Markov Model and Markov Property

Prof. Yan Wang
Woodruff School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.
yan.wang@me.gatech.edu

Learning Objectives

- To understand the concept of independence in probability
- To familiarize the Markov model and Markov property
- To familiarize Chapman-Kolmogorov equation

Independence

- One of the most important concepts defined in probability theory is independence.
- The concept of independence is essential to decompose a complex problem into simpler and manageable components.
- Markov models rely on assumptions of independence.

Definition of Independence

- **Definition 3.1** (*Conditional Independence*). For $A, B, C \in \mathcal{A}$, A is said to be *conditionally independent* with B on C if and only if $p\left(A \cap B \mid C\right) = p\left(A \mid C\right) p\left(B \mid C\right)$.
- **Definition 3.2** (*Independence*). For $A, B \in \mathcal{A}$, A is said to be *independent* with B if and only if $p\left(A \cap B\right) = p\left(A\right) p\left(B\right)$.

Independence can be seen as a special case of Conditional Independence where $C = \Omega$.

Knowledge Accumulation

- **Lemma 3.1**. For $A, B, C \in \mathcal{A}$,

$$p\left(A \cap B \mid C\right) = p\left(A \mid B \cap C\right) p\left(B \mid C\right)$$

Proof.

$$\begin{aligned} p\left(A \cap B \mid C\right) &= p\left(A \cap B \cap C\right) / \, p\left(C\right) \\ &= p\left(A \mid B \cap C\right) p\left(B \cap C\right) / \, p\left(C\right) \\ &= p\left(A \mid B \cap C\right) p\left(B \mid C\right) \end{aligned}$$

Equivalent Views of Independence

- **Theorem 3.2**. For $A, B, C \in \mathcal{A}$,

$$p\left(A \cap B \mid C\right) = p\left(A \mid C\right) p\left(B \mid C\right) \Leftrightarrow p\left(A \mid B \cap C\right) = p\left(A \mid C\right)$$

Proof.

$$p\left(A \cap B \mid C\right) = p\left(A \mid B \cap C\right) p\left(B \mid C\right) = p\left(A \mid C\right) p\left(B \mid C\right)$$

$$\Leftrightarrow \; p\left(A \mid B \cap C\right) = p\left(A \mid C\right)$$

Graphoid Properties

- The most intuitive meaning of 'independence' is that an independence relationship satisfies several *graphoid* properties.
- With X,Y,Z,W as sets of disjoint random variables and "$\perp$" denoting independence and "|" as condition, the axioms of graphoid are:
  - (A1) Symmetry
  - (A2) Decomposition
  - (A3) Weak union
  - (A4) Contraction
  - (A5) Intersection

Graphoid - Symmetry

$$X \perp Y \mid Z \Rightarrow Y \perp X \mid Z$$

- **Remark 3.1**. If knowing Y does not tell us more about X, then similarly knowing X does not tell us more about Y.

Graphoid - Decomposition

$$X \perp (W, Y) \mid Z \Rightarrow X \perp Y \mid Z$$

- **Remark 3.2**. If combined two pieces of information is irrelevant to X, either individual one is also irrelevant to X.

Graphoid - Weak Union

$$X \perp (W, Y) \mid Z \Rightarrow X \perp W \mid (Y, Z)$$

- **Remark 3.3**. Gaining more information about irrelevant *Y* does not affect the irrelevance between X and W.

Graphoid - Contraction

$$\left(X \perp Y \mid Z\right) \wedge \left(X \perp W \mid \left(Y, Z\right)\right) \Rightarrow X \perp \left(W, Y\right) \mid Z$$

- **Remark 3.4**. If two pieces of information X and Y are irrelevant with prior knowledge of Z and X is also irrelevant to a third piece of information W after knowing Y, then X is irrelevant to both W and Y before knowing Y.

