Inference by variable elimination

Factors and Factor Graph

$$f(x) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$

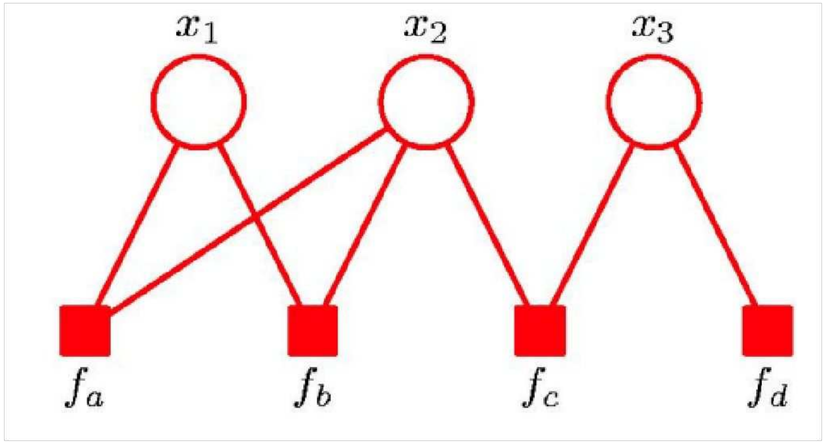


Figure 1: Two factors f_a and f_b of the same variables

They are bipartite since

- Two types of nodes
- All links go between nodes of opposite type

Factors

We assume that we are given a graphical model as a product of factors

$$p\left(x_1, \ldots, x_n\right) = \prod_{c \in C} \phi_c\left(x_c\right)$$

- We can view a factor as a multi-dimensional table assigning a value to each assignment of a set of variables x_c.
- In a Bayesian network, the factors correspond to conditional probability distributions.
- In a Markov Random Field, the factors encode an unnormalized distribution; to compute marginals.

Conditional Independence and Factorization

- Assuming the descriptive features are conditionally independent of each other given MENINGITIS we only need to store four factors:

$$
\begin{aligned}
Factor_1 &: < P(M) > \\
Factor_2 &: < P(h|m), P(h|\neg m) > \\
Factor_3 &: < P(f|m), P(f|\neg m) > \\
Factor_4 &: < P(v|m), P(v|\neg m) >
\end{aligned}
$$

$$P(H, F, V, M) = P(M) \times P(H|M) \times P(F|M) \times P(V|M)$$

Factor Product

A	B	$f_1(A,B)$	B	C	$f_2(B,C)$	A	B	C	$f_3(A,B,C)$
T	T	0.3	T	T	0.2	T	T	T	$0.3 \times 0.2 = 0.06$
T	F	0.7	T	F	0.8	T	T	F	$0.3 \times 0.8 = 0.24$
F	T	0.9	F	T	0.6	T	F	T	$0.7 \times 0.6 = 0.42$
F	F	0.1	F	F	0.4	T	F	F	$0.7 \times 0.4 = 0.28$
						F	T	T	$0.9 \times 0.2 = 0.18$
						F	T	F	$0.9 \times 0.8 = 0.72$
						F	F	T	$0.1 \times 0.6 = 0.06$
						F	F	F	$0.1 \times 0.4 = 0.04$

Factor Marginalization

- Marginalization operation "locally" eliminates a set of variables from a factor.
- If we have a factor $\phi(X, Y)$ over two sets of variables X, Y, marginalizing Y produces a new factor

$$\tau(x) = \sum_{y} \phi(x, y)$$

where the sum is over all joint assignments to the set of variables Y.

Example

a^1	b^1	c^1	0.25
a^1	b^1	c^2	0.35
a^1	b^2	c^1	0.08
a^1	b^2	c^2	0.16
a^2	b^1	c^1	0.05
a^2	b^1	c^2	0.07
a^2	b^2	c^1	0
a^2	b^2	c^2	0
a^3	b^1	c^1	0.15
a^3	b^1	c^2	0.21
a^3	b^2	c^1	0.09
a^3	b^2	c^2	0.18

a^1	c^1	0.33
a^1	c^2	0.51
a^2	c^1	0.05
a^2	c^2	0.07
a^3	c^1	0.24
a^3	c^2	0.39

Figure 2: We are marginalizing out variable B from factor $\phi(A, B, C)$.

Variable Elimination Algorithm

- The variable elimination algorithm repeatedly performs two factor operations: product and marginalization.
- We have been implicitly performing these operations in our chain example.
- The variable elimination algorithm requires an ordering over the variables according to which variables will be "eliminated."
- In our chain example, we took the ordering implied by the DAG.
- Note that different orderings may dramatically alter the running time. For simplicity, let the ordering be fixed.

Variable Elimination Algorithm

We are now ready to formally define the variable elimination (VE) algorithm.

For each variable X_i (ordered according to O):

1. Multiply all factors Φ_i containing X_i
2. Marginalize out X_i to obtain a new factor τ
3. Replace the factors Φ_i with τ

Intuitively, this corresponds to choosing a sum and "pushing it in" as far as possible inside the product of the factors

Examples

$$p(x_n) = \sum_{x_{n-1}} p(x_n \mid x_{n-1}) \sum_{x_{n-2}} p(x_{n-1} \mid x_{n-2}) \cdots \sum_{x_2} p(x_3 \mid x_2)\, \tau(x_2)$$

- Let's try to understand what these steps correspond to in our chain example.

- The chosen ordering was $x_1, x_2, \ldots, x_{n-1}$. Starting with x_1, we collected all the factors involving x_1, which were $p(x_1)$ and $p(x_2 \mid x_1)$.

- We then used them to construct a new factor $\tau(x_2) = \sum_{x_1} p(x_2 \mid x_1)\, p(x_1)$.

