State sequence prediction in imprecise hidden Markov models

Jasper De Bock & Gert de Cooman

BENE@WORK

7 December 2011

Jasper De Bock & Gert de Cooman

Jasper De Bock & Gert de Cooman

UNIVERSITEIT GENT

Research group

SYSTeMS

Jasper De Bock

Gert de Cooman

UNIVERSITEIT GENT

Research group

SYSTeMS

Jasper De Bock

Gert de Cooman

Filip Hermans

Erik Quaeghebeur

Keivan Shariatmadar

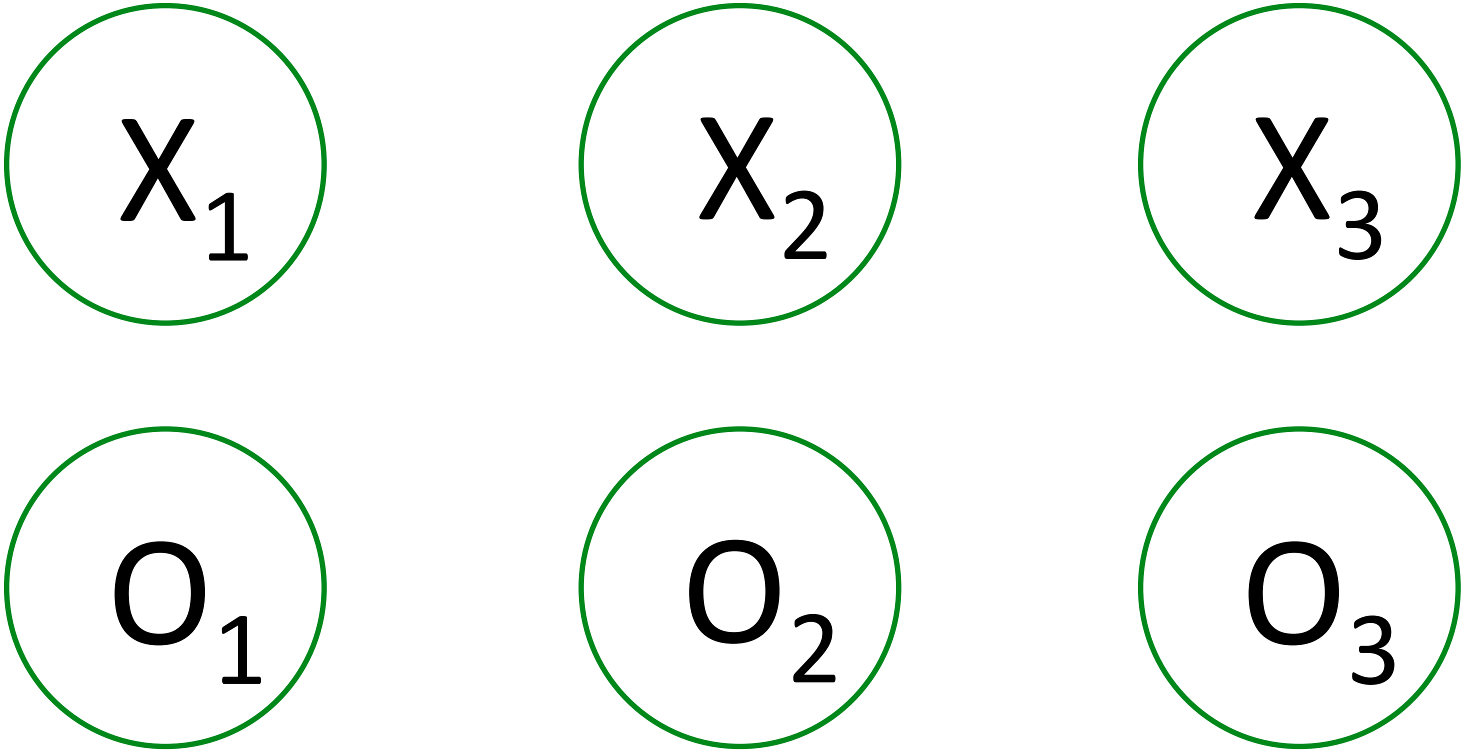
Arthur Van Camp

State sequence prediction in imprecise hidden Markov models

State sequence prediction in imprecise **hidden Markov models**

(Precise) hidden Markov model

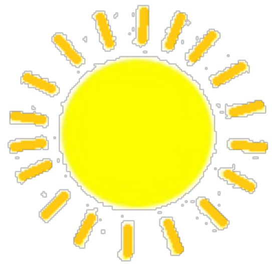
A sequence of hidden state variables

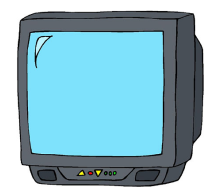


A sequence of observable variables

(Precise) hidden Markov model

A sequence of hidden state variables

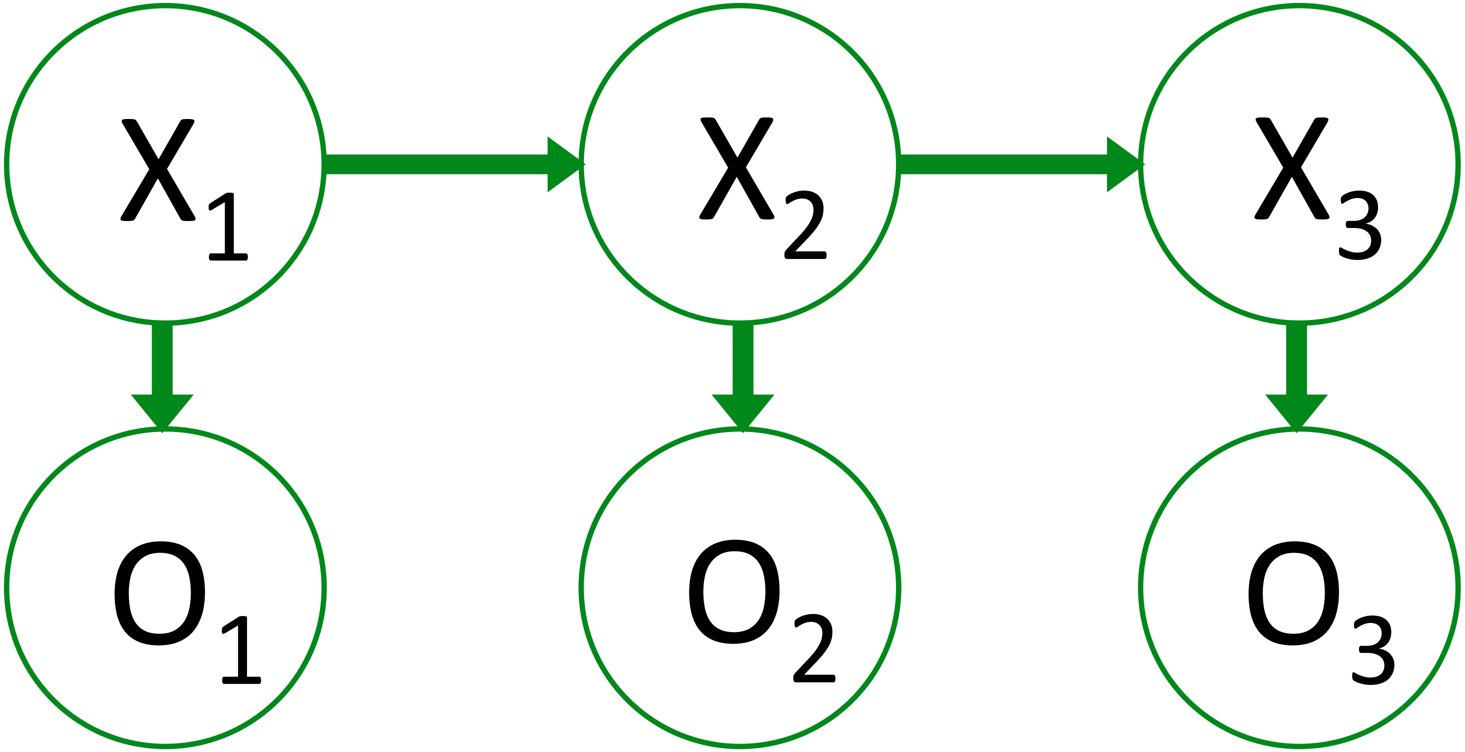
