A new method for learning imprecise hidden Markov models

Arthur Van Camp and Gert de Cooman

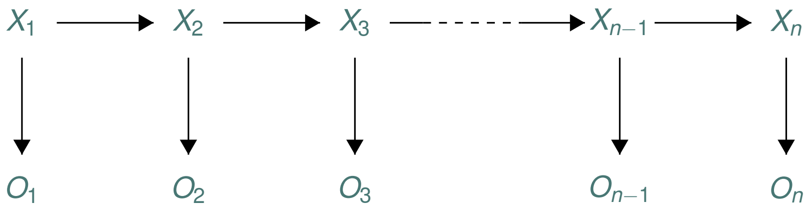
Ghent University, SYSTeMS

Arthur.VanCamp@UGent.be, Gert.deCooman@UGent.be

Imprecise hidden Markov models

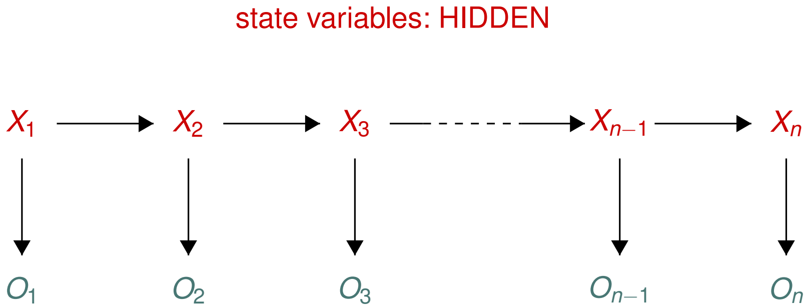
Imprecise hidden Markov model

graphical representation



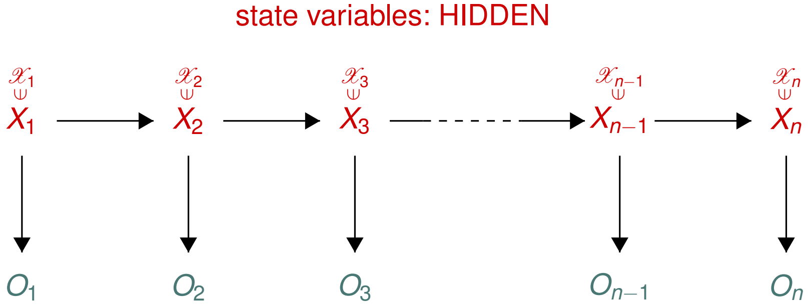
Imprecise hidden Markov model

random variables



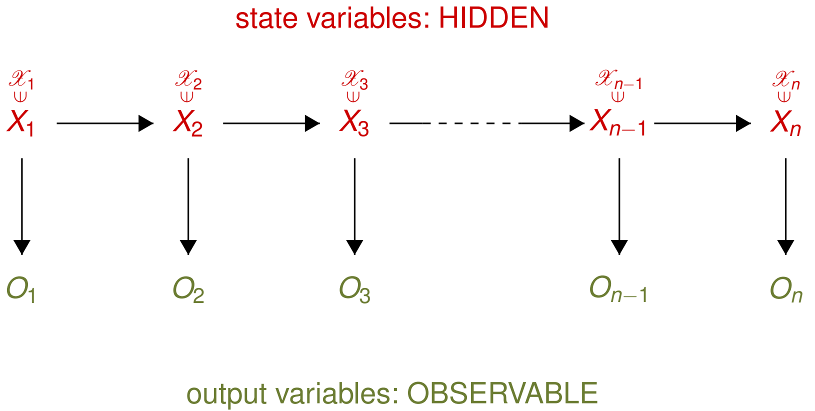
Imprecise hidden Markov model

random variables

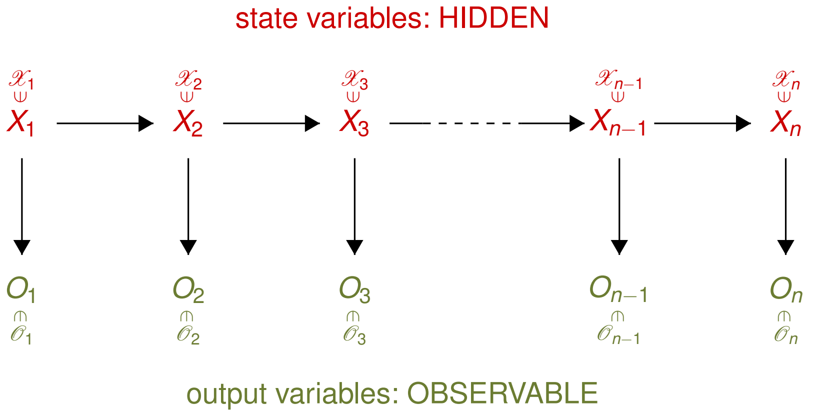


Imprecise hidden Markov model

random variables

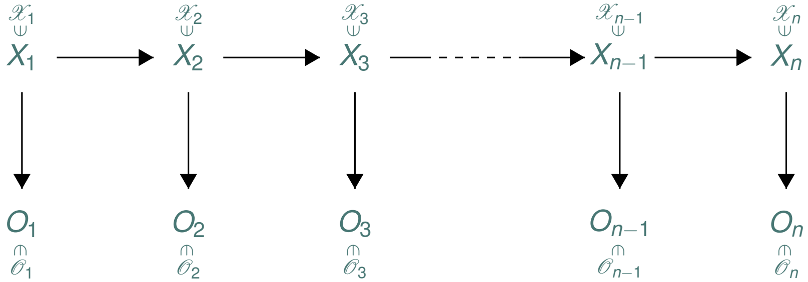


Imprecise hidden Markov model
random variables



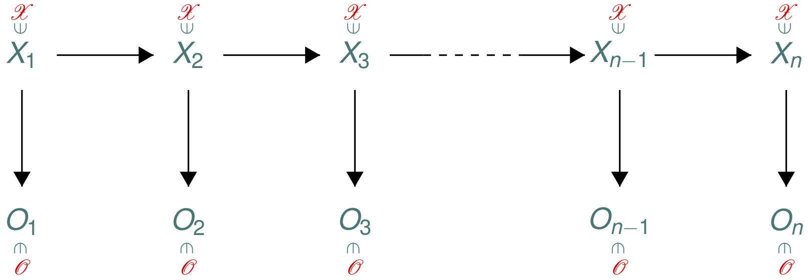
Imprecise hidden Markov model

