Tufts UNIVERSITY

MARKOV MODELS 1

ARTIFICIAL INTELLIGENCE | COMP 131

TODAY ON AI

- Probability theory recap
- Stochastic processes
- Markov models
- Hidden Markov models
- Questions?

- **Conditional probability**: $P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$

- **Product rule**: $P(X,Y) = P(X \mid Y)\, P(Y) = P(Y \mid X)\, P(X)$

- **Chain rule**: $P(X_1, \ldots, X_n) = \prod_i P(X_i \mid X_1, \ldots, X_{i-1})$

- ***X* and *Y* are independent**: iff $P(X,Y) = P(X)\, P(Y)$

- **Bayes rule**: $P(Cause \mid Observation) = \frac{P(Observation \mid Cause) P(Cause)}{P(Observation)}$

- **Conditional independence**: $P(X,Y \mid Z) = P(X \mid Z)\, P(Y \mid Z)$

Often, we want a model able to handle a sequence of observations:

- Speech recognition
- Robot localization
- Classification
- Filtering / Smoothing of sensor data
- State estimate of a system

How do we introduce **time** or **space** in our models?

SECTION 01

Stochastic processes

Let T be a subset of $[0, \infty)$. A family of random variables $\{X_t\}_{t \in T}$, indexed by T is called a **stochastic** (or **random**) **process**.

A **stochastic process** is fundamentally a collection of data points that express the values of the random variables of the process in the period of observation.

When $T \in \mathbb{N}_0$, the process is said to be a **discrete-time process**, and when $T \in \mathbb{R}_+$, it is called a **continuous-time process**.

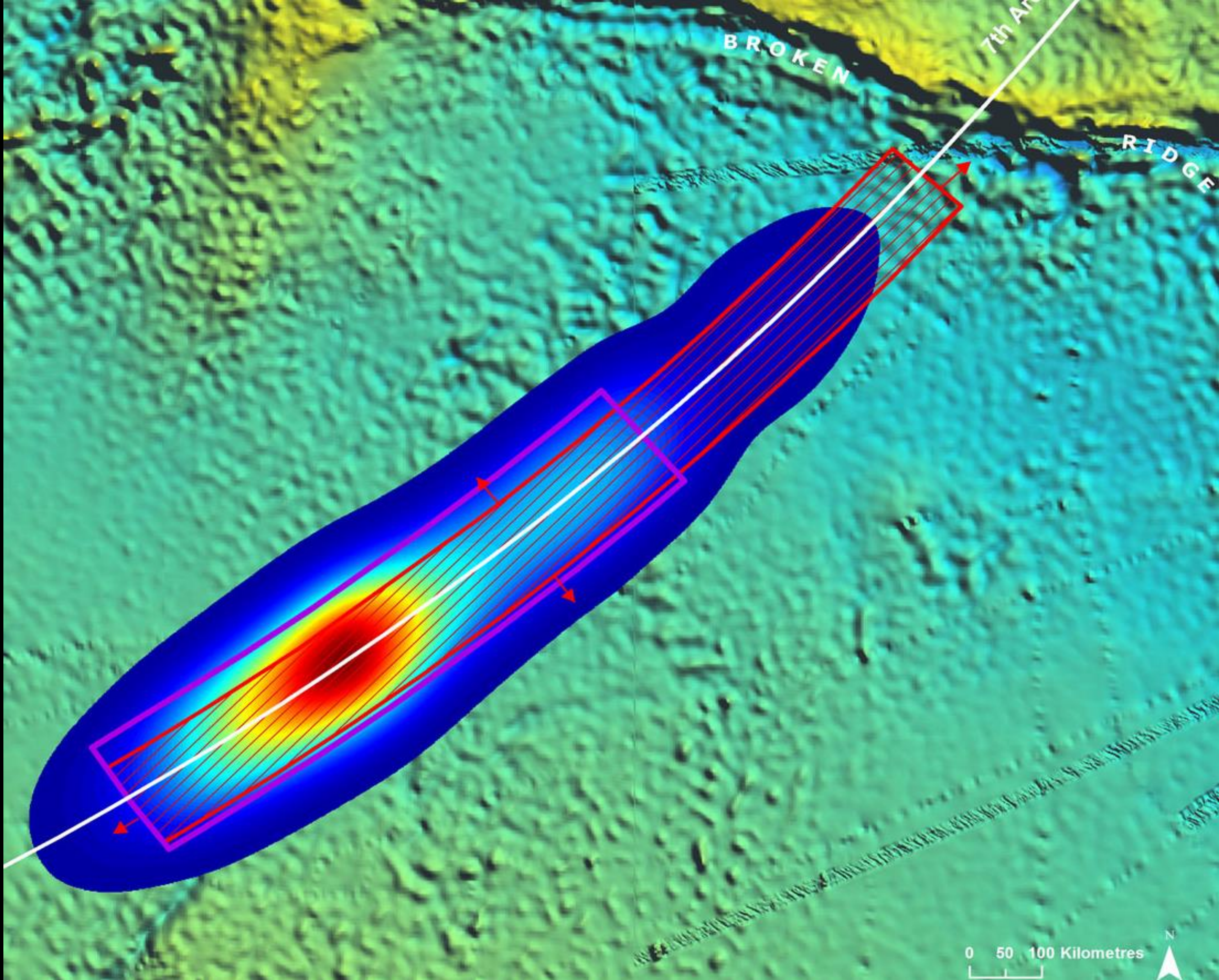
There are different types of stochastic processes:

- **Random walks**: All variables are identically distributed, and the domain is the integer numbers
- **Poisson process**: All variables are identically distributed and independent from each other
- **Markov process**: The variables are dependent with each other with a simple relationship

A patient's heart pulse during surgery is a **stochastic process**.

- It is measured continuously during the interval $[0, T]$
- The process can be considered:
  - A **discrete-time binary process**: the stochastic variable X_t is binary (**F**=no heartbeat, **T**=heartbeat), and $T \in \mathbb{N}_0$
  - A **discrete-time continuous process**: the stochastic variable X_t is continuous (the voltage of the sensor), and $T \in \mathbb{N}_0$
  - A **continuous-time binary process**: the stochastic variable X_t is binary (**F**=no heartbeat, **T**=heartbeat), and $T \in \mathbb{R}_+$
  - A **continuous-time continuous process**: the stochastic variable X_t is continuous (the voltage of the sensor), and $T \in \mathbb{R}_+$

A preliminary analysis of where Malaysia Airlines flight MH370 might have disappeared was done with Markov processes.



SECTION 02

Markov models

Markov models or **Markov chains** are **Markov processes** in which the state space is discrete.

Because of their **discrete nature**, Markov models can be represented with Bayes networks.

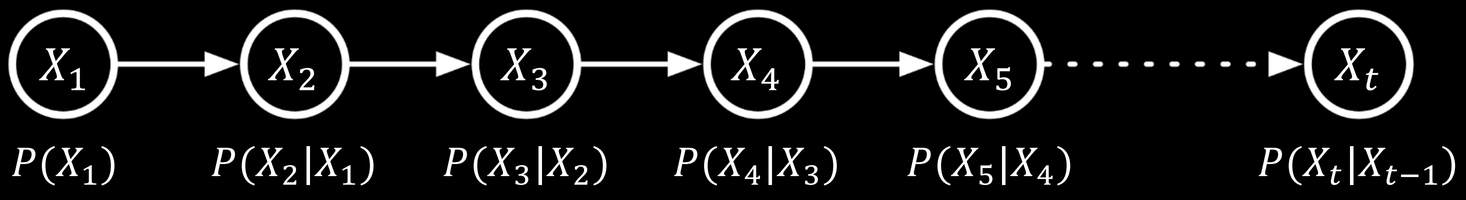
- The order of the Markov model determines how the future is **influenced** by the past:
  - **First-order Markov model**: The future is always independent from the past, given the present

$$P(X_{t+1} | X_{1:t}) = P(X_{t+1} | X_t)$$

  - **M-order Markov model**: The future depends on the present and the most recent past