X =  or  or 

O =  or  or 

A sequence of observable variables

(Precise) hidden Markov model

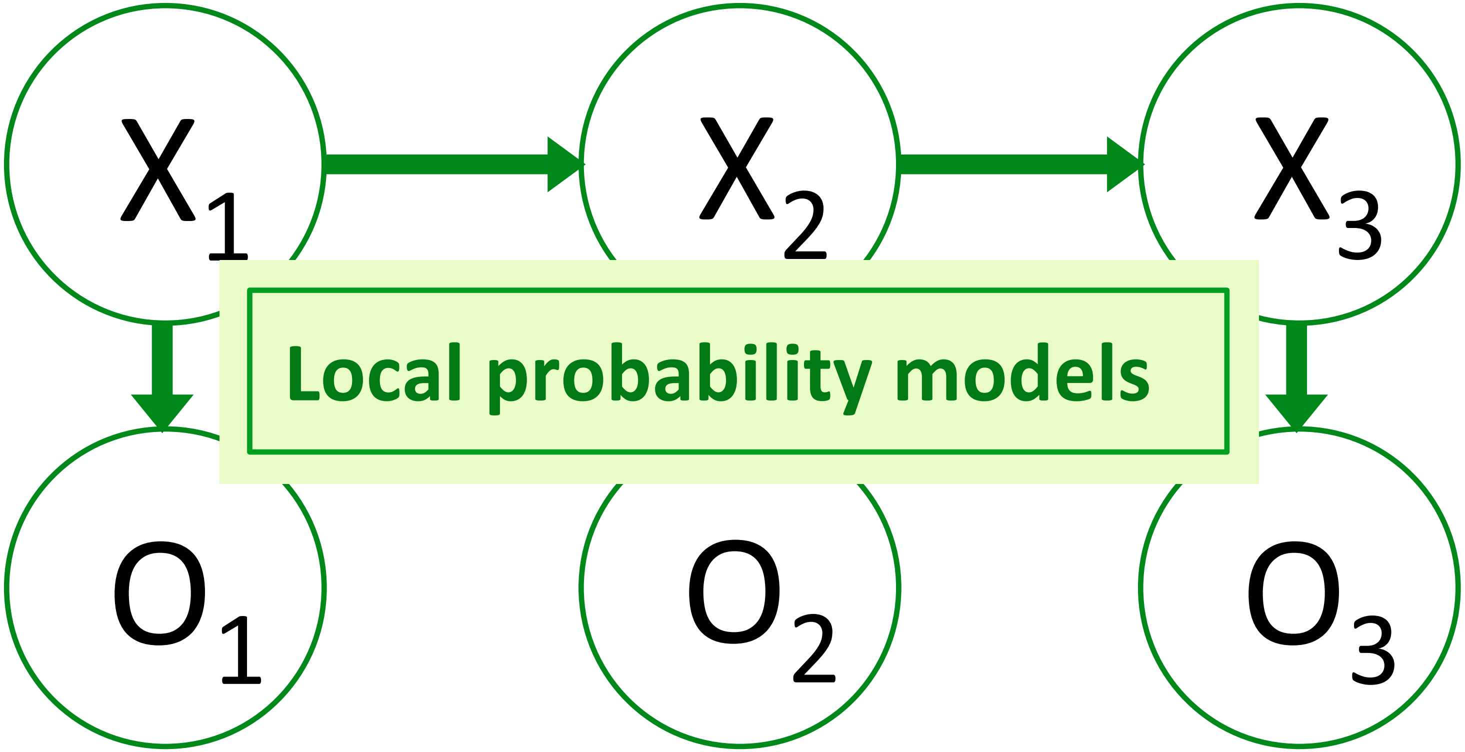
A sequence of hidden state variables



A sequence of observable variables

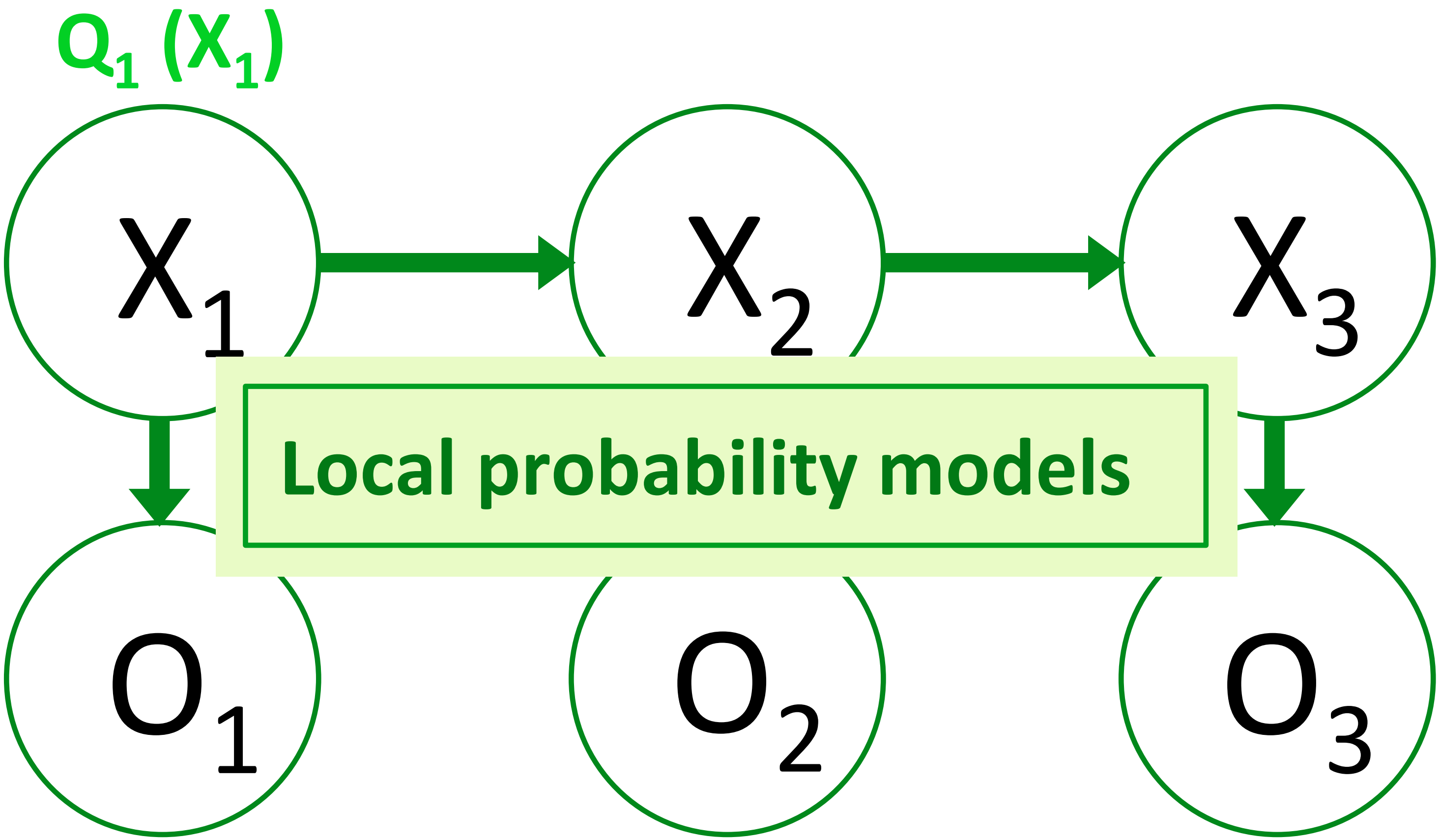
(Precise) hidden Markov model

A sequence of hidden state variables



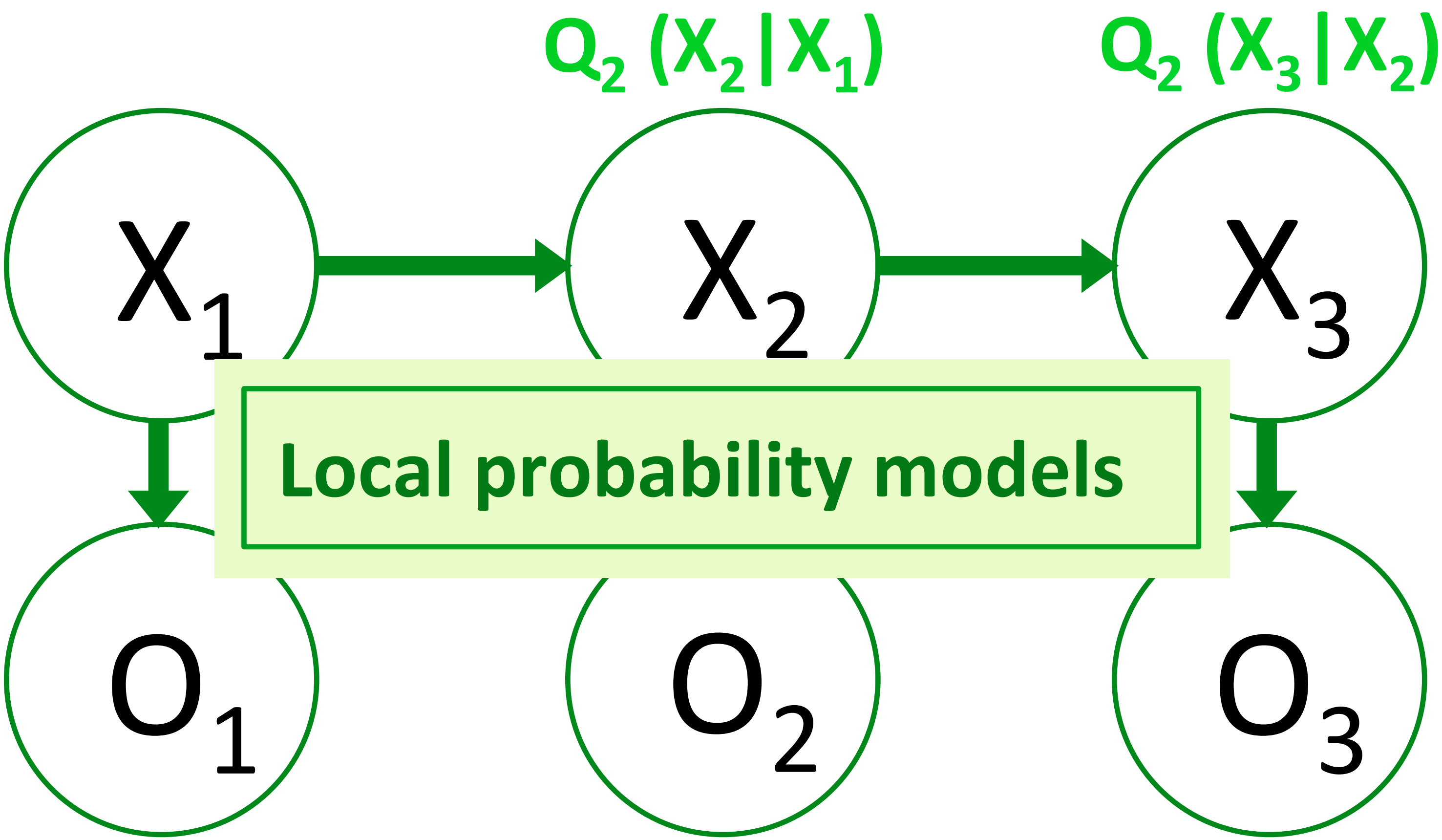
A sequence of observable variables

(Precise) hidden Markov model



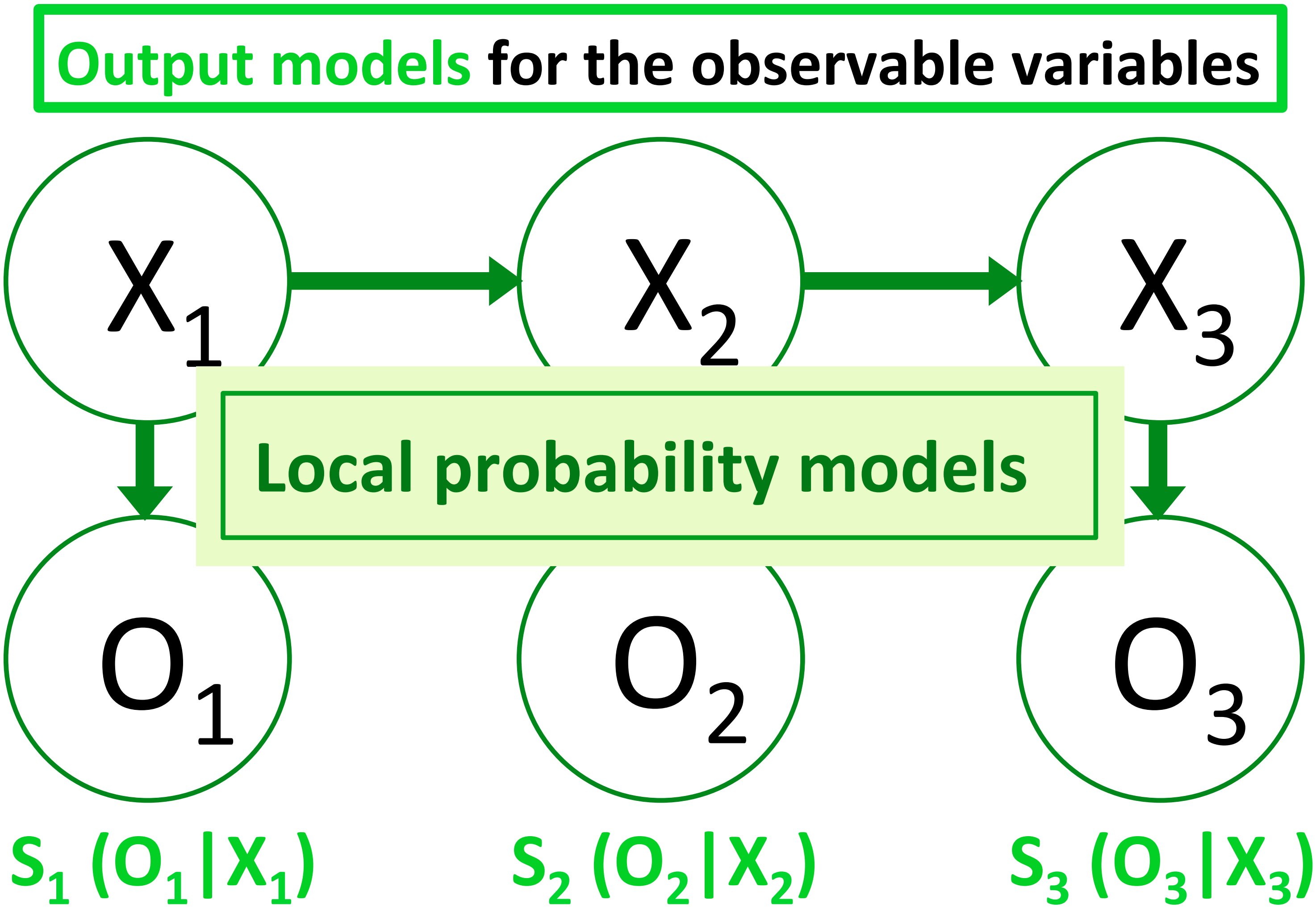
Marginal model for the first hidden variable

(Precise) hidden Markov model

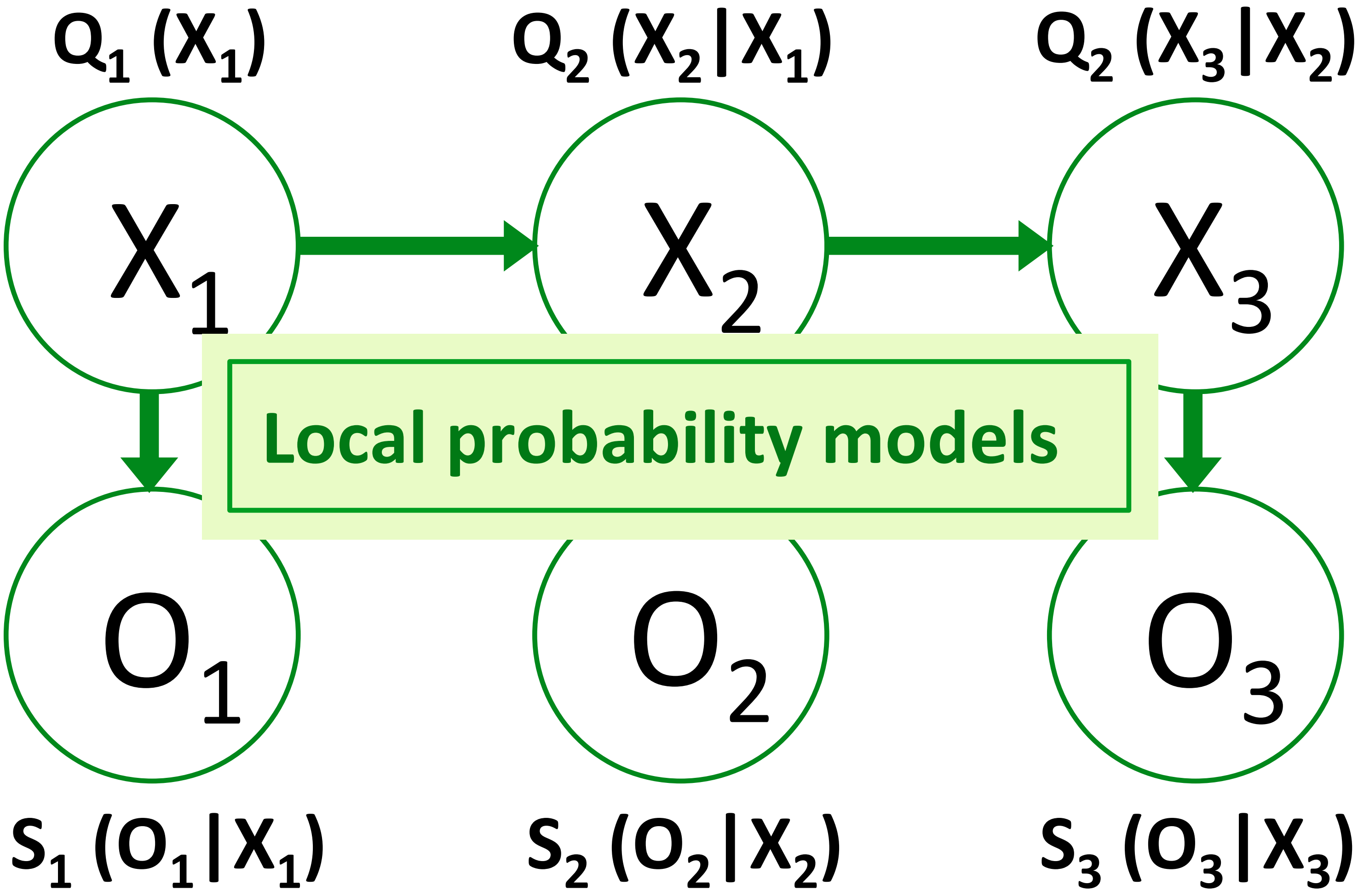


Transition models for the next hidden variables

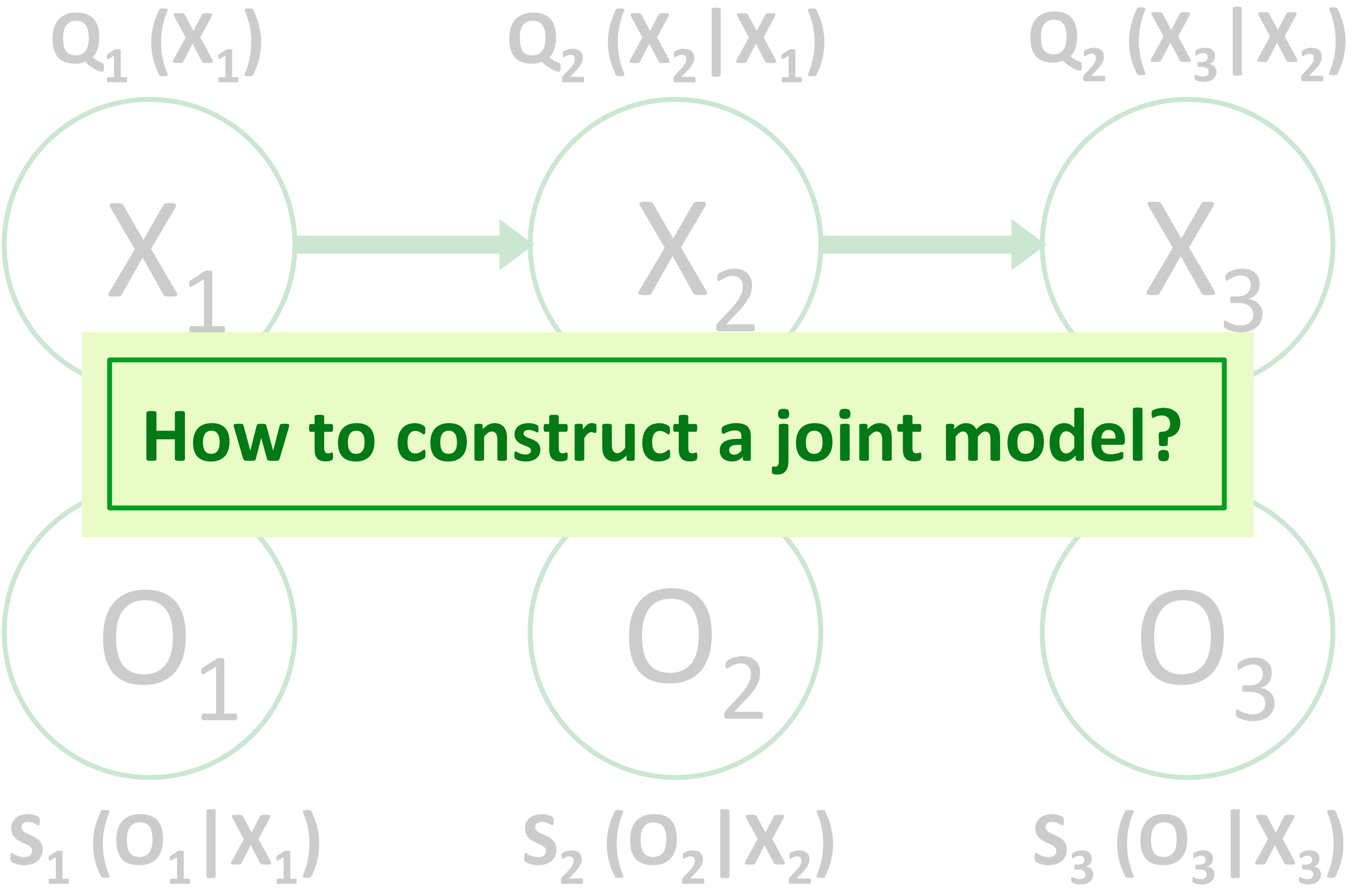
(Precise) hidden Markov model



(Precise) hidden Markov model

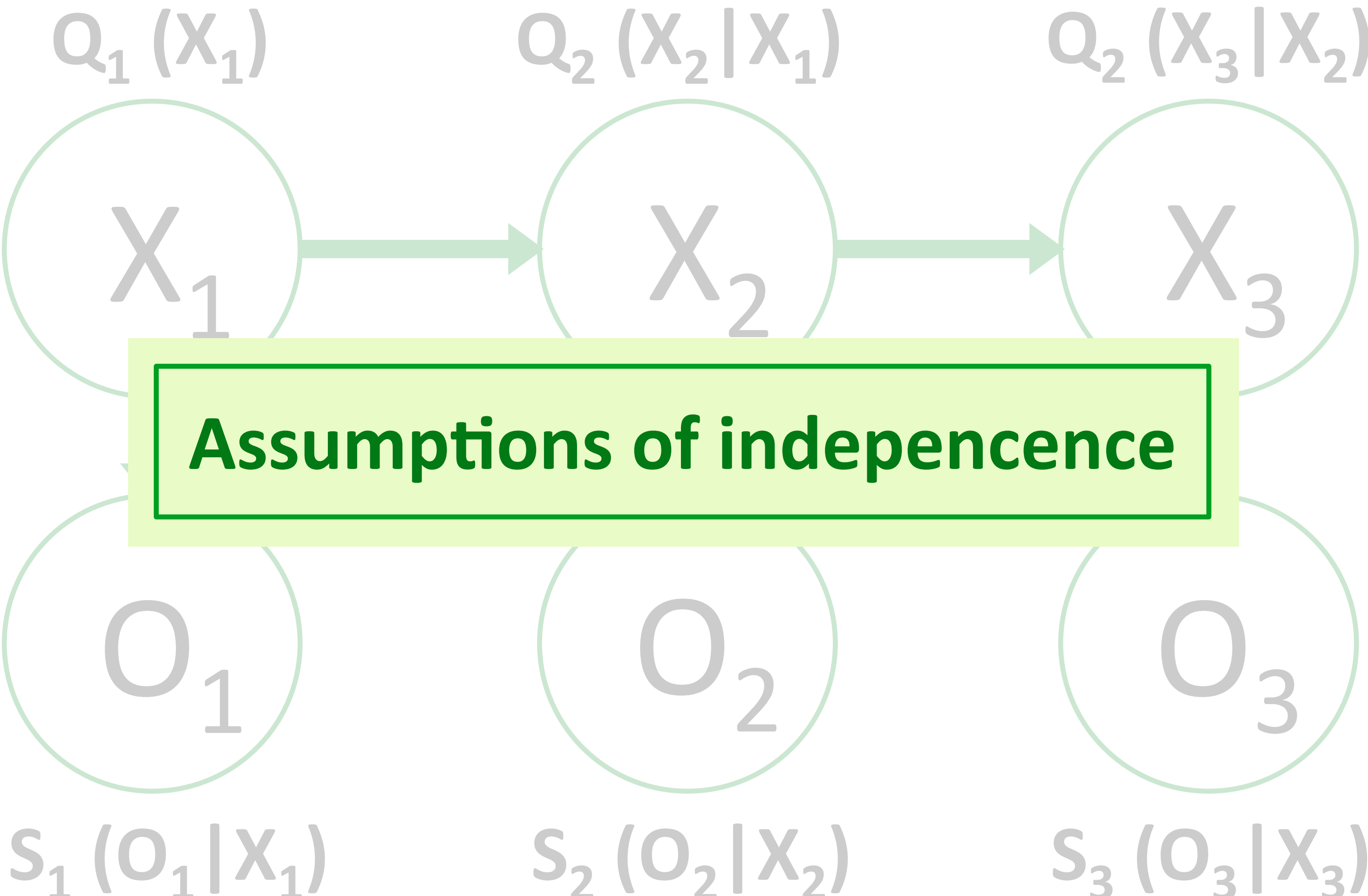


(Precise) hidden Markov model



How to construct a joint model?

