Imprecise Markov chains

From basic theory to applications I

prof. Gert de Cooman

UTOPIÆ
Uncertainty Treatment and Optimisation in Aerospace Engineering
Handling the unknown at the edge of tomorrow
MARIE CURIE ACTIONS

Imprecise discrete-time Markov chains

Precise probability models

Mass functions and expectations

Assume we are uncertain about:

- the value or a variable X
- in a set of possible values $\mathscr{X}$.

This is usually modelled by a probability mass function p on $\mathscr{X}$:

$$p(x) \geq 0 \text{ and } \sum_{x \in \mathscr{X}} p(x) = 1;$$

With p we can associate an expectation operator E_p:

$$E_p(f) := \sum_{x \in \mathscr{X}} p(x) f(x) \text{ where } f\colon \mathscr{X} \to \mathbb{R}.$$

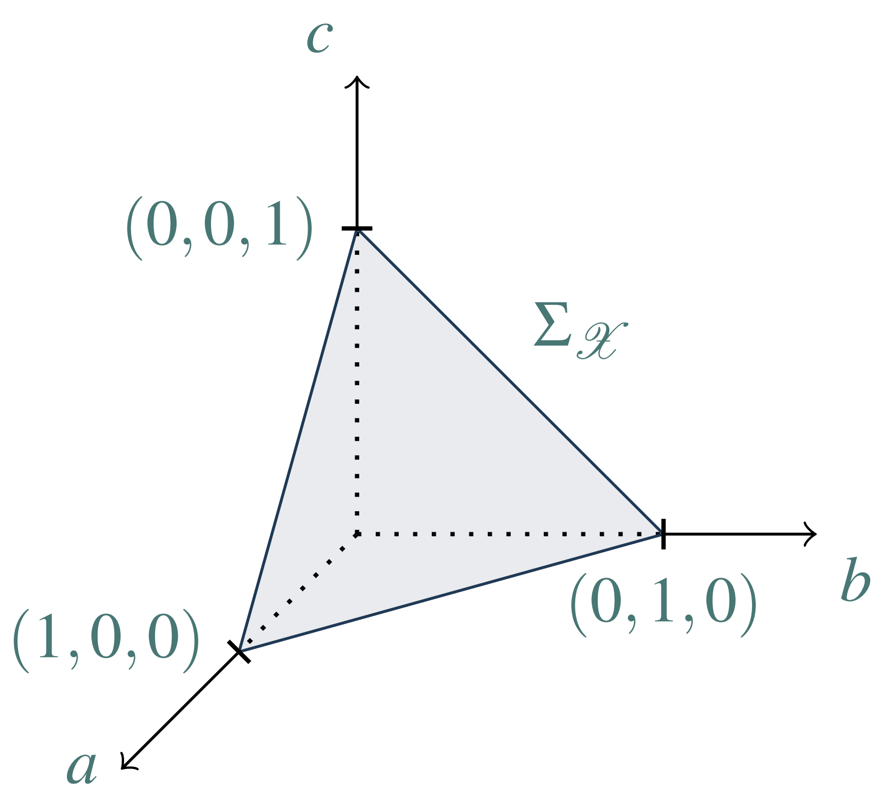
If $A \subseteq \mathscr{X}$ is an event, then its probability is given by

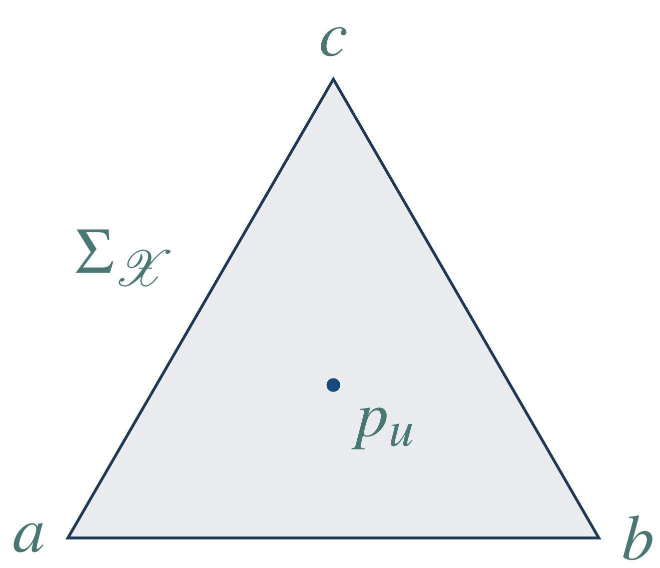
$$P_p(A) = \sum_{x \in A} p(x) = E_p(I_A).$$

The simplex of all probability mass functions

Consider the simplex $\Sigma_{\mathscr{X}}$ of all mass functions on $\mathscr{X}$:

$$\Sigma_{\mathscr{X}} := \left\{ p \in \mathbb{R}_{+}^{\mathscr{X}} : \sum_{x \in \mathscr{X}} p(x) = 1 \right\}.$$