Examples

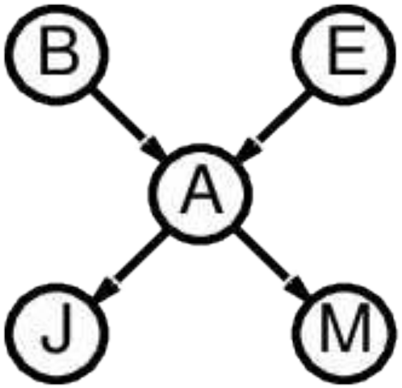
$$p(x_n) = \sum_{x_{n-1}} p(x_n \mid x_{n-1}) \sum_{x_{n-2}} p(x_{n-1} \mid x_{n-2}) \cdots \sum_{x_2} p(x_3 \mid x_2)\, \tau(x_2)$$

- Let's try to understand what these steps correspond to in our chain example.
- The chosen ordering was $x_1, x_2, \ldots, x_{n-1}$. Starting with x_1, we collected all the factors involving x_1, which were $p(x_1)$ and $p(x_2 \mid x_1)$.
- We then used them to construct a new factor $\tau(x_2) = \sum_{x_1} p(x_2 \mid x_1)\, p(x_1)$.
- This can be seen as the results of steps 1 and 2 of the VE algorithm: first we form a large factor $\sigma(x_2, x_1) = p(x_2 \mid x_1)\, p(x_1)$; then we eliminate x_1 from that factor to produce τ.
- Then, we repeat the same procedure for x_2, except that the factors are now $p(x_3 \mid x_2), \tau(x_2)$.

Inference by enumeration

Slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation

Simple query on the burglary network:



$$
\begin{aligned}
&\mathbf{P}(B \mid j, m) \\
&\quad = \mathbf{P}(B, j, m)/P(j, m) \\
&\quad = \alpha \mathbf{P}(B, j, m) \\
&\quad = \alpha \textstyle\sum_e \sum_a \mathbf{P}(B, e, a, j, m)
\end{aligned}
$$

Inference by variable elimination

Variable elimination: carry out summations right-to-left, storing intermediate results, *factors*, to avoid recomputation

Use matrix operations

$$\mathbf{P}(B \mid j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m)$$

Inference by variable elimination

Variable elimination: carry out summations right-to-left, storing intermediate results, *factors*, to avoid recomputation

Use matrix operations

$$\mathbf{P}(B \mid j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m)$$
$$= \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

Inference by variable elimination

Variable elimination: carry out summations right-to-left, storing intermediate results, *factors*, to avoid recomputation

Use matrix operations

$$\mathbf{P}(B \mid j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m)$$
$$= \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$
$$= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)$$

Inference by variable elimination

Variable elimination: carry out summations right-to-left, storing intermediate results, *factors*, to avoid recomputation

Use matrix operations

$$\mathbf{P}(B \mid j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m)$$

$$= \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

$$= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)$$

$$= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_{3,4,5}(A, B, E)$$

Inference by variable elimination

Variable elimination: carry out summations right-to-left, storing intermediate results, *factors*, to avoid recomputation

Use matrix operations

$$\begin{aligned}
\mathbf{P}(B \mid j, m) &= \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m) \\
&= \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)} \\
&= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A) \\
&= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_{3,4,5}(A, B, E) \\
&= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B, E)
\end{aligned}$$

Inference by variable elimination

Variable elimination: carry out summations right-to-left, storing intermediate results, *factors*, to avoid recomputation

Use matrix operations

$$\mathbf{P}(B \mid j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m)$$

$$= \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

$$= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)$$

$$= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_{3,4,5}(A, B, E)$$

$$= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B, E)$$

$$= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_{2,6}(B, E)$$

Inference by variable elimination

Variable elimination: carry out summations right-to-left, storing intermediate results, *factors*, to avoid recomputation

Use matrix operations

$$
\begin{aligned}
\mathbf{P}(B \mid j, m) &= \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m) \\
&= \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)} \\
&= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A) \\
&= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_{3,4,5}(A, B, E) \\
&= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B, E) \\
&= \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_{2,6}(B, E) \\
&= \alpha\, \mathbf{f}_1(B) \times \mathbf{f}_7(B) \\
&= \alpha\, \mathbf{f}_{1,7}(B)
\end{aligned}
$$

Markov Model,
Hidden Markov Model

Introduction

- Board games played with dice: example a game of snakes and ladders or any other game whose moves are determined entirely by dice is a Markov chain.
- Gambling: The probabilities of weather conditions (modeled as either rainy or sunny), given the weather on the preceding day
- A simple weather model: The probabilities of weather conditions (modeled as either rainy or sunny), given the weather on the preceding day.
- Stock Market Trend Prediction can be done using Markov Model.

Overview

Markov Models		Do we have control over the state transitons?	
		NO	**YES**
Are the states completely observable?	**YES**	**Markov Chain**	**MDP** Markov Decision Process
	NO	**HMM** Hidden Markov Model	**POMDP** Partially Observable Markov Decision Process

Introduction

- **Markov chain or Markov process**: It is a probabilistic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event (i.e., memorylessness).

- **Markov model**: It models the state of a system with a random variable that changes through time. The simplest Markov model is the Markov chain.

- **Hidden Markov model**: It is a Markov chain for which the state is only partially observable or noisily observable.

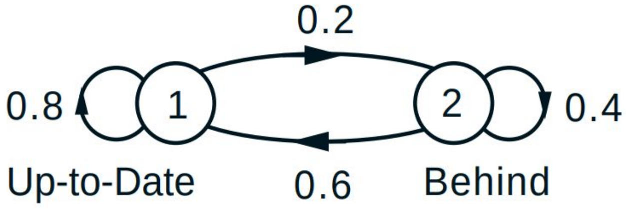
Introduction

- **Markov decision process (MDP)**: It is a Markov chain in which state transitions depend on current ***state*** and an ***action vector*** applied to the system. MDP is used to compute a policy of actions to maximize utility with respect to expected rewards.

- **Partially observable Markov decision process (POMDP)**: It is a MDP in which the state of the system is only partially observed.

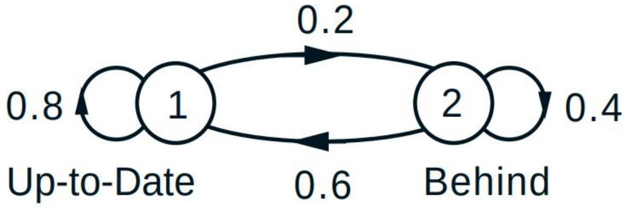
Markov model

Markov Model Example



- Alice can be either up-to-Date or Behind (state 1 or 2) for her classes. The figure shows transition probability graph.
- We assume that probabilities for next week only depends upon present week and do not depend on if she was up-to-date or behind in previous weeks.
- We can model the above scenario using Markov Model (see figure).

