Learning imprecise hidden Markov models

Arthur Van Camp, Gert de Cooman, Jasper De Bock, Erik Quaeghebeur, Filip Hermans
SYSTeMS research group, Ghent University, Belgium
{Arthur.VanCamp,Gert.deCooman,Jasper.DeBock,Erik.Quaeghebeur,Filip.Hermans}@UGent.be

UNIVERSITEIT GENT

Learning precise HMMs

Precise hidden Markov model Consider a stationary precise hidden Markov model (HMM) with $2n$ variables: n hidden states X_k, taking values x_k in a set $\{1,\dots,m\}$ and n observations O_k, taking values o_k. Both the marginal model $p_{X_1}(x_1)$, the transition models $p_{X_k|X_{k-1}}(x_k|x_{k-1})$ and the emission models $p_{O_k|X_k}(o_k|x_k)$ are unknown.

Baum–Welch algorithm Given the observation sequence $(O_1=o_1,\dots,O_n=o_n)$, we can use the Baum–Welch algorithm to obtain a maximum-likelihood estimate of these local models. With this algorithm, the likelihood of the observation sequence converges to a *local* maximum, but it is not guaranteed that we find the *global* maximum.

Expected number of transitions The Baum–Welch algorithm implicitly constructs the expected number of transitions

$$n_{ij} := \sum_{k=2}^{n} p_{X_{k-1},X_k|O_{1:n}}(i,j|o_{1:n})$$

in the whole Markov chain of the HMM.

Learning imprecise HMMs

Imprecise hidden Markov model An imprecise hidden Markov model (iHMM) has the same graphical model but the local models are imprecise.

Using Baum–Welch With the classical Baum–Welch algorithm we obtain *precise* local models. We present a method for learning *imprecise* transition models in an iHMM. We use the expected number of transitions, obtained by the Baum–Welch algorithm after sufficient iterations, to construct imprecise transition models.

Multinomial processes The transitions from a state $X_{k-1}=i$ to a state $X_k=j$ are multinomial processes. An *imprecise Dirichlet model* (IDM) is a convenient model for describing uncertainty about such processes. In order to learn using an IDM, we need the number of transitions and a choice for the pseudocounts s.

Proposed transition model Since the hidden states are unavailable, our method consists in taking the *expected* number of transitions derived from the Baum–Welch algorithm, rather than real counts. We estimate the lower and upper probability for state j conditional on state i by

$$\underline{Q}(\{j\}|i) = \frac{n_{ij}}{s+n_i} \quad \text{and} \quad \overline{Q}(\{j\}|i) = \frac{s+n_{ij}}{s+n_i},$$

where $n_i := \sum_{j=1}^{m} n_{ij}$.

Imprecision

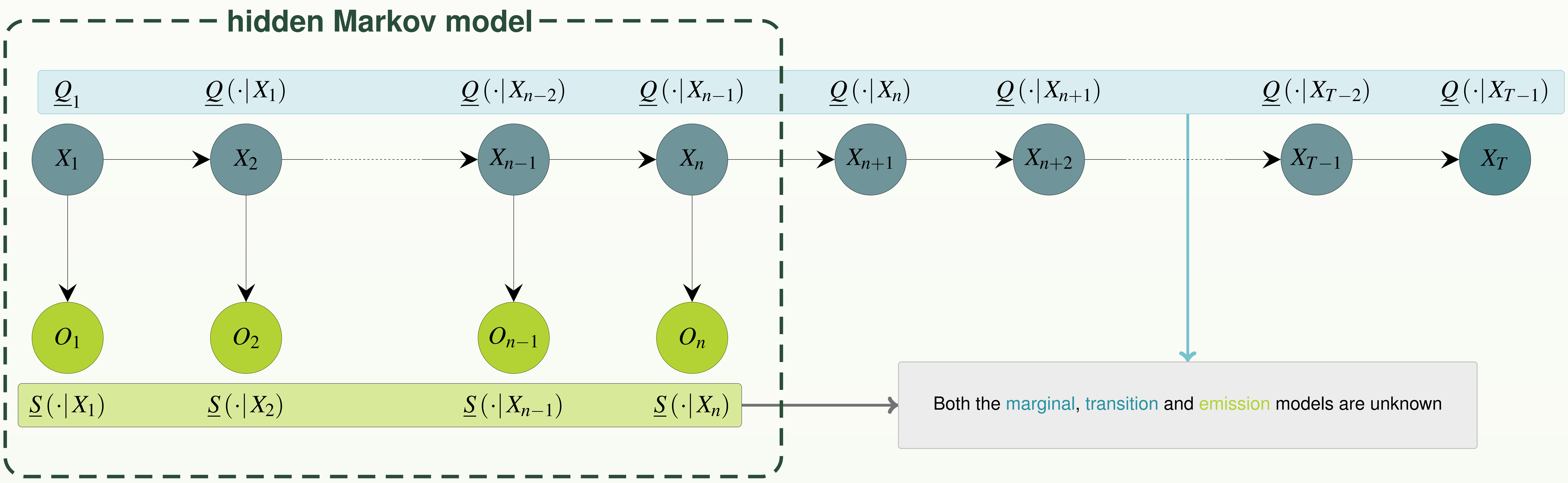
The lower and upper probabilities have the following property: the imprecision increases by increasing number of states m.