Geometrical interpretation of expectation

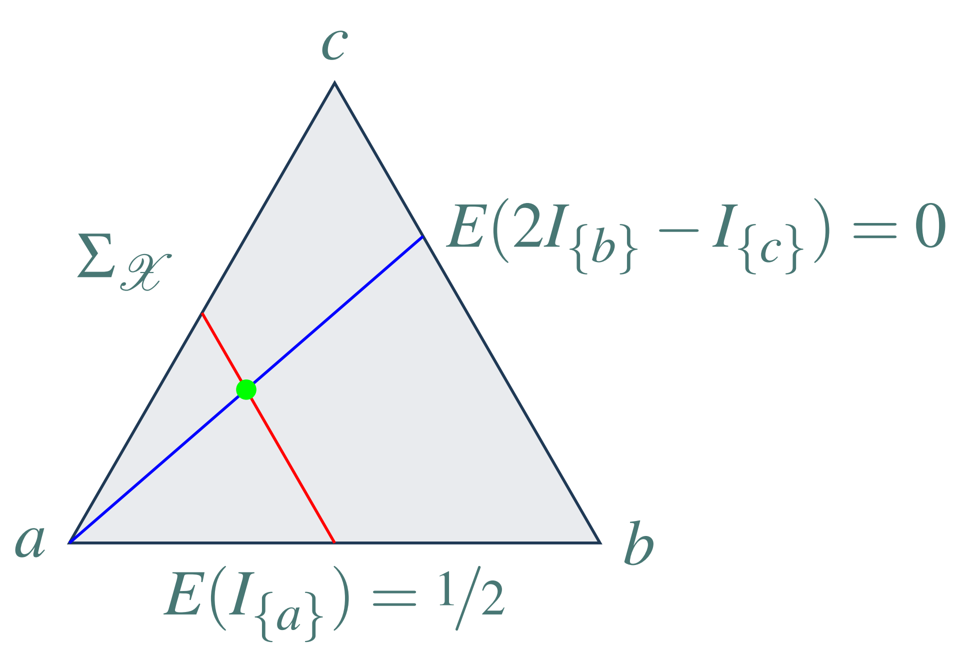
Assessments lead to constraints

Specifying an expectation $E(f)$ for a map $f\colon \mathscr{X} \to \mathbb{R}$

$$\sum_{x \in \mathscr{X}} p(x) f(x) = E(f)$$

imposes a linear constraint on the possible values for p in $\Sigma_{\mathscr{X}}$.

It corresponds to intersecting the simplex $\Sigma_{\mathscr{X}}$ with a hyperplane, whose direction depends on f:



Imprecise probability models

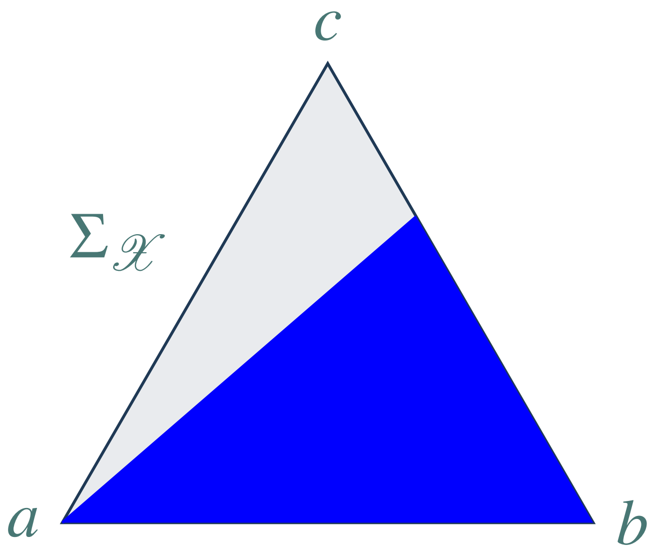
Linear inequality constraints

More flexible assessments

Impose linear inequality constraints on p in $\Sigma_{\mathscr{X}}$:

$$\underline{E}(f) \leq \sum_{x \in \mathscr{X}} p(x) f(x) \quad \text{or} \quad \sum_{x \in \mathscr{X}} p(x) f(x) \leq \overline{E}(f).$$

Corresponds to intersecting $\Sigma_{\mathscr{X}}$ with affine semi-spaces:



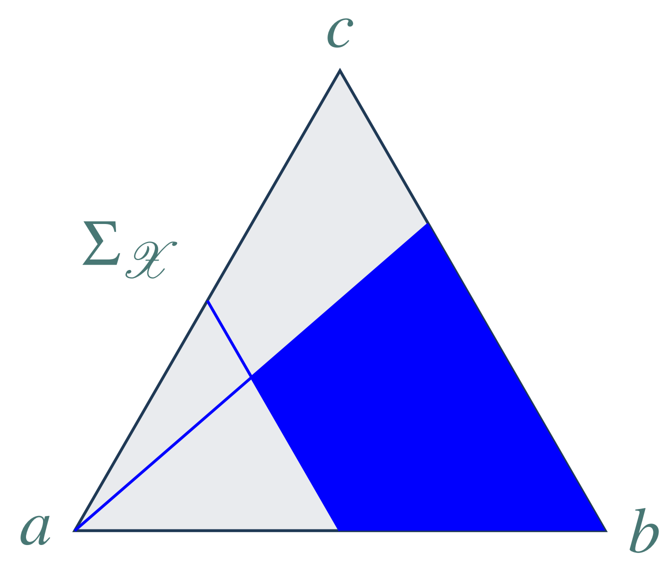
Linear inequality constraints

More flexible assessments

Impose linear inequality constraints on p in $\Sigma_{\mathscr{X}}$:

$$\underline{E}(f) \leq \sum_{x\in\mathscr{X}} p(x)f(x) \quad \text{or} \quad \sum_{x\in\mathscr{X}} p(x)f(x) \leq \overline{E}(f).$$

Corresponds to intersecting $\Sigma_{\mathscr{X}}$ with affine semi-spaces:



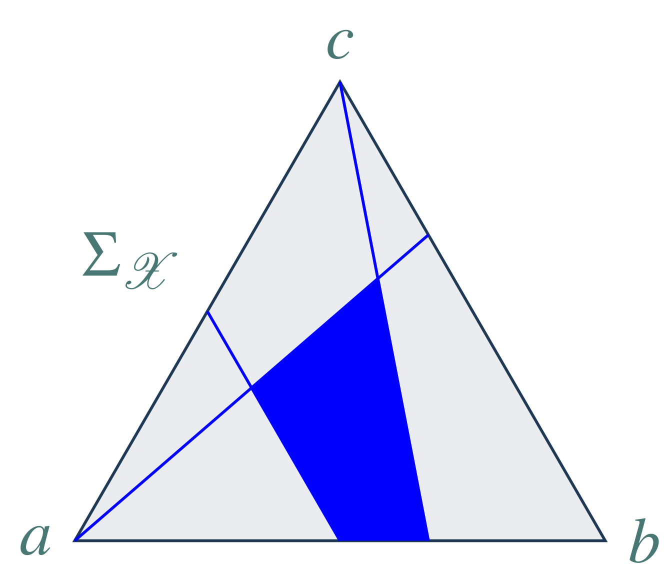
Linear inequality constraints

More flexible assessments

Impose linear inequality constraints on p in $\Sigma_{\mathscr{X}}$:

$$\underline{E}(f) \leq \sum_{x \in \mathscr{X}} p(x) f(x) \quad \text{or} \quad \sum_{x \in \mathscr{X}} p(x) f(x) \leq \overline{E}(f).$$

Corresponds to intersecting $\Sigma_{\mathscr{X}}$ with affine semi-spaces:



