MATH 829: Introduction to Data Mining and Analysis
Hidden Markov Models - Review of Markov chains

Dominique Guillot

Departments of Mathematical Sciences
University of Delaware

May 9, 2016

Markov chains

- Let $S := \{s_1, s_2, \dots\}$ be a countable set.
- A (discrete time) **Markov chain** is a discrete stochastic process $\{X_n : n = 0, 1, \dots\}$ such that
  1. X_n is an S-valued random variable $\forall n \geq 0$.
  2. (Markov Property) For all $i, j, i_0, \dots, i_{n-1} \in S$, and all $n \geq 0$:

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i).$$

Interpretation: Given the present X_n, the future X_{n+1} is independent of the past $(X_0, \dots, X_{n-1})$.

- The elements of S are called the **states** of the Markov chain.
- When $X_n = j$, we say that the process is in state j at time n.

Stationarity and transition probabilities

- A Markov chain is **homogeneous** (or **stationary**) if for all $n \geq 0$ and all $i, j \in S$,

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) =: p(i, j).$$

In other words, the **transition probabilities** do not depend on time.

- We will only consider homogeneous chains in what follows.
- We denote by $P := (p(i,j))_{i,j \in S}$ the **transition matrix** of the chain.
- Note: P is a **stochastic matrix**, i.e.,

$$\forall i, j \in S,\ p(i,j) \geq 0, \quad \text{and} \quad \forall i \in S, \sum_{j \in S} p(i,j) = 1.$$

- Conversely, every stochastic matrix is the transition matrix of some homogeneous discrete time Markov chain.

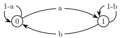
Examples

Example 1: (Two-state Markov chain)

$$S = \{0, 1\}, \quad p(0,1) = a, \quad p(1,0) = b, \quad a, b \in [0,1]$$

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}.$$

We naturally represent P using a transition (or state) diagram:



Interpretation: machine is either broken (0) or working (1) at start of n-th day.

Examples (cont.)

Example 2: (Simple random walk) Let $\xi_1, \xi_2, \xi_3, \dots$ be iid random variables such that $\forall i \geq 1$,

$$\xi_i = \begin{cases} +1 & P(\xi_i = +1) = p \\ 0 & P(\xi_i = 0) = r \\ -1 & P(\xi_i = -1) = q \end{cases},$$

where $p + r + q = 1$, $p, r, q \geq 0$.

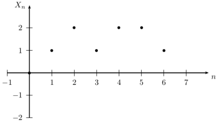
- Let X_0 be an integer valued random variable independent of the ξ_i's.
- Define $\forall n \geq 1$,

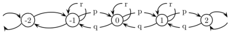
$$X_n = X_0 + \sum_{i=1}^{n} \xi_i.$$

- The process is a **random walk**.

Review of Markov chains (cont.)

- Here $S = \{0, \pm 1, \pm 2, \dots\}$.





Exercise: What is P for that Markov chain?

n-step transitions

Let $\{X_n : n \geq 0\}$ be a Markov chain.

- We define the **initial distribution** of the chain by

$$\mu_0(i) := P(X_0 = i) \qquad (i \in S).$$

- All distributional properties of a (homogeneous) Markov Chain are determined by its initial distribution and transition probability matrix.
- For $n \geq 1$, we define the n-**step transition probability** $p^n(i,j)$ by

$$p^n(i,j) := P(X_n = j | X_0 = i) = P(X_{n+m} = j | X_m = i).$$

Also, define

$$p^0(i,j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

- We define the n-**step transition matrix** by

$$P^{(n)} := (p^n(i,j) : i, j \in S).$$

Chapman-Kolmogorov

Theorem: (The Chapman-Kolmogorov Equations) We have for all $m, n \geq 1$:

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)}.$$

In particular, for all $n \geq 1$,

$$P^{(n)} = P \cdot P^{(n-1)} = \cdots = P^n.$$

Moral: n-step transition probabilities are computed using matrix multiplications.

- Let $\mu_n := (\mu_n(i) : i \in S)$ denote the **distribution of** X_n:

$$\mu_n(i) := P(X_n = i).$$

Proposition: We have

$$\mu_{m+n} = \mu_m P^n, \quad \text{and} \quad \mu_n = \mu_0 P^n.$$

Moral: Distributional computations for Markov Chains are just matrix multiplications.

Reducibility

- **Reducibility:**
  - A state $j \in S$ is said to be **accessible** from $i \in S$ (denotde $i \to j$) if a system started in state i has a non-zero probability of transitioning into state j at some point.
  - A state $i \in S$ is said to **communicate** with state $j \in S$ (denoted $i \leftrightarrow j$) if both $i \to j$ and $j \to i$.

Note: Communication is an equivalence relation.

A Markov chain is said to be **irreducible** if its state space is a single communicating class.

Transience and periodicity

- **Transience:**
  - A state $i \in S$ is said to be **transient** if, given that we start in state i, there is a non-zero probability that we will never return to i.
  - A state is **recurrent** if it is not transient.
  - The **recurrence time** of state $i \in S$ is $T_i := \min\{n \geq 1 : X_n = i \text{ given } X_0 = i\}$.
  - Note: $i \in S$ is recurrent iff $P(T_i < \infty) = 1$.
  - A recurrent state $i \in S$ is **positive recurrent** if $E[T_i] < \infty$.
- **Periodicity:**
  - A state $i \in S$ has period k if

$$k = \gcd\{n > 0 : P(X_n = i | X_0 = i) > 0\}.$$

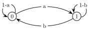
    For example, suppose you start in state i and can only return to i at time $6, 8, 10, 12, \text{etc.}$. Then the period of i is 2.
  - If $k = 1$, then the state is said to be aperiodic.

A Markov chain is **aperiodic** if every state is aperiodic.

Limiting behavior

Limiting behavior of Markov chains: What happens to $p^n(i, j)$ as $n \to \infty$?

Example: (The two-state Markov chain)



If $(a, b) \neq (0, 0)$, we have (exercise):

$$P^n = \frac{1}{a+b}\begin{pmatrix} b & a \\ b & a \end{pmatrix} + \frac{(1-a-b)^n}{a+b}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix}.$$

Thus, if $(a, b) \neq (0, 0)$ and $(a, b) \neq (1, 1)$, then

$$\lim_{n\to\infty} p^n(0,0) = \lim_{n\to\infty} p^n(1,0) = \frac{b}{a+b}$$
$$\lim_{n\to\infty} p^n(0,1) = \lim_{n\to\infty} p^n(1,1) = \frac{a}{a+b}.$$

Thus, the chain has a **limiting distribution**.

The limiting distribution is **independent of the initial state**.

Stationary distribution

Recall: $\mu_{n+1} = \mu_n P$.

A vector $\pi = (\pi(i) : i \in S)$ is said to be a **stationary distribution** for a Markov chain $\{X_n : n \geq 0\}$ if

1. $0 \leq \pi_i \leq 1 \ \forall i \in S$.
2. $\sum_{i \in S} \pi_i = 1$.
3. $\pi = \pi P$, where P is the transition probability matrix of the Markov chain.

Remark: In general, a stationary distribution may not exist or be unique.

Theorem: Let $\{X_n : n \geq 0\}$ be an irreducible and aperiodic Markov chain where each state is positive recurrent. Then

1. The chain has a unique stationary distribution π.
2. For all $i \in S$, $\lim_{n\to\infty} P(X_n = i) = \pi(i)$.
3. $\pi_i = \frac{1}{E[T_i]}$.

$\pi(i)$ can be interpreted as the average proportion of time spent by the chain in state i.