\{X_{n_1} = i_1, X_{n_2} = i_2, \ldots, X_{n_{k-1}} = i_{k-1}, X_{n_k} = i_k\}$$
$$= \mathbb{P}\{X_{n_1} = i_1\}\mathbb{P}\{X_{n_2} = i_2 \mid X_{n_1} = i_1\} \cdots \mathbb{P}\{X_{n_k} = i_k \mid n_{k-1} = i_{k-1}\}$$
$$= p_{i_1}^{(n_1)} p_{i_1 i_2}^{(n_2 - n_1)} \cdots p_{i_{k-1} i_k}^{(n_k - n_{k-1})}$$

Example

Danny's daily lunch choices are a Markov chain w/ states $1 =$ burrito, $2 =$ falafel, $3 =$ pizza and $4 =$ sushi, and with trans. probab.

$$\mathbf{P} = \begin{pmatrix} 0.0 & 0.5 & 0.5 & 0.0 \\ 0.5 & 0.0 & 0.5 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.6 \\ 0.0 & 0.2 & 0.6 & 0.2 \end{pmatrix}$$

On Sunday, he chooses lunch uniformly at random. Find the probab. that he chooses sushi on the following Wednesday and Friday, and pizza on Saturday

- Want $\mathbb{P}\{X_3 = 4, X_5 = 4, X_6 = 3\}$ w/ $X_0 \sim \boldsymbol{\alpha} = (0.25, 0.25, 0.25, 0.25)$
- Have $\mathbb{P}\{X_3 = 4, X_5 = 4, X_6 = 3\} = p_4^{(3)} p_{44}^{(2)} p_{43}$

$$\mathbf{p}^{(3)} = \boldsymbol{\alpha}\mathbf{P}^3 = \boldsymbol{\alpha} \begin{pmatrix} 0.100 & 0.285 & 0.405 & 0.210 \\ 0.225 & 0.160 & 0.405 & 0.210 \\ 0.384 & 0.024 & 0.232 & 0.360 \\ 0.108 & 0.250 & 0.430 & 0.212 \end{pmatrix} = (0.204, 0.180, 0.368, 0.248)$$

$$\mathbf{P}^2 = \begin{pmatrix} 0.45 & 0.00 & 0.25 & 0.30 \\ 0.20 & 0.25 & 0.25 & 0.30 \\ 0.00 & 0.32 & 0.56 & 0.12 \\ 0.34 & 0.04 & 0.22 & 0.40 \end{pmatrix}$$

Long-term behavior: Numerical evidence

- In any stochastic process, the long-term behavior is of interest
- We have seen in examples that $\mathbf{P}^n$ converges to a limiting matrix
- Furthermore, the rows of the limiting matrix are the same
- In Danny's daily lunch process:

$$\mathbf{P} = \begin{pmatrix} 0.0 & 0.5 & 0.5 & 0.0 \\ 0.5 & 0.0 & 0.5 & 0.0 \\ 0.4 & 0.0 & 0.0 & 0.6 \\ 0.0 & 0.2 & 0.6 & 0.2 \end{pmatrix}, \quad \mathbf{P}^2 = \begin{pmatrix} 0.45 & 0.00 & 0.25 & 0.30 \\ 0.20 & 0.25 & 0.25 & 0.30 \\ 0.00 & 0.32 & 0.56 & 0.12 \\ 0.34 & 0.04 & 0.22 & 0.40 \end{pmatrix}, \quad \mathbf{P}^3 = \begin{pmatrix} 0.100 & 0.285 & 0.405 & 0.210 \\ 0.225 & 0.160 & 0.405 & 0.210 \\ 0.384 & 0.024 & 0.232 & 0.360 \\ 0.108 & 0.250 & 0.430 & 0.212 \end{pmatrix}$$

$\ldots$

$$\mathbf{P}^{30} = \begin{pmatrix} 0.222222 & 0.163743 & 0.350877 & 0.263158 \\ 0.222222 & 0.163743 & 0.350877 & 0.263158 \\ 0.222222 & 0.163743 & 0.350877 & 0.263158 \\ 0.222222 & 0.163743 & 0.350877 & 0.263158 \end{pmatrix}$$

$$\mathbf{P}^{31} = \begin{pmatrix} 0.222222 & 0.163743 & 0.350877 & 0.263158 \\ 0.222222 & 0.163743 & 0.350877 & 0.263158 \\ 0.222222 & 0.163743 & 0.350877 & 0.263158 \\ 0.222222 & 0.163743 & 0.350877 & 0.263158 \end{pmatrix}$$

Simulating a Markov chain in R

- Use the function `markov(alpha, P, n)` to simulate $X_0, \ldots, X_n$ starting from $X_0 \sim \alpha$

- In coin flipping model (1 and 2 correspond to 0 and 1):
$1, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1$

- To obtain states 0 and 1: `markov(alpha, P, n, S)`, where `S = matrix(c(0, 1), nrow=1, byrow=T)`: $0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0$

Limiting distribution

Let $X_0, X_1, \ldots$ be a Markov chain w/ transition matrix $\mathbf{P}$. A limiting distribution for the Markov chain is a probability distribution $\boldsymbol{\pi}$ such that, for all i and j,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$$

- Consider the two-state Markov chain w/

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

- Does it have a limiting distribution? If so, what is it?

Example: two-state Markov chain

- Consider the two-state Markov chain w/

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

- Does it have a limiting distribution? If so, what is it?
- Have:

$$\mathbf{P}^n = \begin{pmatrix} \frac{q+p(1-p-q)^n}{p+q} & \frac{p-p(1-p-q)^n}{p+q} \\ \frac{q-q(1-p-q)^n}{p+q} & \frac{p+q(1-p-q)^n}{p+q} \end{pmatrix}$$

- If p and q are not both 0, nor both 1, then $|1-p-q| < 1$ and

$$\mathbf{P}^n = \begin{pmatrix} \frac{q+p(1-p-q)^n}{p+q} & \frac{p-p(1-p-q)^n}{p+q} \\ \frac{q-q(1-p-q)^n}{p+q} & \frac{p+q(1-p-q)^n}{p+q} \end{pmatrix} \to \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \end{pmatrix}$$

Stationary distribution

Let $X_0, X_1, \ldots$ be a Markov chain w/ transition matrix $\mathbf{P}$. A stationary distribution for the Markov chain is a probability distribution $\boldsymbol{\pi}$ such that

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P},$$

i.e.,

$$\pi_j = \sum_i \pi_i p_{ij}, \quad \text{for all } j$$

Result: If $\boldsymbol{\pi}$ is the limiting distribution of a Markov chain then it is a stationary distribution for it