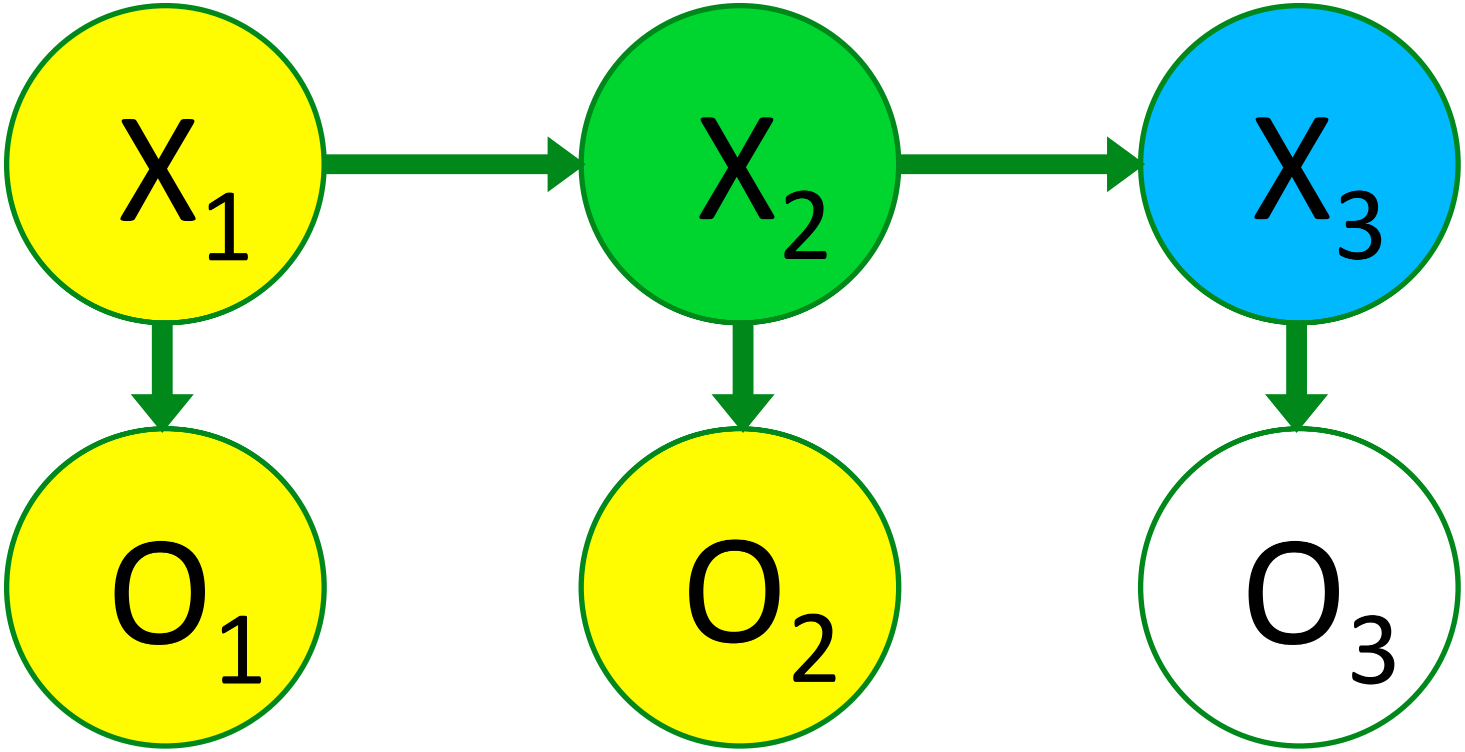
(Precise) hidden Markov model



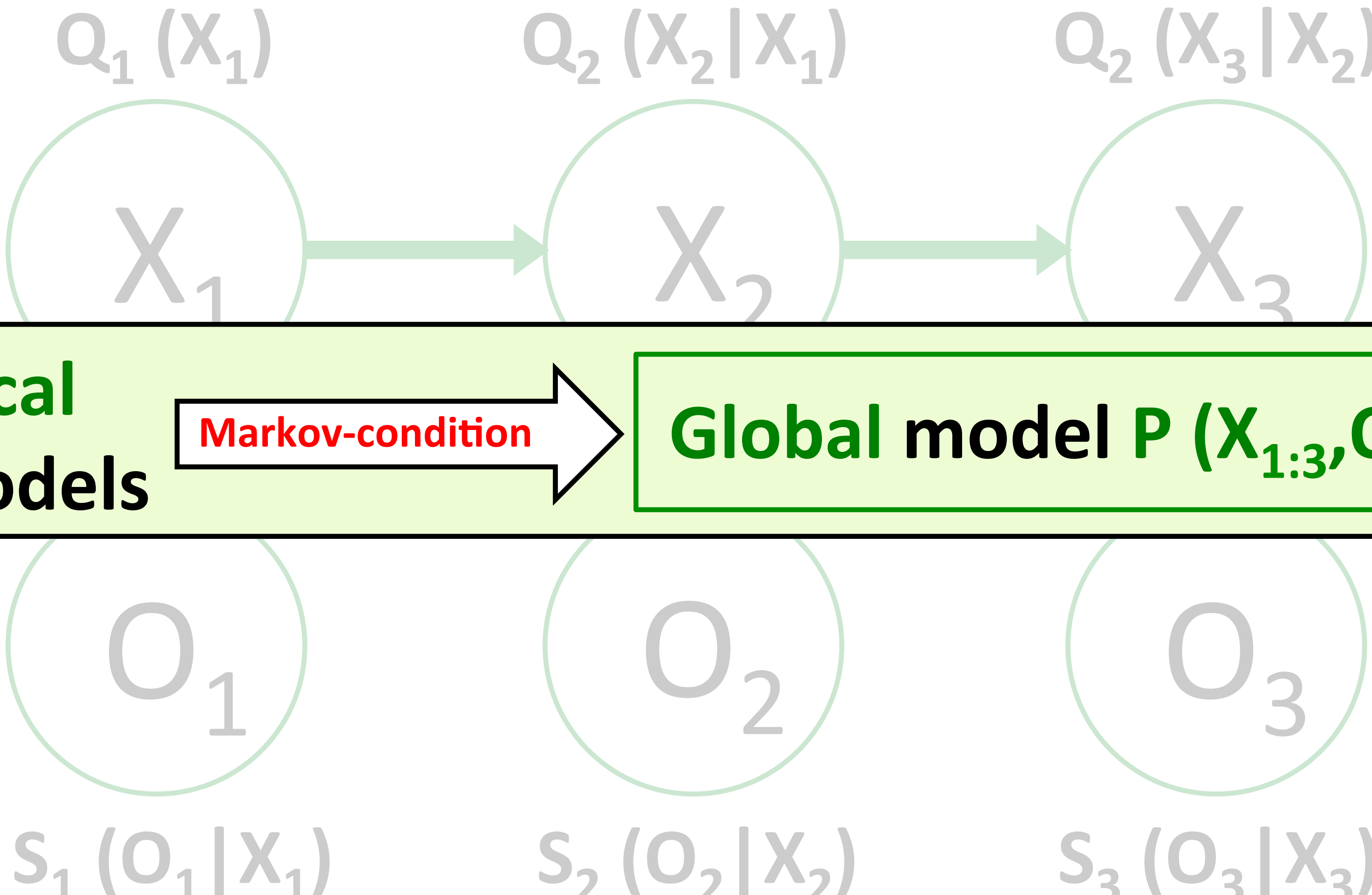
Assumptions of indepencence

(Precise) hidden Markov model

Conditional on its **mother variable**, any **variable** is independent of its **non-parent non-descendants**



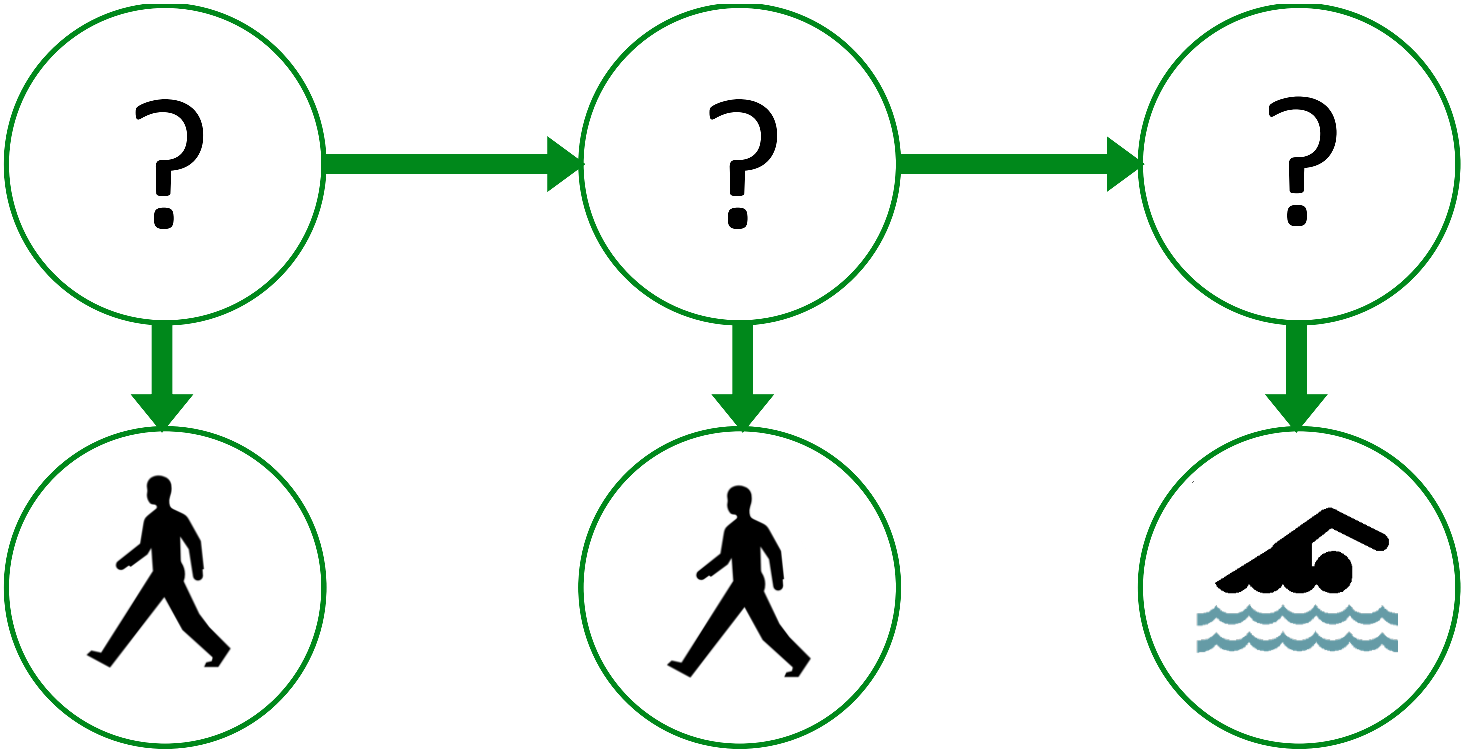
(Precise) hidden Markov model



Local models → Markov-condition → **Global model $P(X_{1:3}, O_{1:3})$**

State sequence prediction in imprecise hidden Markov models

State sequence prediction



State sequence prediction

$Q_1(X_1)$ $Q_2(X_2|X_1)$ $Q_2(X_3|X_2)$

Conditioning the model on the observations

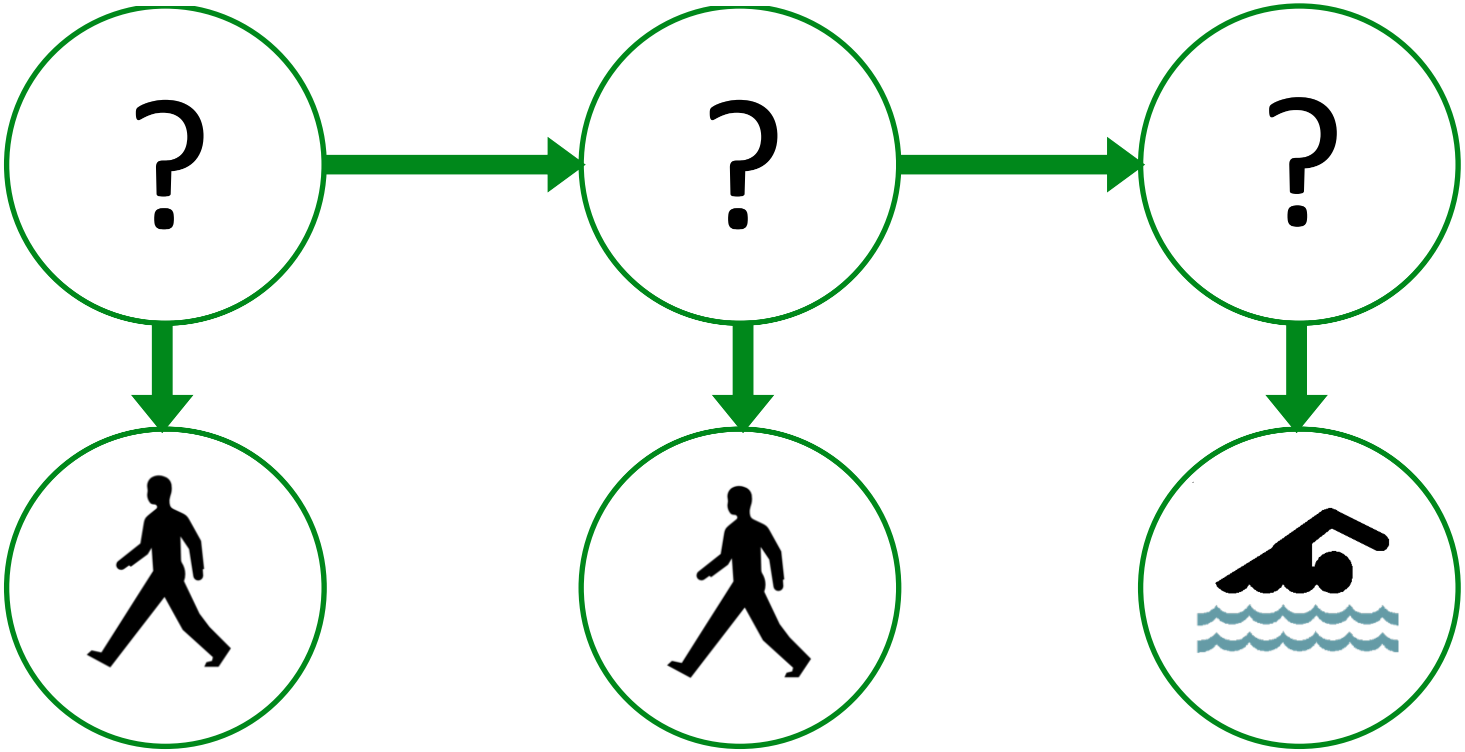
Local models → Markov-condition → **Global model** $P(X_{1:3},O_{1:3})$

