Limit behavior of imprecise Markov chains

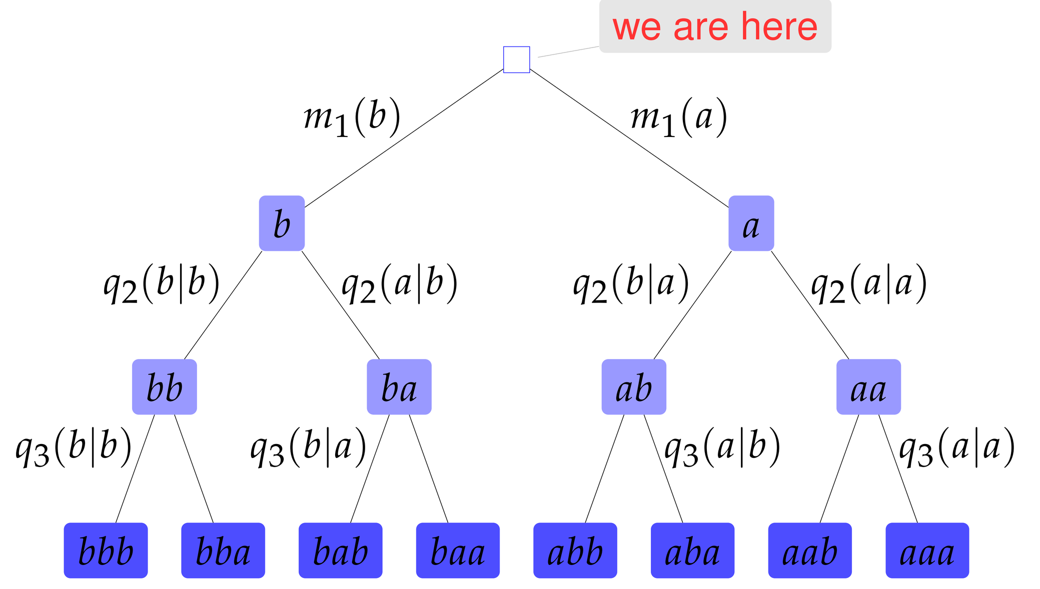
Gert de Cooman, Filip Hermans & Erik Quaeghebeur

SYSTeMS Research Group, EESA Department
Faculty of Engineering, Ghent University
Technologiepark-Zwijnaarde 914, 9052 Zwijnaarde

{Gert.deCooman,Filip.Hermans,Erik.Quaeghebeur}@UGent.be

The classical two-state Markov chain and its event tree

Consider a Markov chain with the *state space* $\{a, b\}$ and *probability mass functions* m_1 (initial) and $q_k(\cdot|a)$, $q_k(\cdot|b)$ (transition step $k-1$).



The link with the classical matrix formalism: the initial probability mass vector and the transition matrix of step $k-1$ respectively are

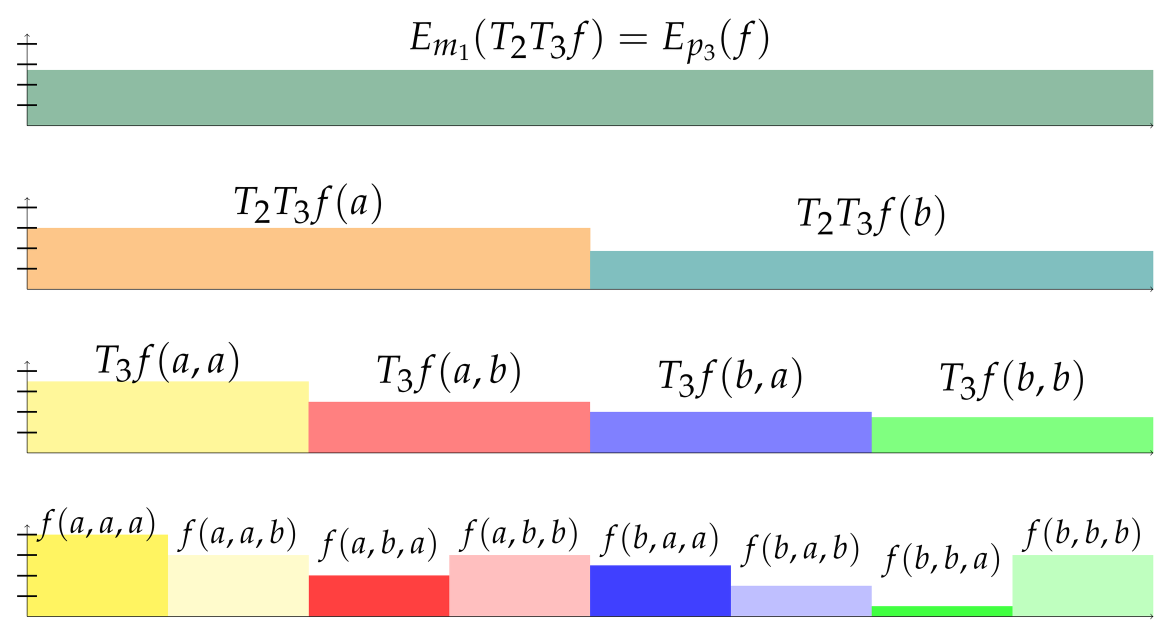
$$Q_k = \begin{bmatrix} q_k(a|a) & q_k(b|a) \\ q_k(a|b) & q_k(b|b) \end{bmatrix} \text{ and } m_1 = \begin{bmatrix} m_1(a) \\ m_1(b) \end{bmatrix}.$$