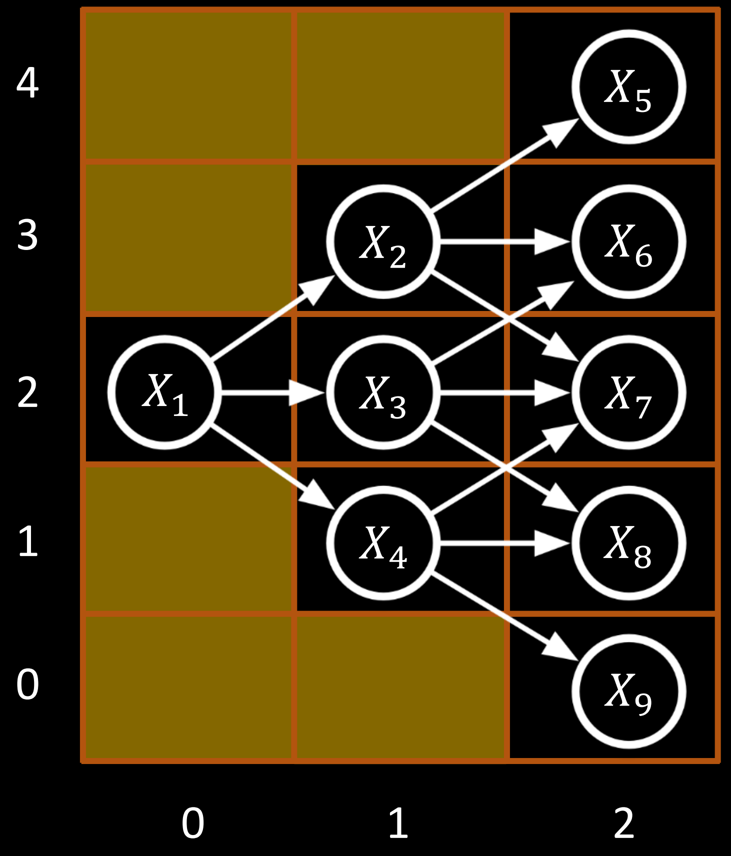
$$P(X_{t+1} | X_{1:t}) = P(X_{t+1} | X_{t:t-m})$$

Value of X at a given step is called **state** of the Markov model. Every state is always tracking the **same random variables**.



SEQUENCES OF EVENTS

N-DIMENSIONAL SPACE



- Primary components of a Markov model are:
  - An **initial prior** specifies where the model starts from. Sometimes is not known
  - A **transition probability** or model dynamics specifies how the state evolves over time

$$X_1 \longrightarrow X_2 \longrightarrow X_3 \dashrightarrow X_T$$

$P(X_1)$ $P(X_2|X_1)$ $P(X_3|X_2)$ $P(X_T|X_{T-1})$

$$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_2)$$

$$P(X_1, \dots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2) \dots P(X_T|X_{T-1})$$

$$= \mathrm{P}(\mathrm{X}_1) \prod_{t=2}^{T} P(X_t|X_{t-1})$$

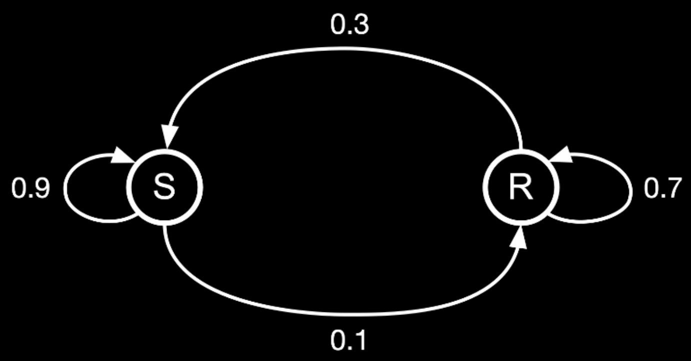
The **stationarity assumption** states that **transition probabilities** do not change with time.