Markov Model Example



$$\begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$$

$$[\boldsymbol{\pi}_1 \quad \boldsymbol{\pi}_2]$$

- Alice can be either up-to-Date or Behind (state 1 or 2) for her classes. The figure shows transition probability graph.
- We assume that probabilities for next week only depends upon present week and do not depend on if she was up-to-date or behind in previous weeks.
- We can model the above scenario using Markov Model.
- The task is to determine the long-term probabilities $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ (steady state)

Specification of Markov Models

- A Markov chain model is specified by identifying
  (a) the set of states $\mathcal{S} = \{1, \ldots, m\}$,
  (b) the set of possible transitions, namely, those pairs (i, j) for which $p_{ij} > 0$, and,
  (c) the numerical values of those p_{ij} that are positive.

A Markov Chain or Markov process

Markov processes are built on the **Markov assumption** that the probability of transitioning to a particular state at the next time-step relies only on the current state, and does not require any knowledge of the history of states that came before that, or

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j \mid X_n = i)$$

$$= p_{ij},$$ (transition probabilities)

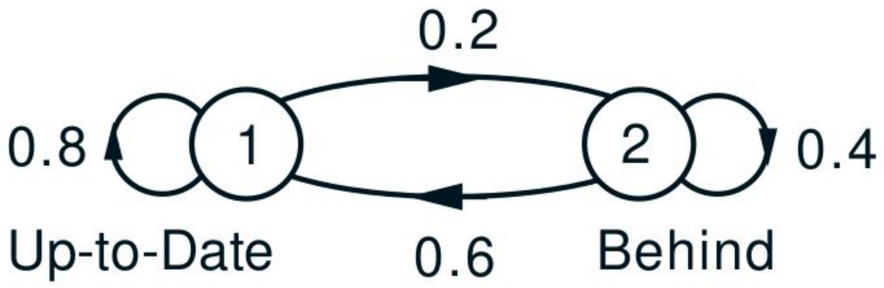
The transition probabilities p_{ij} must be of course nonnegative, and sum to one:

$$\sum_{j=1}^{m} p_{ij} = 1, \qquad \text{for all } i.$$

Transition probability matrix

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

Example



	UpD	B
UpD	0.8	0.2
B	0.6	0.4

n-Step Transition Probabilities

Many Markov chain problems require the calculation of the probability law of the state at some future time, conditioned on the current state. This probability law is captured by the n-**step transition probabilities**, defined by

$$r_{ij}(n) = \mathbf{P}(X_n = j \mid X_0 = i).$$

In words, $r_{ij}(n)$ is the probability that the state after n time periods will be j, given that the current state is i.

n-Step Transition

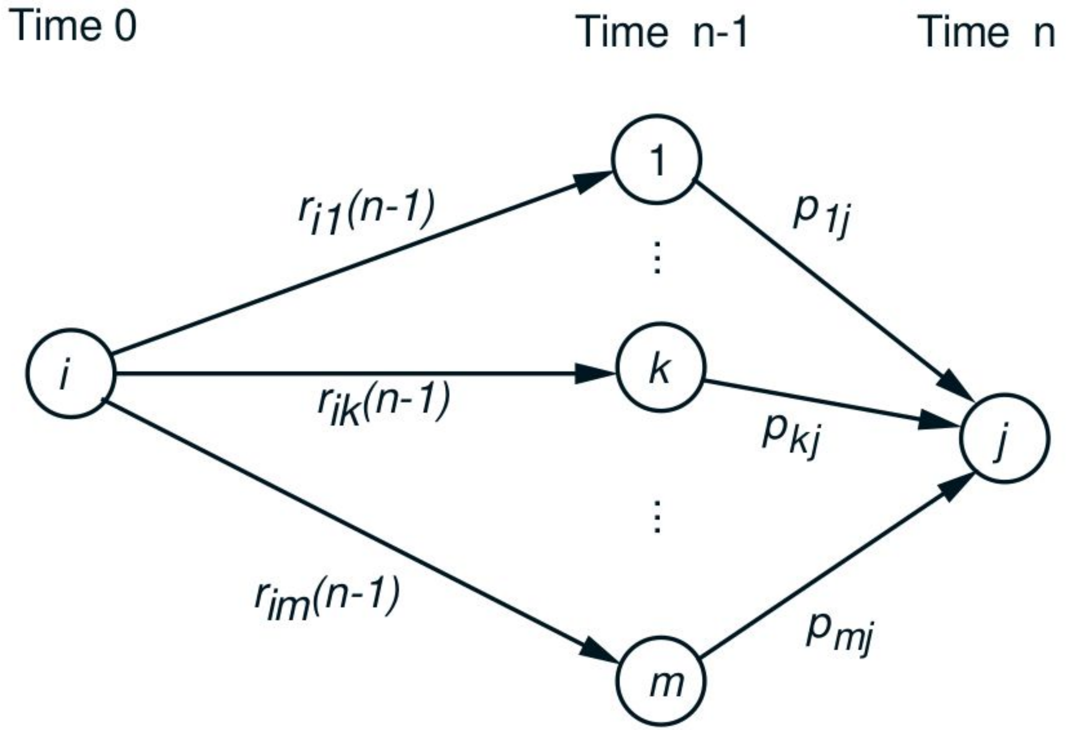
The n-step transition probabilities can be generated by the recursive formula

$$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1)p_{kj}, \qquad \text{for } n > 1, \quad \text{and all } i,\ j,$$