We consider stationary imprecise hidden Markov models

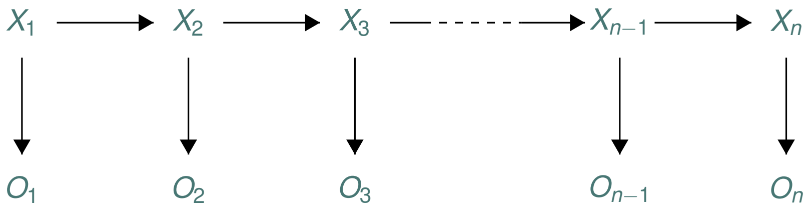


Imprecise hidden Markov model

We consider stationary imprecise hidden Markov models

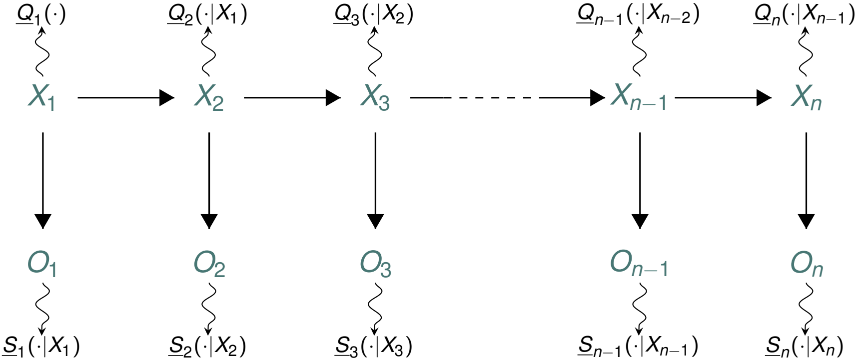


Imprecise hidden Markov model



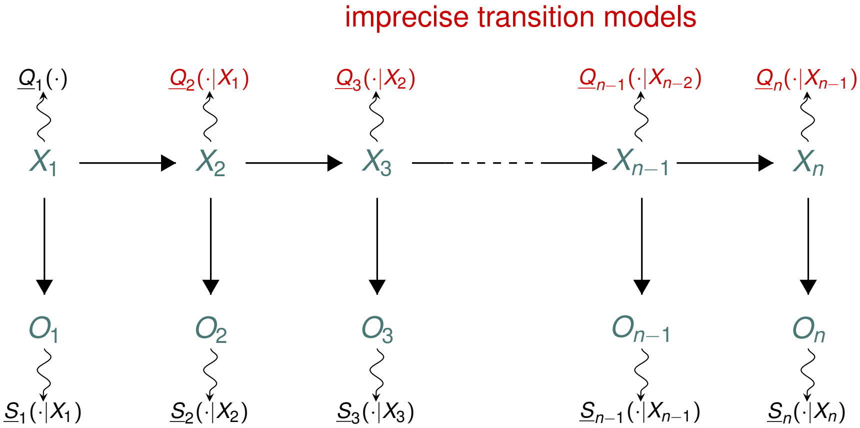
Imprecise hidden Markov model

local uncertainty models in terms of coherent lower previsions



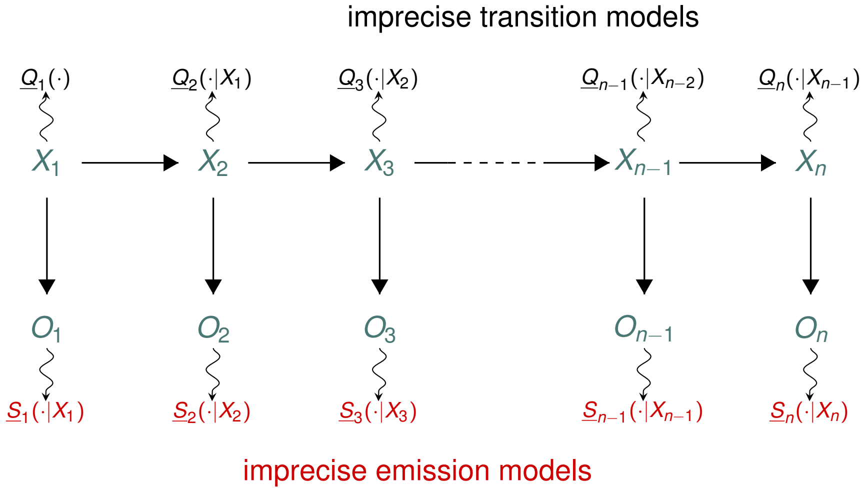
Imprecise hidden Markov model

local uncertainty models in terms of coherent lower previsions



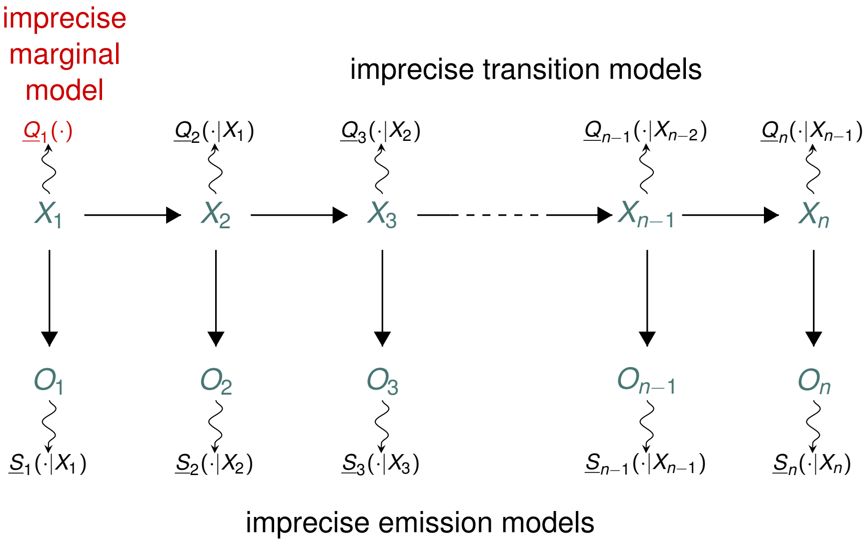
Imprecise hidden Markov model

local uncertainty models in terms of coherent lower previsions



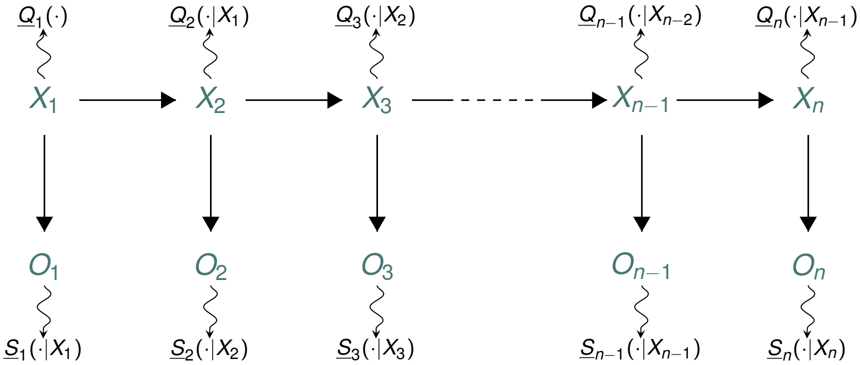
Imprecise hidden Markov model