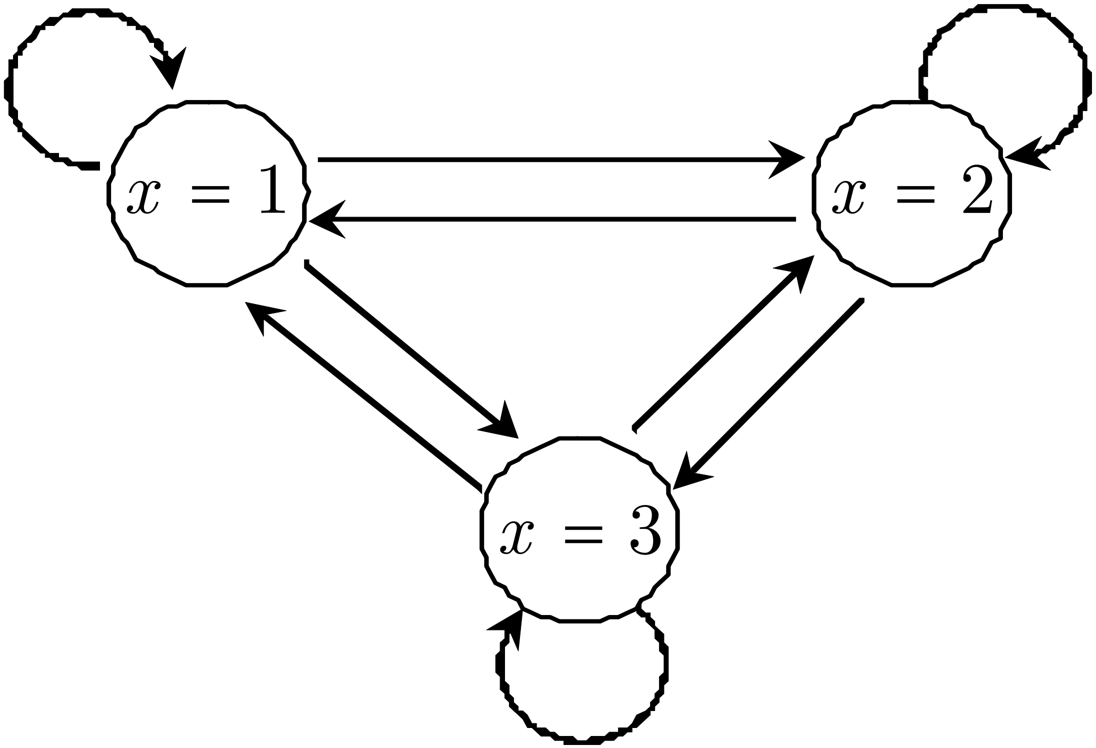
Graphoid - Intersection

$$\left(X \perp W \mid \left(Y,Z\right)\right) \wedge \left(X \perp Y \mid \left(W,Z\right)\right) \Rightarrow X \perp \left(W,Y\right) \mid Z$$

- **Remark 3.5**. If combined information W and Y is relevant to X, then at least either W or Y is relevant to X after learning the other.