↓ Bayes' theorem, $P(O_{1:3}) \neq 0$

Conditional global model $P(X_{1:3}|O_{1:3})$

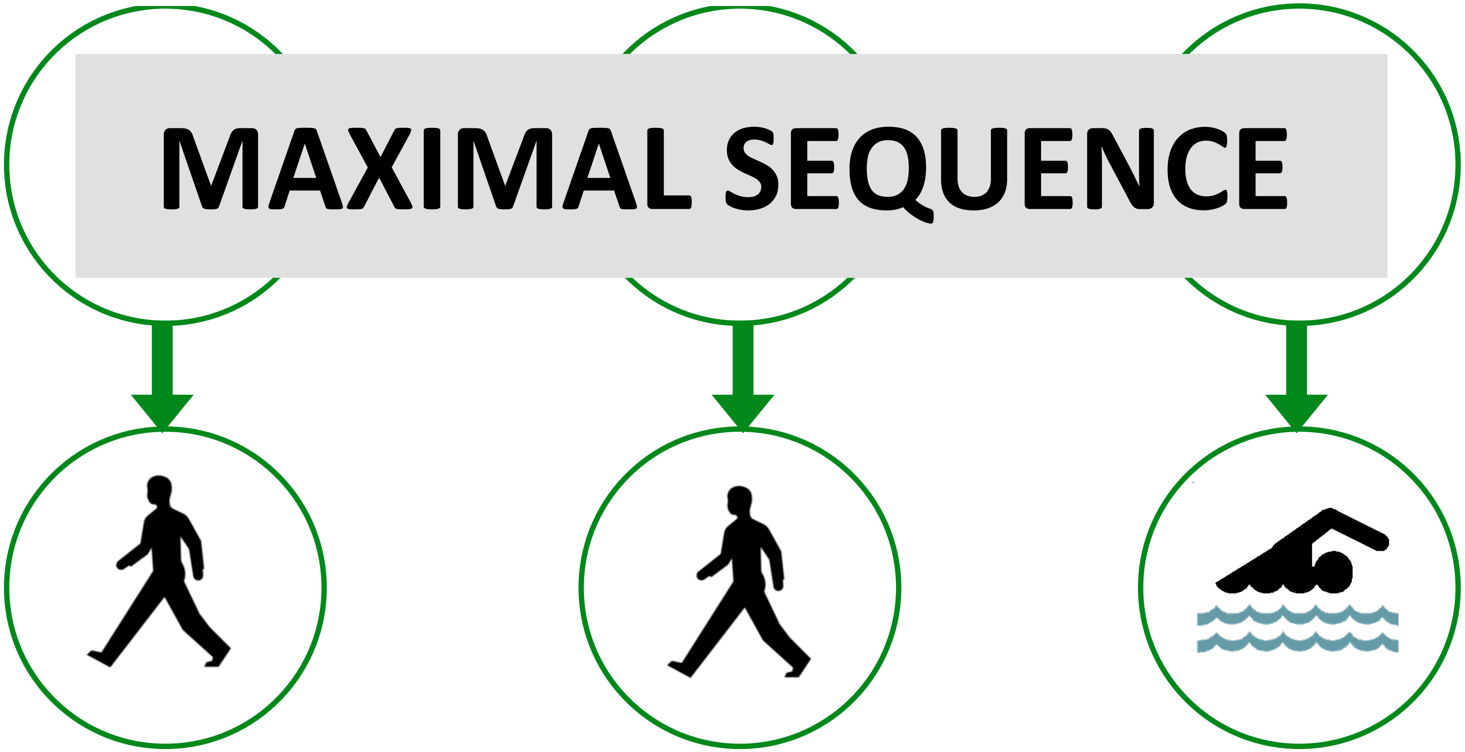
State sequence prediction

P (? ? ? | ) **Largest probability?**



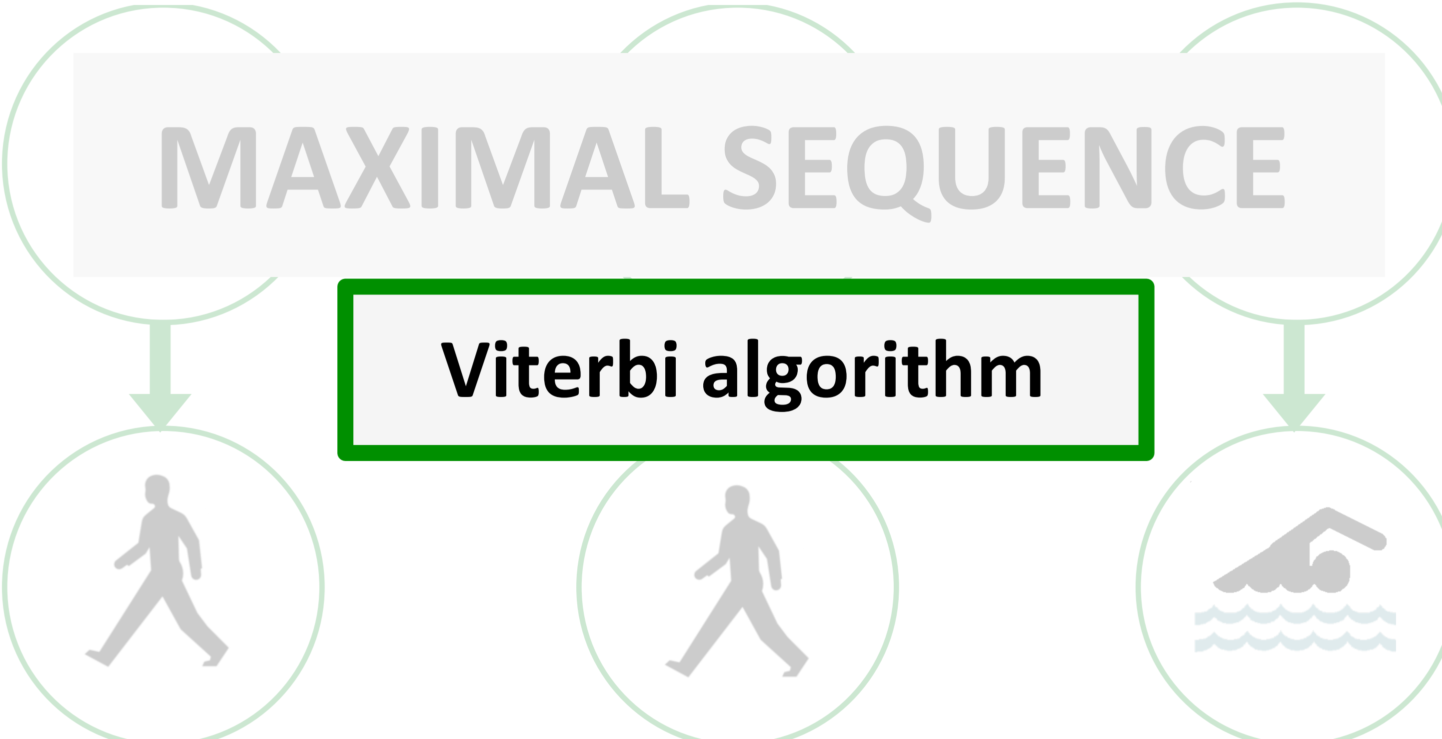
State sequence prediction

P (? ? ? | ) **Largest probability?**



State sequence prediction

P (? ? ? |) Largest probability?

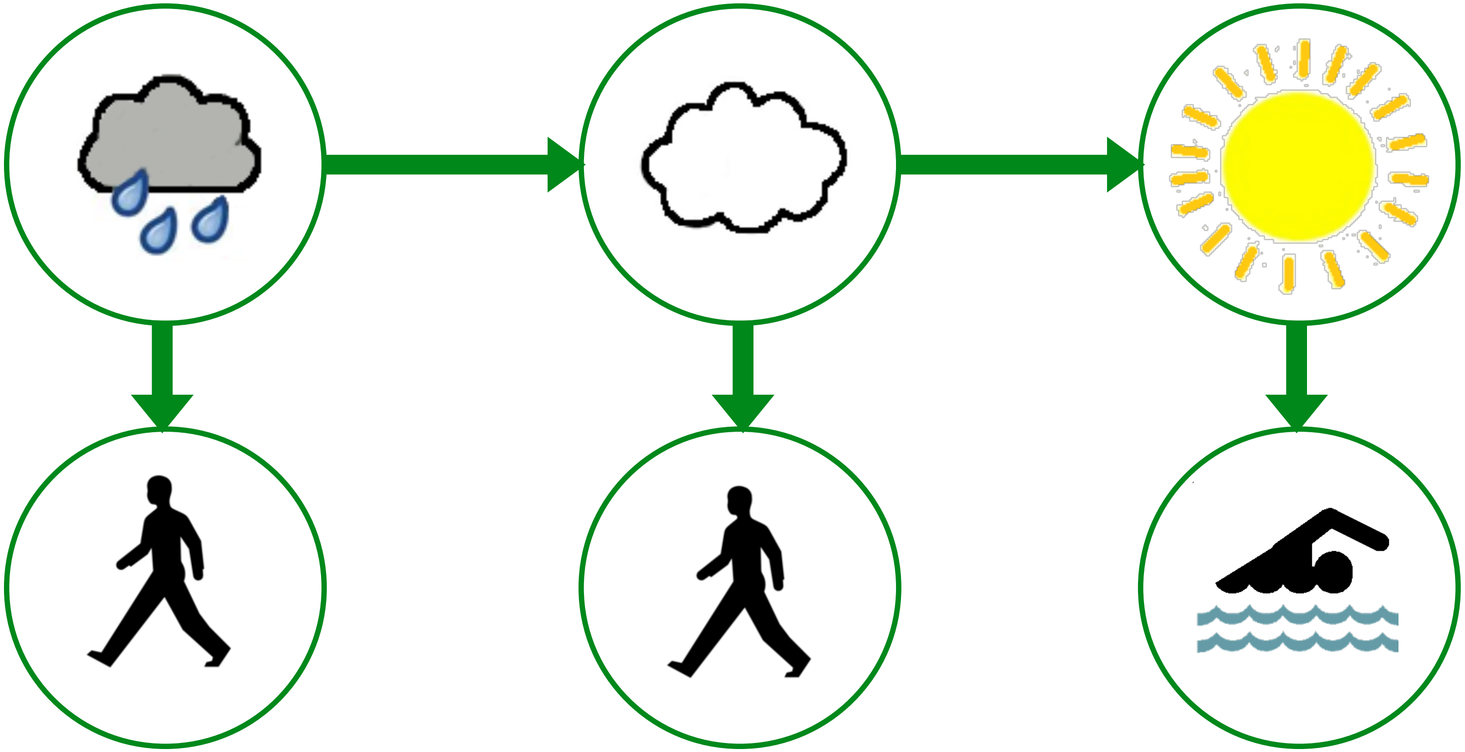


MAXIMAL SEQUENCE

Viterbi algorithm

State sequence prediction

P (? ? ? | ) Largest probability?



State sequence prediction in **imprecise** hidden Markov models

Imprecise probabilities

~~20%?~~ ⇒ [10%,30%]

Imprecise probabilities

~~20%?~~ → [10%,30%]

Credal set M

Closed and convex set of mass functions **p**

Coherent lower previsions

$\mathbf{p} : \mathcal{X} \rightarrow [0,1]$

$\mathbf{f} : \mathcal{X} \rightarrow \mathbb{R}$

→ **Prevision** $\mathbf{P(f) = E_p(f)}$

(expectation functional)

Coherent lower previsions

$\mathbf{p} : \mathcal{X} \rightarrow [0,1]$

$\mathbf{f} : \mathcal{X} \rightarrow \mathbb{R}$

→ **Prevision P(f) = $\mathbf{E_p(f)}$**

(expectation functional)

Credal set M

Closed and convex set of mass functions **p**

Coherent lower previsions

$\mathbf{p} : \mathcal{X} \rightarrow [0,1]$

$\mathbf{f} : \mathcal{X} \rightarrow \mathbb{R}$

→ **Prevision** $\mathbf{P(f) = E_p(f)}$

(expectation functional)

Credal set M

Closed and convex set of mass functions **p**

Lower prevision $\mathbf{\underline{P}(f) = \min\{ E_p(f) : p \in M \}}$

Coherent lower previsions

$\mathbf{p} : \mathcal{X} \rightarrow [0,1]$

$\mathbf{f} : \mathcal{X} \rightarrow \mathbb{R}$

→ **Prevision** $P(f) = E_p(f)$

(expectation functional)

Credal set M

Closed and convex set of mass functions **p**

Lower prevision $\underline{P}(f) = \min\{ E_p(f) : p \in M \}$

Upper prevision $\overline{P}(f) = \max\{ E_p(f) : p \in M \} = -\underline{P}(-f)$

Coherent lower previsions

Indicator function $I_A(X) = \begin{cases} 1 & \text{if } X = A \\ 0 & \text{if } X \neq A \end{cases}$

Probability of A = $\underline{P}(A) = \underline{P}(I_A)$

Coherent lower previsions

Indicator function $I_A(X) = \begin{cases} 1 & \text{if } X = A \\ 0 & \text{if } X \neq A \end{cases}$

Probability of A = $\underline{P}(A) = \underline{P}(I_A)$

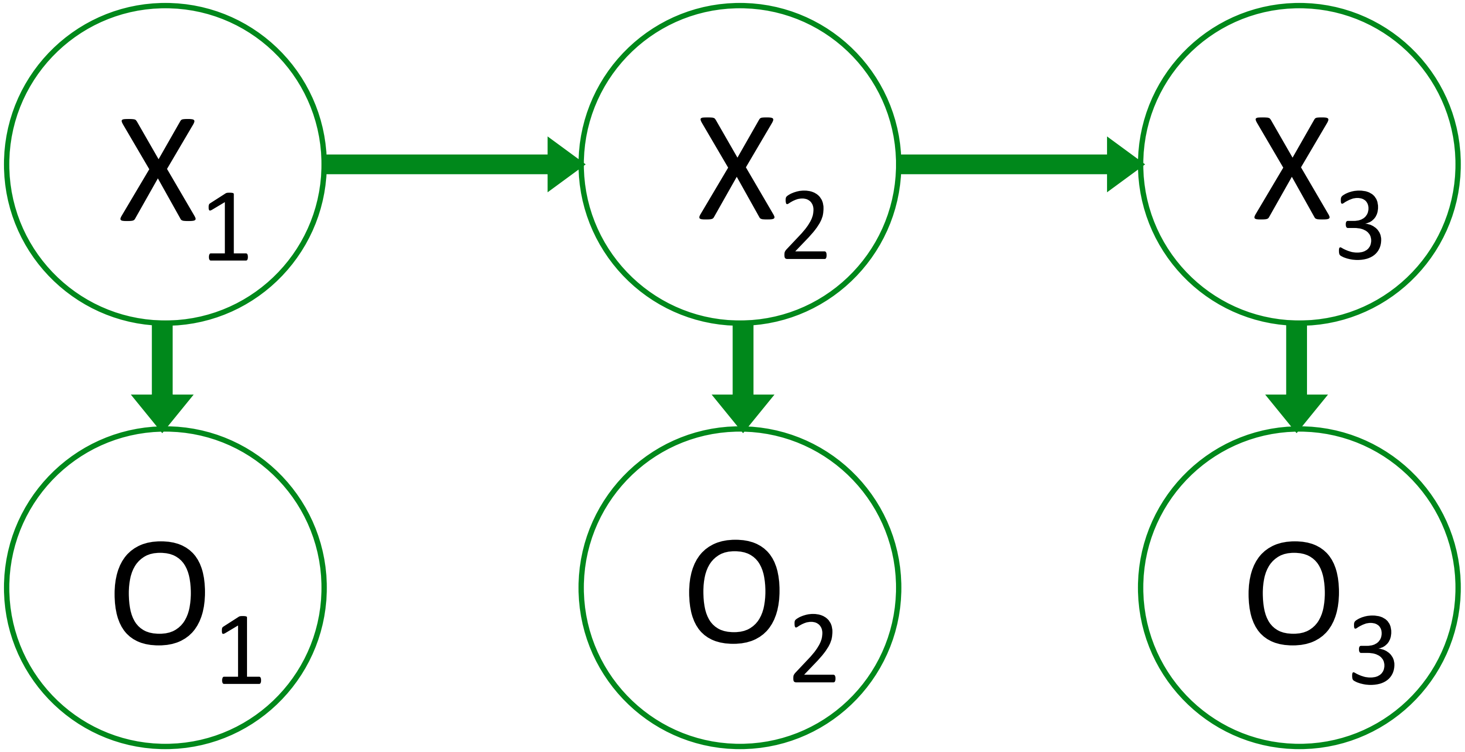
Lower probability of A = $\underline{P}(A) = \underline{P}(I_A)$