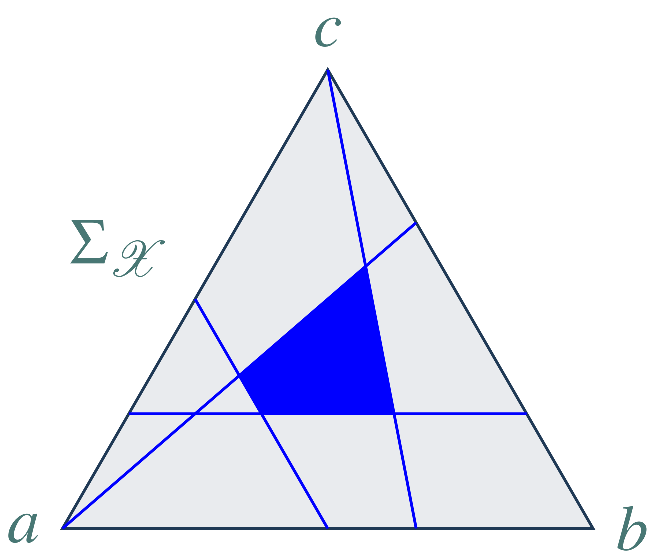
Linear inequality constraints

More flexible assessments

Impose linear inequality constraints on p in $\Sigma_{\mathscr{X}}$:

$$\underline{E}(f) \leq \sum_{x \in \mathscr{X}} p(x)f(x) \quad \text{or} \quad \sum_{x \in \mathscr{X}} p(x)f(x) \leq \overline{E}(f).$$

Corresponds to intersecting $\Sigma_{\mathscr{X}}$ with affine semi-spaces:



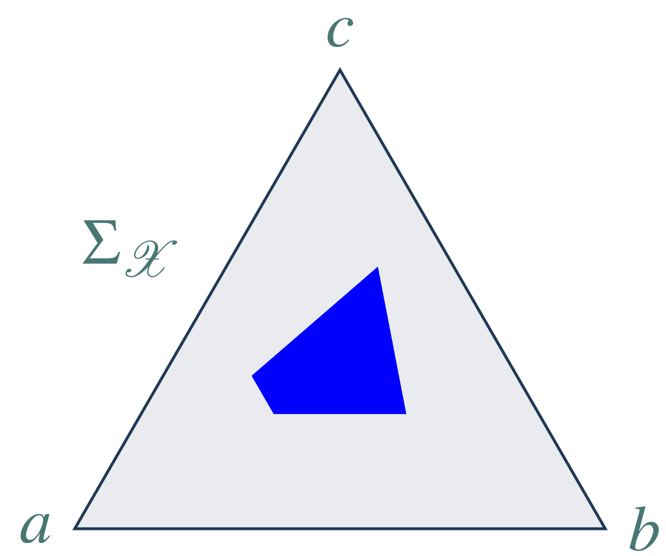
Linear inequality constraints

More flexible assessments

Impose linear inequality constraints on p in $\Sigma_{\mathscr{X}}$:

$$\underline{E}(f) \leq \sum_{x \in \mathscr{X}} p(x) f(x) \quad \text{or} \quad \sum_{x \in \mathscr{X}} p(x) f(x) \leq \overline{E}(f).$$

Corresponds to intersecting $\Sigma_{\mathscr{X}}$ with affine semi-spaces:

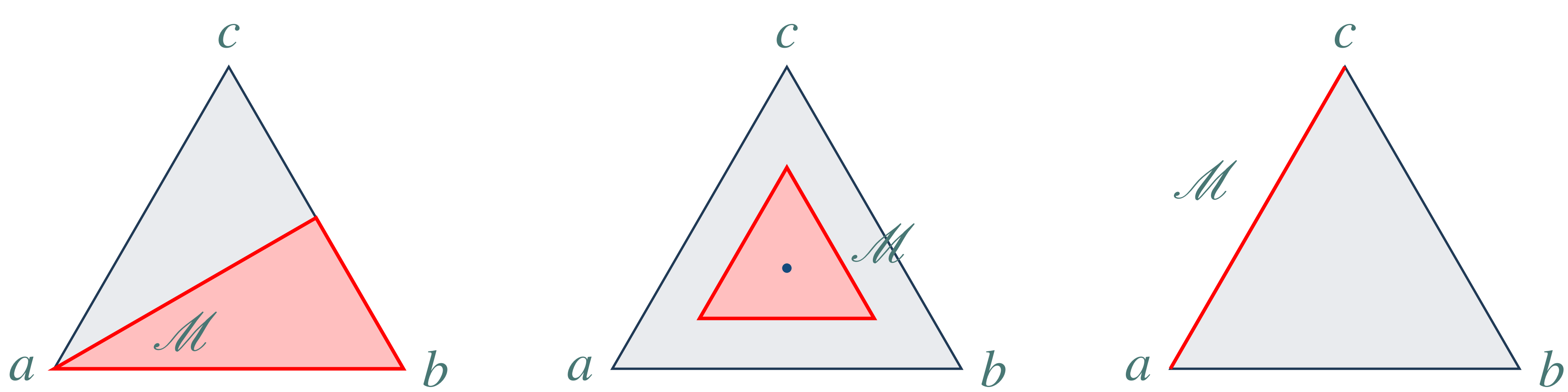


Credal sets

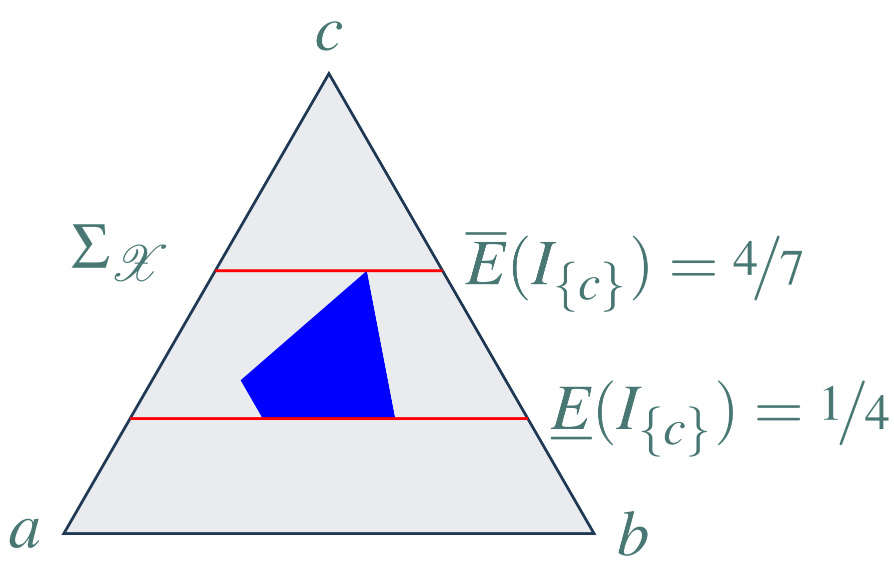
Any such number of assessments leads to a credal set $\mathscr{M}$.