$$\overline{Q}(\{j\}|i) - \underline{Q}(\{j\}|i) = \frac{s+n_{ij}}{s+n_i} - \frac{n_{ij}}{s+n_i} = \frac{s}{s+n_i}$$

Here $n_i = \sum_{j=1}^{m} n_{ij}$ the expected number of times that state i occurs in the $n-1$ variables $X_1,\dots,X_{n-1}$.
If the *number of states m increases*, then in general $\sum_{j=1}^{m} n_{ij}$ will decrease, so the *imprecision increases*.

- With large m, we can know *less* precisely which state occurs, but knowing this state *tells* us much,
- With small m, we can know *more* precisely which state occurs, but knowing this state *doesn't tell* us much.

$\Rightarrow$ **The amount of information we can infer about an iHMM is limited.**



Predicting the earthquake rate

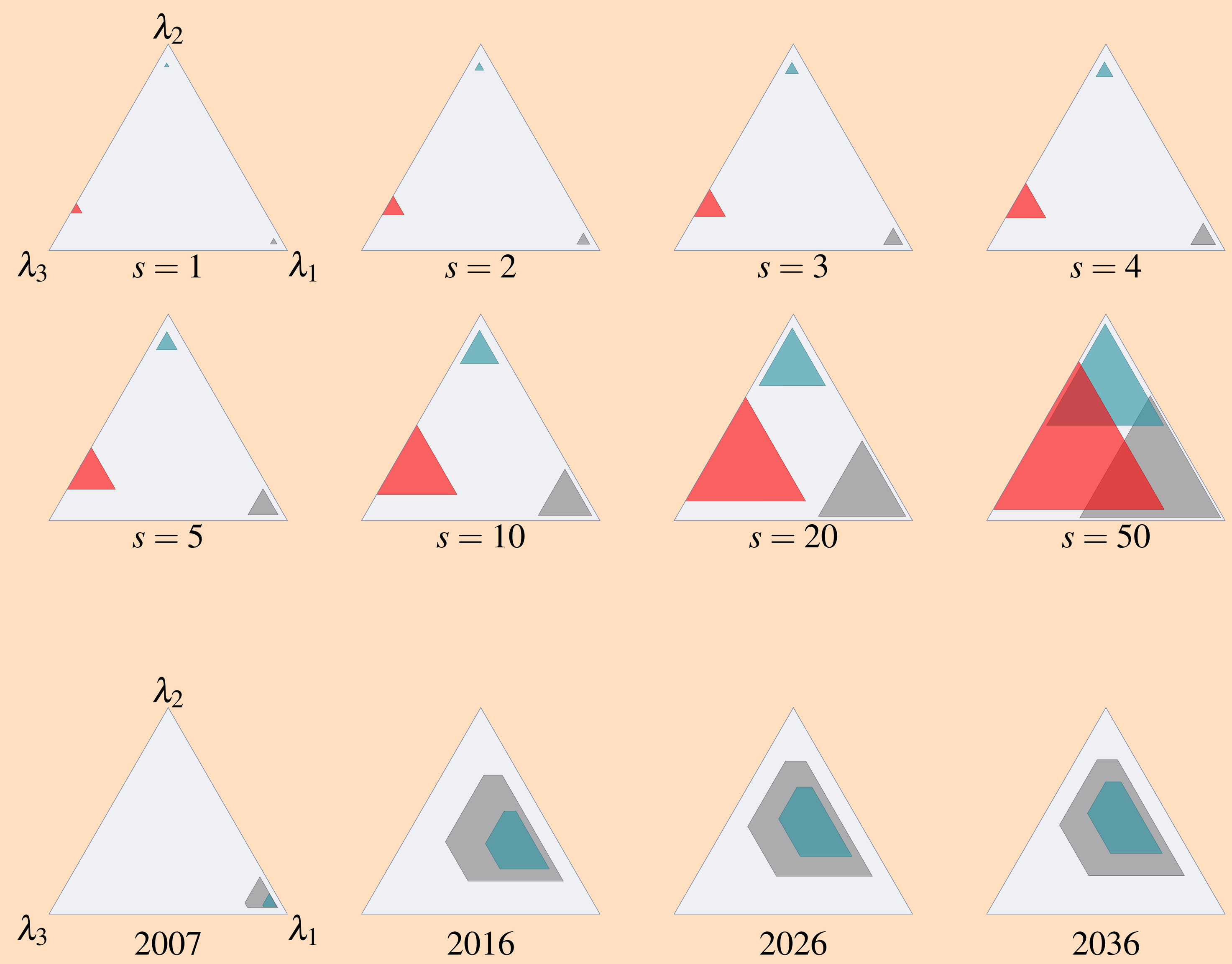
We apply our method to the following problem: based on counted number of annual earthquakes in 107 subsequent years (1900 – 2006), we are interested in predicting the **earthquake rate** in future years.