The transition probability can be represented in four ways:

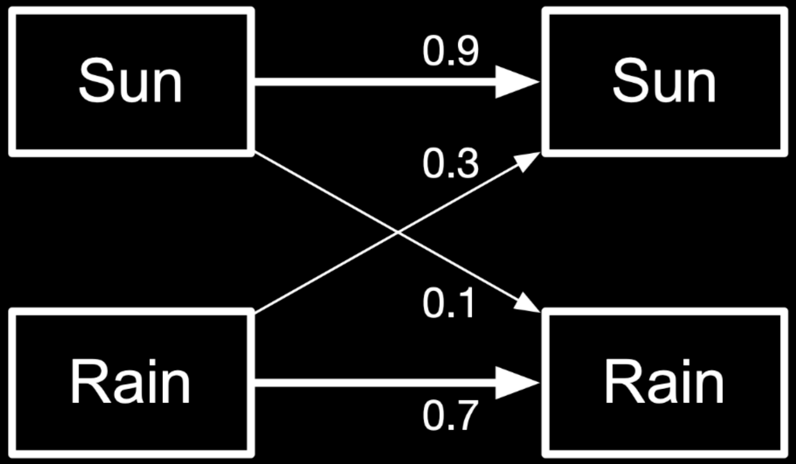
- A transition probability table (CPT):

X_t	X_{t+1}	$P(X_{t+1} \mid X_t)$
sun	sun	0.9
sun	$rain$	0.1
$rain$	sun	0.3
$rain$	$rain$	0.7

- A **finite state machine** with edges that specify the probability of transition:



- A widely used representation is the **Trellis diagram** through time:



- A **matrix representation**:

$$\begin{array}{c c} \begin{array}{r} sun_{t-1} \\ rain_{t-1} \end{array} & \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix} \\ & \quad sun_t \; rain_t \end{array}$$

- **STATE**

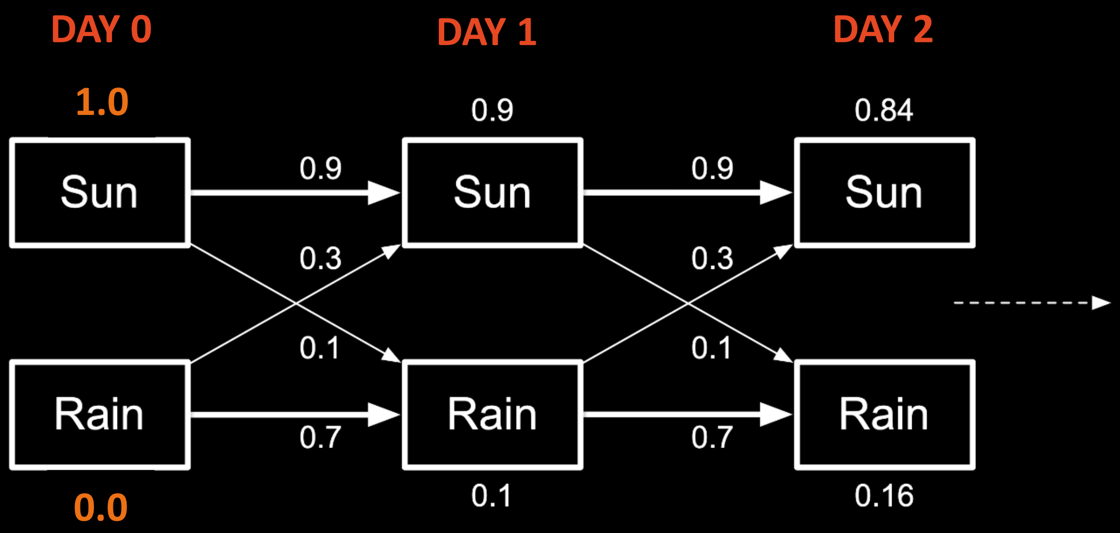
$X = \{rain, sun\}$

- **INITIAL PRIOR**

$P(X_1 = sun) = 1.0$

- **TRANSITION PROBABILITY**

X_t	X_{t+1}	$P(X_{t+1}\|X_t)$
sun	sun	0.9
sun	$rain$	0.1
$rain$	sun	0.3
$rain$	$rain$	0.7



$$P(X_2 = sun) = P(X_2 = sun|X_1 = sun)P(X_1 = sun) + P(X_2 = sun|X_1 = rain)P(X_1 = rain)$$

$$P(X_2 = sun) = \mathbf{0.9}$$

What about after time t?

$$P(X_1) = \boldsymbol{known\ prior}$$

$$P(X_t) = \sum_{X_{t-1}} P(X_{t-1}, X_t) = \sum_{X_{t-1}} P(X_t|X_{t-1})P(X_{t-1})$$

Let P be the transition matrix of Markov chain $\{X_0, X_1, \ldots\}$:

- A state i has **period** $k \geq 1$ if any chain starting at and returning to state i with positive probability must take a number of steps that divisible by k. If $k = 1$, then the state is called **aperiodic**, and if $k > 1$, the state is **periodic**. If all states are aperiodic, then the **Markov chain is aperiodic**.