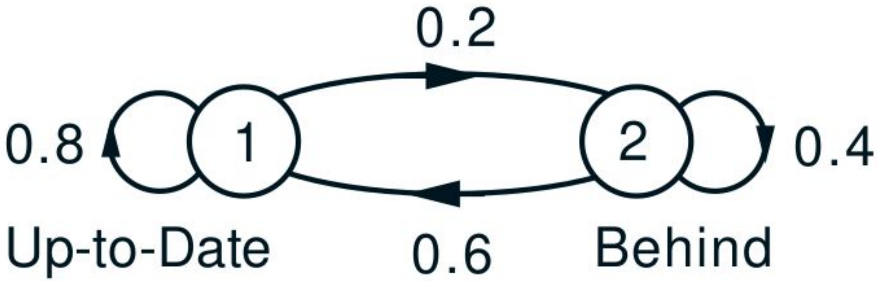
starting with

$$r_{ij}(1) = p_{ij}.$$

n-Step Transition Probabilities



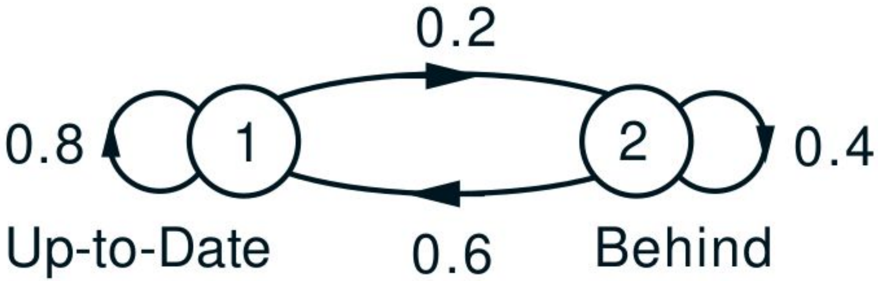
Example



	UpD	B
UpD	0.8	0.2
B	0.6	0.4

$r_{ij}(1)$

Example



	UpD	B
UpD	0.8	0.2
B	0.6	0.4

$r_{ij}(1)$

.76	.24
.72	.28

$r_{ij}(2)$

Example

$$\underbrace{\begin{array}{cc} & \begin{array}{cc} \mathrm{UpD} & \mathrm{B} \end{array} \\ \begin{array}{c} \mathrm{UpD} \\ \mathrm{B} \end{array} & \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \end{array}}_{r_{ij}(1)} \quad \underbrace{\begin{array}{cc} & \begin{array}{cc} \mathrm{UpD} & \mathrm{B} \end{array} \\ \begin{array}{c} \mathrm{UpD} \\ \mathrm{B} \end{array} & \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \end{array}}_{r_{ij}(1)} = \underbrace{\begin{bmatrix} .76 & .24 \\ .72 & .28 \end{bmatrix}}_{r_{ij}(2)}$$

Transition probability matrix

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

Let A be an $N \times N$ matrix. Then

$$(A^2)_{ij} \ = \ \sum_{k=1}^{N} (A)_{ik}(A)_{kj} = \sum_{k=1}^{N} a_{ik}a_{kj}.$$

Example

	UpD	B
UpD	0.8	0.2
B	0.6	0.4

$r_{ij}(1)$

.76	.24
.72	.28

$r_{ij}(2)$

.752	.248
.744	.256

$r_{ij}(3)$

.7504	.2496
.7488	.2512

$r_{ij}(4)$

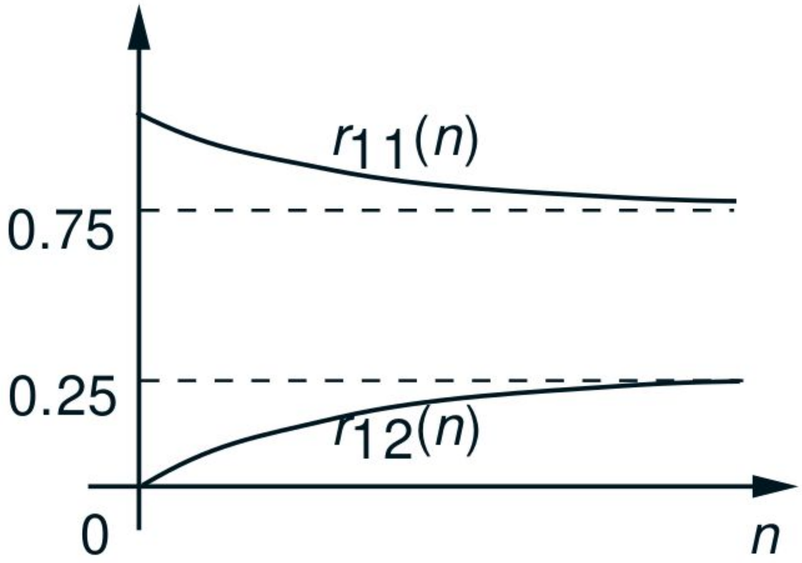
.7501	.2499
.7498	.2502

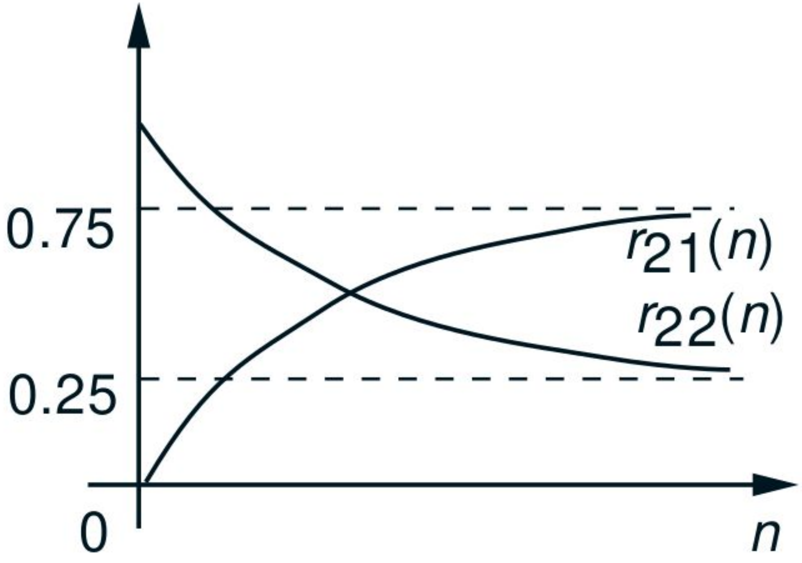
$r_{ij}(5)$

Sequence of n-step transition probability matrices

Example

$$\text{long-term probabilities (steady state)} = \begin{array}{cc} \text{UpD} & \text{B} \\ \boldsymbol{\pi_1} & \boldsymbol{\pi_2} \end{array} = [0.75 \quad 0.25]$$