local uncertainty models in terms of coherent lower previsions



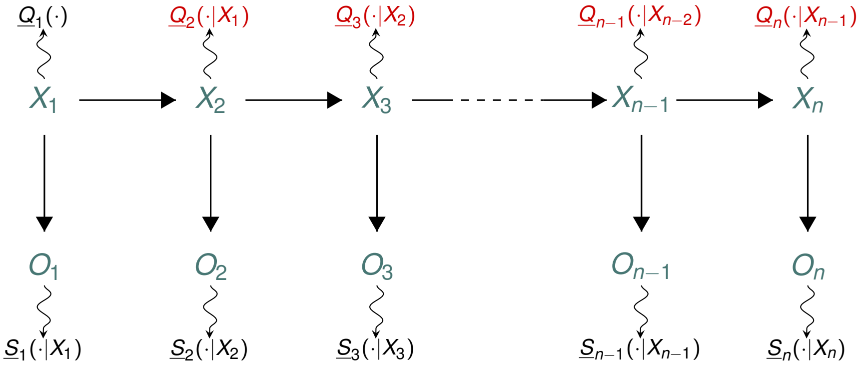
Imprecise hidden Markov model

We consider stationary imprecise hidden Markov models



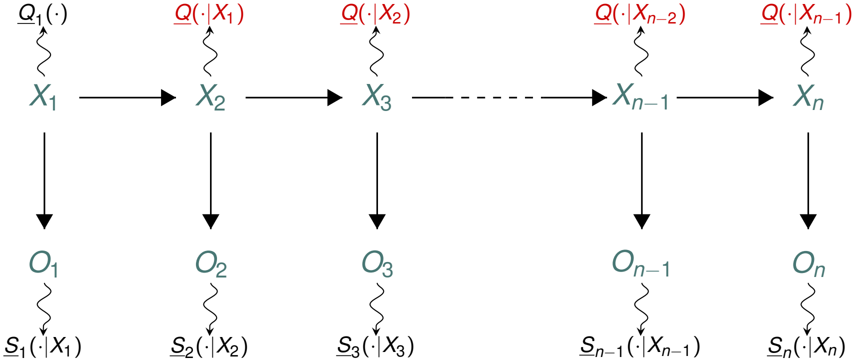
Imprecise hidden Markov model

We consider stationary imprecise hidden Markov models



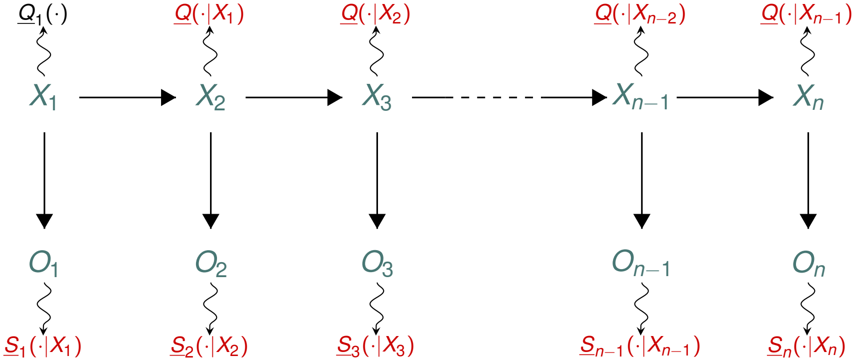
Imprecise hidden Markov model

We consider stationary imprecise hidden Markov models



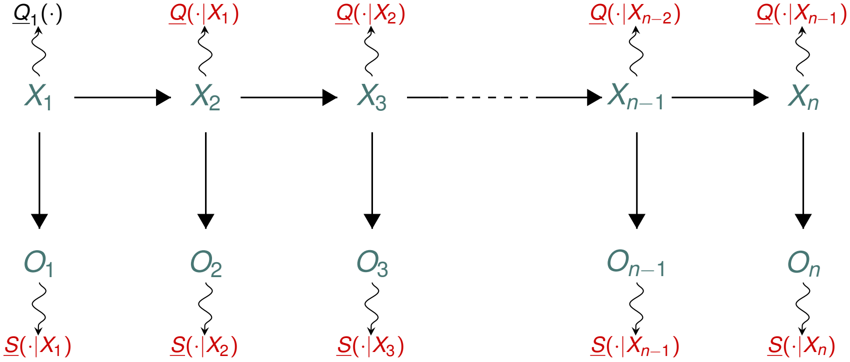
Imprecise hidden Markov model

We consider stationary imprecise hidden Markov models



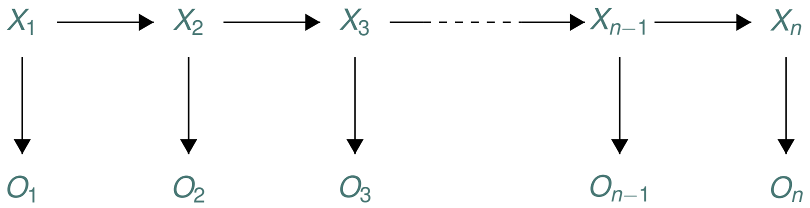
Imprecise hidden Markov model

We consider stationary imprecise hidden Markov models



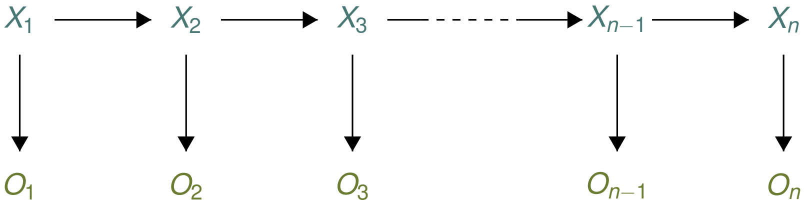
What do we want to do?

