MTH 565

Lectures 1 - 8

Yevgeniy Kovchegov
Oregon State University

Topics:

- Markov chains.
- Stationary distribution.
- Recurrent and transient states.

Markov chains.

Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ with values in a discrete state space S.

The sequence $\{X_t\}_{t=0,1,\ldots}$ is said to be a discrete time Markov chain if it satisfies the following property, known as Markov property:

$$P\big(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1,\ X_0 = i_0\big) \ = \ P(X_{t+1} = j \mid X_t = i)$$

A Markov chain $\{X_t\}_{t=0,1,\ldots}$ is said to be time homogeneous if

$$P(X_{t+1} = j \mid X_t = i) = p(i,j) \qquad \text{for all } t = 0, 1, 2, \ldots$$

Markov chains.

Consider a time homogeneous Markov chain $X_0, X_1, X_2, \ldots$ with a discrete state space S and transition probabilities

$$P(X_{t+1} = j \mid X_t = i) = p(i,j) \qquad \text{for all } t = 0, 1, 2, \ldots$$

Matrix (operator) $P = \Big(p(i,j)\Big)_{i,j\in S}$ is called the transition probability matrix (operator). Then

$$\sum_{j\in S} p(i,j) = 1 \quad \forall i \in S$$

Example. Consider $S = \{0, 1\}$ (two states) and

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

Birth-and-death chain. Consider state space

$$S = \{0, 1, 2, \dots\}$$

and a Markov chain $\{X_t\}_{t=0,1,\dots}$ on S with transition probabilities

$$p(i, i+1) = p_i, \quad p(i, i-1) = q_i, \quad \text{and} \quad p(i, i) = r_i$$

satisfying $q_0 = 0$ and $q_i + r_i + p_i = 1 \quad \forall i$

$$P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & \ddots \\ 0 & q_2 & r_2 & p_2 & \ddots \\ 0 & 0 & q_3 & r_3 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

This is a Markov chain with only nearest neighbor transitions.

Distribution of X_t.

Consider a time homogeneous Markov chain $\{X_t\}_{t=0,1,\dots}$ on a discrete state space S.

Define the distribution of X_t as a row vector of probability mass functions

$$\mu_t = \Big(P(X_t = i)\Big)_{i\in S}$$

ordered in the same way as enumeration of S used in $P = \Big(p(i,j)\Big)_{i,j\in S}$.

Proposition.

$$\mu_{t+1} = \mu_t P, \qquad \text{and therefore} \qquad \mu_t = \mu_0 P^t.$$

Distribution of X_t.

Define the distribution of X_t as a row vector of probability mass functions

$$\mu_t = \Big(P(X_t = i)\Big)_{i\in S}$$

ordered in the same way as enumeration of S used in $P = \Big(p(i,j)\Big)_{i,j\in S}$.

Proposition.

$$\mu_{t+1} = \mu_t P, \qquad \text{and therefore} \qquad \mu_t = \mu_0 P^t.$$

Proof. Denote by $\mu_t(j)$ the j-th coordinate of μ_t. Then

$$\mu_{t+1}(j) = P(X_{t+1} = j) = \sum_{i\in S} P(X_t = i)P(X_{t+1} = j \,|\, X_t = i) = \sum_{i\in S} \mu_t(i)p(i,j)$$

□

Chapman-Kolmogorov equation.

Consider a time homogeneous Markov chain $\{X_t\}_{t=0,1,\ldots}$ on a discrete state space S. For an integer $s \geq 0$, let

$$p_s(i,j) = P(X_{t+s} = j \mid X_t = i) \quad \forall t \geq 0$$

be the s step transition probabilities, and let

$$P_s = \Big(p_s(i,j)\Big)_{i,j\in S}$$

Observe that $P_0 = I$ and $P_1 = P$.

Chapman-Kolmogorov equation.

$$P_{s+t} = P_s\,P_t \quad \forall s,t \geq 0, \qquad \text{and therefore} \qquad P_t = P^t$$

Proof. $p_{s+t}(i,j) = \sum_{k\in S} p_s(i,k)p_t(k,j)$ follows from

$$P(X_{s+t} = j \mid X_0 = i) = \sum_{k\in S} P(X_s = k \,|X_0 = i)P(X_{s+t} = j\,|X_s = k) \quad \square$$

Stationary distribution.