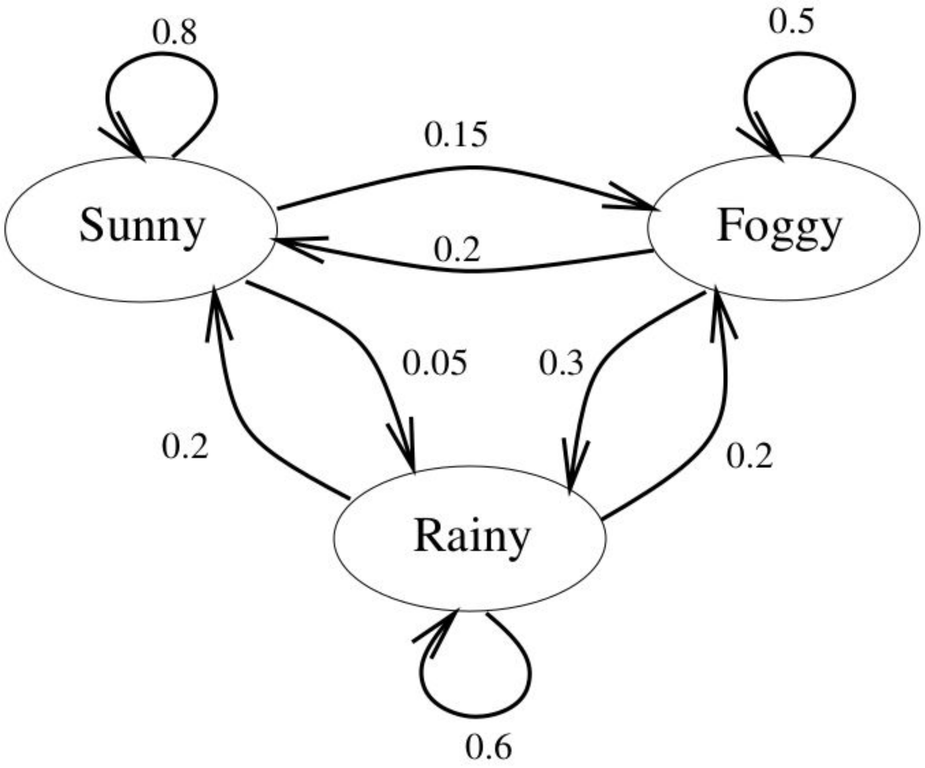
n-step transition probabilities as a function of the number *n* of transitions

Weather Example

		Tomorrow's Weather		
		Sunny	Rainy	Foggy
Today's Weather	Sunny	0.8	0.05	0.15
	Rainy	0.2	0.6	0.2
	Foggy	0.2	0.3	0.5

Table 1: Probabilities of Tomorrow's weather based on Today's Weather

Weather Example



Example: *Given today is sunny, what is the probability that tomorrow is sunny and the day after is rainy?*

$$P\left(w_2 = \text{ Sun },w_3 = \text{ Rain } \mid w_1 = \text{ Sun }\right)$$

Example: *Given today is sunny, what is the probability that tomorrow is sunny and the day after is rainy?*

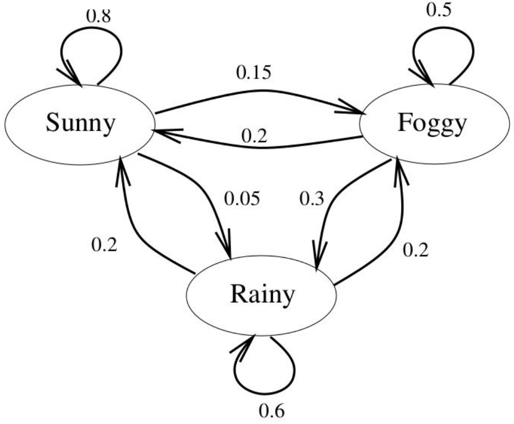
$$P(w_2 = \text{ Sun }, w_3 = \text{ Rain } \mid w_1 = \text{ Sun }) = P(w_3 = \text{ Rain } \mid w_2 = \text{ Sun }, w_1 = \text{ Sun }) * P(w_2 = \text{ Sun } \mid w_1 = \text{ Sun })$$

Example: *Given today is sunny, what is the probability that tomorrow is sunny and the day after is rainy?*

$$
\begin{aligned}
P(w_2 = \text{ Sun }, w_3 = \text{ Rain } \mid w_1 = \text{ Sun }) = & P(w_3 = \text{ Rain } \mid w_2 = \text{ Sun }, w_1 = \text{ Sun }) * \\
& P(w_2 = \text{ Sun } \mid w_1 = \text{ Sun }) \\
= & P(w_3 = \text{ Rain } \mid w_2 = \text{ Sun }) * \\
& P(w_2 = \text{ Sun } \mid w_1 = \text{ Sun })
\end{aligned}
$$

Example: *Given today is sunny, what is the probability that tomorrow is sunny and the day after is rainy?*

$$
\begin{aligned}
P(w_2 = \text{ Sun }, w_3 = \text{ Rain } \mid w_1 = \text{ Sun }) =& P(w_3 = \text{ Rain } \mid w_2 = \text{ Sun }, w_1 = \text{ Sun }) * \\
& P(w_2 = \text{ Sun } \mid w_1 = \text{ Sun }) \\
=& P(w_3 = \text{ Rain } \mid w_2 = \text{ Sun }) * \\
& P(w_2 = \text{ Sun } \mid w_1 = \text{ Sun }) \\
=& (0.05)(0.8) \\
=& 0.04
\end{aligned}
$$



Example: *Given today is foggy, what's the probability that it will be rainy two days from now?*

Example: *Given today is foggy, what's the probability that it will be rainy two days from now?*

There are three ways to get from foggy today to rainy two days from now: {foggy, foggy, rainy }, { foggy, rainy, rainy }, and { foggy, sunny, rainy }. Therefore, lets sum over these paths:

Example: *Given today is foggy, what's the probability that it will be rainy two days from now?*

There are three ways to get from foggy today to rainy two days from now: {foggy, foggy, rainy }, { foggy, rainy, rainy }, and { foggy, sunny, rainy }. Therefore, lets sum over these paths:

$$P(w_3 = \text{ Rain } \mid w_1 = \text{ Fog })$$

Example: *Given today is foggy, what's the probability that it will be rainy two days from now?*