The probability mass vector m_n at time n is

$$m_n^T = m_1^T Q_2 \cdots Q_n.$$

Expectation of a function after some transition steps

Consider a real-valued function f on $\{a, b\}^3$. Its *expectation* in different parts of the event tree is (here under uniform distributions)



Let r be a probability mass function on $\{a, b\}^\ell$ and g a real-valued function on $\{a, b\}^\ell$, then the expectation of g under r is

$$E_r(g) = \sum_{x \in \{a,b\}^\ell} g(x) r(x).$$

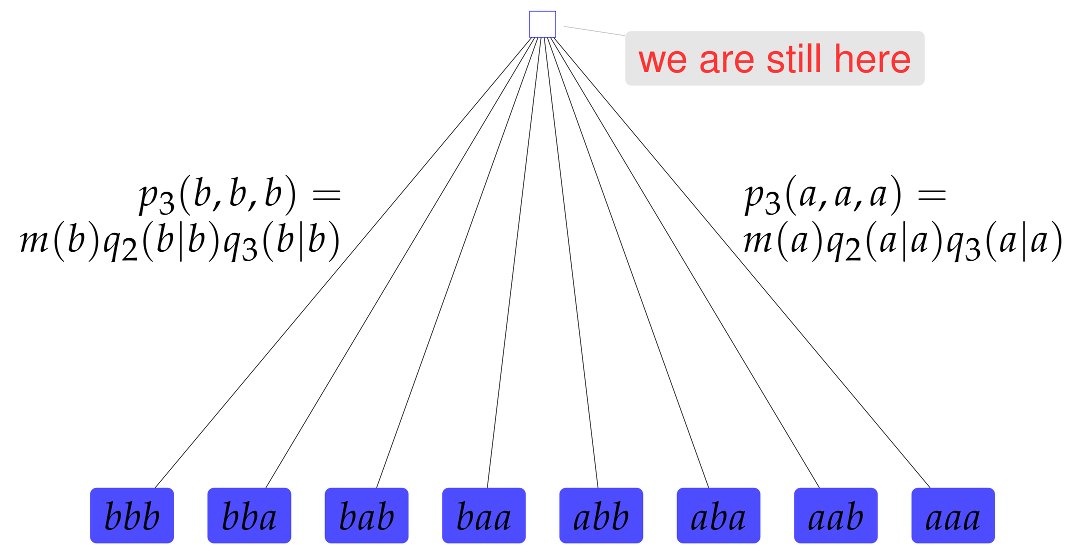
The *transition operator* T_ℓ is defined for every real-valued function h on $\{a, b\}^{\ell+1}$ and all x in $\{a, b\}^\ell$ by

$$T_\ell h(x) = E_{q_\ell(\cdot|x_\ell)}(h(x, \cdot)) = h(x, a) q_\ell(a|x_\ell) + h(x, b) q_\ell(b|x_\ell).$$

The joint distribution of a Markov chain after some transition steps

The *joint mass function* p_n after n steps is defined for all x in $\{a, b\}^n$ by

$$p_n(x) = m_1(x_1) \prod_{k=2}^{n} q_k(x_k|x_{k-1}).$$



Transition and expectation operators can also be used for calculating the joint mass function:

$$p_n(x) = E_{p_n}(I_{\{x\}}) = E_{m_1}(T_2 \cdots T_n I_{\{x\}}),$$

where I_A is the *indicator function* of the set $A \subseteq \{a, b\}^n$, which is one on that set and zero elsewhere. This also illustrates that expectation operators are in some sense more general that mass functions and can replace them.