For a homogeneous Markov chain with the transition probability matrix $P = \Big(p(i,j)\Big)_{i,j\in S}$, the stationary distribution (aka 'equilibrium distribution') π is defined as follows:

$$\pi P = \pi \qquad \Leftrightarrow \qquad \sum_{i\in S} \pi(i)p(i,j) = \pi(j).$$

Thus $\sum_i \pi(i)p(i,j) = \pi(j)\sum_i p(j,i)$, and for any state $j \in S$,

$$\sum_{i:\ i\neq j} \pi(i)p(i,j) = \sum_{i:\ i\neq j} \pi(j)p(j,i).$$

Thus when restated in terms of traffic flow, the influx to the state j is equal to outflow from j, for each j. Thus the distribution stays unchanged.

Stationary distribution.

Consider a homogeneous Markov chain over a discrete state space S with transition matrix P.

Definition. A homogeneous Markov chain is said to be irreducible if for any pair $x, y \in S$, there exists an integer $k \geq 1$ such that $p_k(x, y) > 0$.

Definition. A homogeneous Markov chain is said to be aperiodic if

$$\mathbf{gcd}\{k \geq 1 \ : \ p_k(x, x) > 0\} = 1 \qquad \forall x \in S$$

Notice that if the Markov chain is irreducible, then

$$\mathbf{gcd}\{k \geq 1 \ : \ p_k(x, x) > 0\} = \mathbf{gcd}\{k \geq 1 \ : \ p_k(y, y) > 0\}$$

for any pair $x, y \in S$.

Irreducible Markov chains.

Definition. A homogeneous Markov chain is said to be irreducible if for any pair $x, y \in S$, there exists an integer $k \geq 1$ such that $p_k(x, y) > 0$.

Example (reducible). Consider $S = \{1, 2, 3, 4, 5\}$ and

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.2 & 0.5 \\ 0 & 0 & 0.5 & 0.3 & 0.2 \\ 0 & 0 & 0.2 & 0.5 & 0.3 \end{pmatrix}$$

$\pi' = (1/2,\ 1/2,\ 0,\ 0,\ 0)$ and $\pi'' = (0,\ 0,\ 1/3,\ 1/3,\ 1/3)$

are both stationary distributions, as well as all

$$\pi = \lambda \pi' + (1 - \lambda) \pi'' \qquad \lambda \in [0, 1].$$

Aperiodic Markov chains.

Definition. A homogeneous Markov chain is said to be aperiodic if

$$\mathbf{gcd}\{k \geq 1 \, : \, p_k(x,x) > 0\} = 1 \qquad \forall x \in S$$

Example (periodic). Consider $S = \{0,1\}$ (two states) and

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Here

$$\mathbf{gcd}\{k \geq 1 \, : \, p_k(x,x) > 0\} = 2 \qquad \forall x \in S$$

The stationary distribution $\pi = (0.5, 0.5)$ is unique.

However, $\lim_{t\to\infty} p_t(x,y)$ does not exist.

Stationary distribution.

The following is version of ergodicity result for Markov chains over a finite state space S.

Theorem (Ergodicity). Consider an irreducible homogeneous Markov chain over a finite state space S. Then there exists a unique stationary distribution π. Furthermore, if the Markov chain is aperiodic, then

$$\lim_{t\to\infty} p_t(x,y) = \pi(y) \qquad \forall x, y \in S.$$

In other words, for any distribution μ_0 of X_0,

$$\mu_t = \mu_0 P^t \longrightarrow \pi \qquad \text{as} \qquad t \to \infty.$$

Example. Consider $S = \{0, 1\}$ (two states) and

$$P = \begin{pmatrix} 0.2 & 0.8 \\ 1 & 0 \end{pmatrix}$$

Stationary distribution.

Majorization proof of convergence. Consider an irreducible and aperiodic homogeneous Markov chain over a finite state space S. Let π denote the unique stationary distribution.

We will prove that for any distribution μ_0 of X_0,

$$\mu_t = \mu_0 P^t \longrightarrow \pi \qquad \text{as} \qquad t \to \infty.$$

Without loss of generality let $S = \{1, 2, \ldots, d\}$.