Discrete-Time Markov Chain

- State transition diagram

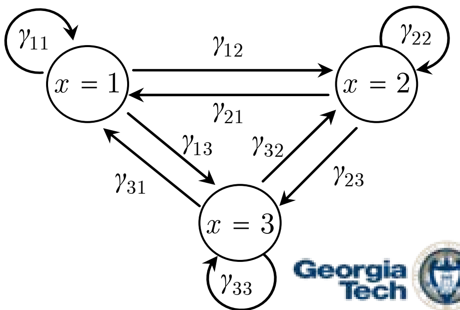


Markov Property

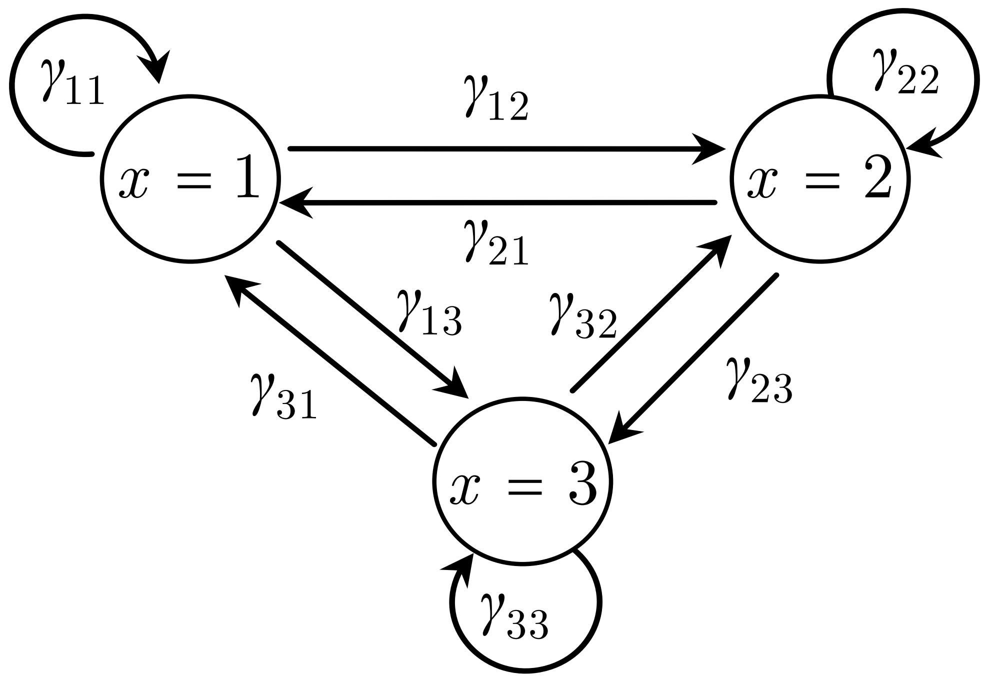
- A Markov chain represents a Markov process of state transitions, where the "memoryless" Markov property is assumed.

$$P\left(x^{(t+1)} \mid x^{(t)}, \ldots, x^{(0)}\right) = P\left(x^{(t+1)} \mid x^{(t)}\right)$$

- Loosely speaking, the future state of a random variable $x^{(t+1)}$ at time *t*+1 only depends on its current state $x^{(t)}$, not the complete transition history.



Transition Probability

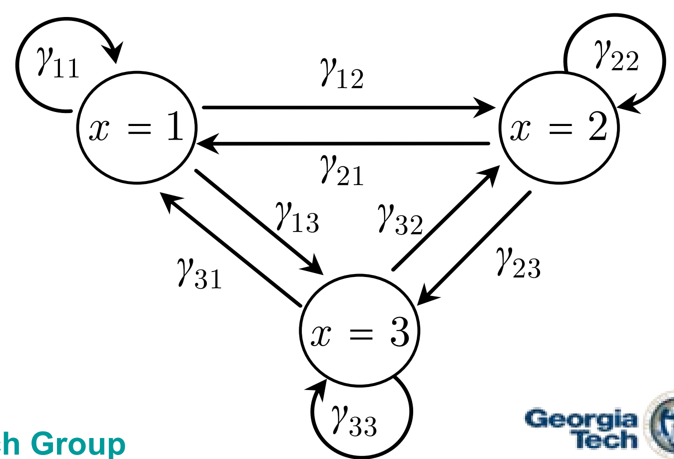


$$P\left(x^{(t+1)} = j \mid x^{(t)} = i\right) = \gamma_{ij} \quad \left(i, j = 1, 2, 3\right)$$

Transition Matrix

$$\Gamma = \left(\gamma_{ij}\right)_{3\times3} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix}$$