There are three ways to get from foggy today to rainy two days from now: {foggy, foggy, rainy }, { foggy, rainy, rainy }, and { foggy, sunny, rainy }. Therefore, lets sum over these paths:

$$
\begin{aligned}
P(w_3 = \text{ Rain } \mid w_1 = \text{ Fog })& \\
&= P(w_2 = \text{ Fog }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) + \\
&\quad P(\qquad\qquad\qquad\qquad\qquad\qquad) + \\
&\quad P(\qquad\qquad\qquad\qquad\qquad\qquad) +
\end{aligned}
$$

Example: *Given today is foggy, what's the probability that it will be rainy two days from now?*

There are three ways to get from foggy today to rainy two days from now: {foggy, foggy, rainy }, { foggy, rainy, rainy }, and { foggy, sunny, rainy }. Therefore, lets sum over these paths:

$$
\begin{aligned}
P\left(w_3 = \text{ Rain } \mid w_1 = \text{ Fog }\right) & \\
&= P\left(w_2 = \text{ Fog }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }\right) + \\
&\quad P\left(w_2 = \text{ Rain}, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }\right) + \\
&\quad P\left(w_2 = \text{ Sun }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }\right) +
\end{aligned}
$$

Example: *Given today is foggy, what's the probability that it will be rainy two days from now?*

There are three ways to get from foggy today to rainy two days from now: {foggy, foggy, rainy }, { foggy, rainy, rainy }, and { foggy, sunny, rainy }. Therefore, lets sum over these paths:

$$
\begin{aligned}
P(w_3 = \text{ Rain } \mid w_1 = \text{ Fog })\\
&= P(w_2 = \text{ Fog }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) + \\
&\quad P(w_2 = \text{ Rain}, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) + \\
&\quad P(w_2 = \text{ Sun }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) + \\
&= P(w_3 = \text{ Rain } \mid w_2 = \text{ Fog })\, P(w_2 = \text{ Fog } \mid w_1 = \text{ Fog }) + \\
&\quad P(\qquad\qquad\qquad\qquad) + \\
&\quad P(\qquad\qquad\qquad\qquad)
\end{aligned}
$$

Example: *Given today is foggy, what's the probability that it will be rainy two days from now?*

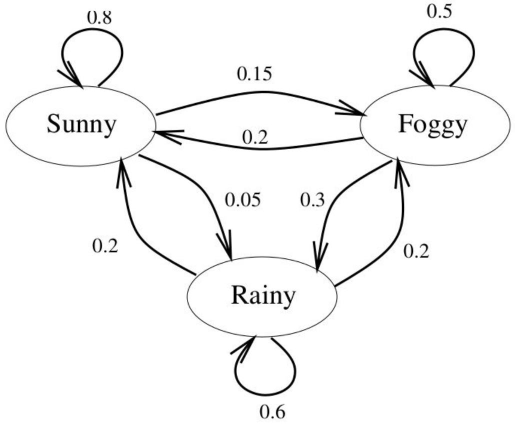
There are three ways to get from foggy today to rainy two days from now: {foggy, foggy, rainy }, { foggy, rainy, rainy }, and { foggy, sunny, rainy }. Therefore, lets sum over these paths:

$$
\begin{aligned}
P(w_3 = \text{ Rain } \mid w_1 = \text{ Fog })\\
&= P(w_2 = \text{ Fog }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) + \\
&\quad P(w_2 = \text{ Rain}, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) + \\
&\quad P(w_2 = \text{ Sun }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) + \\
&= P(w_3 = \text{ Rain } \mid w_2 = \text{ Fog })\, P(w_2 = \text{ Fog } \mid w_1 = \text{ Fog }) + \\
&\quad P(w_3 = \text{ Rain } \mid w_2 = \text{ Rain })\, P(w_2 = \text{ Rain } \mid w_1 = \text{ Fog }) + \\
&\quad P(w_3 = \text{ Rain } \mid w_2 = \text{ Sun })\, P(w_2 = \text{ Sun } \mid w_1 = \text{ Fog })
\end{aligned}
$$

Example: *Given today is foggy, what's the probability that it will be rainy two days from now?*

There are three ways to get from foggy today to rainy two days from now: {foggy, foggy, rainy }, { foggy, rainy, rainy }, and { foggy, sunny, rainy }. Therefore, lets sum over these paths:

$$
\begin{aligned}
&P(w_3 = \text{ Rain } \mid w_1 = \text{ Fog })\\
&= P(w_2 = \text{ Fog }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) +\\
&\quad P(w_2 = \text{ Rain}, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) +\\
&\quad P(w_2 = \text{ Sun }, w_3 = \text{ Rain } \mid w_1 = \text{ Fog }) +\\
&= P(w_3 = \text{ Rain } \mid w_2 = \text{ Fog })\, P(w_2 = \text{ Fog } \mid w_1 = \text{ Fog }) +\\
&\quad P(w_3 = \text{ Rain } \mid w_2 = \text{ Rain })\, P(w_2 = \text{ Rain } \mid w_1 = \text{ Fog }) +\\
&\quad P(w_3 = \text{ Rain } \mid w_2 = \text{ Sun })\, P(w_2 = \text{ Sun } \mid w_1 = \text{ Fog })\\
&= (0.3)(0.5) + (0.6)(0.3) + (0.05)(0.2) = 0.34
\end{aligned}
$$