Can show that irreducibility and aperiodicity imply

$$\exists m \in \mathbb{N} \quad \text{such that} \quad P^m > 0, \quad \text{i.e.} \quad P^m(i,j) > 0 \;\; \forall i, j \in S.$$

Write

$$P^m = \begin{bmatrix} — & r_1 & — \\ — & r_2 & — \\ \vdots & \vdots & \vdots \\ — & r_d & — \end{bmatrix}, \quad \text{where row vectors } r_j > 0 \text{ (positive coord.)}$$

Majorization proof of convergence (continued).

$\exists m \in \mathbb{N}$ such that $P^m > 0$, i.e. $P^m(i,j) > 0 \ \ \forall i,j \in S$.

Write

$$P^m = \begin{bmatrix} — & r_1 & — \\ — & r_2 & — \\ \vdots & \vdots & \vdots \\ — & r_d & — \end{bmatrix}, \quad \text{where row vectors } r_j > 0 \ \text{(positive coord.)}$$

Thus, there exists $\epsilon \in (0,1)$ such that

$$\epsilon\pi \le r_j \qquad \text{(coordinatewise)} \qquad \forall j \in S.$$

Claim: for any distribution μ_0 and $k = 1, 2, \ldots,$

$$\mu_{km} = \big(1 - (1-\epsilon)^k\big)\pi + \nu_k,$$

where $\nu_k \ge 0$ (coordinatewise) and $\sum_{i=1}^{d} \nu_k(i) = (1-\epsilon)^k$.

Majorization proof of convergence (continued).

$\exists \epsilon \in (0,1)$ s.t. $\epsilon\pi \leq r_j$ (coordinatewise) $\forall j \in S$.

Claim: for any distribution μ_0 and $k = 1, 2, \ldots,$

$$\mu_{km} = \big(1 - (1-\epsilon)^k\big)\pi + \nu_k,$$

where $\nu_k \geq 0$ (coordinatewise) and $\sum_{i=1}^{d} \nu_k(i) = (1-\epsilon)^k$.

<u>Case $k = 1$:</u>

$$\mu_m = \mu_0 P^m = \sum_{j=1}^{d} \mu_0(j) r_j = \epsilon\pi \sum_{j=1}^{d} \mu_0(j) + \sum_{j=1}^{d} \mu_0(j)(r_j - \epsilon\pi)$$

Hence, $\mu_m = \epsilon\pi + \nu_1,$ where $\nu_1 = \sum_{j=1}^{d} \mu_0(j)(r_j - \epsilon\pi)$

satisfies $\nu_1 \geq 0$ (coordinatewise) and $\sum_{i=1}^{d} \nu_1(i) = 1 - \epsilon$.

Case $k = 1$: $\mu_m = \mu_0 P^m = \epsilon\pi + \nu_1$.

Induction step: Suppose $\mu_{km} = \big(1 - (1-\epsilon)^k\big)\pi + \nu_k$, where $\nu_k \geq 0$ (coordinatewise) and $\sum_{i=1}^{d} \nu_k(i) = (1-\epsilon)^k$.

$$\mu_{(k+1)m} = \mu_{km} P^m = \big(1 - (1-\epsilon)^k\big)\pi + (1-\epsilon)^k \mu_0' P^m,$$

where $\mu_0' = (1-\epsilon)^{-k}\nu_k$ is a distribution.

By case $k = 1$, we have $\mu_0' P^m = \epsilon\pi + \nu_1'$, where $\nu_1' \geq 0$ (coordinatewise) and $\sum_{i=1}^{d} \nu_1'(i) = 1 - \epsilon$. Thus,

$$\mu_{(k+1)m} = \big(1 - (1-\epsilon)^k\big)\pi + (1-\epsilon)^k(\epsilon\pi + \nu_1') = \big(1 - (1-\epsilon)^{k+1}\big)\pi + \nu_{k+1},$$

where $\nu_{k+1} = (1-\epsilon)^k \nu_1'$ satisfies

$$\nu_{k+1} \geq 0 \quad \text{and} \quad \sum_{i=1}^{d} \nu_{k+1}(i) = (1-\epsilon)^k \sum_{i=1}^{d} \nu_1'(i) = (1-\epsilon)^{k+1}.$$

We established the following claim.