Upper probability of A = $\overline{P}(A) = \overline{P}(I_A)$

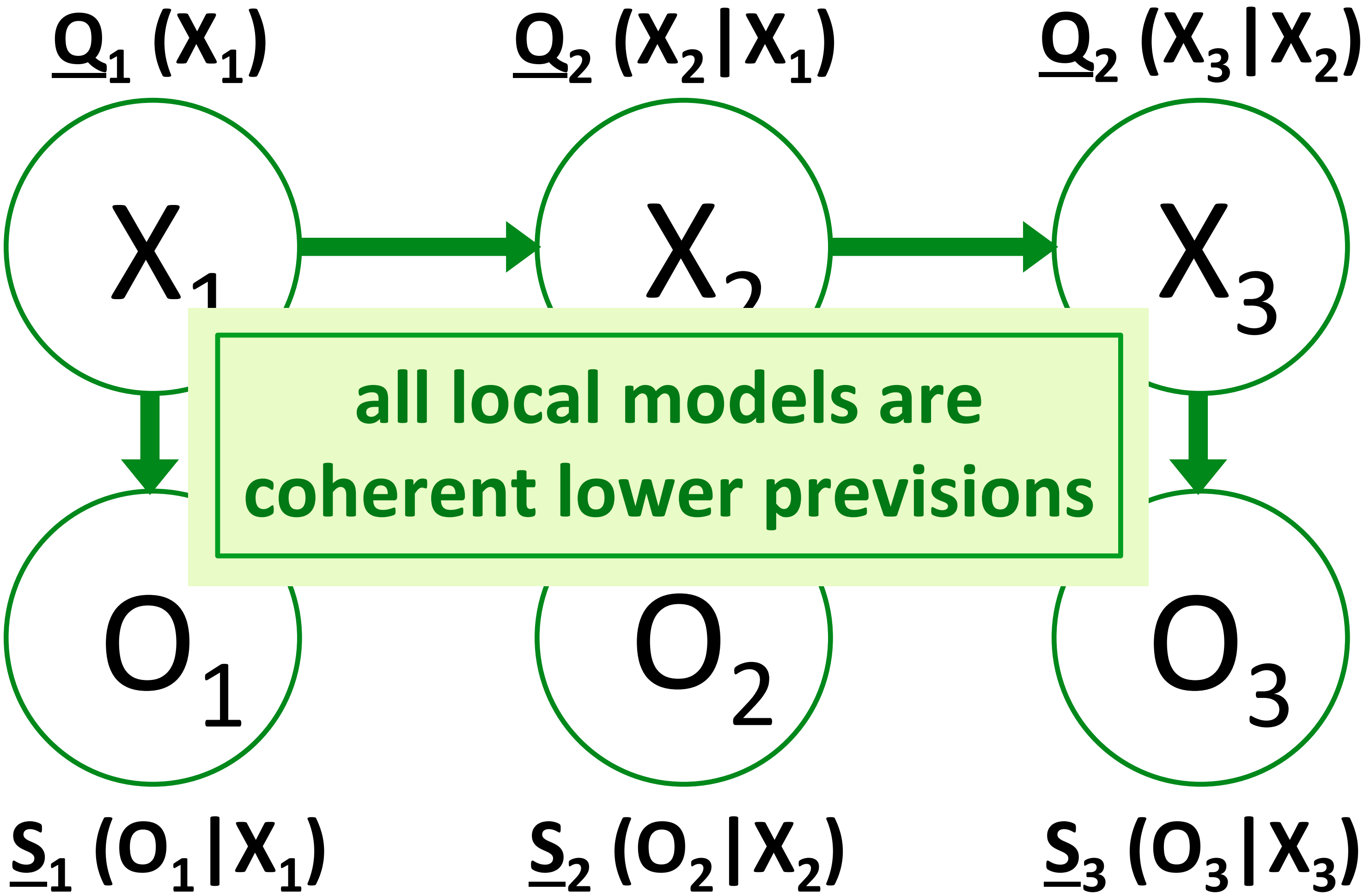
State sequence prediction in

imprecise hidden Markov models

Imprecise hidden Markov model

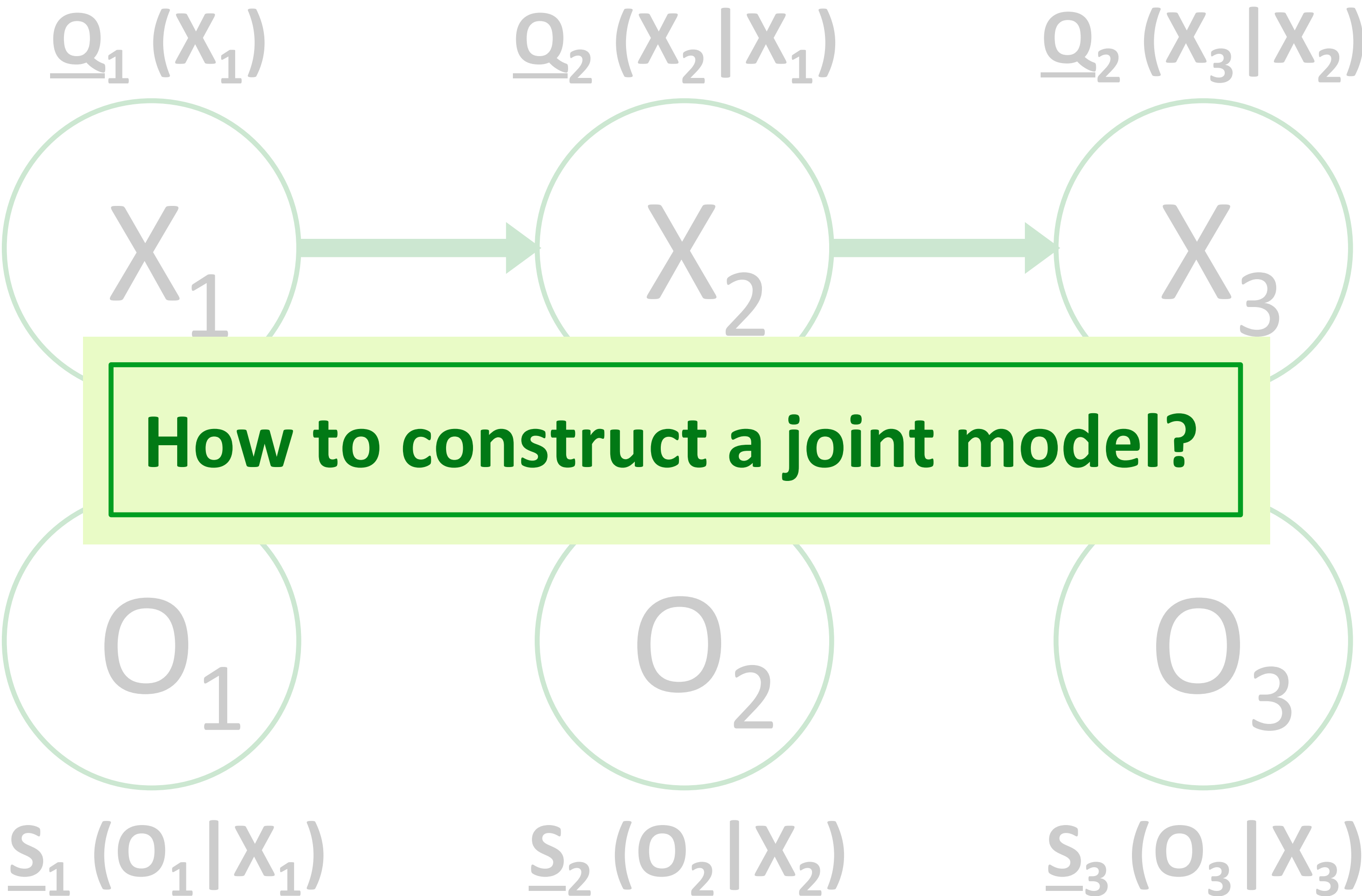


Imprecise hidden Markov model



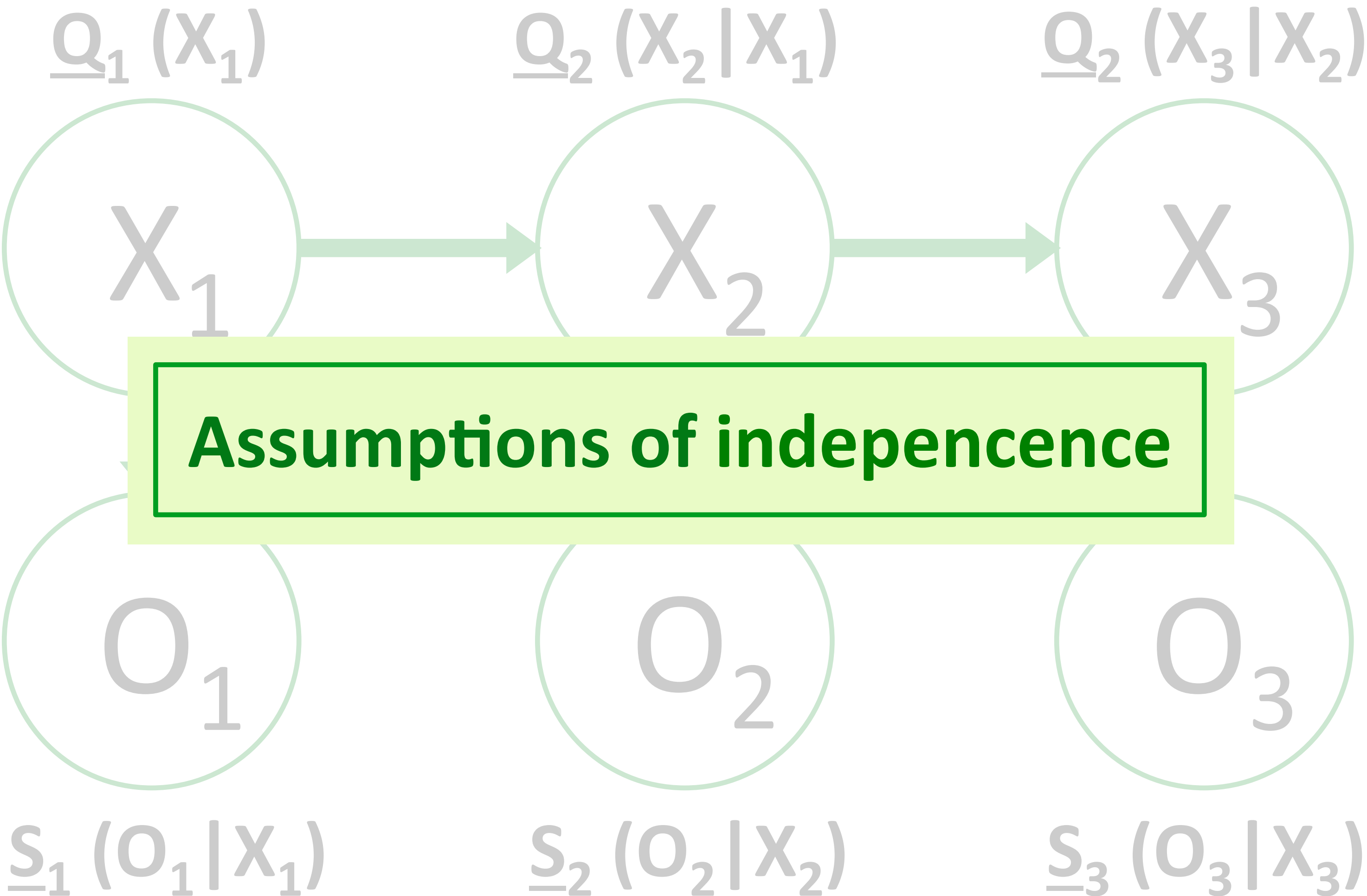
all local models are coherent lower previsions

Imprecise hidden Markov model



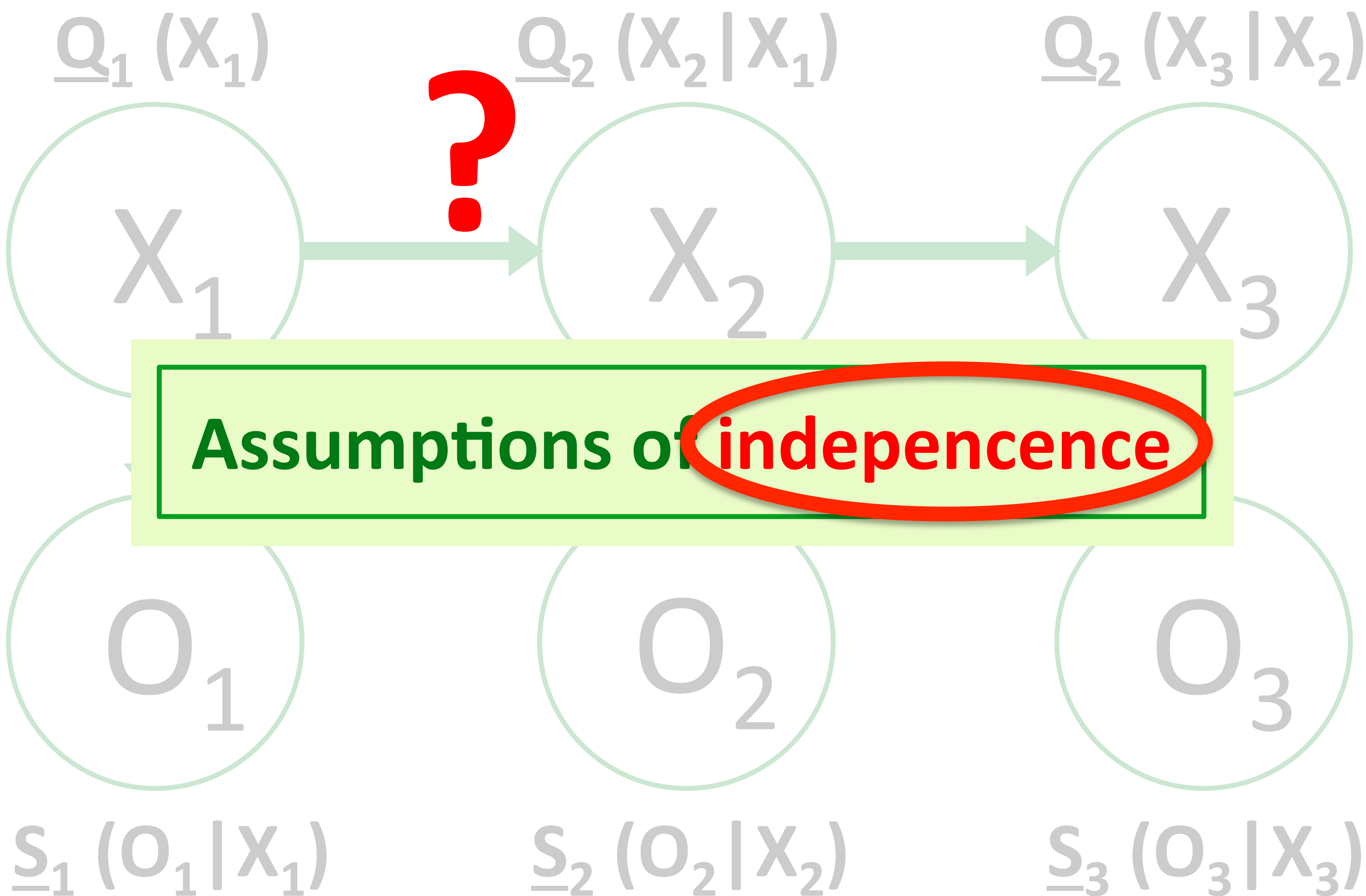
How to construct a joint model?

Imprecise hidden Markov model



Assumptions of indepencence

Imprecise hidden Markov model



Assumptions of indepencence

Imprecise hidden Markov model

Precise indepencence

P(A|B) = P(A)

Imprecise notions of indepencence

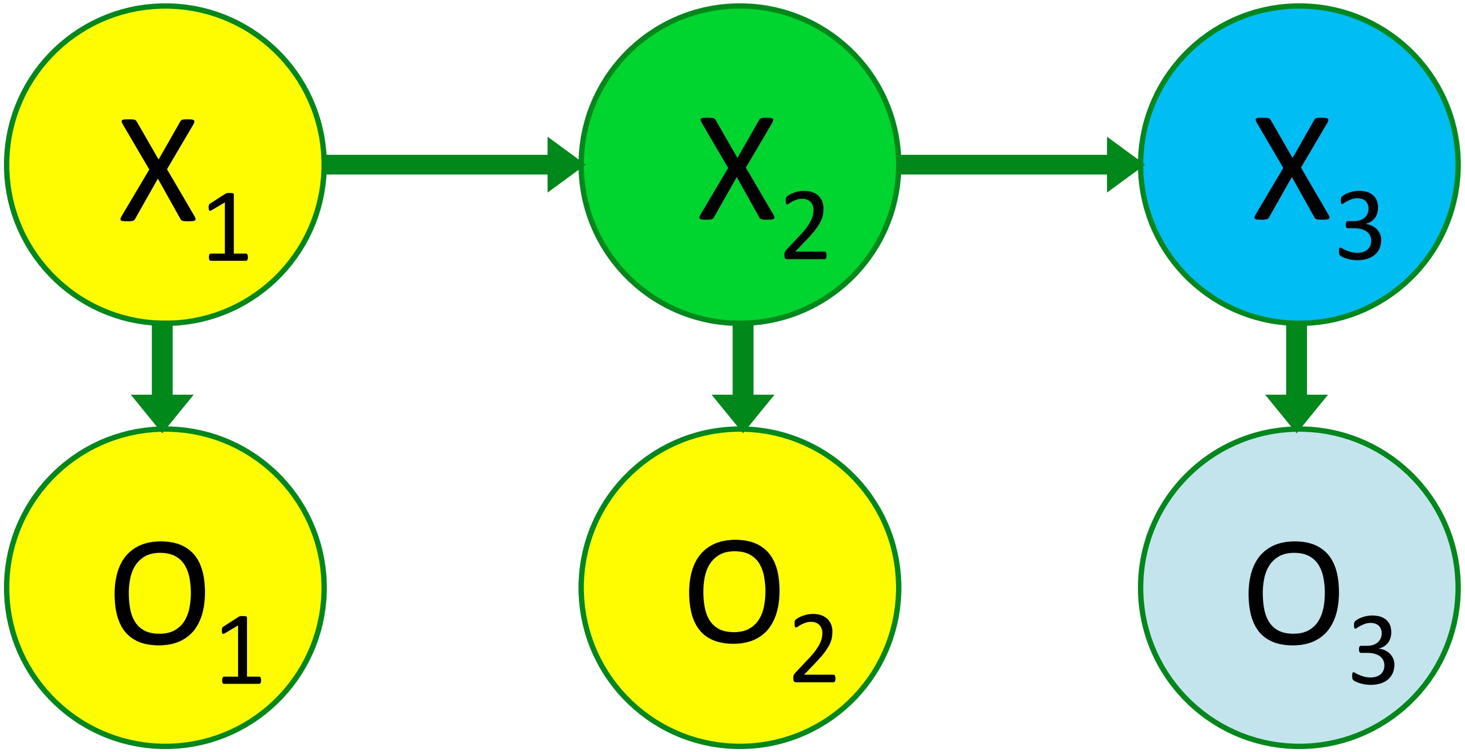
Strong independence

P(A|B) = P(A) for every extreme point in the credal set of $\underline{P}$

Epistemic irrelevance (a weaker notion)

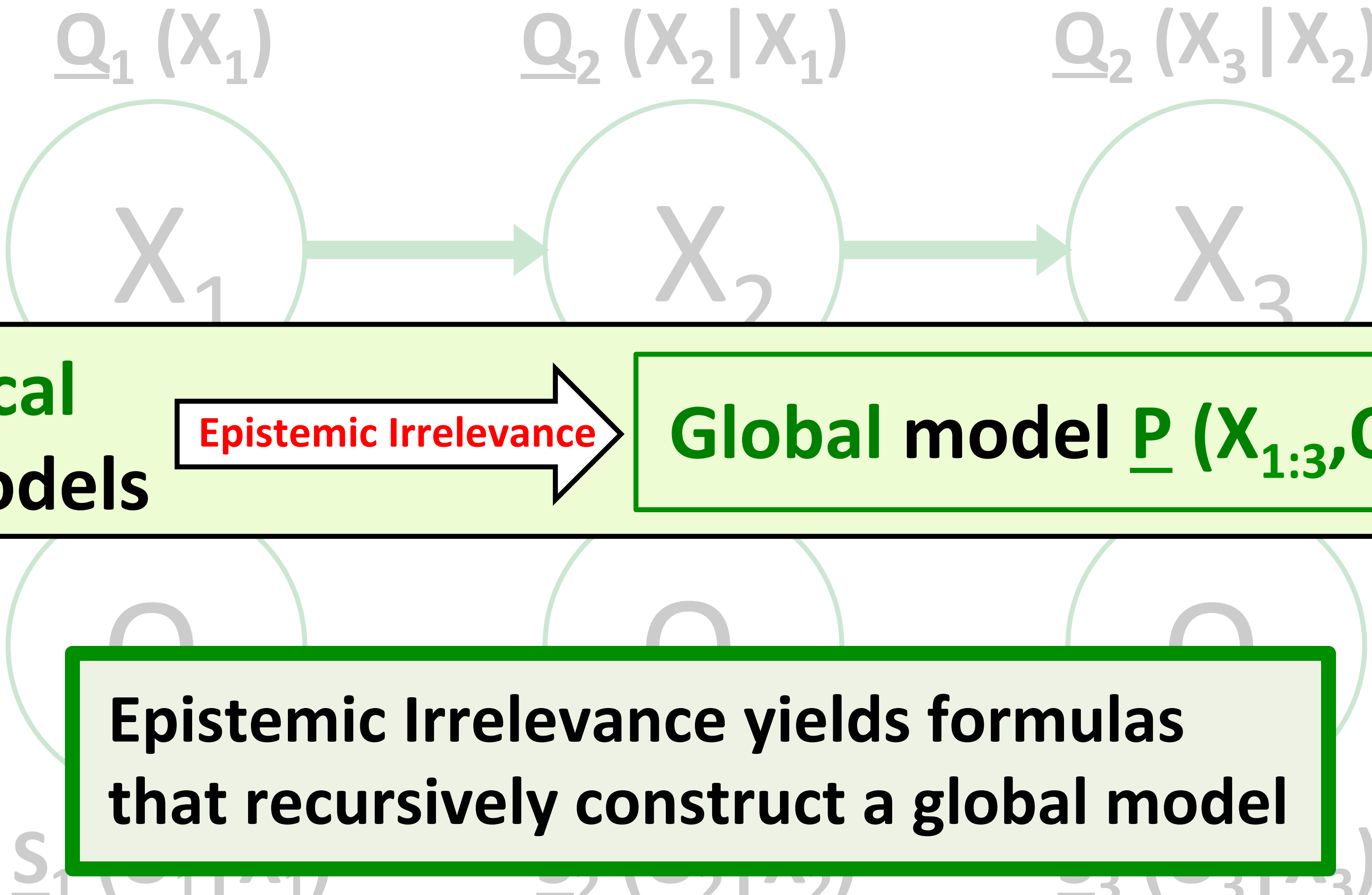
$\underline{P}$(A|B) = $\underline{P}$(A)