Our problem of interest



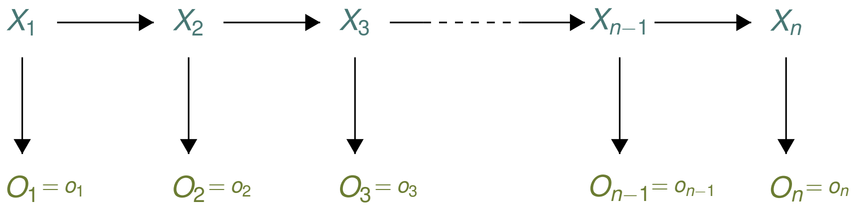
Our problem of interest

Suppose we know the output sequence



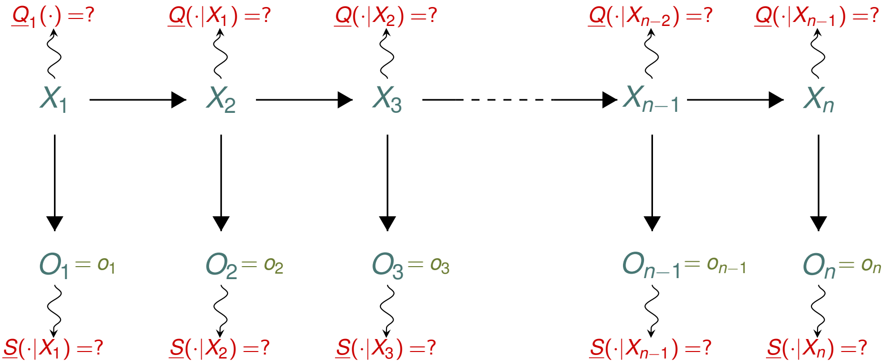
Our problem of interest

Suppose we know the output sequence: $O_{1:n} = o_{1:n} \in \mathscr{O}^n$



Our problem of interest

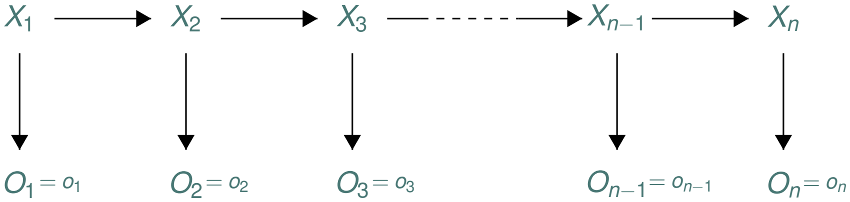
Suppose we know the output sequence: $O_{1:n} = o_{1:n} \in \mathscr{O}^n$,
we want to estimate the unknown local uncertainty models.



How could you learn the local models if the state sequence were known?

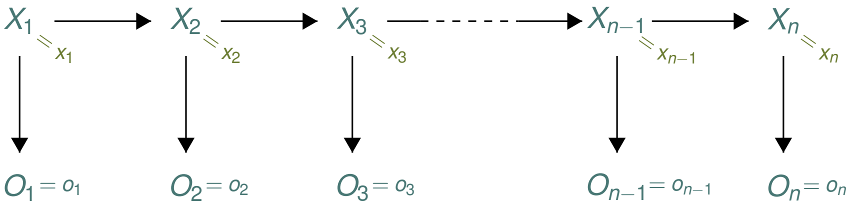
An easier problem

We know the output sequence: $O_{1:n} = o_{1:n} \in \mathscr{O}^n$.



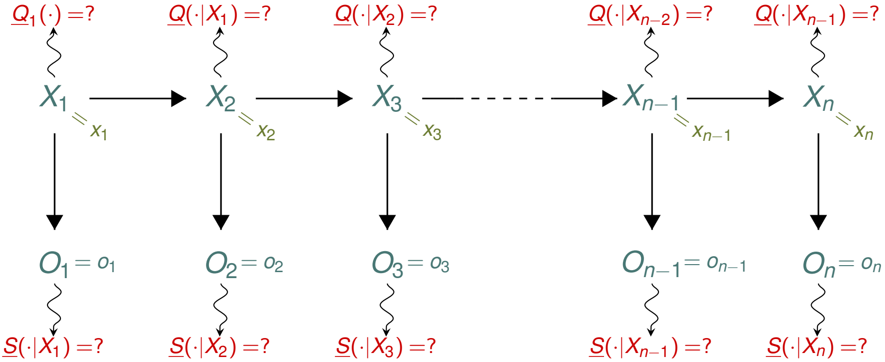
An easier problem

We know the output sequence: $O_{1:n} = o_{1:n} \in \mathscr{O}^n$.
Suppose we know in addition also the state sequence: $X_{1:n} = x_{1:n} \in \mathscr{X}^n$

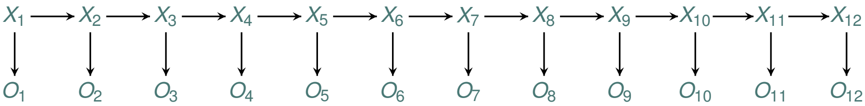


An easier problem

We know the output sequence: $O_{1:n} = o_{1:n} \in \mathscr{O}^n$.
Suppose we know in addition also the state sequence: $X_{1:n} = x_{1:n} \in \mathscr{X}^n$,
how can we learn local models now?