Claim: for any distribution μ_0 and $k = 1, 2, \ldots,$

$$\mu_{km} = \mu_0 P^{km} = \big(1 - (1-\epsilon)^k\big)\pi + \nu_k,$$

where $\nu_k \geq 0$ (coordinatewise) and $\sum_{i=1}^{d} \nu_k(i) = (1-\epsilon)^k.$

Thus, since $\epsilon \in (0, 1)$,

$$\lim_{k\to\infty} \mu_0 P^{km} = \lim_{k\to\infty} \Big[\big(1 - (1-\epsilon)^k\big)\pi + \nu_k\Big] = \pi.$$

for any distribution μ_0.

Hence, for any $r = 0, 1, \ldots, m-1$, we have

$$\lim_{k\to\infty} \mu_{km+r} = \lim_{k\to\infty} \mu_r P^{km} = \pi.$$

Therefore, we have proven that for any distribution μ_0,

$$\lim_{t\to\infty} \mu_t = \pi.$$

Stationary distribution.

Example. Consider $S = \{0, 1\}$ (two states) and

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \qquad p, q \in (0, 1)$$

Eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = 1 - p - q$.

Matrix theory:
For an eigenvalue λ_i of P, the **left eigenvector** u_i is a nonzero row vector satisfying $u_i P = \lambda_i u_i$.

The **right eigenvector** v_i is a nonzero column vector satisfying $P v_i = \lambda_i v_i$.

Spectral Theorem: for 'reversible' Markov chain, one can construct biorthogonal ($u_i v_j = 0$ for $i \neq j$) pair of left and right eigenbases $\{u_i\}_{i \in S}$ and $\{v_i\}_{i \in S}$, and

$$P = \sum_i \lambda_i \frac{v_i u_i}{u_i v_i}, \quad \text{and therefore,} \quad P^t = \sum_i \lambda_i^t \frac{v_i u_i}{u_i v_i}$$

Stationary distribution.

Example. Consider $S = \{0, 1\}$ (two states) and

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \qquad p, q \in (0, 1)$$

Eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = 1 - p - q$.

Left eigenvectors: $u_1 = \pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ and $u_2 = (1, -1)$

Right eigenvectors: $v_1 = \mathbf{1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} p \\ -q \end{pmatrix}$

Spectral Theorem:

$$P^t = \sum_i \lambda_i^t \frac{v_i u_i}{u_i v_i} = \mathbf{1}\pi + (1-p-q)^t \frac{v_2 u_2}{p+q}$$

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

Stationary distribution.

Example. Consider $S = \{0,1\}$ (two states) and

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \qquad p, q \in (0,1)$$

Eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = 1-p-q$.

Left eigenvectors: $u_1 = \pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ and $u_2 = (1, -1)$

Right eigenvectors: $v_1 = \mathbf{1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} p \\ -q \end{pmatrix}$

Spectral Theorem:

$$P^t = \sum_i \lambda_i^t \frac{v_i u_i}{u_i v_i} = \mathbf{1}\pi + (1-p-q)^t \frac{v_2 u_2}{p+q} \longrightarrow \mathbf{1}\pi = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}$$

as $t \to \infty$. Hence, $\lim_{t\to\infty} p_t(x,y) = \pi(y) \quad \forall x, y \in S$.

Stationary distribution.

Theorem (Ergodicity). Consider an irreducible homogeneous Markov chain over a finite state space S. Then there exists a unique stationary distribution π. Furthermore, if the Markov chain is aperiodic, then

$$\lim_{t\to\infty} p_t(x,y) = \pi(y) \qquad \forall x,y \in S.$$

In other words, for any distribution μ_0 of X_0,

$$\mu_t = \mu_0 P^t \longrightarrow \pi \qquad \text{as} \qquad t \to \infty.$$

The **linear algebraic proof** uses the Perron-Frobenius Theorem to prove the above version of ergodicity for Markov chains with finite state space S.

Perron-Frobenius Theorem. If $A = \big(a_{i,j}\big) \in \mathbb{R}^{d\times d}$ is an irreducible nonnegative matrix, then there exists a real eigenvalue $\rho \geq 0$, called Perron-Frobenius eigenvalue, such that

- $|\lambda| \leq \rho$ for any other eigenvalue λ.

- ρ is simple (i.e., its left and right eigenspaces are one-dimensional), and it has left and right eigenvectors (u and v) whose coordinates are all positive.