Definition

A credal set $\mathscr{M}$ is a convex closed subset of $\Sigma_{\mathscr{X}}$.



Lower and upper expectations



Equivalent model

Consider the set $\mathscr{L}(\mathscr{X}) = \mathbb{R}^{\mathscr{X}}$ of all real-valued maps on $\mathscr{X}$. We define two real functionals on $\mathscr{L}(\mathscr{X})$: for all $f\colon \mathscr{X} \to \mathbb{R}$

$$\boxed{\begin{array}{l}\underline{E}_{\mathscr{M}}(f) = \min\{E_p(f)\colon p \in \mathscr{M}\} \text{ lower expectation} \\ \overline{E}_{\mathscr{M}}(f) = \max\{E_p(f)\colon p \in \mathscr{M}\} \text{ upper expectation.}\end{array}}$$

Observe that

$$\boxed{\overline{E}_{\mathscr{M}}(f) = -\underline{E}_{\mathscr{M}}(-f).}$$

Basic properties of upper expectations

Definition

We call a real functional $\overline{E}$ on $\mathscr{L}(\mathscr{X})$ an upper expectation if it satisfies the following properties:
for all f and g in $\mathscr{L}(\mathscr{X})$ and all real $\lambda \geq 0$:

1. $\overline{E}(f) \leq \max f$ [boundedness];
2. $\overline{E}(f+g) \leq \overline{E}(f) + \overline{E}(g)$ [sub-additivity];
3. $\overline{E}(\lambda f) = \lambda \overline{E}(f)$ [non-negative homogeneity].

Theorem

A real functional $\overline{E}$ is an upper expectation if and only if it is the upper envelope of some credal set $\mathscr{M}$.

Proof.

Use $\mathscr{M} = \left\{ p \in \Sigma_{\mathscr{X}} : (\forall f \in \mathscr{L}(\mathscr{X}))(E_p(f) \leq \overline{E}(f)) \right\}$. $\square$

Discrete-time uncertain processes

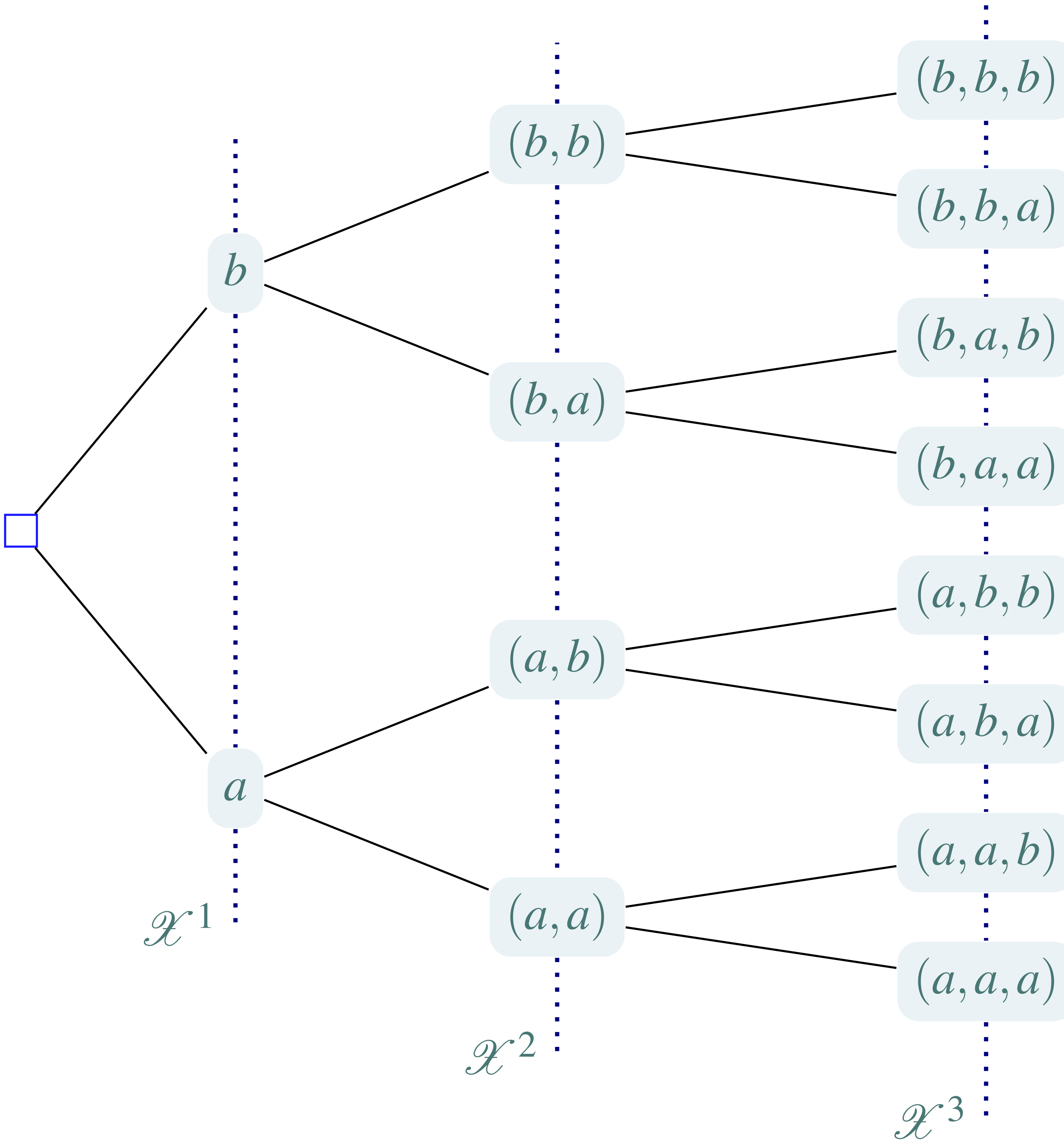
Precise probability trees

We consider an uncertain process with variables $X_1, X_2, \ldots, X_n, \ldots$ assuming values in a finite set of states $\mathscr{X}$.

