ECE 250: Stochastic Processes: Week #9

Instructor: Behrouz Touri

Outline:

- Discrete Time, Discrete State Markov Chains
- Homogeneous Markov Chains
- Aperiodicity, Irreducibility, and Ergodicity

Informal Definition and Examples

- Markov Chain: A **random process** that the probability distribution given past **only depends** on the information from the latest time.
- Almost all the discussed examples were Markov chains:
  a. *Simple Random Walk (discrete-time continuous space)*:
  $$\Pr(X_{k+1} \in A \mid X_k, \ldots, X_1) = \Pr(X_{k+1} \in A \mid X_k).$$
  More generally
  $$\Pr(X_{k+1} \in A \mid X_{k_i}, \ldots, X_{k_1}) = \Pr(X_{k+1} \in A \mid X_{k_i}),$$
  for any $k_1 < k_1 < \ldots < k_i < k$.
  b. *Polar Code Dynamics (discrete-time discrete-space (infinitely many though))*: The same relationship as above holds.
  c. *Random Walk on a Graph (discrete-time finite-space)*: Given a (directed or undirected) graph $G = (V, E)$ where $V = \{1, \ldots, d\}$, we define a random walk on G started at vertex i, to be a discrete-time random process $\{X_k\}$ on V (i.e., $\Pr(X_k \in V) = 1$) with:
     a. $\Pr(X_1 = i) = 1$.
     b. For any k,
     $$\Pr(X_{k+1} = j \mid X_k = i_k, X_{k-1} = i_{k-1}, \ldots) = \frac{1}{d_{i_k}},$$
     where d_ℓ is the number of out-neighbors of node ℓ.
- There are other variations: continuous-time discrete-space (pandemics on continuous time), and continuous-time continuous-space (Brownian motion).
- In this course: we focus on discrete-time discrete-space Markov chains.

DT-DS Markov Chains: Formal Definition

- **Definition**: We say that a (DT) random process $\{X_k\}$ is a Markov chain over a discrete-space if

  1. X_ks are all discrete random variables, i.e., $\mathbf{Pr}(X_k \in S) = 1$ for all k and a countable set S, and
  2. for all $k \geq 1$, and all $s_1, \ldots, s_k \in S$, and all $s \in S$:

$$\Pr(X_{k+1} = s \mid X_k = s_k, \ldots, X_1 = s_1) = \Pr(X_{k+1} = s \mid X_k = s_k). \tag{1}$$

- S is called the state space and each $s \in S$ is called a state. Relation (2) is called *Markov property*.

- If S is finite, $\{X_k\}$ is called a discrete-time finite-state Markov chain.

- Condition (2.) can be replaced (and is equivalent) to the following stronger condition: 2'. for all $i \geq 1$, all $1 \leq k_1 < k_2 < \ldots < k_i \leq k$, and all $s_1, \ldots, s_i, s \in S$:

$$\Pr(X_{k+1} = s \mid X_{k_i} = s_i, \ldots, X_{k_1} = s_1) = \Pr(X_{k+1} = s \mid X_{k_i} = s_i). \tag{2}$$

DT-DS Markov Chains

- From this point on assume S is a countable set with elements, $S = \{1, \ldots, d\}$. Unless otherwise stated, all the following discussions hold for $d = \infty$ (but countable) but for convenient we assume d is finite.
- For any k, let π_k be the (marginal) probability mass function X_k, i.e.,

$$\pi_k(i) = \Pr(X_k = i).$$

  Note that the vector π_k is non-negative and $\sum_{i=1}^{d} \pi_k(i) = 1$. Such a vector is called a stochastic (sometimes probability) vector. It is convenient to assume that π_k is a **row** vector.
- For any $1 \leq k < n$, define the matrix (array)

$$P_{k,n}(i,j) = \Pr(X_n = j \mid X_k = i).$$

- $P_{k,n}$ is called the transition matrix of the MC from time k to time n.
- We also (naturally) define $P_{k,k} := I$, where I is the $d \times d$ identity matrix.

DT-DS Markov Chains: Properties of Transition Matrices

- Definition: We say that a $d \times d$ matrix A is a row-stochastic matrix if (i) A is non-negative, and (ii) $A\mathbf{1} = \mathbf{1}$ (or each row sums up to one).
- Properties of the transition matrices:
  - **Row-stochastic**: For any $k \leq n$, $P_{k,n}$ is a row-stochastic matrix: The non-negativeness follows from the definition. Also, each row adds up to one:
  $$\sum_{j=1}^{d} P_{k,n}(i,j) = \sum_{j=1}^{d} \Pr(X_n = j \mid X_k = i) = 1.$$
  - For any $k \leq n$, we have:
  $$\pi_n = \pi_k P_{k,n}.$$
  This follow from the fact:
  $$\begin{aligned} \pi_n(j) = \Pr(X_n = j) &= \sum_{i=1}^{d} \Pr(X_n = j, X_k = i) \\ &= \sum_{i=1}^{d} \Pr(X_n = j \mid X_k = i) \Pr(X_k = i) \\ &= [\pi_k P_{k,n}]_j. \end{aligned}$$

DT-DS Markov Chains: Properties of Transition Matrices cont.