Imprecise hidden Markov model



Conditional on its **mother variable**, the **non-parent non-descendants** of any **variable** in the tree are **epistemically irrelevant** to this **variable** and its **descendants**

Imprecise hidden Markov model

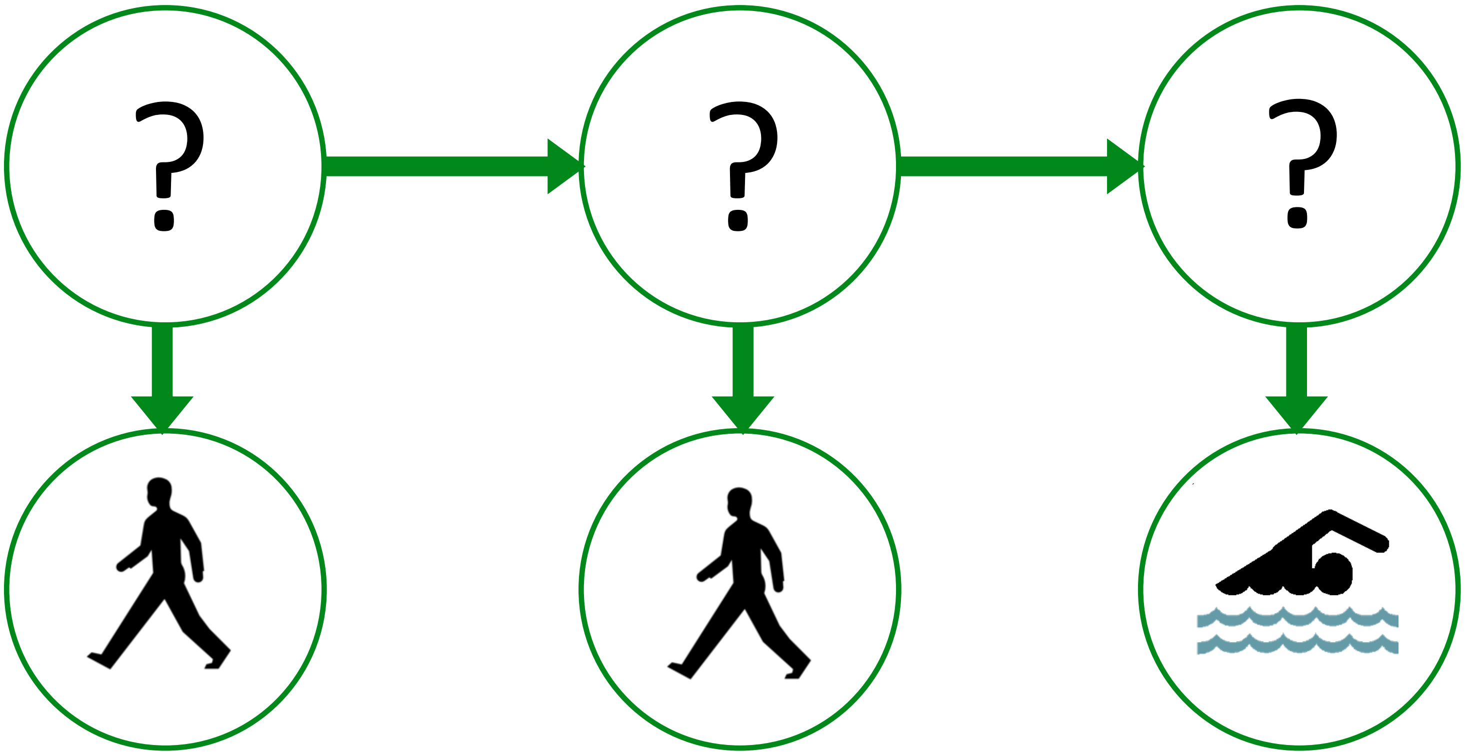


Local models → Epistemic Irrelevance → **Global model $\underline{P}\,(X_{1:3},O_{1:3})$**

Epistemic Irrelevance yields formulas that recursively construct a global model

State sequence prediction in imprecise hidden Markov models

(Imprecise) state sequence prediction



(Imprecise) state sequence prediction

Conditioning the model on the observations

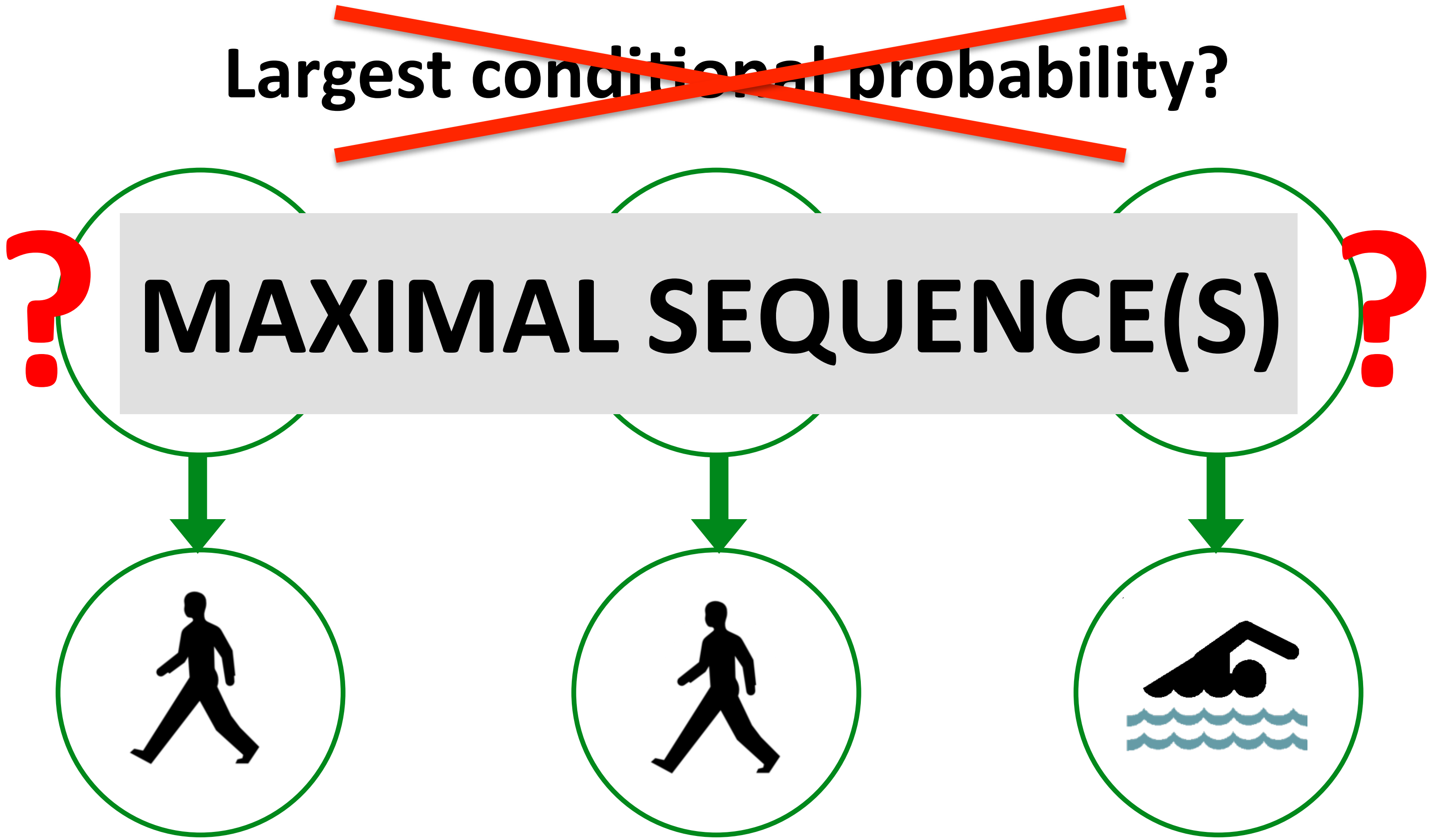
Local models → Epistemic Irrelevance → **Global model** $\underline{P}(X_{1:3},O_{1:3})$

Regular Extension

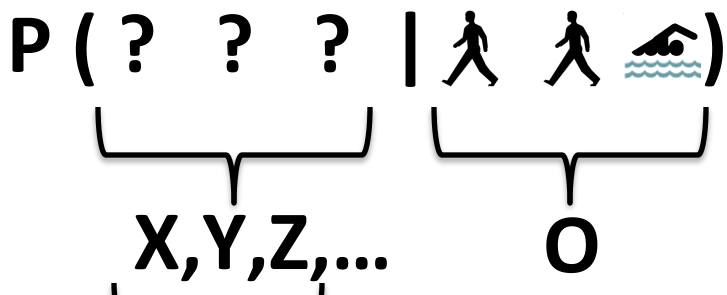
$\overline{P}(O_{1:3}) \neq 0$

Conditional global model $\underline{P}(X_{1:3}|O_{1:3})$

(Imprecise) state sequence prediction



(Imprecise) state sequence prediction



Largest probability?

PRECISE

Total ordering:

$Y > X$ if $P(Y|O) > P(X|O)$

$\Updownarrow$

$P(I_Y|O) > P(I_X|O) \Leftrightarrow P(I_Y - I_X|O) > 0$

(Imprecise) state sequence prediction

Total ordering:

PRECISE

$Y > X$ if $P(I_Y - I_X | O) > 0$

Bayes' theorem ⇕ $P(O) \neq 0$

$P(I_O[I_Y - I_X]) > 0$

Maximal sequence

= the sequence with the highest probability

= The highest sequence in this ordering

(Imprecise) state sequence prediction

Partial ordering:

Y > X if $\underline{P}(I_Y - I_X|O) > 0$

IMPRECISE

$P(I_Y - I_X|O) > 0$

For every P in $\underline{P}$

Maximal sequence

The undominated sequence(s) in this partial ordering

(Imprecise) state sequence prediction

Partial ordering:

$Y > X$ if $\underline{P}(I_Y - I_X | O) > 0$

IMPRECISE

$\overline{P}(O) \neq 0$

$\Updownarrow$

$\underline{P}(I_O[I_Y - I_X]) > 0$

IMPORTANT RESULT!
Doesn't hold in general
Only for HMMs

Maximal sequence

The undominated sequence(s) in this partial ordering

(Imprecise) state sequence prediction

Partial ordering:

$Y > X$ if $\underline{P}(I_O[I_Y - I_X]) > 0$

IMPRECISE

Maximal sequence

The undominated sequence(s) in this partial ordering

X is maximal $\Leftrightarrow$ For all Y : $Y \not> X$

$\Leftrightarrow$ For all Y : $\underline{P}(I_O[I_Y - I_X]) \leq 0$

(Imprecise) state sequence prediction

X is maximal $\Leftrightarrow$ **For all Y : $\underline{P}(I_O[I_Y - I_X]) \leq 0$**

How can we determine the set of maximal sequences efficiently?

(Imprecise) state sequence prediction

X is maximal $\Leftrightarrow$ **For all Y : $\underline{P}(I_O[I_Y - I_X]) \leq 0$**

How can we determine the set of maximal sequences efficiently?

EstiHMM: an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

(Imprecise) state sequence prediction

How can we determine the set of maximal sequences efficiently?

Trick nr. 1 **Using the joint model instead of the conditional one**

$Y > X$ if $\underline{P}(I_Y - I_X | O) > 0$

$\overline{P}(O) \neq 0$ ⇕

$\underline{P}(I_O[I_Y - I_X]) > 0$

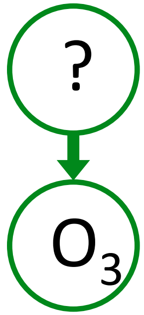
IMPORTANT RESULT!
Doesn't hold in general
Only for HMMs

(Imprecise) state sequence prediction

How can we determine the set of maximal sequences efficiently?

Trick nr. 2 **Working recursively**

Principle of optimality (Bellman)



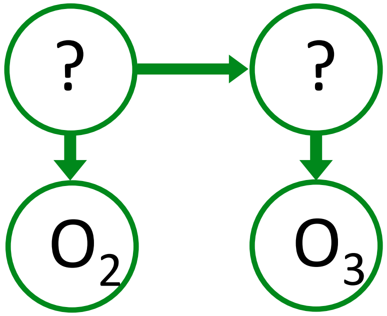
(Imprecise) state sequence prediction

How can we determine the set of maximal sequences efficiently?

Trick nr. 2

Working recursively

Principle of optimality (Bellman)

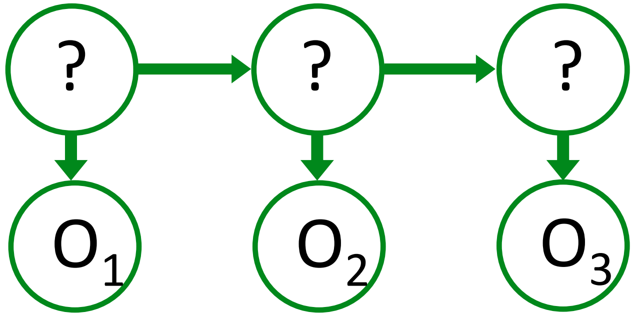


(Imprecise) state sequence prediction

How can we determine the set of maximal sequences efficiently?

Trick nr. 2 **Working recursively**

Principle of optimality (Bellman)



(Imprecise) state sequence prediction

How can we determine the set of maximal sequences efficiently?

Trick nr. 3

Reformulating the criterion of maximality

X is maximal $\Leftrightarrow$ **For all Y**: $\underline{P}(I_O[I_Y - I_X]) \leq 0$

$\Leftrightarrow$ $\alpha_k^{\mathrm{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n})$.

How can we determine the set of maximal sequences efficiently?

Trick nr. 4 Storing solutions efficiently

6 maximal sequences for a binary HMM of length 8:

Two state values: **0** or **1**

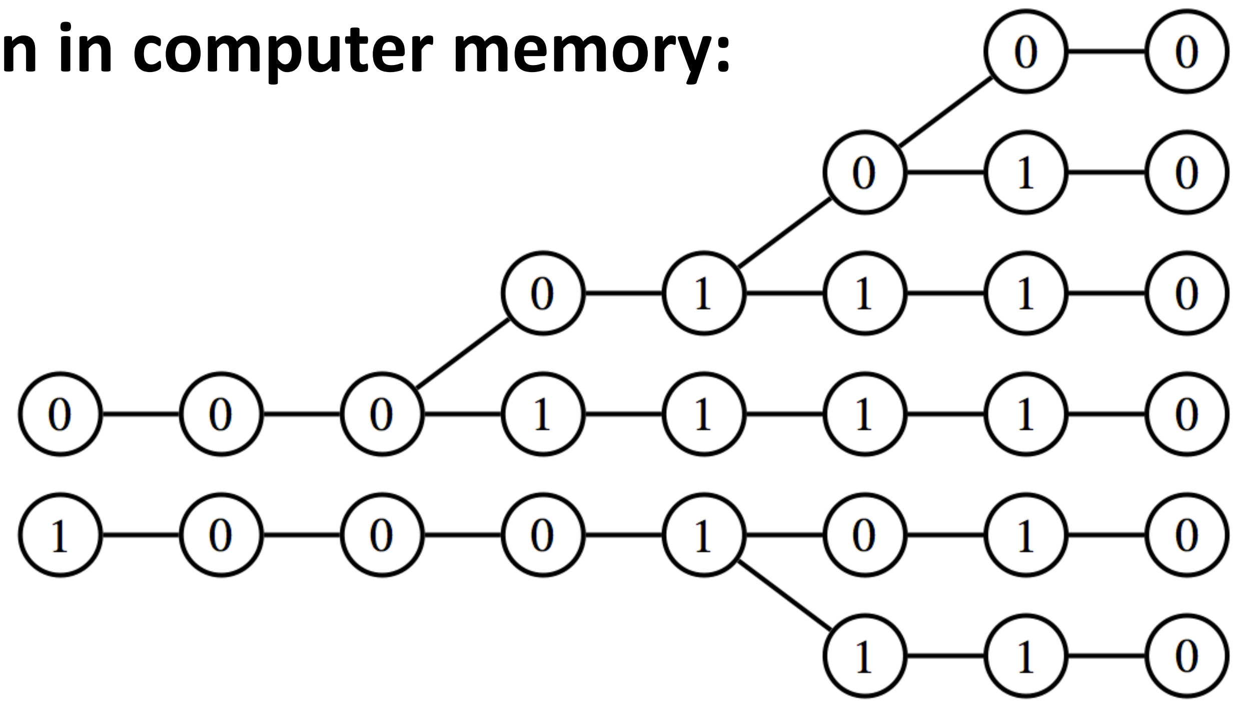
$$\{00001000, 00001010, 0001110, 00011110, 1001010, 10001110\}$$

(Imprecise) state sequence prediction

Trick nr. 4 — Storing solutions efficiently

$$\{00001000, 00001010, 0001110, 00011110, 1001010, 10001110\}$$

Representation in computer memory:

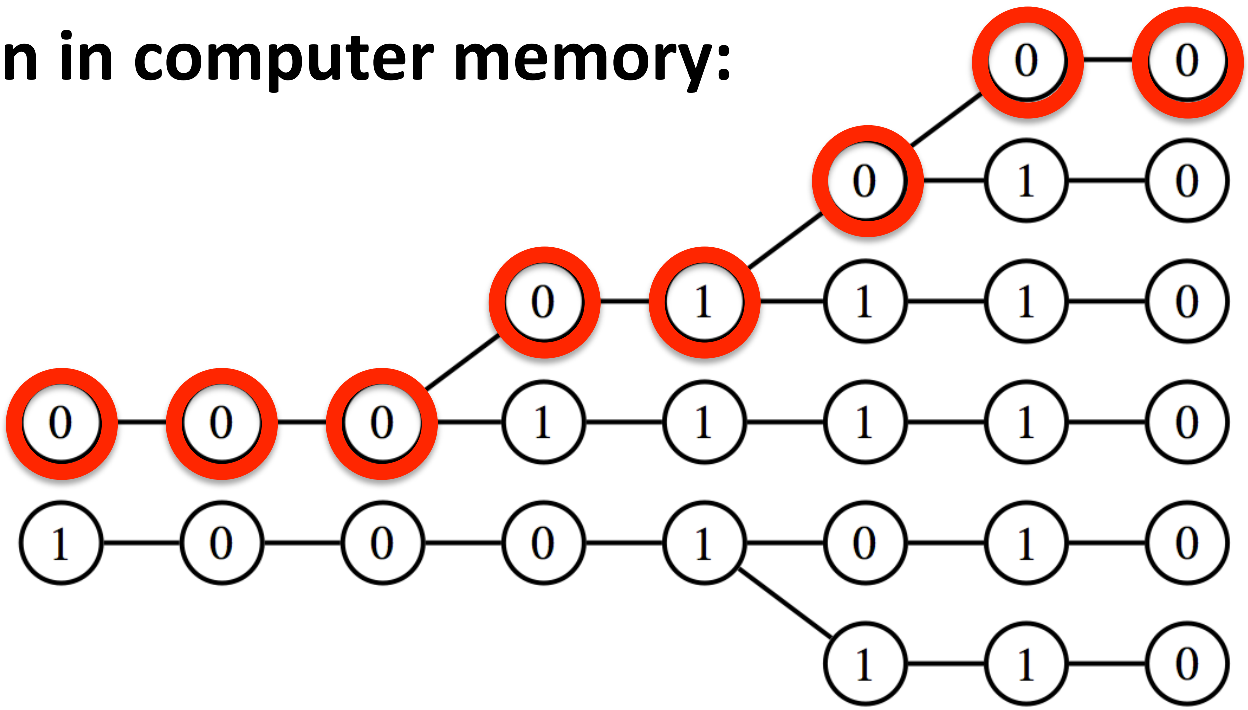


(Imprecise) state sequence prediction

Trick nr. 4 **Storing solutions efficiently**

{00001000, 00001010, 00001110, 00011110, 10001010, 10001110}

Representation in computer memory:

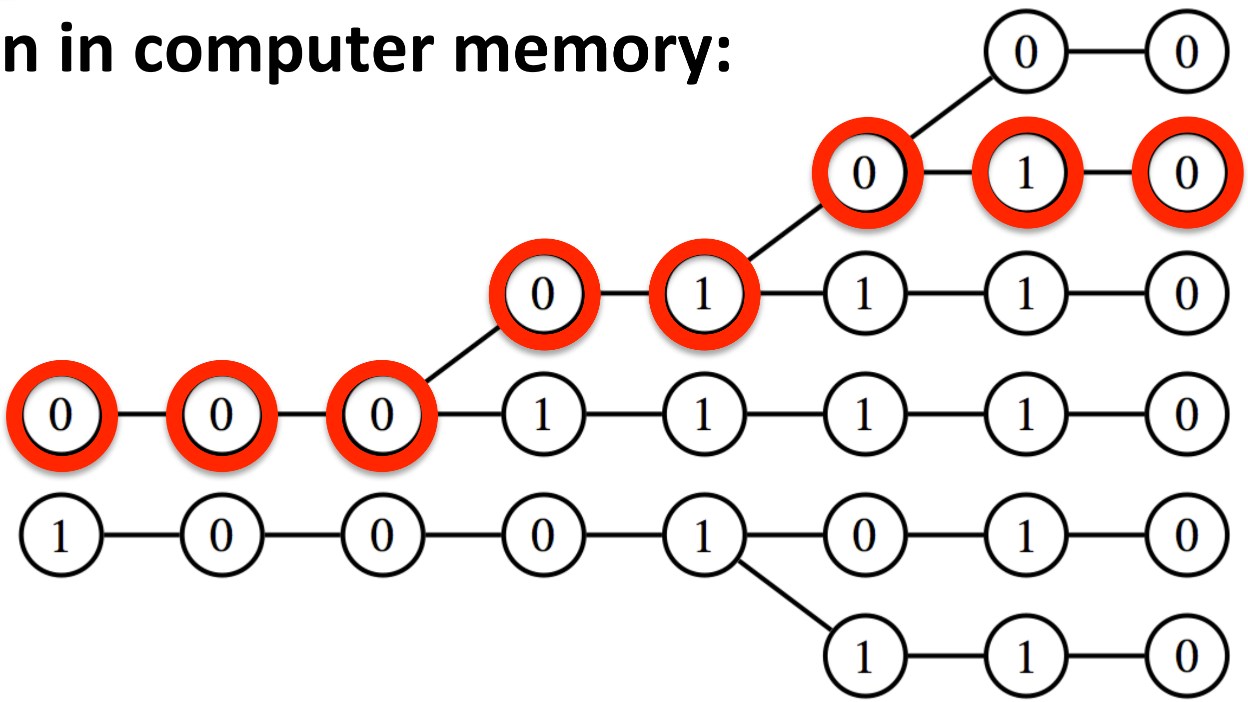


(Imprecise) state sequence prediction

Trick nr. 4 **Storing solutions efficiently**

{00001000, 00001010, 00001110, 00011110, 10001010, 10001110}

Representation in computer memory:

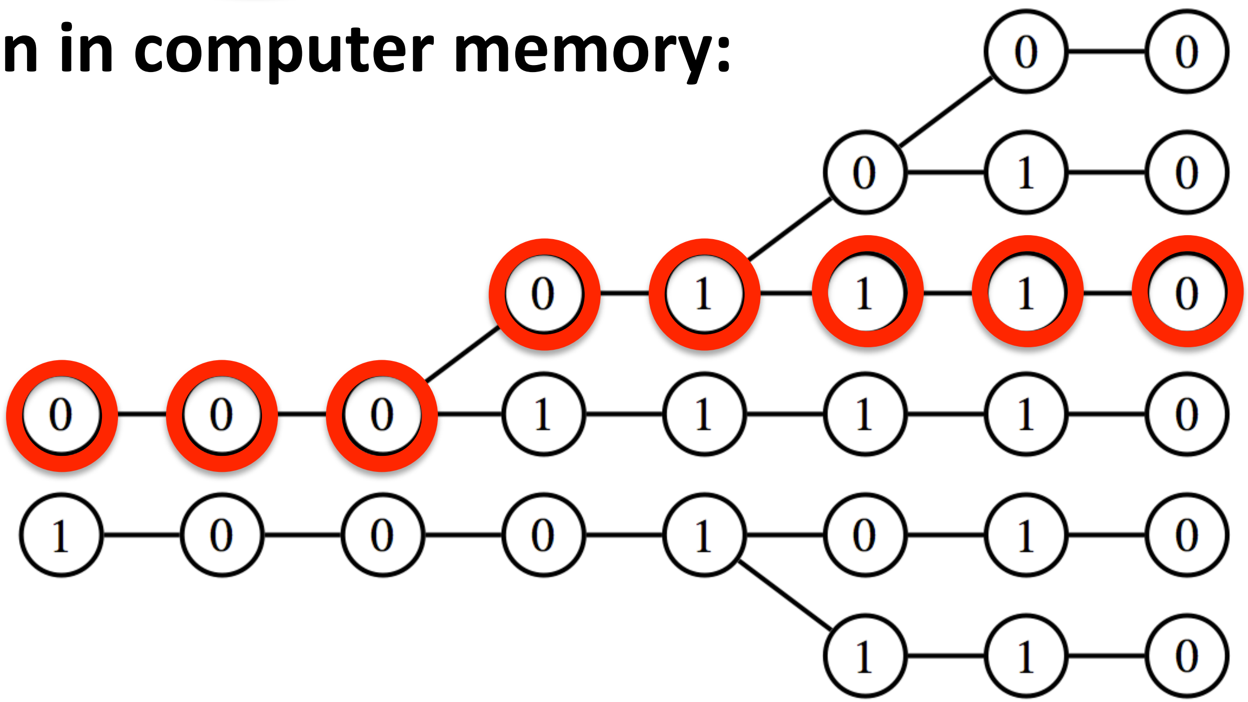


(Imprecise) state sequence prediction

Trick nr. 4 **Storing solutions efficiently**

$$\{00001000, 00001010, 00001110, 00011110, 10001010, 10001110\}$$

Representation in computer memory:

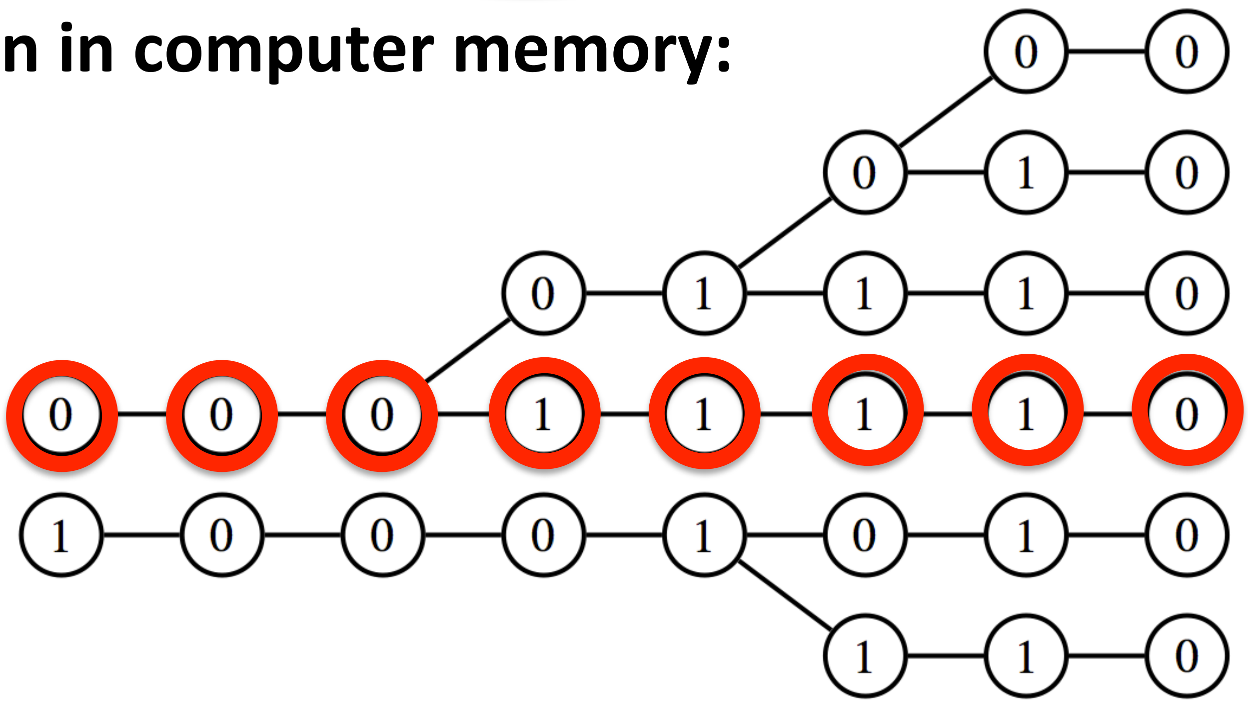


(Imprecise) state sequence prediction

Trick nr. 4 **Storing solutions efficiently**

$\{00001000, 00001010, 00001110, 00011110, 10001010, 10001110\}$

Representation in computer memory:

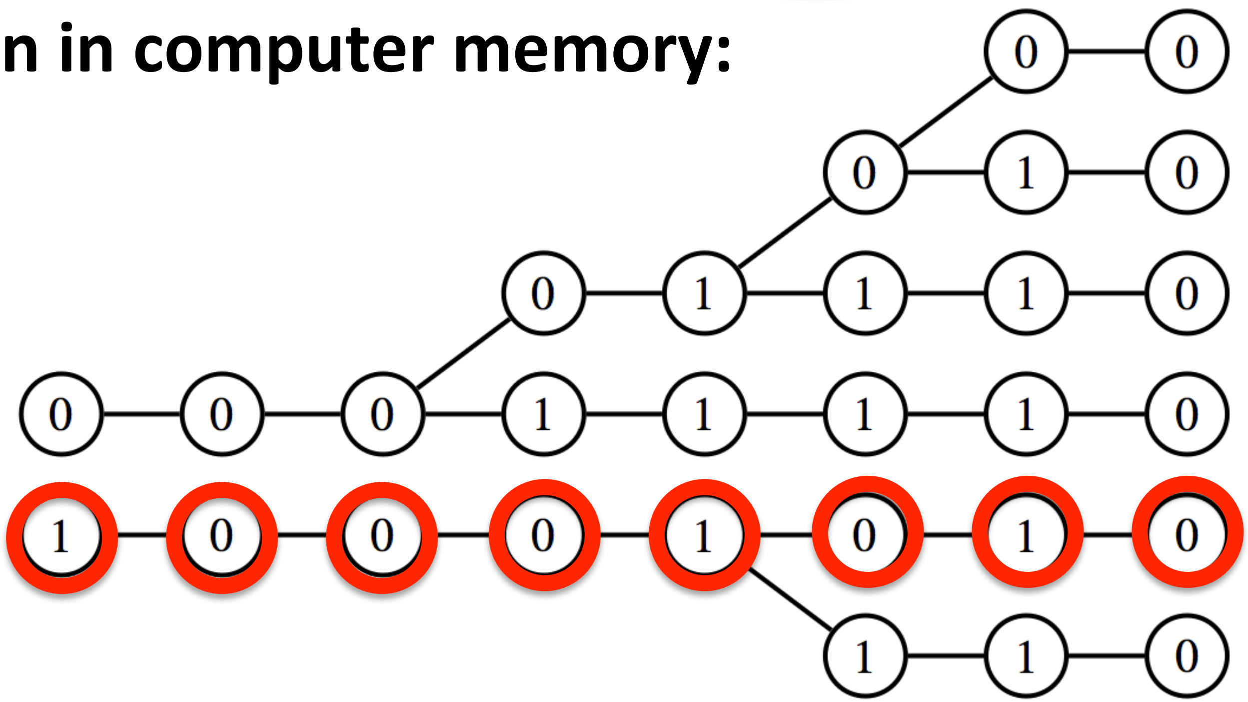


(Imprecise) state sequence prediction

Trick nr. 4 **Storing solutions efficiently**

$\{00001000, 00001010, 00001110, 00011110, 10001010, 10001110\}$

Representation in computer memory:

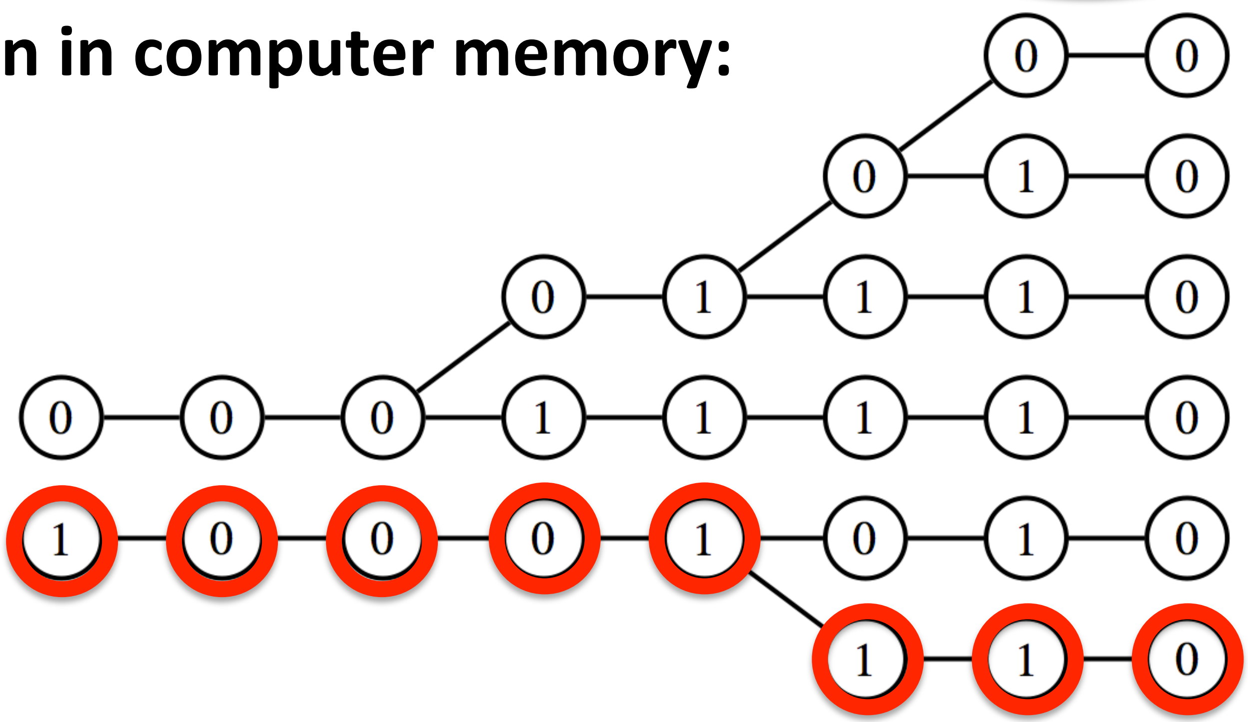


(Imprecise) state sequence prediction

Trick nr. 4 **Storing solutions efficiently**

$\{00001000, 00001010, 00001110, 00011110, 10001010, 10001110\}$

Representation in computer memory:



(Imprecise) state sequence prediction

EstiHMM: an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

Computational complexity

(Imprecise) state sequence prediction

EstiHMM: an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

Computational complexity

- **Theoretical analysis**

(Imprecise) state sequence prediction

EstiHMM: an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

Computational complexity

- **Theoretical analysis** **(Empirical confirmation)**

(Imprecise) state sequence prediction

EstiHMM: an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

Computational complexity

- **Theoretical analysis (Empirical confirmation)**
- **Linear in the number of maximal sequences**

(Imprecise) state sequence prediction

EstiHMM: an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

Computational complexity

- **Theoretical analysis (Empirical confirmation)**
- **Linear in the number of maximal sequences**
- **Quadratic in the length of the HMM**

(Imprecise) state sequence prediction

EstiHMM: an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

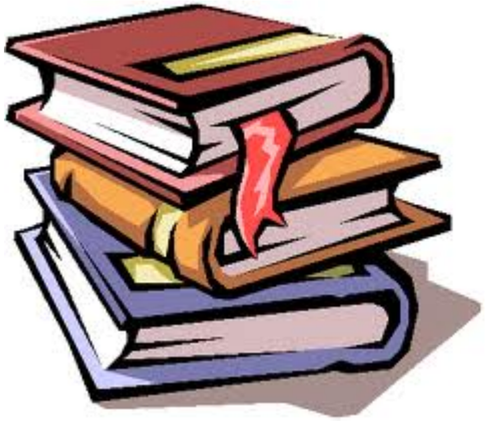
Computational complexity

- **Theoretical analysis (Empirical confirmation)**
- **Linear in the number of maximal sequences**
- **Quadratic in the length of the HMM**
- **Cubic in the number of possible states**

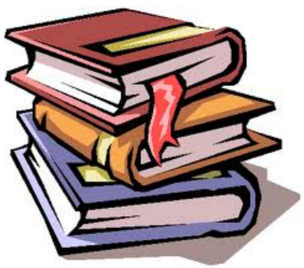

"What can you do with it?"