- A state is **recurrent** if the Markov chain will eventually return to it. A recurrent state is known as **positive recurrent** if it is expected to return within a finite number of steps, and **null recurrent** otherwise.

- A state is **transient** if the Markov chain will never see it again.

- A state is **ergodic** if it is **positive recurrent** and **aperiodic**. A **Markov chain is ergodic** if all its states are.

Ergodic Markov chains converge to a **stationary distribution P_∞**:

- Influence of the initial distribution gets less and less over time
- P_∞ does not depend on the initial distribution
- It is defined as follow: $P_\infty(X) = P_{\infty+1}(X) = \sum_{X_{t-1}} P(X_t|X_{t-1})P_\infty(X_{t-1})$

- For the example:

$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

$$P_\infty(sun) = 0.9\ P_\infty(sun) + 0.3\ P_\infty(rain)$$
$$P_\infty(rain) = 0.1\ P_\infty(sun) + 0.7\ P_\infty(rain)$$

$$\begin{cases} P_\infty(sun) = 3\ P_\infty(rain) \\ P_\infty(rain) = \frac{1}{3} P_\infty(sun) \\ P_\infty(sun) + P_\infty(rain) = 1 \end{cases}$$

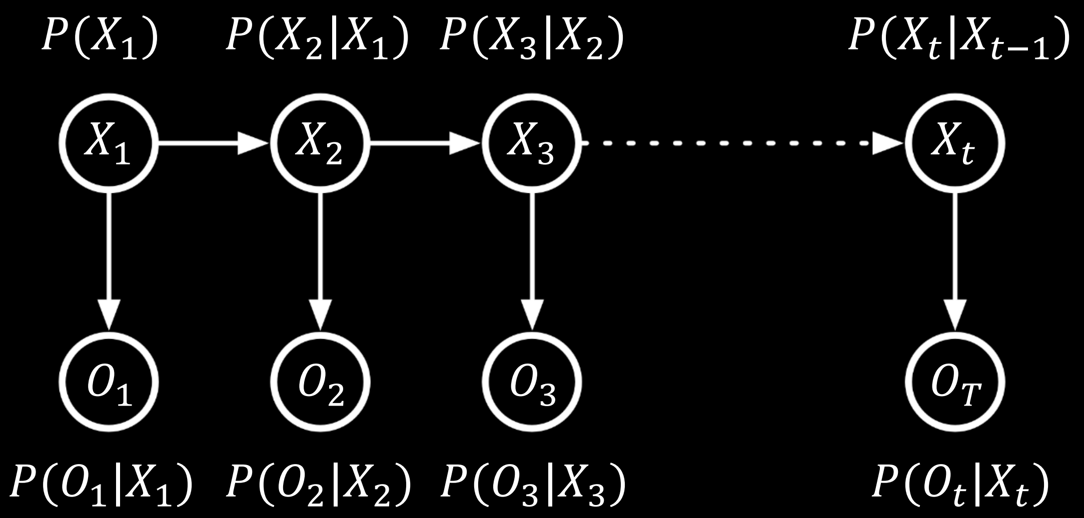
$$P_\infty(sun) = \frac{3}{4}$$
$$P_\infty(rain) = \frac{1}{4}$$

SECTION 03

Hidden Markov models

- Speech recognition HMMs:
  - Observations are acoustic signals (continuous valued)
  - States are specific positions in specific words (so, tens of thousands)
- Classification:
  - Observations are sensor readings
  - States are the classes of the samples
- Robot tracking:
  - Observations are range readings (continuous)
  - States are positions on a map (continuous)