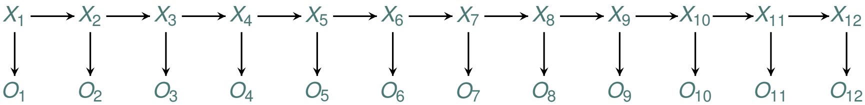


Solution



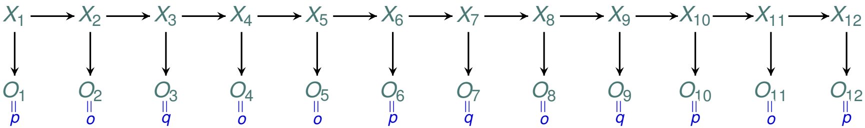
X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12}
O_1 O_2 O_3 O_4 O_5 O_6 O_7 O_8 O_9 O_{10} O_{11} O_{12}

Solution



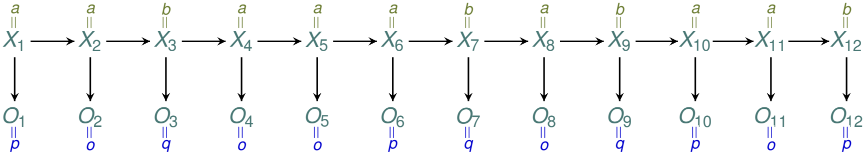
Suppose $\mathscr{X} = \{a, b\}$ and $\mathscr{O} = \{o, p, q\}$.

Solution



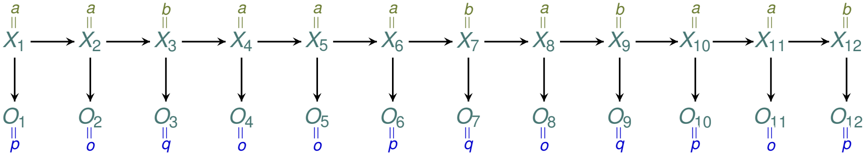
Suppose $\mathscr{X} = \{a, b\}$ and $\mathscr{O} = \{o, p, q\}$.

Solution



Suppose $\mathscr{X} = \{a, b\}$ and $\mathscr{O} = \{o, p, q\}$.

Solution



Suppose $\mathscr{X} = \{a, b\}$ and $\mathscr{O} = \{o, p, q\}$.

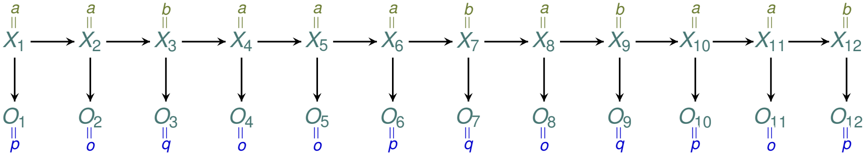
With (known) hidden state sequence $x_{1:n}$ and output sequence $o_{1:n}$ ($x, y \in \mathscr{X}$ and $z \in \mathscr{O}$):

n_x : number of times a state x is reached,

$n_{x,y}$: number of times a state transition from x to y takes place,

$n_{x,z}$: number of times a state x emits an output z.

Solution



Suppose $\mathscr{X} = \{a, b\}$ and $\mathscr{O} = \{o, p, q\}$.

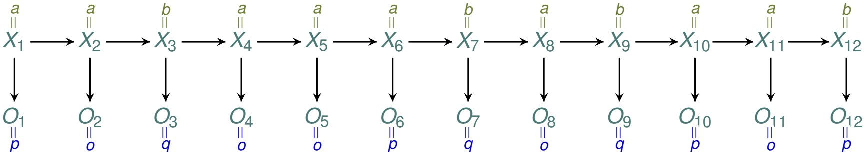
With (known) hidden state sequence $x_{1:n}$ and output sequence $o_{1:n}$ ($x, y \in \mathscr{X}$ and $z \in \mathscr{O}$):

$$\begin{aligned} n_x &: \text{number of times a state } x \text{ is reached,} \\ n_{x,y} &: \text{number of times a state transition from } x \text{ to } y \text{ takes place,} \\ n_{x,z} &: \text{number of times a state } x \text{ emits an output } z. \end{aligned}$$

Here:

$$\left.\begin{aligned} n_a = 8, n_b = 4, \\ n_{a,a} = 4, n_{a,b} = 4, n_{b,a} = 3, n_{b,b} = 0, \\ n_{a,o} = 5, n_{a,p} = 3, n_{a,q} = 0, \\ n_{b,o} = 0, n_{b,p} = 1, n_{b,q} = 3. \end{aligned}\right\}$$

Solution



Suppose $\mathscr{X} = \{a, b\}$ and $\mathscr{O} = \{o, p, q\}$.

With (known) hidden state sequence $x_{1:n}$ and output sequence $o_{1:n}$ ($x, y \in \mathscr{X}$ and $z \in \mathscr{O}$):

n_x : number of times a state x is reached,

$n_{x,y}$: number of times a state transition from x to y takes place,

$n_{x,z}$: number of times a state x emits an output z.

Here:

$$\left.\begin{array}{r} n_a = 8, n_b = 4, \\ n_{a,a} = 4, n_{a,b} = 4, n_{b,a} = 3, n_{b,b} = 0, \\ n_{a,o} = 5, n_{a,p} = 3, n_{a,q} = 0, \\ n_{b,o} = 0, n_{b,p} = 1, n_{b,q} = 3. \end{array}\right\}$$

With these counts, how can we build local models?

Imprecise Dirichlet model

We use the imprecise Dirichlet model (IDM) to compute estimates for the local models. If $n(A)$ is the number of occurences of an event A in N experiments, then the lower and upper probability of A according to an IDM are defined as

$$\underline{P}(A) = \frac{n(A)}{s+N} \quad \text{and} \quad \overline{P}(A) = \frac{s+n(A)}{s+N}.$$

$s > 0$ is the number of pseudo-counts, which is an inverse measure of the speed of convergence to a precise model.

Imprecise Dirichlet model

We use the imprecise Dirichlet model (IDM) to compute estimates for the local models. If $n(A)$ is the number of occurences of an event A in N experiments, then the lower and upper probability of A according to an IDM are defined as

$$\underline{P}(A) = \frac{n(A)}{s+N} \quad \text{and} \quad \overline{P}(A) = \frac{s+n(A)}{s+N}.$$

$s > 0$ is the number of pseudo-counts, which is an inverse measure of the speed of convergence to a precise model.