- $\min\limits_i \sum\limits_{j=1}^{d} a_{i,j} \;\leq\; \rho \;\leq\; \max\limits_i \sum\limits_{j=1}^{d} a_{i,j}$.

- A has exactly h eigenvalues in $\mathbb{C}$ with absolute value ρ:

$$\rho\, e^{2\pi i k/h} \qquad k = 0, \ldots, h-1,$$

where h is the period of A.

- $\lim\limits_{n\to\infty} \frac{1}{\rho^n} A^n = \frac{vu}{uv}$, where $uA = \rho u$ and $Av = \rho v$.

Stationary distribution.

Matrix A is irreducible if $\forall i,j\ \exists m \in \mathbb{N}$ such that $\left(A^m\right)_{i,j} > 0$.

Consider an irreducible homogeneous Markov chain over a finite state space S. Then P is an irreducible nonnegative matrix. Hence, Perron-Frobenius Theorem applies.

The Perron-Frobenius eigenvalue ρ of P satisfies

$$1 = \min_i \sum_{j=1}^{d} p(i,j) \le \rho \le \max_i \sum_{j=1}^{d} p(i,j) = 1.$$

Hence, $\rho = 1$.

Finally, $\rho = 1$ is a simple eigenvalue, and it has left and right eigenvectors with all positive coordinates:

There exists a unique distribution π such that $\pi P = \pi$.

We know that $P\mathbf{1} = \mathbf{1}$ as the rows of P add up to 1.

Stationary distribution.

Consider an irreducible homogeneous Markov chain over a finite state space S. Then P is an irreducible nonnegative matrix. Hence, Perron-Frobenius Theorem applies with $\rho = 1$.

There exists a unique distribution π such that $\pi P = \pi$.

We know that $P\mathbf{1} = \mathbf{1}$ as the rows of P add up to 1.

For an irreducible homogeneous Markov chain, matrix P has period

$$h = \mathbf{gcd}\{k \geq 1 \, : \, p_k(x,x) > 0\} \quad \text{same for all} \quad x \in S.$$

By Perron-Frobenius Theorem, P has exactly h eigenvalues with absolute value $\rho = 1$:

$$e^{2\pi i k/h} \qquad k = 0, \ldots, h-1.$$

Stationary distribution.

For an irreducible homogeneous Markov chain, matrix P has period

$$h = \mathbf{gcd}\{k \geq 1 \ : \ p_k(x,x) > 0\} \quad \text{same for all} \quad x \in S.$$

By Perron-Frobenius Theorem, P has exactly h eigenvalues with absolute value $\rho = 1$:

$$e^{2\pi i k/h} \qquad k = 0, \ldots, h-1.$$

If P is aperiodic ($h = 1$),

$$\lim_{t\to\infty} P^t = \frac{\mathbf{1}\pi}{\pi\mathbf{1}} = \mathbf{1}\pi.$$

If $h > 1$, P^t does not have a limit as $t \to \infty$.

Recurrent and transient states.

We will use the following notations:

$P_x(A) = P(A \,|\, X_0 = x)$ and $E_x[Y] = E[Y \,|\, X_0 = x]$.

For $x \in S$, consider the first hitting time

$$T_x = \min\{n \geq 1 \,:\, X_n = x\}.$$

Definition. A state $x \in S$ is said to be recurrent if

$$P_x(T_x < \infty) = 1.$$

A recurrent state $x \in S$ is positive recurrent if

$$E_x[T_x] < \infty.$$

Otherwise it is null recurrent.

Definition. A state $x \in S$ is said to be transient if

$$P_x(T_x < \infty) < 1.$$

Stationary distribution.

Definition. A state $x \in S$ is said to be positive recurrent if

$$E_x[T_x] < \infty$$

The following is a version of ergodicity theorem for a general discrete state space S.

Theorem (Ergodicity). Consider an irreducible homogeneous Markov chain over a discrete state space S. If all of its states are positive recurrent, then there exists a unique stationary distribution π such that

$$\pi(x) = \frac{1}{E_x[T_x]}.$$

Furthermore, if the Markov chain is aperiodic,

$$\lim_{t \to \infty} p_t(x, y) = \pi(y) \qquad \forall x, y \in S.$$