Markov chains 4

Mihir Arjunwadkar

Centre for Modeling and Simulation
Savitribai Phule Pune University

PMF over the state space $\mathcal{X}$

Probability of observing state i at time t,

$$p_i(t) = P(X_t = i)$$

PMF over the state space $\mathcal{X} = \{1, \ldots, k\}$ at time t:

$$p(t) = (p_1(t), \ldots, p_k(t))$$

with

$$\sum_{i=1}^{k} p_i(t) = 1$$

Interpretation

- Imagine a very large *ensemble* (i.e., collection) of identical systems described/modeled by the same Markov chain (TPM).
- Each system evolves independently under the same *dynamics* described by the TPM.
- Take a snapshot of the ensemble at time t.
- Then $p_i(t) \approx$ the proportion of systems in state i.

Evolution of the PMF over the state space $\mathcal{X}$

How does $p(t)$ evolve over one time step?

Total probability of state i at time $t+1$:

$$
\begin{aligned}
p_i(t+1) &= P(X_{t+1}=i) = \sum_{j=1}^{k} P(X_{t+1}=i|X_t=j) \times P(X_t=j) \\
&= \sum_{j=1}^{k} \mathcal{P}_{ji} \times p_j(t) = \sum_{j=1}^{k} p_j(t) \times \mathcal{P}_{ji} = (p(t) \times \mathcal{P})_i.
\end{aligned}
$$

Hence

$$p(t+1) = p(t) \times \mathcal{P}.$$

Note

- $p(t) = (p_1(t), \ldots, p_k(t))$ is interpreted as a *row* vector.
- The equation is *linear* in p.
- $p(t) \times \mathcal{P}$ is a row vector - square matrix multiplication.

Evolution of the PMF over the state space $\mathcal{X}$

How does $p(t)$ evolve over time $t = 1, 2, \ldots$?

$$\begin{aligned} p(t) &= p(t-1) \times \mathcal{P} \\ &= (p(t-2) \times \mathcal{P}) \times \mathcal{P} = p(t-2) \times \mathcal{P}^2 \\ &= \ldots \\ &= p(1) \times \mathcal{P}^{t-1}. \end{aligned}$$

Note

- State space $\mathcal{X} = \{1, \ldots, k\}$, and $t = 1, 2, \ldots$.
- $p(t) = (p_1(t), \ldots, p_k(t))$ ::
  row vector representing the PMF over $\mathcal{X}$ at time t.
- The equation is *linear* in p.

Evolution of the PMF: Numerical example

$$
\begin{aligned}
\mathcal{P} &= \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \\
p(1) &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \\
p(2) &= \begin{bmatrix} 0.5000000 & 0.2500000 & 0.2500000 \end{bmatrix} \\
p(3) &= \begin{bmatrix} 0.4375000 & 0.1875000 & 0.3750000 \end{bmatrix} \\
p(4) &= \begin{bmatrix} 0.4062500 & 0.2031250 & 0.3906250 \end{bmatrix} \\
p(5) &\approx \begin{bmatrix} 0.4023438 & 0.1992188 & 0.3984375 \end{bmatrix} \\
p(6) &\approx \begin{bmatrix} 0.4003906 & 0.2001953 & 0.3994141 \end{bmatrix} \\
p(7) &\approx \begin{bmatrix} 0.4001465 & 0.1999512 & 0.3999023 \end{bmatrix} \\
p(8) &\approx \begin{bmatrix} 0.4000244 & 0.2000122 & 0.3999634 \end{bmatrix} \\
p(9) &\approx \begin{bmatrix} 0.4000092 & 0.1999969 & 0.3999939 \end{bmatrix} \\
&\vdots \\
p(\infty) &= \begin{bmatrix} 2/5 & 1/5 & 2/5 \end{bmatrix}
\end{aligned}
$$

Evolution of the PMF: Numerical example

$$
\begin{aligned}
\mathcal{P} &= \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \\
p(1) &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \\
p(2) &= \begin{bmatrix} 0.5000000 & 0.0000000 & 0.5000000 \end{bmatrix} \\
p(3) &= \begin{bmatrix} 0.3750000 & 0.2500000 & 0.3750000 \end{bmatrix} \\
p(4) &= \begin{bmatrix} 0.4062500 & 0.1875000 & 0.4062500 \end{bmatrix} \\
p(5) &\approx \begin{bmatrix} 0.3984375 & 0.2031250 & 0.3984375 \end{bmatrix} \\
p(6) &\approx \begin{bmatrix} 0.4003906 & 0.1992188 & 0.4003906 \end{bmatrix} \\
p(7) &\approx \begin{bmatrix} 0.3999023 & 0.2001953 & 0.3999023 \end{bmatrix} \\
p(8) &\approx \begin{bmatrix} 0.4000244 & 0.1999512 & 0.4000244 \end{bmatrix} \\
p(9) &\approx \begin{bmatrix} 0.3999939 & 0.2000122 & 0.3999939 \end{bmatrix} \\
&\vdots \\
p(\infty) &= \begin{bmatrix} 2/5 & 1/5 & 2/5 \end{bmatrix}
\end{aligned}
$$