Now, we use the quantities n_x, $n_{x,y}$ and $n_{x,z}$ (with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) to estimate the imprecise transition and emission models:

Imprecise Dirichlet model

We use the imprecise Dirichlet model (IDM) to compute estimates for the local models. If $n(A)$ is the number of occurences of an event A in N experiments, then the lower and upper probability of A according to an IDM are defined as

$$\underline{P}(A) = \frac{n(A)}{s+N} \quad \text{and} \quad \overline{P}(A) = \frac{s+n(A)}{s+N}.$$

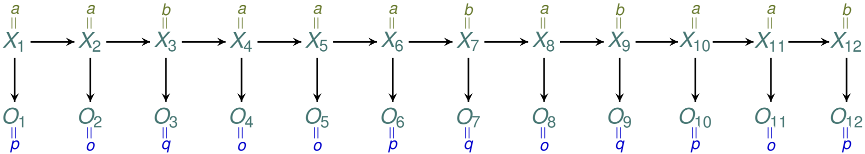
$s > 0$ is the number of pseudo-counts, which is an inverse measure of the speed of convergence to a precise model.

Now, we use the quantities n_x, $n_{x,y}$ and $n_{x,z}$ (with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) to estimate the imprecise transition and emission models:

$$\underline{Q}(\{y\}|x) = \frac{n_{x,y}}{s+\sum_{y^* \in \mathscr{X}} n_{x,y^*}} \quad \text{and} \quad \overline{Q}(\{y\}|x) = \frac{s+n_{x,y}}{s+\sum_{y^* \in \mathscr{X}} n_{x,y^*}},$$

$$\underline{S}(\{z\}|x) = \frac{n_{x,z}}{s+n_x} \quad \text{and} \quad \overline{S}(\{z\}|x) = \frac{s+n_{x,z}}{s+n_x}.$$

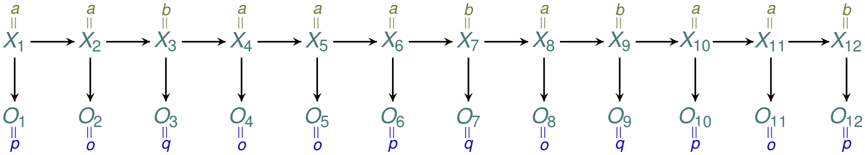
Example



(with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$):

$$\underline{Q}(\{y\}|x) = \frac{n_{x,y}}{s + \sum_{y^* \in \mathscr{X}} n_{x,y^*}}, \overline{Q}(\{y\}|x) = \frac{s + n_{x,y}}{s + \sum_{y^* \in \mathscr{X}} n_{x,y^*}}, \underline{S}(\{z\}|x) = \frac{n_{x,z}}{s + n_x}, \overline{S}(\{z\}|x) = \frac{s + n_{x,z}}{s + n_x}.$$

Example



(with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$):

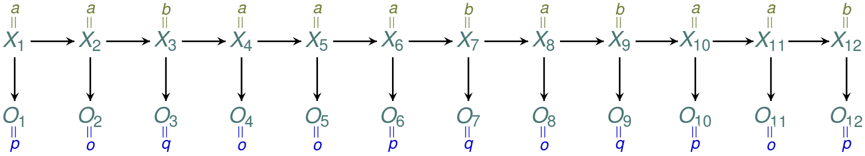
$$\underline{Q}(\{y\}|x) = \frac{n_{x,y}}{s + \sum_{y^* \in \mathscr{X}} n_{x,y^*}}, \overline{Q}(\{y\}|x) = \frac{s + n_{x,y}}{s + \sum_{y^* \in \mathscr{X}} n_{x,y^*}}, \underline{S}(\{z\}|x) = \frac{n_{x,z}}{s + n_x}, \overline{S}(\{z\}|x) = \frac{s + n_{x,z}}{s + n_x}.$$

Here, with $s = 2$:

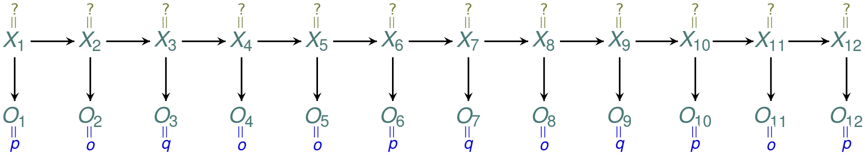
$\underline{Q}(\{a\}|a) = 2/5$, $\overline{Q}(\{a\}|a) = 3/5$, $\underline{Q}(\{b\}|a) = 2/5$, $\overline{Q}(\{b\}|a) = 3/5$,
$\underline{Q}(\{a\}|b) = 3/5$, $\overline{Q}(\{a\}|b) = 1$, $\underline{Q}(\{b\}|b) = 0$, $\overline{Q}(\{b\}|b) = 2/5$,
$\underline{S}(\{o\}|a) = 1/2$, $\overline{S}(\{o\}|a) = 7/10$, $\underline{S}(\{o\}|b) = 0$, $\overline{S}(\{o\}|b) = 1/3$,
$\underline{S}(\{p\}|a) = 3/10$, $\overline{S}(\{p\}|a) = 1/2$, $\underline{S}(\{p\}|b) = 1/6$, $\overline{S}(\{p\}|b) = 1/2$,
$\underline{S}(\{q\}|a) = 0$, $\overline{S}(\{q\}|a) = 1/5$, $\underline{S}(\{q\}|b) = 1/5$, $\overline{S}(\{q\}|b) = 3/5$.

But the state sequence is hidden...

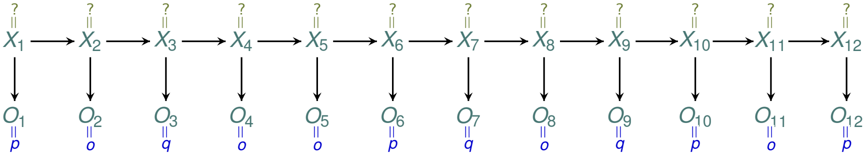
We are almost there



We are almost there

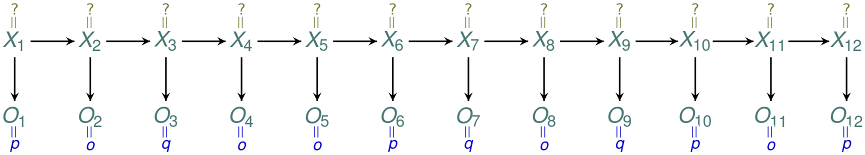


We are almost there



The state sequence $x_{1:n} \in \mathscr{X}^n$ is hidden, so it is a random variable $X_{1:n}$.

