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

$$p(A \cap B \mid C) = p(A \mid C)p(B \mid C).$$

- **Definition 3.2** (*Independence*). For $A, B \in \mathcal{A}$, $A$ is said to be *independent* with $B$ if and only if

$$p(A \cap B) = p(A)p(B).$$

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

Lemma 3.1. For $A, B, C \in A$, 

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

Proof.

$$p(A \cap B \mid C) = p(A \cap B \cap C) / p(C)$$

$$= p(A \mid B \cap C) p(B \cap C) / p(C)$$

$$= p(A \mid B \cap C) p(B \mid C)$$
Equivalent Views of Independence

Theorem 3.2. For $A, B, C \in A$,

\[ p(A \cap B \mid C) = p(A \mid C)p(B \mid C) \iff p(A \mid B \cap C) = p(A \mid C) \]

Proof.

\[ p(A \cap B \mid C) = p(A \mid B \cap C)p(B \mid C) = p(A \mid C)p(B \mid C) \]

\[ \iff p(A \mid B \cap C) = p(A \mid C) \]
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

\[(X \perp Y \mid Z) \land (X \perp W \mid (Y, Z)) \Rightarrow X \perp (W, Y) \mid Z\]

**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

\[(X \perp W \mid (Y, Z)) \land (X \perp Y \mid (W, Z)) \Rightarrow X \perp (W, Y) \mid Z\]

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
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

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

\[
\Gamma = \begin{pmatrix}
\gamma_{11} & \gamma_{12} & \gamma_{13} \\
\gamma_{21} & \gamma_{22} & \gamma_{23} \\
\gamma_{31} & \gamma_{32} & \gamma_{33}
\end{pmatrix}
\]

where \( \sum_{j=1}^{3} \gamma_{ij} = 1 \) \( (i = 1, 2, 3) \)
State Vector

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

where \( P(x^{(t)} = i) = \delta_i \) \( (i = 1, 2, 3) \)
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 $\Gamma(t) = \begin{pmatrix} \gamma_{ij}^{(t)} \end{pmatrix}$ 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 $\Gamma(t) = \begin{pmatrix} \gamma_{ij}^{(t)} \end{pmatrix}$ is called **aperiodic** if for all states $x \in \Omega$, $\gcd \left\{ t : \gamma_{ij}^{t} > 0 \right\} = 1$.
  
  “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

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

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

Then we have

\[
\left[P\left(x_{(t+h)}^{(t+h)} = j \mid x^{(0)} = i\right) - P\left(x^{(t)} = j \mid x^{(0)} = i\right)\right] / h \\
= \sum_{k \neq j} \left[P\left(x^{(t)} = k \mid x^{(0)} = i\right)P\left(x_{(t+h)}^{(t+h)} = j \mid x^{(t)} = k\right) / h \\
- P\left(x^{(t)} = j \mid x^{(0)} = i\right)P\left(x_{(t+h)}^{(t+h)} = k \mid x^{(t)} = j\right) / h\right]
\]
Forward Differential Kolmogorov Equation

Define the jump rate as
\[ q_{k \leftarrow j}^{(t)} = P\left(x^{(t+h)} = k \mid x^{(t)} = j\right) / h \quad (j \neq k) \]

Then
\[
P'(x^{(t)} = j \mid x^{(0)} = i) = \sum_{k \neq j} \left[ q_{j \leftarrow k}^{(t)} P\left(x^{(t)} = k \mid x^{(0)} = i\right) - q_{k \leftarrow j}^{(t)} P\left(x^{(t)} = j \mid x^{(0)} = i\right) \right]
\]
Summary

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