Imprecise Markov chains: using convex sets of probabilities

Every probability mass function can be seen as a point of some *unit simplex*

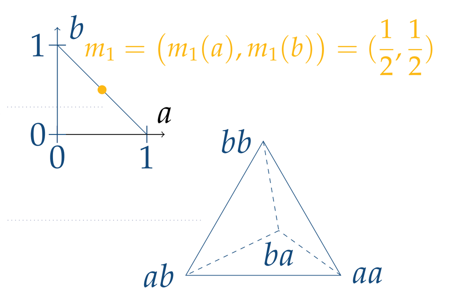
$$\Sigma_{2^\ell} = \left\{ r \in [0,1]^{2^\ell} : \sum_{x \in \{a,b\}^\ell} r(x) = 1 \right\}.$$

For example:

$$m_1, q_k(\cdot|a), q_k(\cdot|b) \in \Sigma_2,$$

$$p_2 \in \Sigma_4,$$

$$p_3 \in \Sigma_8.$$

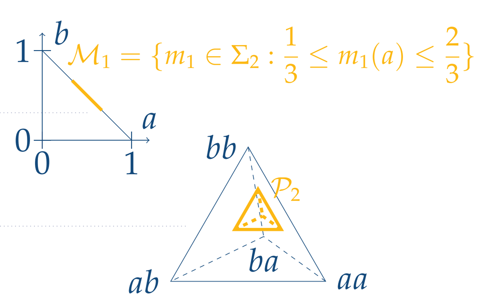


For imprecise Markov chains, the assumption is that we only know these mass functions belong to a *credal set*: a closed convex subset of the unit simplex. For example: (introducing new notation)

$$\mathcal{M}_1, \mathcal{Q}_k^a, \mathcal{Q}_k^b \subseteq \Sigma_2,$$

$$\mathcal{P}_2 \subseteq \Sigma_4,$$

$$\mathcal{P}_3 \subseteq \Sigma_8.$$



Expectation & transition operators of an imprecise Markov chain

For imprecise Markov chains, the linear expectation operator is replaced by a superlinear *lower expectation* and a sublinear *upper expectation*, respectively the minimum and maximum expectation over all probability mass functions in the credal set. Let $\mathcal{R} \subseteq \Sigma_{2^\ell}$ be a credal set on $\{a, b\}^\ell$, then

$$\underline{E}_{\mathcal{R}}(g) = \min_{r \in \mathcal{R}} E_r(g) \geq \sum_{x \in \{a,b\}^\ell} g(x) \min_{r \in \mathcal{R}} r(x) = \sum_{x \in \{a,b\}^\ell} g(x) \underline{E}_{\mathcal{R}}(I_{\{x\}}),$$

$$\overline{E}_{\mathcal{R}}(g) = \max_{r \in \mathcal{R}} E_r(g) \leq \sum_{x \in \{a,b\}^\ell} g(x) \max_{r \in \mathcal{R}} r(x) = \sum_{x \in \{a,b\}^\ell} g(x) \overline{E}_{\mathcal{R}}(I_{\{x\}}).$$

These can be used to define the upper and lower probability mass functions corresponding to $\mathcal{M}_1$ and $\mathcal{Q}_k^c$ ($c \in \{a, b\}$), whose value in any x in $\{a, b\}$ is given by

$$\underline{m}_1(x) = \underline{E}_{\mathcal{M}_1}(I_{\{x\}}), \qquad \underline{q}_1(x|c) = \underline{E}_{\mathcal{Q}_k^c}(I_{\{x\}}),$$

$$\overline{m}_1(x) = \overline{E}_{\mathcal{M}_1}(I_{\{x\}}), \qquad \overline{q}_1(x|c) = \overline{E}_{\mathcal{Q}_k^c}(I_{\{x\}}).$$