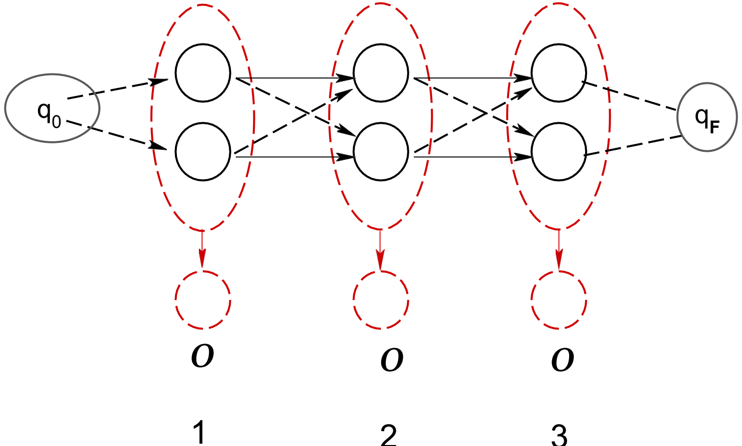


Hidden Markov Model

HMM components

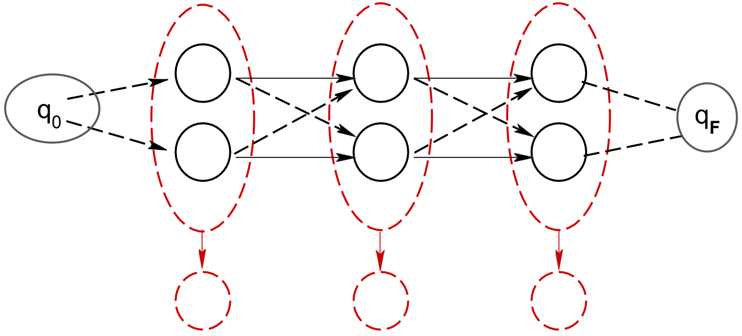
An **HMM** is specified by the following components:

$Q = q_1 q_2 \ldots q_N$	a set of N **states**
$A = a_{11} a_{12} \ldots a_{n1} \ldots a_{nn}$	a **transition probability matrix** A, each a_{ij} representing the probability of moving from state i to state j, s.t. $\sum_{j=1}^{n} a_{ij} = 1 \quad \forall i$



HMM components

$O = o_1 o_2 \ldots o_T$	a sequence of T **observations**,
$B = b_i(o_t)$	A sequence of **observation likelihoods:**, also called **emission probabilities**, each expressing the probability of an observation o_t being generated from a state i.

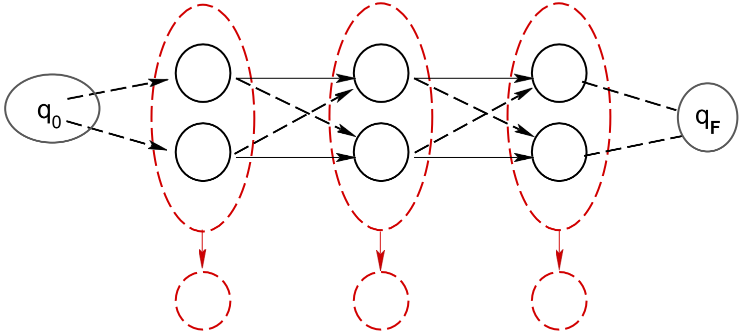


o o o

1 2 3

HMM components

q_0, q_F a special **start state** and **end (final) state** which are not associated with observations, together with transition probabilities $a_{01}a_{02}..a_{0n}$ out of the start state and $a_{1F}a_{2F}...a_{nF}$ into the end state.



o_1 o_2 o_3

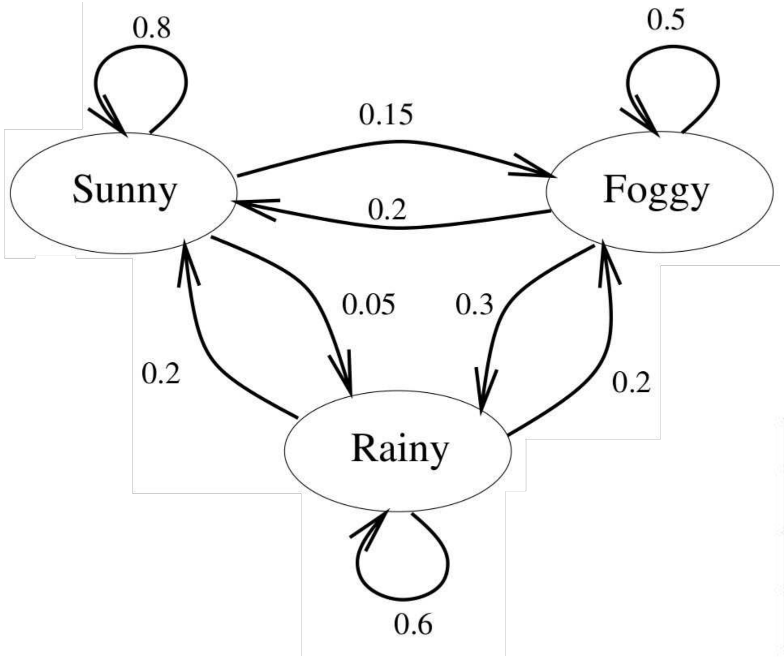
Weather Example for HMM

- Suppose you are locked in some secret underground room
- You want to know whether it is raining today
- But, your only access to the outside world occurs each morning, when you see the caretaker coming in with, or without an umbrella.

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Table 2: Probabilities of Seeing an Umbrella Based on the Weather

Weather Example for HMM



	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Table 2: Probabilities of Seeing an Umbrella Based on the Weather

Example: *Suppose the day you were locked in it was sunny. The next day, the caretaker carried an umbrella into the room. Assuming that the prior probability of the caretaker carrying an umbrella on any day is 0.5 , what's the probability that the second day was rainy?*

$$
\begin{aligned}
P(w_2 = \text{ Rain } \mid w_1 = \text{ Sun}, u_2 = \text{ True }) &= \frac{P(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \mathrm{T})}{P(w_1 = \text{ Sun } \mid u_2 = \mathrm{T})} \\
(u_2 \text{ and } w_1 \text{ independent }) &= \\
(\text{ Bayes' Rule }) &= \\
(\text{ Markov assumption }) &= \\
&=
\end{aligned}
$$

Example: *Suppose the day you were locked in it was sunny. The next day, the caretaker carried an umbrella into the room. Assuming that the prior probability of the caretaker carrying an umbrella on any day is 0.5 , what's the probability that the second day was rainy?*

$$
\begin{aligned}
P(w_2 = \text{ Rain } \mid w_1 = \text{ Sun}, u_2 = \text{ True }) &= \frac{P(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \text{T})}{P(w_1 = \text{ Sun } \mid u_2 = \text{T})} \\
(u_2 \text{ and } w_1 \text{ independent }) &= \frac{P(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \text{T})}{P(w_1 = \text{ Sun })} \\
(\text{ Bayes' Rule }) &= \\
(\text{ Markov assumption }) &= \\
&=
\end{aligned}
$$