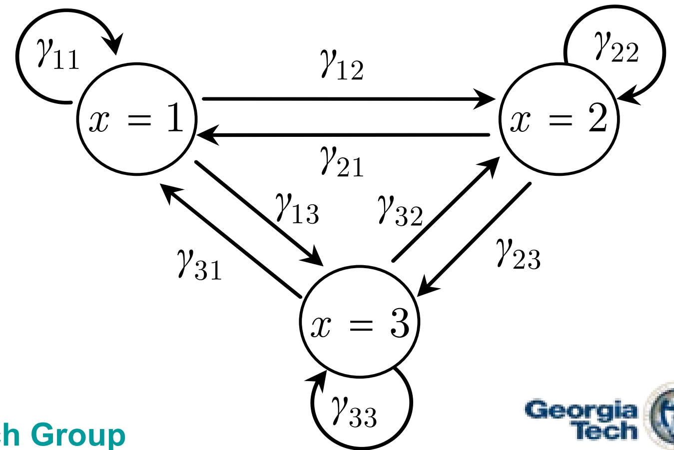
where $\sum_{j=1}^{3} \gamma_{ij} = 1 \quad \left(i = 1, 2, 3\right)$



State Vector

$$\Delta^{(t)} = \left(\delta_1^{(t)}, \delta_2^{(t)}, \delta_3^{(t)} \right)$$

where $P(x^{(t)} = i) = \delta_i \quad \left(i = 1, 2, 3\right)$



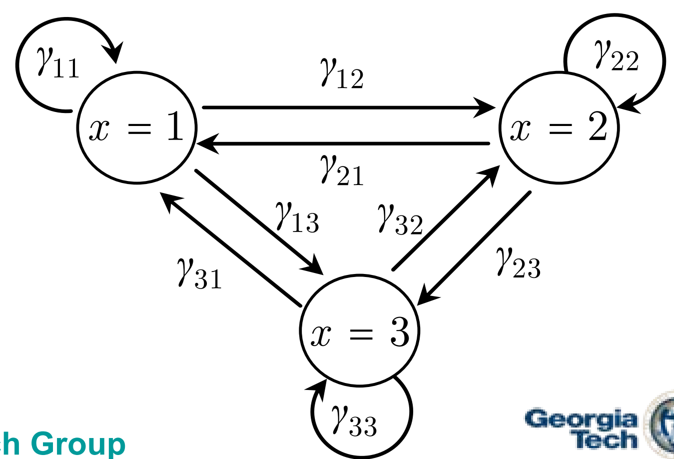
State Transition

- State update

$$\Delta^{(t+1)} = \Delta^{(t)}\Gamma$$

- Stationary distribution

$$\Delta = \Delta\Gamma$$



Extensions of Discrete-Time Markov Chain

- When the transition matrix $\Gamma^{(t)}$ is not constant, it is called a *non-homogeneous Markov chain*.
- When the time is not discrete, it is called *continuous-time Markov chain*.

Ergodicity

- **Definition 3.3**. A Markov chain $\Gamma^{(t)} = \left(\gamma_{ij}^{(t)}\right)$ is called *irreducible* if for all states $x \in \Omega$, there exists a time *t* such that $\gamma_{ij}^{(t)} > 0 \; (\forall i, j)$

  "every state is eventually reachable"

- **Definition 3.4**. A Markov chain $\Gamma^{(t)} = \left(\gamma_{ij}^{(t)}\right)$ is called *aperiodic* if for all states $x \in \Omega$,
  $$\gcd\left\{t : \gamma_{ij}^{t} > 0\right\} = 1 \quad .$$

  "it doesn't get caught in cycles"

- **Definition 3.5**. A Markov chain is called *ergodic* if it is both irreducible and aperiodic.