Lower and upper transition operators are defined analogously as before:

$$\underline{T}_\ell h(x) = \underline{E}_{\mathcal{Q}_\ell^{x_\ell}}(h(x, \cdot)),$$

$$\overline{T}_\ell h(x) = \overline{E}_{\mathcal{Q}_\ell^{x_\ell}}(h(x, \cdot)).$$

Lower and upper expectation and transition operators are related by *conjugacy*:

$$\underline{E}_{\mathcal{R}}(g) = -\overline{E}_{\mathcal{R}}(-g), \qquad \underline{T}_\ell h = -\overline{T}_\ell(-h).$$

Defining the joint distribution of an imprecise Markov chain

The *lower and upper joint mass functions* satisfy

$$\underline{p}_n(x) = \underline{E}_{\mathcal{P}_n}(I_{\{x\}}) = \underline{m}_1(x_1) \prod_{k=2}^{n} \underline{q}_k(x_k|x_{k-1}),$$

$$\overline{p}_n(x) = \overline{E}_{\mathcal{P}_n}(I_{\{x\}}) = \overline{m}_1(x_1) \prod_{k=2}^{n} \overline{q}_k(x_k|x_{k-1}).$$

However, because $\underline{E}_{\mathcal{P}_n}$ and $\overline{E}_{\mathcal{P}_n}$ are respectively super- and subadditive, these mass functions cannot be used to calculate arbitrary lower and upper expectations. Luckily, the approach using transition operators can be generalized from the classical case:

$$\underline{E}_{\mathcal{P}_n}(f) = \underline{E}_{\mathcal{M}_1}(\underline{T}_2 \cdots \underline{T}_n f),$$

$$\overline{E}_{\mathcal{P}_n}(f) = \overline{E}_{\mathcal{M}_1}(\overline{T}_2 \cdots \overline{T}_n f).$$

The *backpropagation* of f (i.e., $f \to \overline{T}_n f \to \overline{T}_{n-1}\overline{T}_n f \to \ldots$) is linear in the number of transition steps. The complexity of each transition step and of the calculation of the lower or upper expectation depends on the nature of the credal sets $\mathcal{M}_1$ and $\mathcal{Q}_k^c$. In some practically interesting cases it is the same as in the classical case (quadratic in the number of states).

Convergence properties of imprecise Markov chains

From now on we assume the transition operator is stationary. Consequently, we can write the k-th backpropagation of f as $\overline{T}^k f$. We work with the upper transition operator because it leads to the strongest results.

Because the decreasing sequence $\{\max(\overline{T}^k f) : k \in \mathbb{N}\}$ dominates the increasing sequence $\{\min(\overline{T}^k f) : k \in \mathbb{N}\}$, both have a limit. This means that the degree of imprecision $\max(\overline{T}^k f) - \min(\overline{T}^k f)$ converges to a constant:

$$\max(f) \geq \max(\overline{T} f) \geq \ldots \geq \lim_{k \to \infty} \max(\overline{T}^k f) \geq$$

$$\lim_{k \to \infty} \min(\overline{T}^k f) \geq \ldots \geq \max(\overline{T} f) \geq \min(f)$$

If the series $\max(\overline{T}^k f)$ ($\min(\overline{T}^k f)$) were (essentially) strictly decreasing (increasing), the degree of imprecision would be zero, hence, the map $\overline{T}$ would converge to a constant gamble. A condition that assures such (essentially) strict monotonicity is regularity.

Definition 1 *A transition operator $\overline{T}$ is regular if there is an $\ell > 0$ such that $\overline{T}^\ell(I_{\{x\}}) > 0$ for every element x of the state-space.*

Intuitively, this means that it is not impossibe to get from any given state into any other state.