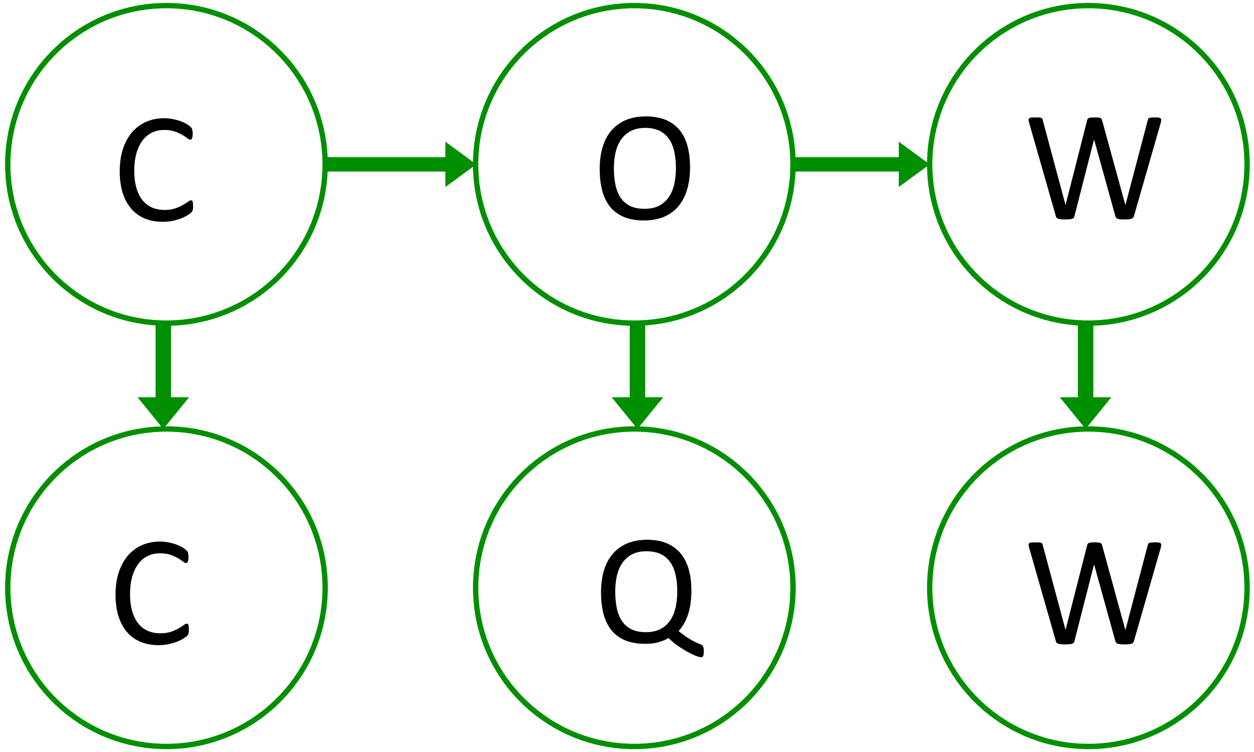
An application

Optical character recognition software

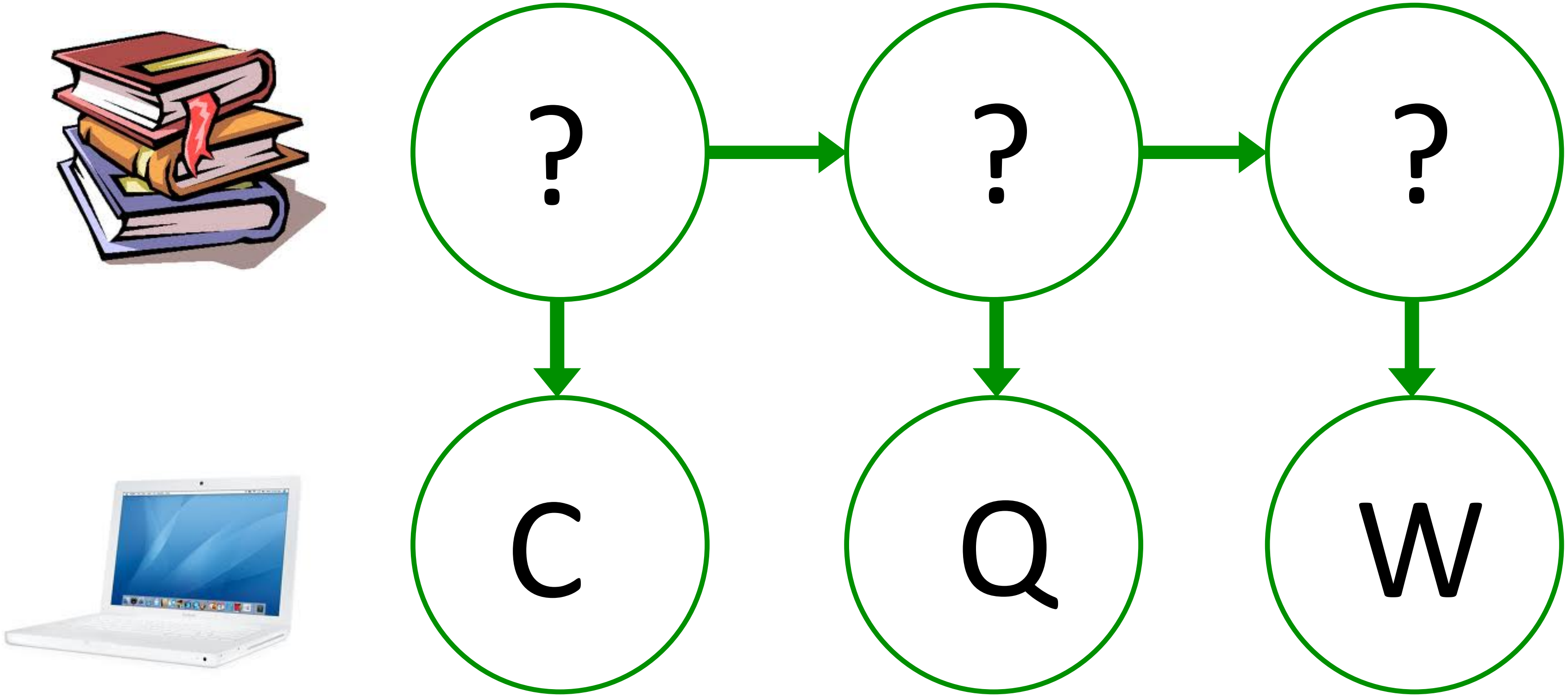


An application



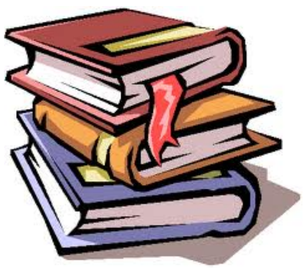


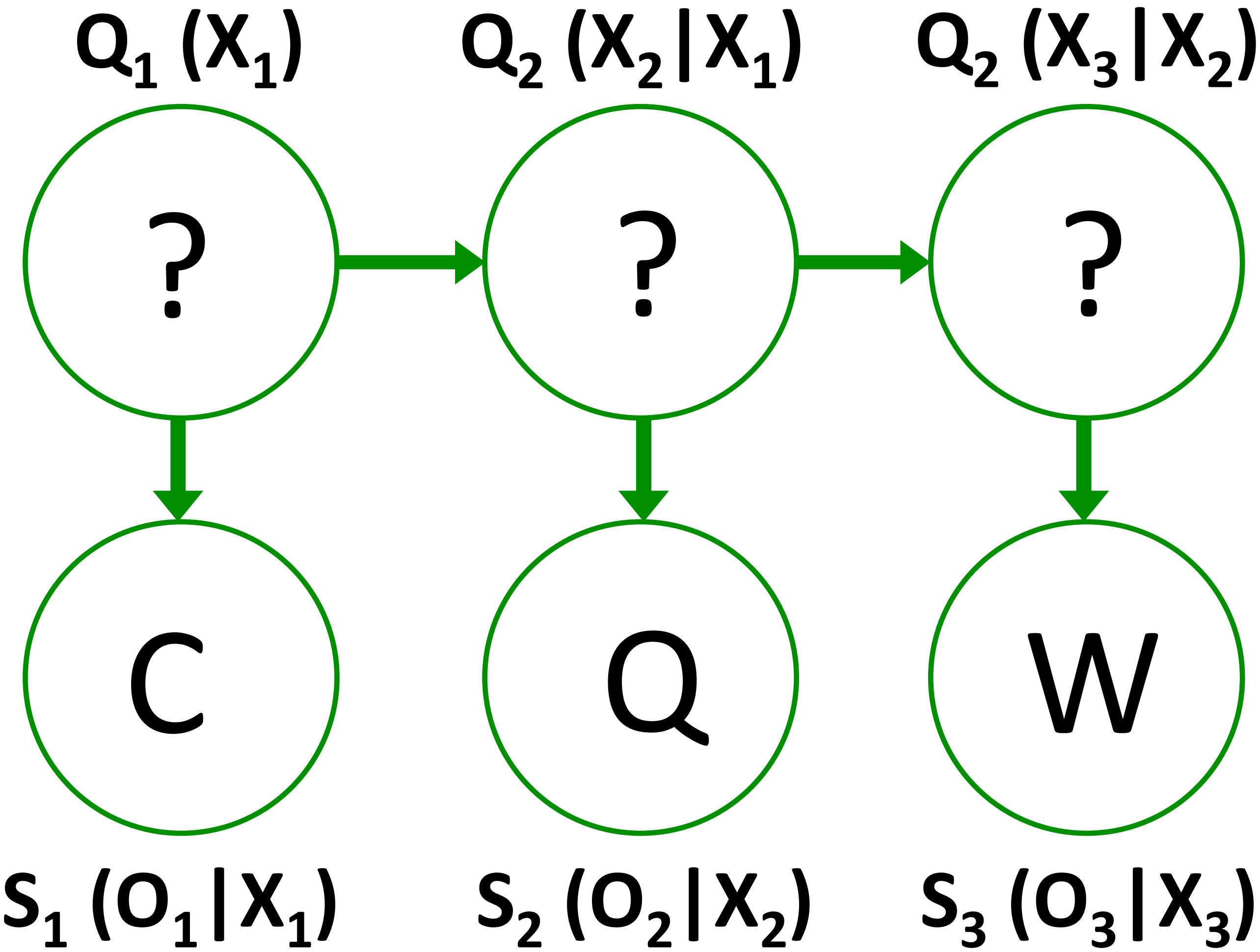
An application



An application

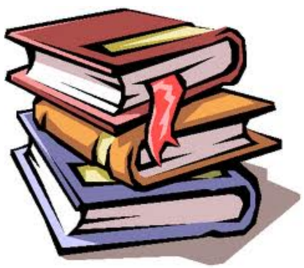
Viterbi algorithm

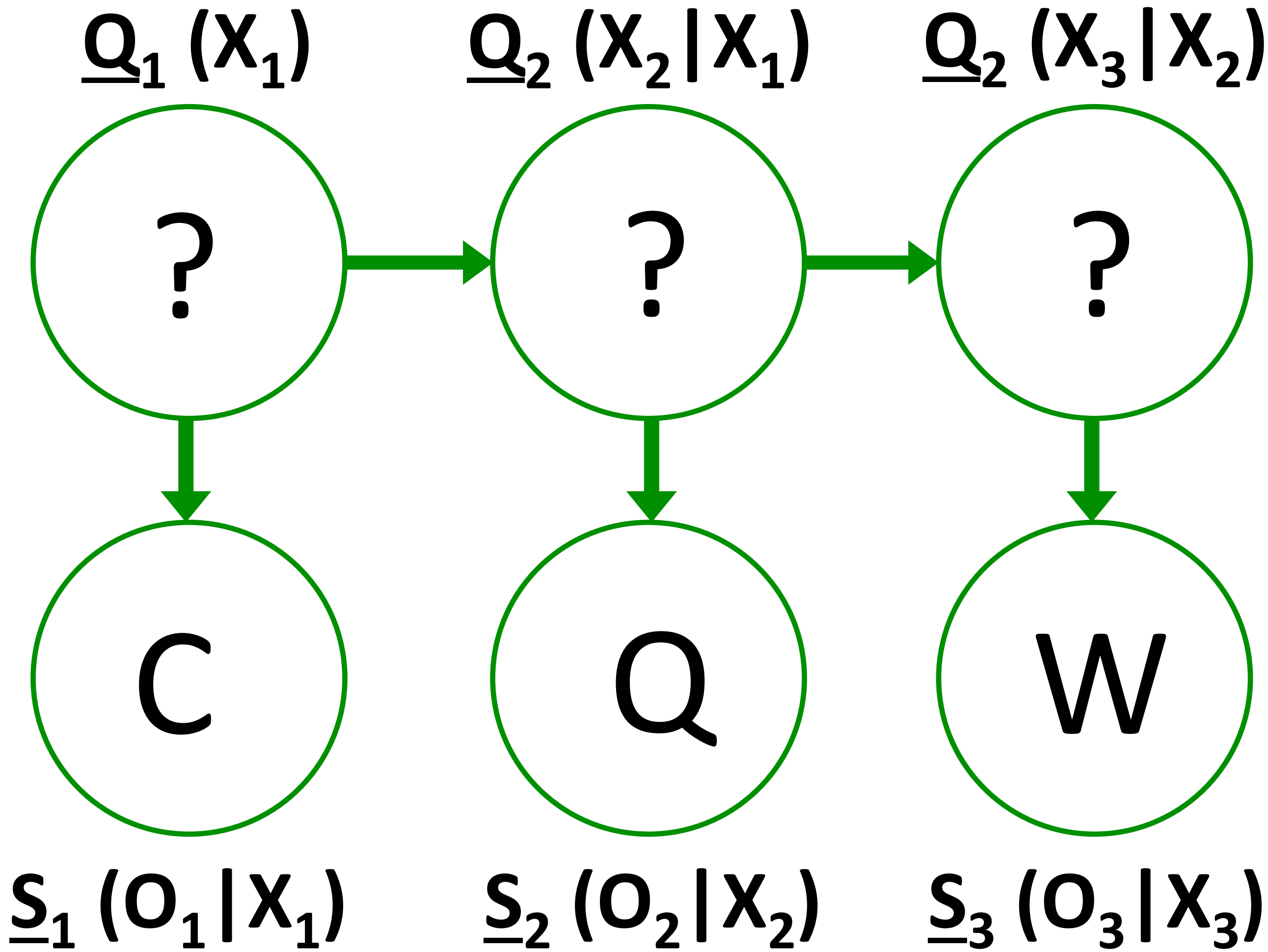




An application

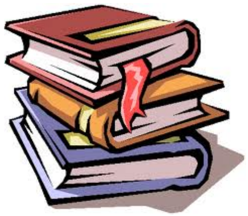
EstiHMM algorithm





An application

La Divina Commedia



original

VITA

correctly read

digital

VITA

Solution Viterbi

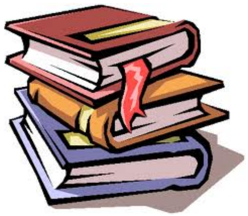
VITA

Solution(s) EstiHMM-algoritme

VITA

An application

La Divina Commedia



original

CON

incorrectly read

digital

CCN

Solution Viterbi

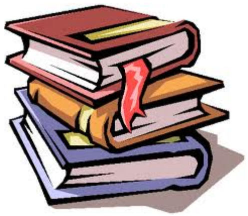
CON

Solution(s) EstiHMM-algoritme

CON

An application

La Divina Commedia



original

EH

correctly read

digital

EH

Solution Viterbi

EN

Solution(s) EstiHMM-algoritme

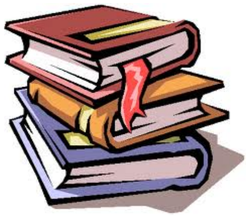
CH

EH

EN

An application

La Divina Commedia



original

IO

incorrectly read

digital

ZO

Solution Viterbi

LO

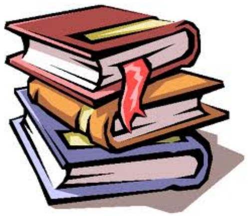
Solution(s) EstiHMM-algoritme

LO

IO

An application

La Divina Commedia



original

CHE

incorrectly read

digital

CNE

Solution Viterbi

ONE

Solution(s) EstiHMM-algoritme

CBE **CHE**

CNE **CZE**

ONE

Thanks for your attention!

Questions?