This leads to a standard event tree with nodes

$$s = (x_1, x_2, \ldots, x_n), \quad x_k \in \mathscr{X}, \quad n \geq 0.$$

Precise probability trees



Precise probability trees

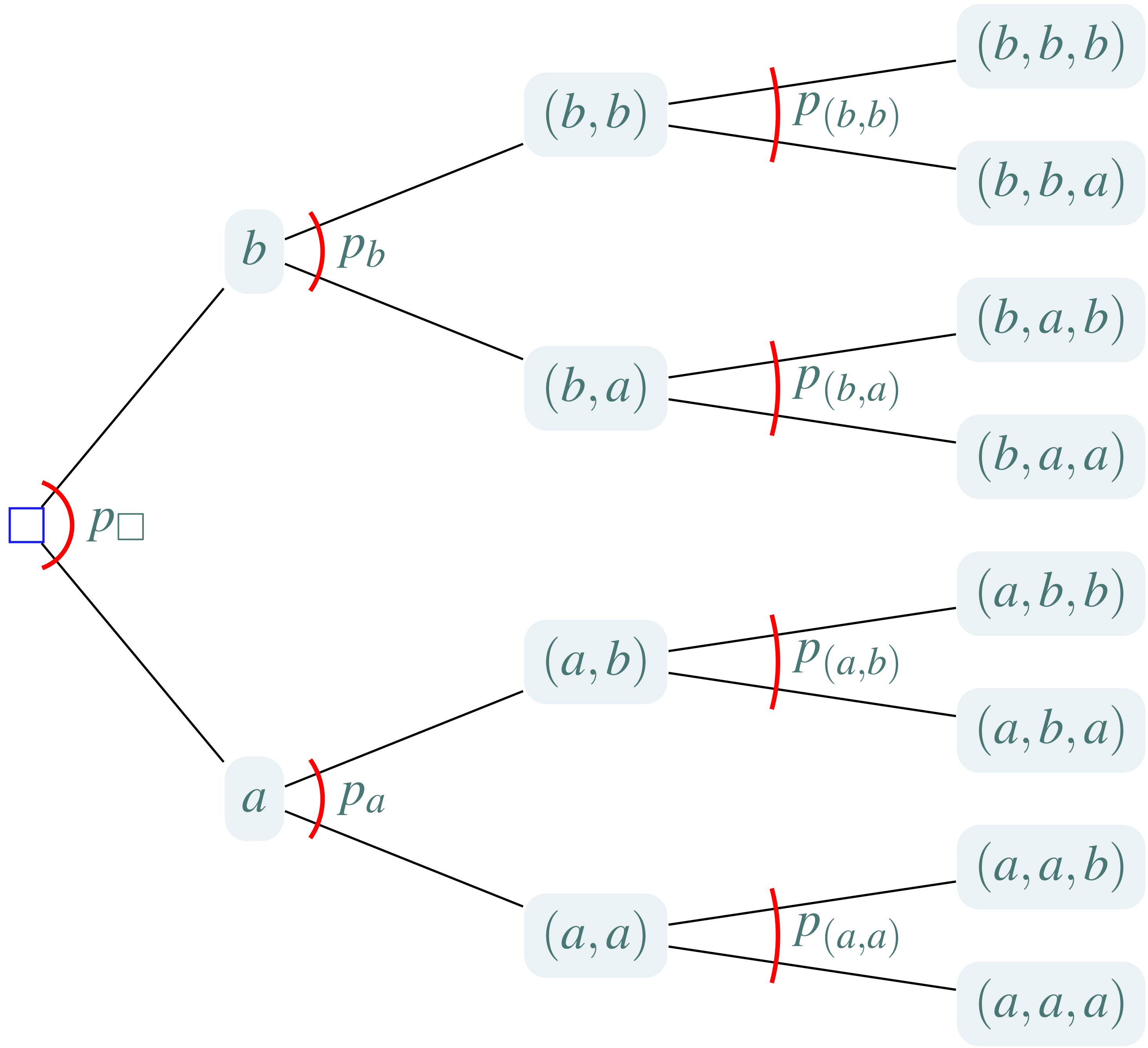
We consider an uncertain process with variables $X_1, X_2, \ldots, X_n, \ldots$ assuming values in a finite set of states $\mathscr{X}$.

This leads to a standard event tree with nodes

$$s = (x_1, x_2, \ldots, x_n), \quad x_k \in \mathscr{X}, \quad n \geq 0.$$

The standard event tree becomes a probability tree by attaching to each node s a local probability mass function p_s on $\mathscr{X}$ with associated expectation operator E_s.

Precise probability trees



Calculating global expectations from local ones

Consider a function $g\colon \mathscr{X}^n \to \mathbb{R}$ of the first n variables:

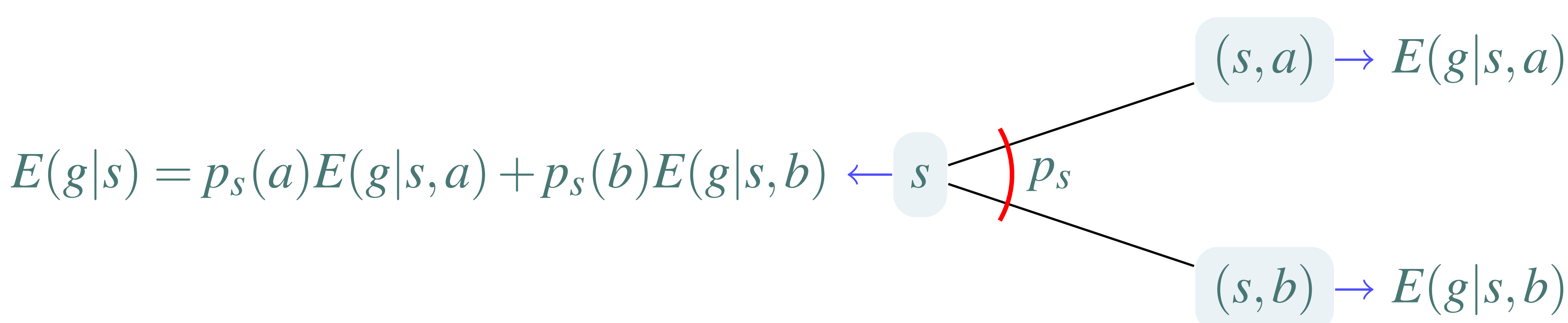
$$g = g(X_1, X_2, \ldots, X_n)$$

We want to calculate its expectation $E(g|s)$ in $s = (x_1, \ldots, x_k)$.