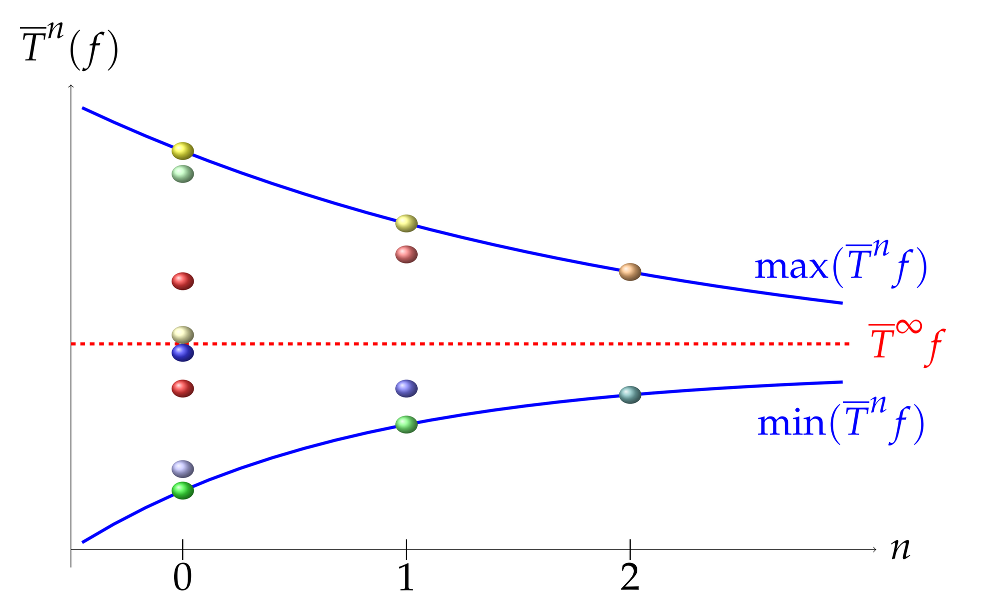
Convergence theorem for regular imprecise Markov chains

We can prove a Perron-Frobenius theorem for the nonlinear operator $\overline{T}$:

Theorem 1 *If the transition operator $\overline{T}$ is regular, then the sequence $\overline{T}^n$, $n \in \mathbb{N}$ converges pointwise to some operator $\overline{T}^\infty$ and for any real-valued function f, $\overline{T}^\infty f = \mu_f$ is some constant. So for the limit distribution we find that*

$$\lim_{n \to \infty} \overline{E}_{\mathcal{P}_n}(f) = \overline{E}_{\mathcal{M}_1}(\overline{T}^\infty f) = \mu_f;$$

so its value is independent of the initial distribution $\mathcal{M}_1$, as in the classical case.



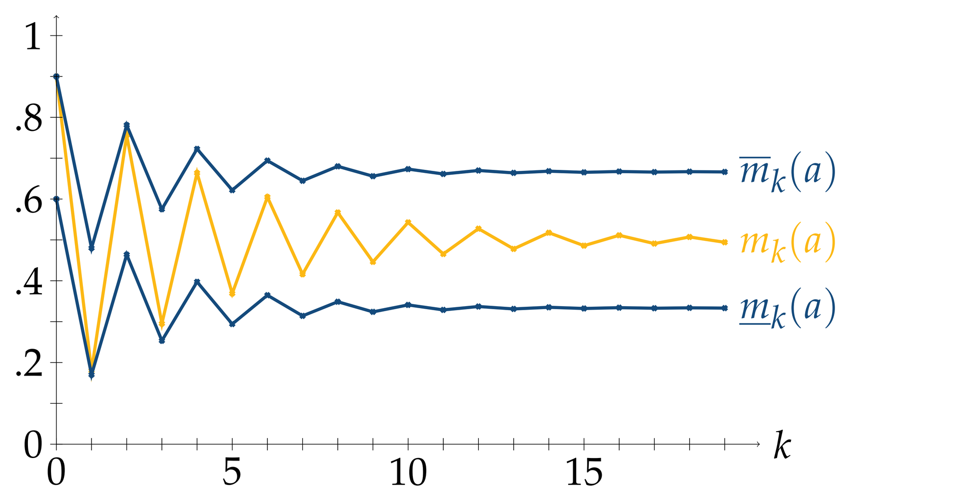
An illustrative result

We have a classical and an imprecise Markov chain defined by

$$m_1(a) = .9 \qquad [\underline{m}_1(a) \ \ \overline{m}_1(a)] = [.6 \ \ .9]$$

$$\begin{bmatrix} q(a|a) \\ q(a|b) \end{bmatrix} = \begin{bmatrix} .1 \\ .9 \end{bmatrix} \qquad \begin{bmatrix} \underline{q}(a|a) & \overline{q}(a|a) \\ \underline{q}(a|b) & \underline{q}(a|b) \end{bmatrix} = \begin{bmatrix} .1 & .2 \\ .8 & .9 \end{bmatrix}$$

This results in the following evolution of the probability mass vectors:



[The data for this plot was generated by a Matlab-program written by a Master's thesis student, Stefaan Dhaenens.]