Example: *Suppose the day you were locked in it was sunny. The next day, the caretaker carried an umbrella into the room. Assuming that the prior probability of the caretaker carrying an umbrella on any day is 0.5 , what's the probability that the second day was rainy?*

$$
\begin{aligned}
P(w_2 = \text{ Rain } \mid w_1 = \text{ Sun}, u_2 = \text{ True }) &= \frac{P(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \text{T})}{P(w_1 = \text{ Sun } \mid u_2 = \text{T})} \\
(u_2 \text{ and } w_1 \text{ independent }) &= \frac{P(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \text{T})}{P(w_1 = \text{ Sun })} \\
(\text{ Bayes' Rule }) &= \frac{P(u_2 = \text{T} \mid w_1 = \text{ Sun}, w_2 = \text{ Rain })\, P(w_2 = \text{ Rain}, w_1 = \text{ Sun })}{P(w_1 = \text{ Sun })\, P(u_2 = \text{T})} \\
(\text{ Markov assumption }) &= \\
&=
\end{aligned}
$$

Example: *Suppose the day you were locked in it was sunny. The next day, the caretaker carried an umbrella into the room. Assuming that the prior probability of the caretaker carrying an umbrella on any day is 0.5 , what's the probability that the second day was rainy?*

$$\begin{aligned}
P\left(w_2 = \text{ Rain } \mid w_1 = \text{ Sun}, u_2 = \text{ True }\right) &= \frac{P\left(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \mathrm{T}\right)}{P\left(w_1 = \text{ Sun } \mid u_2 = \mathrm{T}\right)} \\
\left(u_2 \text{ and } w_1 \text{ independent }\right) &= \frac{P\left(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \mathrm{T}\right)}{P\left(w_1 = \text{ Sun }\right)} \\
(\text{ Bayes' Rule }) &= \frac{P\left(u_2 = \mathrm{T} \mid w_1 = \text{ Sun}, w_2 = \text{ Rain }\right) P\left(w_2 = \text{ Rain}, w_1 = \text{ Sun }\right)}{P\left(w_1 = \text{ Sun }\right) P\left(u_2 = \mathrm{T}\right)} \\
(\text{ Markov assumption }) &= \frac{P\left(u_2 = \mathrm{T} \mid w_2 = \text{ Rain }\right) P\left(w_2 = \text{ Rain}, w_1 = \text{ Sun }\right)}{P\left(w_1 = \text{Sun}\right) P\left(u_2 = \mathrm{T}\right)}
\end{aligned}$$

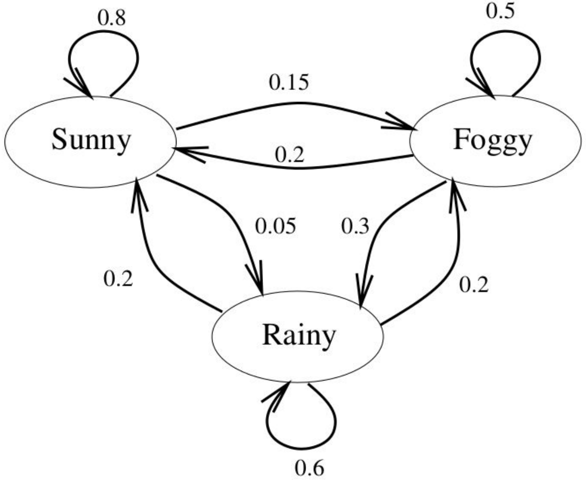
Example: *Suppose the day you were locked in it was sunny. The next day, the caretaker carried an umbrella into the room. Assuming that the prior probability of the caretaker carrying an umbrella on any day is 0.5 , what's the probability that the second day was rainy?*

$$
\begin{aligned}
P\left(w_2 = \text{ Rain } \mid \right. & \\
\left. w_1 = \text{ Sun}, u_2 = \text{ True }\right) &= \frac{P\left(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \mathrm{T}\right)}{P\left(w_1 = \text{ Sun } \mid u_2 = \mathrm{T}\right)} \\
\left(u_2 \text{ and } w_1 \text{ independent }\right) &= \frac{P\left(w_2 = \text{ Rain}, w_1 = \text{ Sun } \mid u_2 = \mathrm{T}\right)}{P\left(w_1 = \text{ Sun }\right)} \\
(\text{ Bayes' Rule }) &= \frac{P\left(u_2 = \mathrm{T} \mid w_1 = \text{ Sun}, w_2 = \text{ Rain }\right) P\left(w_2 = \text{ Rain}, w_1 = \text{ Sun }\right)}{P\left(w_1 = \text{ Sun }\right) P\left(u_2 = \mathrm{T}\right)} \\
(\text{ Markov assumption }) &= \frac{P\left(u_2 = \mathrm{T} \mid w_2 = \text{ Rain }\right) P\left(w_2 = \text{ Rain}, w_1 = \text{ Sun }\right)}{P\left(w_1 = \text{Sun}\right) P\left(u_2 = \mathrm{T}\right)} \\
&= \frac{P\left(u_2 = \mathrm{T} \mid w_2 = \text{ Rain }\right) P\left(w_2 = \text{ Rain } \mid w_1 = \text{ Sun }\right) P\left(w_1 = \text{ Sun }\right)}{P\left(w_1 = \text{ Sun }\right) P\left(u_2 = \mathrm{T}\right)}
\end{aligned}
$$

$$(\text{ Cancel } : P(\text{ Sun })) = \frac{P\left(u_2 = \mathrm{T} \mid w_2 = \text{ Rain }\right) P\left(w_2 = \text{ Rain } \mid w_1 = \text{ Sun }\right)}{P\left(u_2 = \mathrm{T}\right)}$$

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Table 2: Probabilities of Seeing an Umbrella Based on the Weather



$$(\text{ Cancel } : P(\text{ Sun })) = \frac{P(u_2 = \text{T} \mid w_2 = \text{ Rain }) P(w_2 = \text{ Rain } \mid w_1 = \text{ Sun })}{P(u_2 = \text{T})}$$

$$= \frac{(0.8)(0.05)}{0.5} = .08$$

	Probability of Umbrella
Sunny	0.1
Rainy	0.8
Foggy	0.3

Table 2: Probabilities of Seeing an Umbrella Based on the Weather