Theorem (Law of Iterated Expectation)

Suppose we know $E(g|s,x)$ for all $x \in \mathscr{X}$, then we can calculate $E(g|s)$ by backwards recursion using the local model p_s:

$$E(g|s) = \underbrace{E_s}_{\text{local}}(E(g|s,\cdot)) = \sum_{x\in\mathscr{X}} p_s(x)E(g|s,x).$$



Calculating global expectations from local ones

All expectations $E(g|x_1,\ldots,x_k)$ in the tree can be calculated from the local models as follows:

1. start in the final cut $\mathscr{X}^n$ and let:

$$E(g|x_1,x_2,\ldots,x_n) = g(x_1,x_2,\ldots,x_n);$$

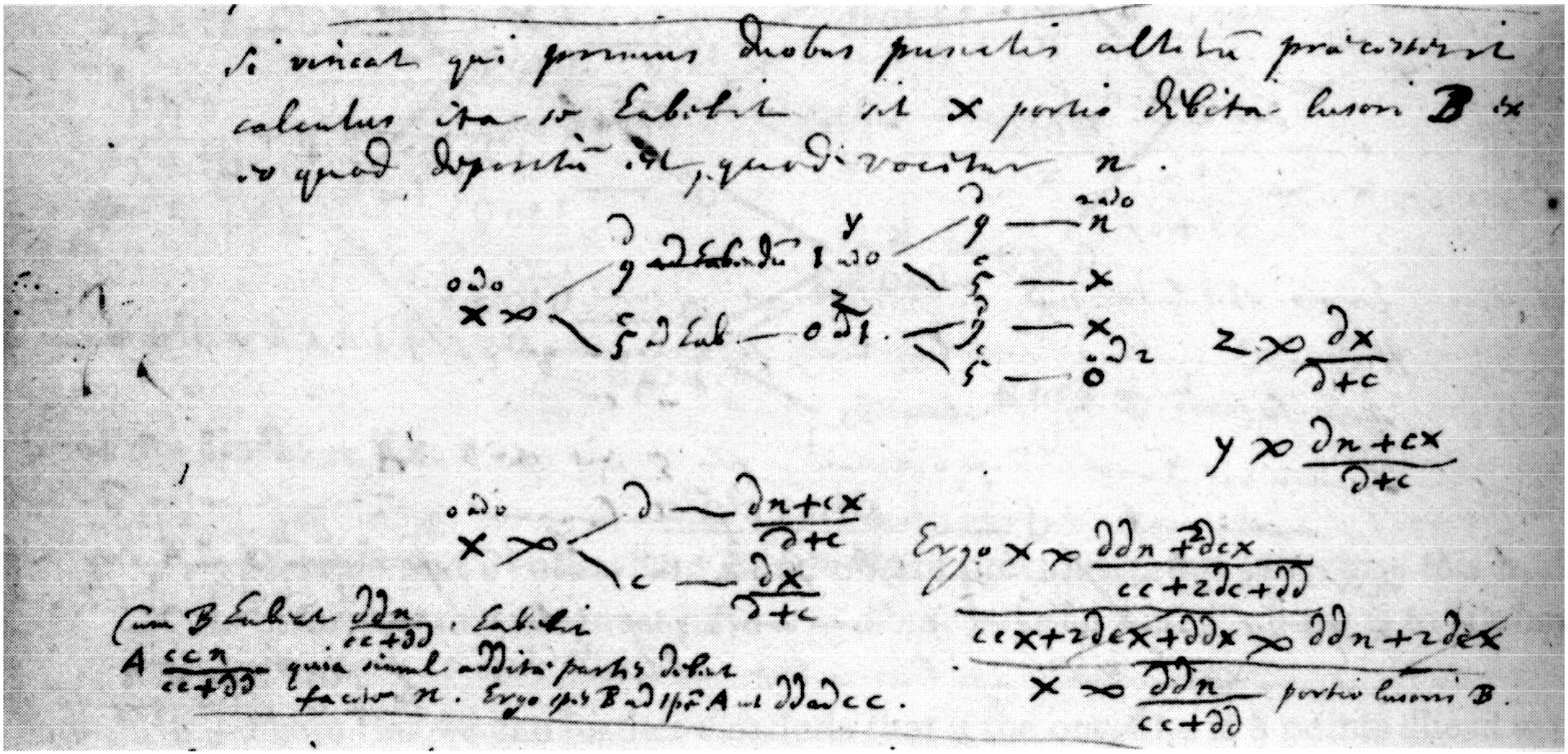
2. do backwards recursion using the Law of Iterated Expectation:

$$E(g|x_1,\ldots,x_k) = \underbrace{E_{(x_1,\ldots,x_k)}}_{\text{local}}(E(g|x_1,\ldots,x_k,\cdot))$$

3. go on until you get to the root node $\square$, where:

$$E(g|\square) = E(g).$$

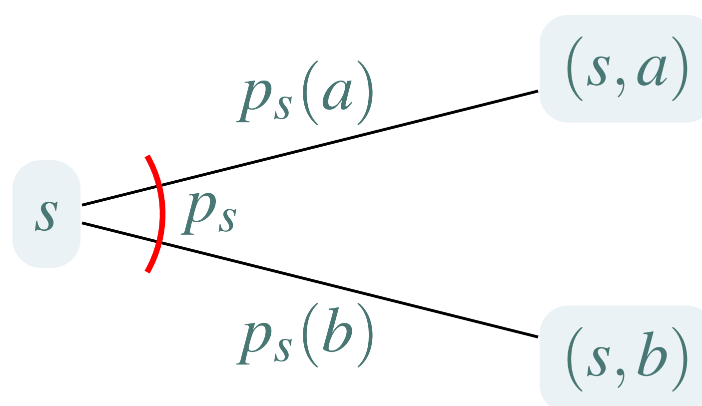
Christiaan Huygens (1656–1657)



Sets of mass functions

Major restrictive assumption

Until now, we have assumed that we have sufficient information in order to specify, in each node s, a probability mass function p_s on the set $\mathscr{X}$ of possible values for the next state.