Evolution of the PMF: Numerical example

$$\mathcal{P} = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

$$p(1) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$\begin{aligned}
p(2) &= \begin{bmatrix} 0.4166667 & 0.1666667 & 0.4166667 \end{bmatrix} \\
p(3) &= \begin{bmatrix} 0.3958333 & 0.2083333 & 0.3958333 \end{bmatrix} \\
p(4) &= \begin{bmatrix} 0.4010417 & 0.1979167 & 0.4010417 \end{bmatrix} \\
p(5) &= \begin{bmatrix} 0.3997396 & 0.2005208 & 0.3997396 \end{bmatrix} \\
p(6) &= \begin{bmatrix} 0.4000651 & 0.1998698 & 0.4000651 \end{bmatrix} \\
p(7) &= \begin{bmatrix} 0.3999837 & 0.2000326 & 0.3999837 \end{bmatrix} \\
p(8) &= \begin{bmatrix} 0.4000041 & 0.1999919 & 0.4000041 \end{bmatrix} \\
p(9) &= \begin{bmatrix} 0.3999990 & 0.2000020 & 0.3999990 \end{bmatrix} \\
&\vdots \\
p(\infty) &= \begin{bmatrix} 2/5 & 1/5 & 2/5 \end{bmatrix}
\end{aligned}$$

Suppose

- $\mathcal{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$ with $0 \leq p, q \leq 1$.
- $p(1) = \begin{bmatrix} \alpha & 1-\alpha \end{bmatrix}$ with $0 \leq \alpha \leq 1$.

It can be shown that[1]

- $\mathcal{P}^t = \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1-p-q)^t}{p+q} \begin{bmatrix} p & -p \\ -q & q \end{bmatrix}$
- $\lim_{t\to\infty} \mathcal{P}^t = \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix}$
- $\lim_{t\to\infty} p(t) = \frac{1}{p+q} \begin{bmatrix} q & p \end{bmatrix}$

[1] www.probabilitycourse.com/chapter11/11_2_6_stationary_and_limiting_distributions.php

Definition. A PMF π over the state space $\mathcal{X}$ is called *limiting* PMF if $\pi_i = \lim_{t\to\infty} p_i(t)$ for every $1 \leq i \leq k$; i.e., if the sequence of PMFs

$$\begin{aligned} p(1) &= (p_1(1), \ldots, p_k(1)) \\ p(2) &= (p_1(2), \ldots, p_k(2)) \\ &\vdots \end{aligned}$$

converges to $\pi = (\pi_1, \ldots, \pi_k)$.

Markov matrix $\mathcal{P}$	limiting PMF π
$\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$	$\begin{bmatrix} 2/5 & 1/5 & 2/5 \end{bmatrix}$
$\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$	$\frac{1}{p+q}\begin{bmatrix} q & p \end{bmatrix}$ for $p+q>0$

Given

- the linearity of the PMF evolution equation

$$p(t) = p(1)\mathcal{P}^{t-1}$$

- that $p(1)$ is an arbitrary PMF,

the only way in which a PMF $\pi = (\pi_1, \ldots, \pi_k)$ can be a limiting distribution is if

$$\lim_{t \to \infty} \mathcal{P}^t = \begin{bmatrix} \pi_1 & \pi_2 & \ldots & \pi_k \\ \pi_1 & \pi_2 & \ldots & \pi_k \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \ldots & \pi_k \end{bmatrix}.$$

Not every Markov chain necessarily has a limiting PMF

See examples in

https://galton.uchicago.edu/~yibi/teaching/stat317/2014/Lectures/Lecture4_6up.pdf

Stationary PMF of a Markov chain

Definition. A PMF π over the state space $\mathcal{X}$ is called *stationary* if

$$\pi = \pi \mathcal{P}.$$

Stationarity $\implies$ π is a left eigenvector of $\mathcal{P}$ with eigenvalue 1.

Markov matrix $\mathcal{P}$	stationary PMF π
$\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$	$\begin{bmatrix} 2/5 & 1/5 & 2/5 \end{bmatrix}$
$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$
$\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$	$\frac{1}{p+q}\begin{bmatrix} q & p \end{bmatrix}$ for $p+q>0$
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} p & 1-p \end{bmatrix}$ for every $0 \le p \le 1$ :: non-unique!

Stationary PMF need not be unique

$$\mathcal{P} = \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 \end{bmatrix}$$

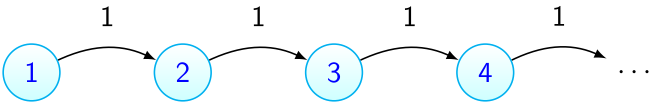
has a 2-fold degenerate unit eigenvalue. Linear combinations $\pi = w_1\pi_1 + w_2\pi_2$ of

$$\begin{aligned} \pi_1 &= \begin{bmatrix} 3/4 & 0 & 1/4 \end{bmatrix} \\ \pi_2 &= \begin{bmatrix} 2/3 & 0 & 1/3 \end{bmatrix} \end{aligned}$$

with $w_1, w_2 \geq 0, w_1 + w_2 = 1$ are all stationary PMFs of $\mathcal{P}$.