We assume that

- the earth can be in 3 different seismic states λ_1, λ_2 and λ_3,
- in each state, the emission of earthquakes is a **Poisson process**: $\quad p_{O_i}(o_i|\boldsymbol{\lambda}) = e^{-\lambda}\dfrac{\lambda^{o_i}}{o_i!}.$

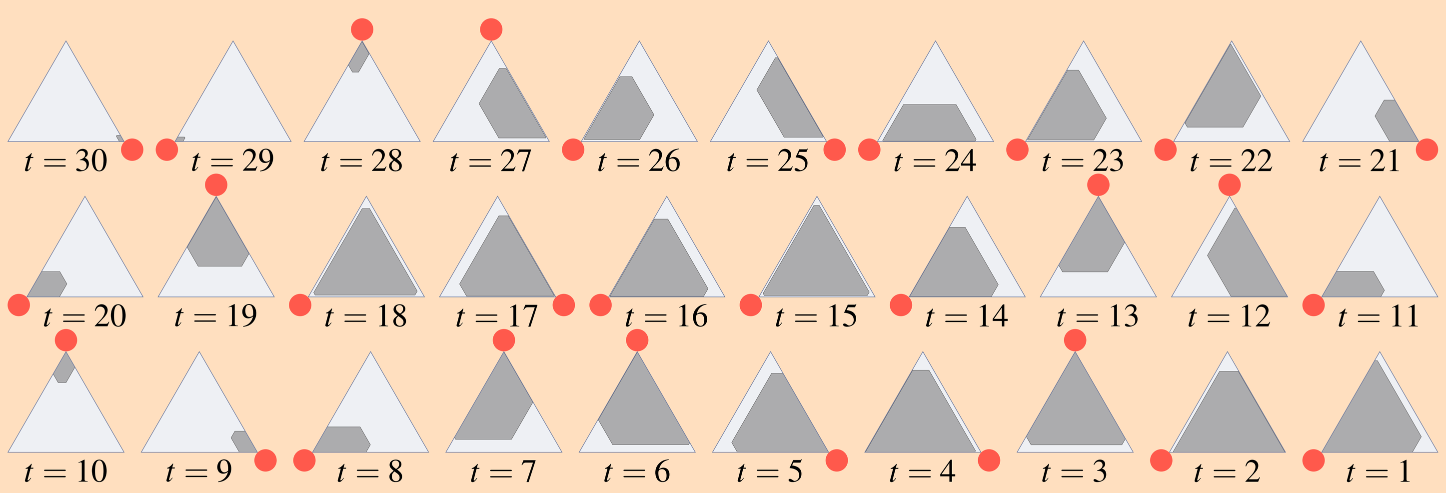
Transition model The credal sets in the upper eight simplices on the right represent, for different values of the pseudocounts s, the transition models. The gray credal set represents $\underline{Q}(\cdot|\lambda_1)$, the blue credal set represents $\underline{Q}(\cdot|\lambda_2)$ and the red credal set represents $\underline{Q}(\cdot|\lambda_3)$

Prediction With the transition models learned with our method, we predicted the earthquake rate in the years 2007, 2016, 2026 and 2036. We did this in two cases: the pseudocounts $s=2$ and $s=5$. The lower four simplices on the right show conservative approximations (the smallest hexagons with vertices parallel with the vertices of the simplex) for the credal sets representing the global model $\underline{R}_{X_T}(\cdot|o_{1:n})$, updated to the observation sequence. The gray credal set represents the updated global model with $s=5$ and the blue credal set represents the updated global model with $s=2$. As expected, the global model for $s=5$ include the global model for $s=2$.



Dilation

Learning imprecise probability models in an iHMM, like our method does, is necessary before being able to make inferences from such a model, e.g., with the MePICTIr algorithm. *Dilation* appears here as the increase of the imprecision of the inferences when the target node X_T goes to the first state X_1. The interpretation of this phenomenon is not yet clear. We did some experiments to estimate the dilation in an iHMM with $n=50$.



	$p_{X_1}(\cdot)$	$p_{X_i\|X_{i-1}}(\cdot\|a)$	$p_{X_i\|X_{i-1}}(\cdot\|b)$	$p_{X_i\|X_{i-1}}(\cdot\|c)$	$p_{O_i\|X_i}(\cdot\|a)$	$p_{O_i\|X_i}(\cdot\|b)$	$p_{O_i\|X_i}(\cdot\|c)$
a	0, 3	0, 1	0, 1	**0, 8**	**0, 8**	0, 1	0, 1
b	0, 3	**0, 8**	0, 1	0, 1	0, 1	**0, 8**	0, 1
c	0, 4	0, 1	**0, 8**	0, 1	0, 1	0, 1	**0, 8**

The 30 simplices represent conservative approximations of the updated global model $\underline{R}_{X_T}(\cdot|o_{1:n})$. The red dots indicate the observations. The local models are linear-vacuous mixtures, of which the precise components are given in the table above.