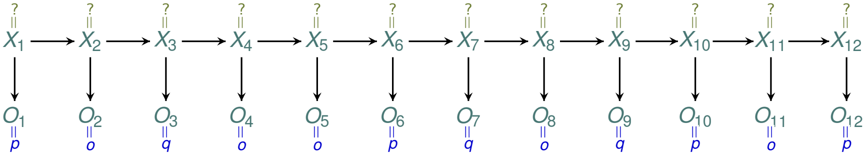
We are almost there



The state sequence $x_{1:n} \in \mathscr{X}^n$ is hidden, so it is a random variable $X_{1:n}$.

(with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) n_x, $n_{x,y}$ and $n_{x,z}$ are random variables N_x, $N_{x,y}$ and $N_{x,z}$.

We are almost there



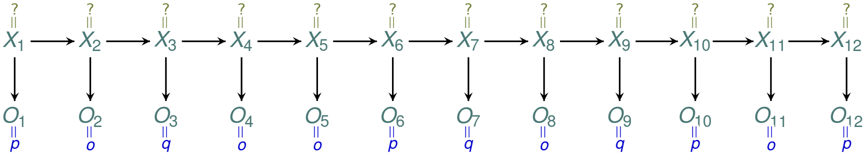
The state sequence $x_{1:n} \in \mathscr{X}^n$ is hidden, so it is a random variable $X_{1:n}$.

(with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) n_x, $n_{x,y}$ and $n_{x,z}$ are random variables N_x, $N_{x,y}$ and $N_{x,z}$.

Idea: instead of using real counts, use estimates:

$$\hat{n}_x,$$
$$\hat{n}_{x,y},$$
$$\hat{n}_{x,z}.$$

We are almost there



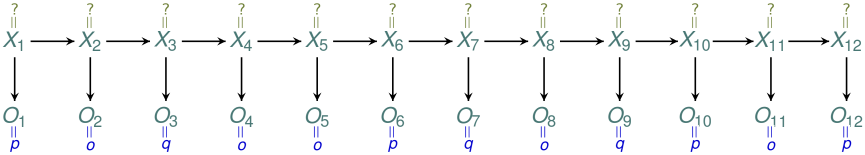
The state sequence $x_{1:n} \in \mathscr{X}^n$ is hidden, so it is a random variable $X_{1:n}$.

(with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) n_x, $n_{x,y}$ and $n_{x,z}$ are random variables N_x, $N_{x,y}$ and $N_{x,z}$.

Idea: instead of using real counts, use expected counts:

$$\hat{n}_x = E(N_x | o_{1:n}, \theta^*),$$
$$\hat{n}_{x,y} = E(N_{x,y} | o_{1:n}, \theta^*),$$
$$\hat{n}_{x,z} = E(N_{x,z} | o_{1:n}, \theta^*).$$

We are almost there



The state sequence $x_{1:n} \in \mathscr{X}^n$ is hidden, so it is a random variable $X_{1:n}$.

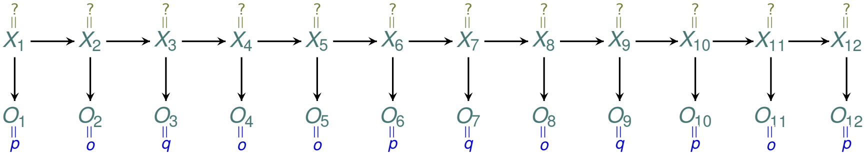
(with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) n_x, $n_{x,y}$ and $n_{x,z}$ are random variables N_x, $N_{x,y}$ and $N_{x,z}$.

Idea: instead of using real counts, use expected counts

$$\hat{n}_x = E(N_x | o_{1:n}, \theta^*),$$
$$\hat{n}_{x,y} = E(N_{x,y} | o_{1:n}, \theta^*),$$
$$\hat{n}_{x,z} = E(N_{x,z} | o_{1:n}, \theta^*).$$

$o_{1:n}$ is the known output sequence, and θ^* represents the model parameter.

We are almost there



The state sequence $x_{1:n} \in \mathscr{X}^n$ is hidden, so it is a random variable $X_{1:n}$.

(with $x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) n_x, $n_{x,y}$ and $n_{x,z}$ are random variables N_x, $N_{x,y}$ and $N_{x,z}$.

Idea: instead of using real counts, use expected counts

$$\begin{aligned} \hat{n}_x &= E(N_x | o_{1:n}, \theta^*), \\ \hat{n}_{x,y} &= E(N_{x,y} | o_{1:n}, \theta^*), \\ \hat{n}_{x,z} &= E(N_{x,z} | o_{1:n}, \theta^*). \end{aligned}$$

$o_{1:n}$ is the known output sequence, and θ^* represents the model parameter. We can calculate θ^* with the Baum–Welch algorithm, so the idea makes sense.

Estimated local models

With known state sequence $x_{1:n}$ ($x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) :

$$\underline{Q}(\{y\}|x) = \frac{n_{x,y}}{s + \sum_{y^* \in \mathscr{X}} n_{x,y^*}} \text{ and } \overline{Q}(\{y\}|x) = \frac{s + n_{x,y}}{s + \sum_{y^* \in \mathscr{X}} n_{x,y^*}},$$

$$\underline{S}(\{z\}|x) = \frac{n_{x,z}}{s + n_x} \text{ and } \overline{S}(\{z\}|x) = \frac{s + n_{x,z}}{s + n_x}.$$

Estimated local models

With unknown state sequence $X_{1:n}$ ($x, y \in \mathscr{X}$ and $z \in \mathscr{O}$) :

$$\underline{Q}(\{y\}|x) = \frac{\hat{n}_{x,y}}{s + \sum_{y^* \in \mathscr{X}} \hat{n}_{x,y^*}} \text{ and } \overline{Q}(\{y\}|x) = \frac{s + \hat{n}_{x,y}}{s + \sum_{y^* \in \mathscr{X}} \hat{n}_{x,y^*}},$$

$$\underline{S}(\{z\}|x) = \frac{\hat{n}_{x,z}}{s + \hat{n}_x} \text{ and } \overline{S}(\{z\}|x) = \frac{s + \hat{n}_{x,z}}{s + \hat{n}_x}.$$

Example: predicting future earthquake rates

We want to predict future earthquake rates, based on number of earthquakes in previous years.

Example: predicting future earthquake rates

We want to predict future earthquake rates, based on number of earthquakes in previous years.