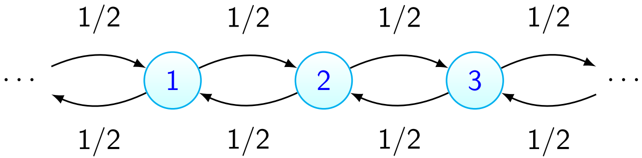
Example from Sec. 6.2 in http://wwwf.imperial.ac.uk/~ejm/M3S4/NOTES3.pdf
https://www.quora.com/In-what-case-do-Markov-Chains-not-have-a-stationary-distribution
https://galton.uchicago.edu/~yibi/teaching/stat317/2014/Lectures/Lecture4_6up.pdf

Not every Markov chain necessarily has a stationary PMF



https://www.quora.com/In-what-case-do-Markov-Chains-not-have-a-stationary-distribution

Not every Markov chain necessarily has a stationary PMF



https://galton.uchicago.edu/~yibi/teaching/stat317/2014/Lectures/Lecture4_6up.pdf

Relationship between stationary and limiting PMFs

Limiting PMFs are stationary, but not vice versa

https://galton.uchicago.edu/~yibi/teaching/stat317/2014/Lectures/Lecture4_6up.pdf

Markov chains with a unique stationary limiting PMF

A class of Markov chains called *ergodic* (*irreducible*) or *regular* chains.

For details:

- *Introduction to Probability* by Grinstead & Snell, American Mathematical Society (1997) Sec. 11.3
- http://wwwf.imperial.ac.uk/~ejm/M3S4/NOTES3.pdf
- https://www.math.ucdavis.edu/~gravner/MAT135A/resources/lecturenotes.pdf

This class is at the heart of Markov chain Monte Carlo (MCMC) methods.

Connection with sampling and "simulation"

Markov chain Monte Carlo (MCMC) methods

Sampling from (and integrating with respect to) an arbitrary high-dimensional PDF / PMF f is done by setting up a Markov chain that has f as its unique stationary limiting distribution.

Change in the probability of ith state between t and $t+1$

$$
\begin{aligned}
p_i(t+1) - p_i(t) &= \sum_{j=1}^{k} p_j(t)\mathcal{P}_{ji} - p_i(t) \times 1 \\
&= \sum_{j=1}^{k} p_j(t)\mathcal{P}_{ji} - p_i(t) \times \left(\sum_{j=1}^{k} \mathcal{P}_{ij} \right) \\
&= \sum_{j=1}^{k} p_j(t)\mathcal{P}_{ji} - \sum_{j=1}^{k} p_i(t)\mathcal{P}_{ij} \\
&= (\text{Net inflow into state } i) - (\text{Net outflow from state } i).
\end{aligned}
$$

If p is stationary, then $\textsc{lhs} = 0$ for every state i; i.e., net inflow into and net outflow out of state i are balanced.

Master equation and stationarity

Same argument, but starting with stationarity this time. Suppose $\pi = \pi\mathcal{P}$. Then

$$\begin{aligned} \pi_i &= \sum_{j=1}^{k} \pi_j \mathcal{P}_{ji} \\ 1 \times \pi_i &= \sum_{j=1}^{k} \pi_j \mathcal{P}_{ji} \\ \left(\sum_{j=1}^{k} \mathcal{P}_{ij}\right) \times \pi_i &= \sum_{j=1}^{k} \pi_j \mathcal{P}_{ji} \\ \sum_{j=1}^{k} \mathcal{P}_{ij}\pi_i &= \sum_{j=1}^{k} \pi_j \mathcal{P}_{ji} \\ \text{Net probability outflow from state } i &= \text{Net probability inflow into state } i \end{aligned}$$

Stationarity of π can thus be viewed as balance of inflow and outflow of probabilities. Rearranging the above, one gets the *Master equation*:

$$\sum_{j=1}^{k} \left(\mathcal{P}_{ij}\pi_i - \pi_j \mathcal{P}_{ji}\right) = 0 \text{ for each } i.$$

Master equation

$$\sum_{j=1}^{k} (\mathcal{P}_{ij}\pi_i - \pi_j\mathcal{P}_{ji}) = 0 \text{ for each } i.$$

One possible way in which the Master equation will hold true is if each term above is separately $= 0$; i.e.,

$$\mathcal{P}_{ij}\pi_i - \pi_j\mathcal{P}_{ji} = 0 \text{ for each pair of states } i, j.$$

This is called the condition of *detailed balance*: It is *sufficient* to ensure stationarity of f, but not *necessary*.

This condition is at the heart of Markov chain Monte Carlo (MCMC) methods.