- Properties of the transition matrices cont.:
  - **Semigroup property**: For any $k \leq m \leq n$, we have:

$$P_{k,n} = P_{k,m}P_{m,n}.$$

To show this, let i, j being fixed. Then, we have

$$\begin{aligned}
P_{k,n}(i,j) = \Pr(X_n = j \mid X_k = i) &= \sum_{\ell=1}^{d} \Pr(X_n = j, X_m = \ell \mid X_k = i) \\
&= \sum_{\ell=1}^{d} \Pr(X_n = j \mid X_m = \ell, X_k = i) \Pr(X_m = \ell \mid X_k = i) \\
\text{(by Markov property)} &= \sum_{\ell=1}^{d} \Pr(X_n = j \mid X_m = \ell) \Pr(X_m = \ell \mid X_k = i) \\
&= \sum_{\ell=1}^{d} P_{k,m}(i,\ell) P_{m,n}(\ell,j) \\
&= [P_{k,m}P_{m,n}]_{i,j}.
\end{aligned}$$

In probability, this property is widely known as *Chapman-Kolmogorov* equation.

  - For DT-DS Markov chains, the second property, and the Chapman-Kolmogorov property imply:

$$\pi_k = \pi_1 P_{1,k} = \pi_1 P_{1,2} P_{2,k} = \cdots = \pi_1 P_{1,2} P_{2,3} \cdots P_{k-1,k}.$$

DT-DS Homogeneous Markov Chains

Definition: We say that a DT-DS Markov chain $\{X_k\}$ is a (time-)homogeneous Markov chain if $P_{1,2} = P_{m,m+1}$ does not depend on m.

- Denote $P := P_{m,m+1}$. P is called the one-step transition matrix of the underlying Homogeneous Markov chain.
- P is a row-stochastic matrix.
- For Homogeneous Markov chains, we have $P_{m,n} = P^{n-m}$.
- A (DT FS) Homogeneous Markov chain can be viewed as a random walk on a weighted directed graph with d vertices where the weights are given by the matrix P.
- Abusing an abuse of notation, P is also called the transition matrix for a Homogeneous Markov chain.
- **Definition**: We say that a stochastic vector $\pi^* \in \mathbb{R}^d$ is a stationary distribution for a Homogeneous[1] Markov chain if $\pi^* = \pi^* P$.
- Interpretation: Note that if $\pi_k = \pi^*$ for some time k, then $\pi_n = \pi^*$ for all $n \geq k$ and hence, the term *stationary distribution*.
- A stationary distribution π^* is then a **stochastic** vector that is a (left) eigenvector corresponding to eigenvalue $\lambda = 1$ for the one step transition matrix P.

[1] The same definition holds for time-inhomogeneous Markov chains: we say that (a stochastic vector) π^* is a stationary distribution if $\pi^* P_{m,n} = \pi^*$ for any $m \leq n$.

Existence of Stationary Distribution

Theorem 1. *Every DT finite and Homogeneous Markov chain admits a stationary distribution π^*.*

Proof.

- Let P be the one-step transition matrix of the Markov chain.
- Define the function $f : \mathbb{R}^d \to \mathbb{R}^d$ by $f(x) = xP$.
- Let $S = \{x \in \mathbb{R}^d \mid x_i \geq 0, x\mathbf{1} = 1\}$ be the set of stochastic vectors in $\mathbb{R}^d$.
- This function is continuous and maps any stochastic vector π to a stochastic vector $\hat{\pi} = \pi P$ as, (i) $\hat{\pi}$ is non-negative, and (ii) we have

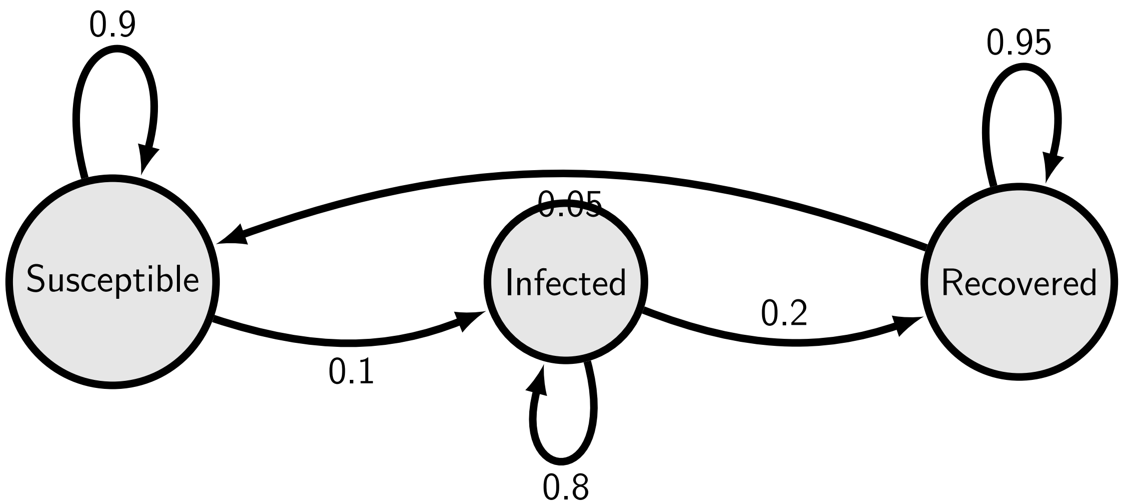
$$\hat{\pi}\mathbf{1} = (\pi P)\mathbf{1} = \pi(P\mathbf{1}) = \pi\mathbf{1} = 1.$$