Assumptions:

- Earth can be in 3 possible seismic states λ_1, λ_2 and λ_3,

Example: predicting future earthquake rates

We want to predict future earthquake rates, based on number of earthquakes in previous years.

Assumptions:

- Earth can be in 3 possible seismic states λ_1, λ_2 and λ_3,
- occurrence of earthquakes in a year depends on the seismic state in that year,

Example: predicting future earthquake rates

We want to predict future earthquake rates, based on number of earthquakes in previous years.

Assumptions:

- Earth can be in 3 possible seismic states λ_1, λ_2 and λ_3,
- occurrence of earthquakes in a year depends on the seismic state in that year,
- Earth in state λ emits O earthquakes in a year.

Example: predicting future earthquake rates

We want to predict future earthquake rates, based on number of earthquakes in previous years.

Assumptions:

- Earth can be in 3 possible seismic states λ_1, λ_2 and λ_3,
- occurrence of earthquakes in a year depends on the seismic state in that year,
- Earth in state λ emits O earthquakes in a year.

We model our problem as an imprecise hidden Markov model.

Example: predicting future earthquake rates

We want to predict future earthquake rates, based on number of earthquakes in previous years.

Assumptions:

- Earth can be in 3 possible seismic states: $\mathscr{X} = \{\lambda_1, \lambda_2, \lambda_3\}$,
- occurrence of earthquakes in a year depends on the seismic state in that year,
- Earth in state λ emits O earthquakes in a year.

We model our problem as an imprecise hidden Markov model.

Example: predicting future earthquake rates

We want to predict future earthquake rates, based on number of earthquakes in previous years.

Assumptions:

- Earth can be in 3 possible seismic states: $\mathscr{X} = \{\lambda_1, \lambda_2, \lambda_3\}$,
- occurrence of earthquakes in a year depends on the seismic state in that year,
- Earth in state λ emits O earthquakes in a year: $\mathscr{O} = \mathbb{N} \cup \{0\}$ and emission model $S(o|\lambda)$ is a Poisson process, represented by the precise probability mass function $p(o|\lambda) = \frac{e^{-\lambda}\lambda^o}{o!}$.

We model our problem as an imprecise hidden Markov model.

Example: predicting future earthquake rates

We want to predict future earthquake rates, based on number of earthquakes in previous years.

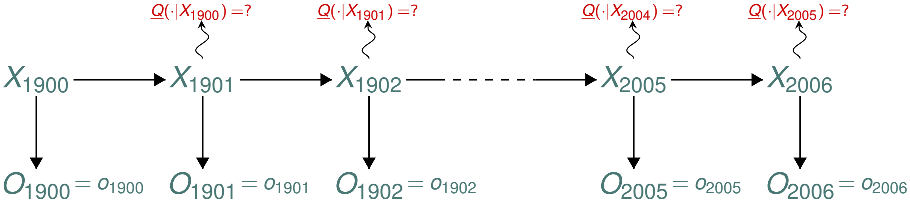
Assumptions:

- Earth can be in 3 possible seismic states: $\mathscr{X} = \{\lambda_1, \lambda_2, \lambda_3\}$,
- occurrence of earthquakes in a year depends on the seismic state in that year,
- Earth in state λ emits O earthquakes in a year: $\mathscr{O} = \mathbb{N} \cup \{0\}$ and emission model $S(o|\lambda)$ is a Poisson process, represented by the precise probability mass function $p(o|\lambda) = \frac{e^{-\lambda}\lambda^o}{o!}$.

We model our problem as an imprecise hidden Markov model.

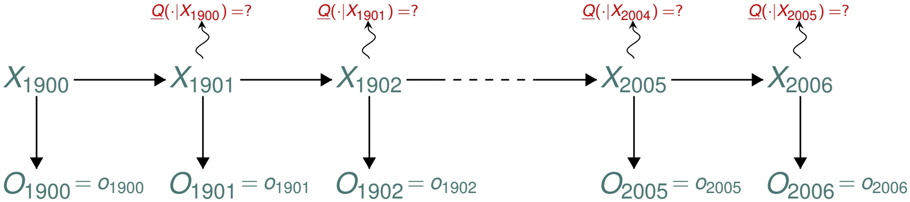
Our observation: number of earthquakes from 1900 to 2006

Example: learned model

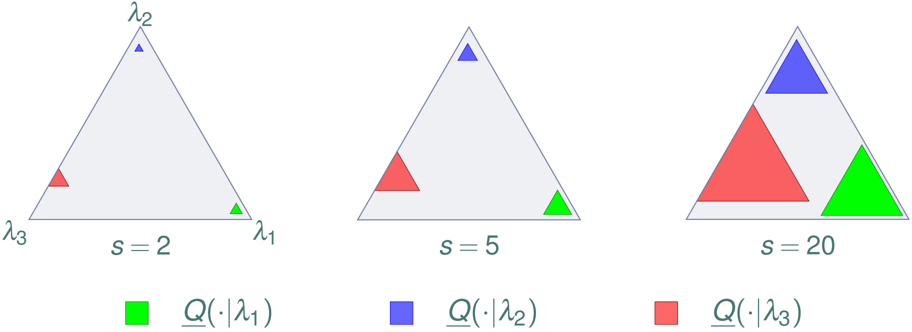


Based on the data, we learn the (imprecise) transition model.

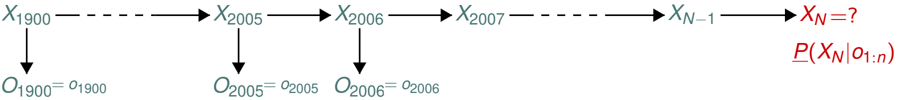
Example: learned model



Based on the data, we learn the (imprecise) transition model.

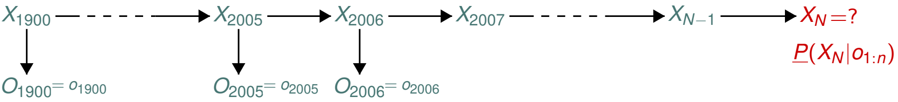


Example: predicting earthquake rates



With the learned imprecise hidden Markov model, we predict future earthquake rates. We use the MePiCTIr algorithm (de Cooman et al., 2010).

Example: predicting earthquake rates



With the learned imprecise hidden Markov model, we predict future earthquake rates. We use the MePiCTIr algorithm (de Cooman et al., 2010).