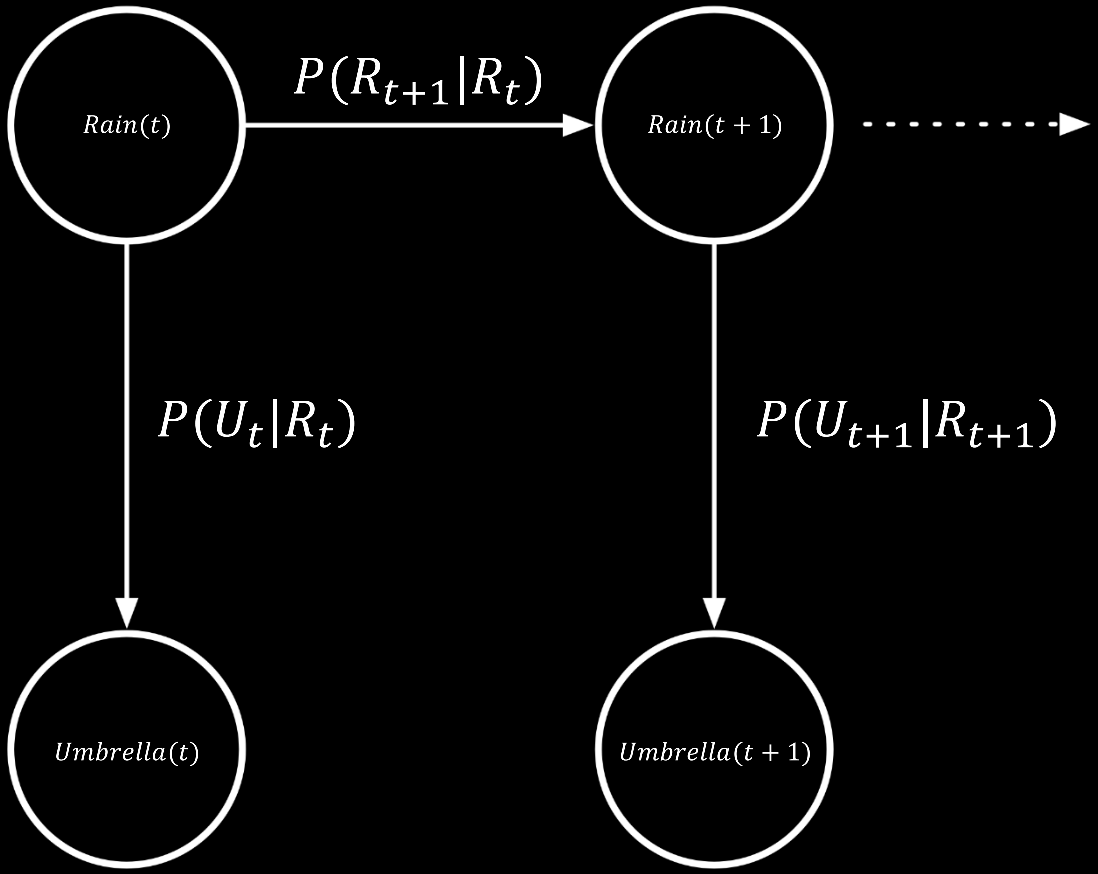
Hidden Markov models (HMMs) describe an underlying hidden state conditionally dependent to some observed evidence:



$$X_t \perp X_1, \dots, X_{t-2} \mid X_{t-1}$$

$$P(X_1, O_1, \dots, X_t, O_t) = P(X_1)P(O_1|X_1) \prod_{i=2}^{t} P(X_i|X_{i-1})P(O_i|X_i)$$

- In addition to the definitions for MMs, there is an **Emission CPT**
- Evidence variables are not independent because correlate via the hidden states



TRANSITION CPT

R_t	R_{t+1}	$P(R_{t+1}\|R_t)$
$\neg rain$	$\neg rain$	0.7
$\neg rain$	$rain$	0.3
$rain$	$\neg rain$	0.3
$rain$	$rain$	0.7

EMISSION CPT

R_t	U_t	$P(U_t\|R_t)$
$rain$	$umbrella$	0.9
$rain$	$\neg umbrella$	0.1
$\neg rain$	$umbrella$	0.2
$\neg rain$	$\neg umbrella$	0.8

Hidden Markov models

EXAMPLE: THE WEATHER AND THE GRAD STUDENT

READINGS

Chapter 15, 16, and 17

QUESTIONS ?

Tufts UNIVERSITY

ARTIFICIAL INTELLIGENCE

COMP 131

FABRIZIO SANTINI

VERSION 4.1