Chapman-Kolmogorov Equation

- Suppose there are a total of K states

$$P\left(x^{\left(t+s\right)}=j \mid x^{(0)}=i\right)$$

$$=\sum_{k=1}^{K} P\left(x^{\left(t+s\right)}=j \mid x^{\left(s\right)}=k\right) P\left(x^{\left(s\right)}=k \mid x^{(0)}=i\right)$$

Forward Differential Kolmogorov Equation

- For a small time step h

$$P\left(x^{(t+h)} = j \mid x^{(0)} = i\right)$$
$$= \sum_{k=1}^{K} P\left(x^{(t+h)} = j \mid x^{(t)} = k\right) P\left(x^{(t)} = k \mid x^{(0)} = i\right)$$

$$= \sum_{k \neq j} P\left(x^{(t+h)} = j \mid x^{(t)} = k\right) P\left(x^{(t)} = k \mid x^{(0)} = i\right)$$
$$+ P\left(x^{(t+h)} = j \mid x^{(t)} = j\right) P\left(x^{(t)} = j \mid x^{(0)} = i\right)$$

Forward Differential Kolmogorov Equation

$$P\left(x^{(t+h)} = j \mid x^{(0)} = i\right) - P\left(x^{(t)} = j \mid x^{(0)} = i\right)$$
$$= \sum_{k \neq j} P\left(x^{(t+h)} = j \mid x^{(t)} = k\right) P\left(x^{(t)} = k \mid x^{(0)} = i\right)$$
$$+P\left(x^{(t)} = j \mid x^{(0)} = i\right)\left[P\left(x^{(t+h)} = j \mid x^{(t)} = j\right) - 1\right]$$

$$= \sum_{k \neq j} P\left(x^{(t+h)} = j \mid x^{(t)} = k\right) P\left(x^{(t)} = k \mid x^{(0)} = i\right)$$
$$-P\left(x^{(t)} = j \mid x^{(0)} = i\right) \sum_{k \neq j} P\left(x^{(t+h)} = k \mid x^{(t)} = j\right)$$

Forward Differential Kolmogorov Equation

$$P\left(x^{(t+h)}=j \mid x^{(0)}=i\right)-P\left(x^{(t)}=j \mid x^{(0)}=i\right)$$

$$=\sum_{k\neq j}\left[\begin{array}{l}P\left(x^{(t+h)}=j \mid x^{(t)}=k\right)P\left(x^{(t)}=k \mid x^{(0)}=i\right)\\ -P\left(x^{(t+h)}=k \mid x^{(t)}=j\right)P\left(x^{(t)}=j \mid x^{(0)}=i\right)\end{array}\right]$$

Then we have

$$\left[P\left(x^{(t+h)}=j \mid x^{(0)}=i\right)-P\left(x^{(t)}=j \mid x^{(0)}=i\right)\right]/h$$

$$=\sum_{k\neq j}\left[\begin{array}{l}P\left(x^{(t)}=k \mid x^{(0)}=i\right)P\left(x^{(t+h)}=j \mid x^{(t)}=k\right)/h\\ -P\left(x^{(t)}=j \mid x^{(0)}=i\right)P\left(x^{(t+h)}=k \mid x^{(t)}=j\right)/h\end{array}\right]$$

Forward Differential Kolmogorov Equation

- Define the jump rate as

$$q_{k\leftarrow j}^{(t)} = P\left(x^{(t+h)} = k \mid x^{(t)} = j\right) / h \qquad (j \neq k)$$

- Then

$$\boxed{\begin{aligned} &P'\left(x^{(t)} = j \mid x^{(0)} = i\right) \\ &= \sum_{k\neq j}\left[\begin{aligned} &q_{j\leftarrow k}^{(t)} P\left(x^{(t)} = k \mid x^{(0)} = i\right) \\ &-q_{k\leftarrow j}^{(t)} P\left(x^{(t)} = j \mid x^{(0)} = i\right) \end{aligned}\right] \end{aligned}}$$

Summary

- Independence is one of the essential properties in probability calculus
- "Memoryless" Markov property simplifies inference in state transition

Memoryless