More general uncertainty models

We consider credal sets as more general uncertainty models: closed convex subsets of $\Sigma_{\mathscr{X}}$.

Imprecise probability trees

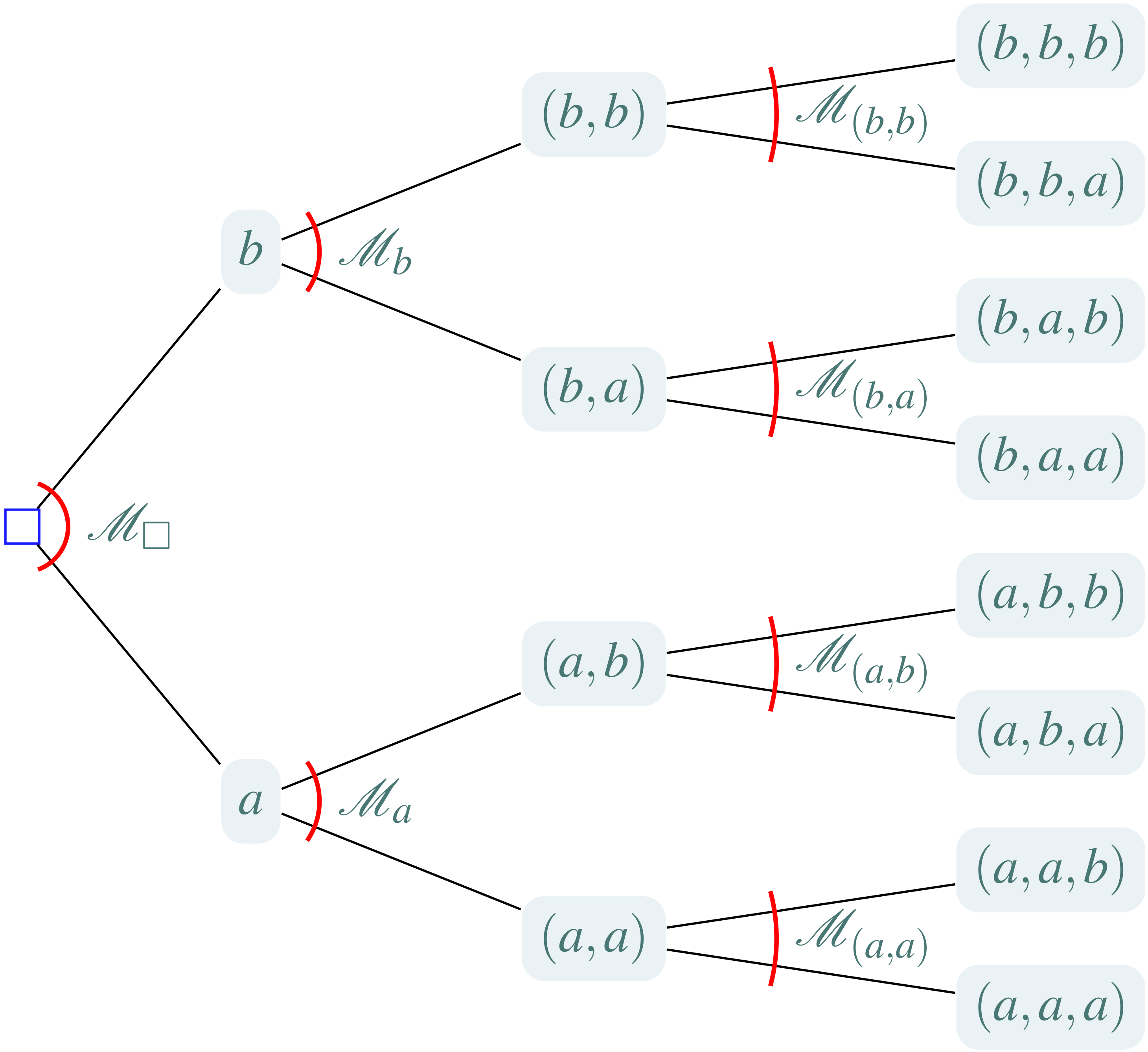
Definition and interpretation

Definition

An imprecise probability tree is a probability tree where in each node s the local uncertainty model is an imprecise probability model $\mathscr{M}_s$, or equivalently, its associated upper expectation $\overline{E}_s$:

$$\overline{E}_s(f) = \max\{E_p(f) \colon p \in \mathscr{M}_s\} \text{ for all real maps } f \text{ on } \mathscr{X}.$$