- Brouwer fixed-point theorem asserts that: *Any continuous function f that maps a bounded, closed, and convex set A to itself, has a fixed point, i.e., there exists a point $x^* \in A$, such that $f(x^*) = x^*$.*
- S is a bounded, closed, and convex set, and hence, by Brouwer fixed-point theorem, there exists a stochastic vector $\pi^* \in S$ such that $\pi^* = \pi^* P$.

Example

- Let $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$.
- Solving for $(u, v)P = (u, v)$ with $v = 1 - u$, we get $u = \frac{2}{5}$ and $v = \frac{3}{5}$. Is this unique?
- What about $P = I$?
- Fundamental questions in the theory of (homogeneous) Markov chains:
  - *Uniqueness*: Is the stationary distribution unique?
  - *Ergodicity*: When unique, under what conditions, $\pi_k \to \pi^*$?
  - *Mixing time*: How fast does it converge to π^*?
  - *Occupation Probability*: How often do we spend time on a given state?

DT-FS Markov Chains as Random Walks over Weighted Graphs



- Consider a Markov chain on state space S with the (one step) transition matrix P.
- Consider a directed weighted graph $G = (V, E, P)$ where
  - $V = S = \{1, \ldots, d\}$,
  - $E = \{(i, j) \mid P_{ij} > 0\}$, and
  - P_{ij} is the weight of edge i, j.
- Then the Markov chain can be viewed as a random walk on this weighted graph.

Uniqueness, Irreducibility, Aperiodicity, and Ergodicity

- We say that a matrix P is irreducible if for any i, j, $[P^{k_{ij}}]_{ij} > 0$ for some $k_{ij} \geq 0$.
- Graph theoretic interpretation: P is irreducible if there is a directed path between any two nodes on the graph.
- We define the period γ_i of a state i, to be[2] $gcd(k \mid [P^k]_{ii} > 0)$.
- Graph theoretic interpretation: gcd of all loops from state i.
- We say that a non-negative matrix P is aperiodic if $\gamma_i = 1$ for all i.
- A (homoegeneous) Markov chain with the transition matrix P is said to be irreducible (aperiodic) if P is irreducible (aperiodic).

  Theorem 2. *A DT-FS Homogeneous Markov chain with the transition matrix P admits a unique stationary distribution π^* with strictly positive entries if P is irreducible. If further, P is aperiodic, then the Markov chain is ergodic, i.e., for any initial distribution π_1, $\lim_{k\to\infty} \pi_k = \pi^*$.*

- Proof utilizes Perron-Frobenious Theorem for non-negative matrices.
- In fact, if P is aperiodic and irreducible, then

$$P^k \to \mathbf{1}\pi^* = A = \begin{pmatrix} \text{———————} \pi^* \text{———————} \\ \text{———————} \pi^* \text{———————} \\ \vdots \\ \text{———————} \pi^* \text{———————} \end{pmatrix}.$$

- What is the probabilistic meaning of the stationary distribution?

  Theorem 3. *Suppose that $\{X_k\}$ is a DT-FS homogeneous and ergodic Markov chain with the stationary distribution π^*. Then*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \mathbf{1}_{X_k=i}}{n} = \pi^*(i) \quad \textit{almost surely.}$$

---

[2] gcd stands for greatest common divisor.

Application: Page-Rank Algorithm

- Original idea of Google search ranking: Model a browsing person as a random walker over the graph of internet!
- Let $G = (V, E)$ where $d =$ number of webpages and there is a node for each webpage.
- $(i, j) \in E$ if i has a link to j.
- Then a person can be *modeled* as a random walker on G where

$$P_{ij} = \begin{cases} \frac{1}{d_i} & j \in \mathcal{N}_i \\ 0 & \text{otherwise.} \end{cases}$$

- Problem with this? Corresponding Markov chain is not irreducible.
- Now let us add a small reset probability, i.e., consider a Markov chain with one-step transition matrix

$$\hat{P} = (1 - a)P + aJ,$$

  where $a \in (0, 1)$ is a small reset parameter and J is the $d \times d$ matrix with all elements being $1/d$.
- Then a Markov chain with the transition matrix $\hat{P}$ is irreducible and aperiodic (why?).
- Therefore, it is ergodic, has a unique stationary distribution π^*, and $\pi_k \to \pi^*$ as $k \to \infty$.
- More importantly *average visit percentage of state (webpage)* i *by time* $k \to \pi_i^*$!
- Therefore, webpage i is superios to j if $\pi_i^* > \pi_j^*$.
- How does Google find π^*? Power method!