Definition and interpretation



Definition and interpretation

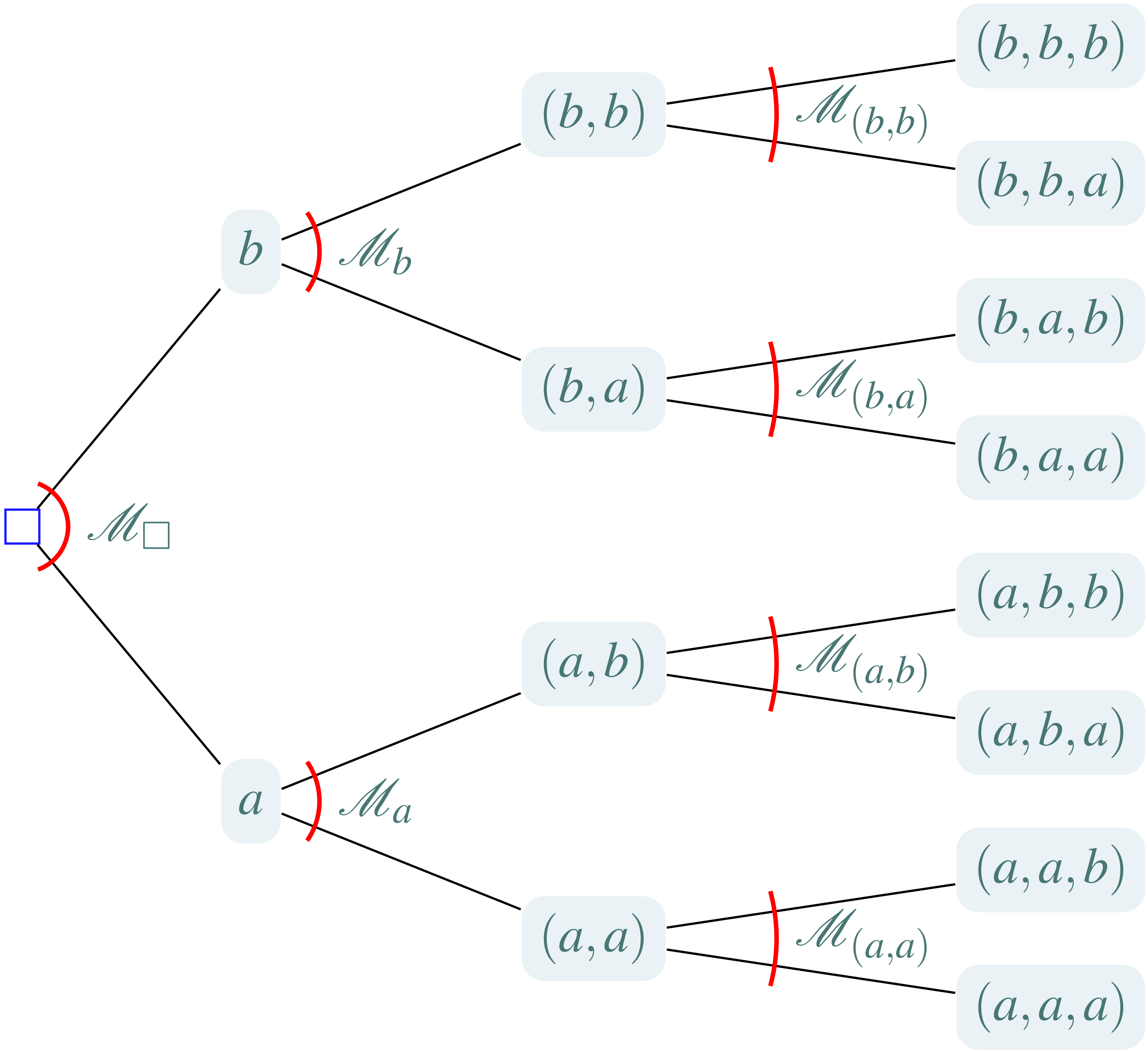
Definition

An imprecise probability tree is a probability tree where in each node s the local uncertainty model is an imprecise probability model $\mathscr{M}_s$, or equivalently, its associated upper expectation $\overline{E}_s$:

$$\overline{E}_s(f) = \max\{E_p(f) \colon p \in \mathscr{M}_s\} \text{ for all real maps } f \text{ on } \mathscr{X}.$$

An imprecise probability tree can be seen as an infinity of compatible precise probability trees: choose in each node s a probability mass function p_s from the set $\mathscr{M}_s$.

Definition and interpretation



Associated lower and upper expectations

For each real map $g = g(X_1, \ldots, X_n)$, each node $s = (x_1, \ldots, x_k)$, and each such compatible precise probability tree, we can calculate the expectation

$$E(g|x_1, \ldots, x_k)$$

using the backwards recursion method described before.

By varying over each compatible probability tree, we get a closed real interval:

$$[\underline{E}(g|x_1, \ldots, x_k), \overline{E}(g|x_1, \ldots, x_k)]$$

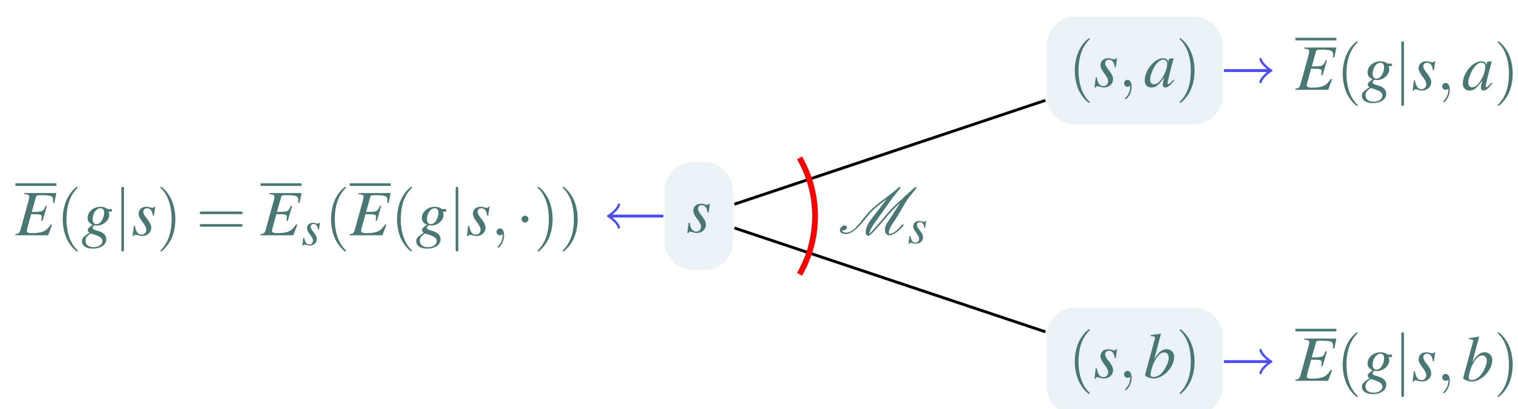
We want a better, more efficient method to calculate these lower and upper expectations $\underline{E}(g|x_1, \ldots x_k)$ and $\overline{E}(g|x_1, \ldots, x_k)$.

The Law of Iterated Expectation

Theorem (Law of Iterated Expectation)

Suppose we know $\overline{E}(g|s,x)$ *for all* $x \in \mathscr{X}$ *, then we can calculate* $\overline{E}(g|s)$ *by backwards recursion using the local model* $\overline{E}_s$*:*

$$\overline{E}(g|s) = \underbrace{\overline{E}_s}_{\text{local}}(\overline{E}(g|s,\cdot)) = \max_{p_s \in \mathscr{M}_s} \sum_{x \in \mathscr{X}} p_s(x)\,\overline{E}(g|s,x).$$



The complexity of calculating the $\overline{E}(g|s)$, as a function of n, is therefore essentially the same as in the precise case!

Imprecise Markov chains

Precise Markov chains: definition

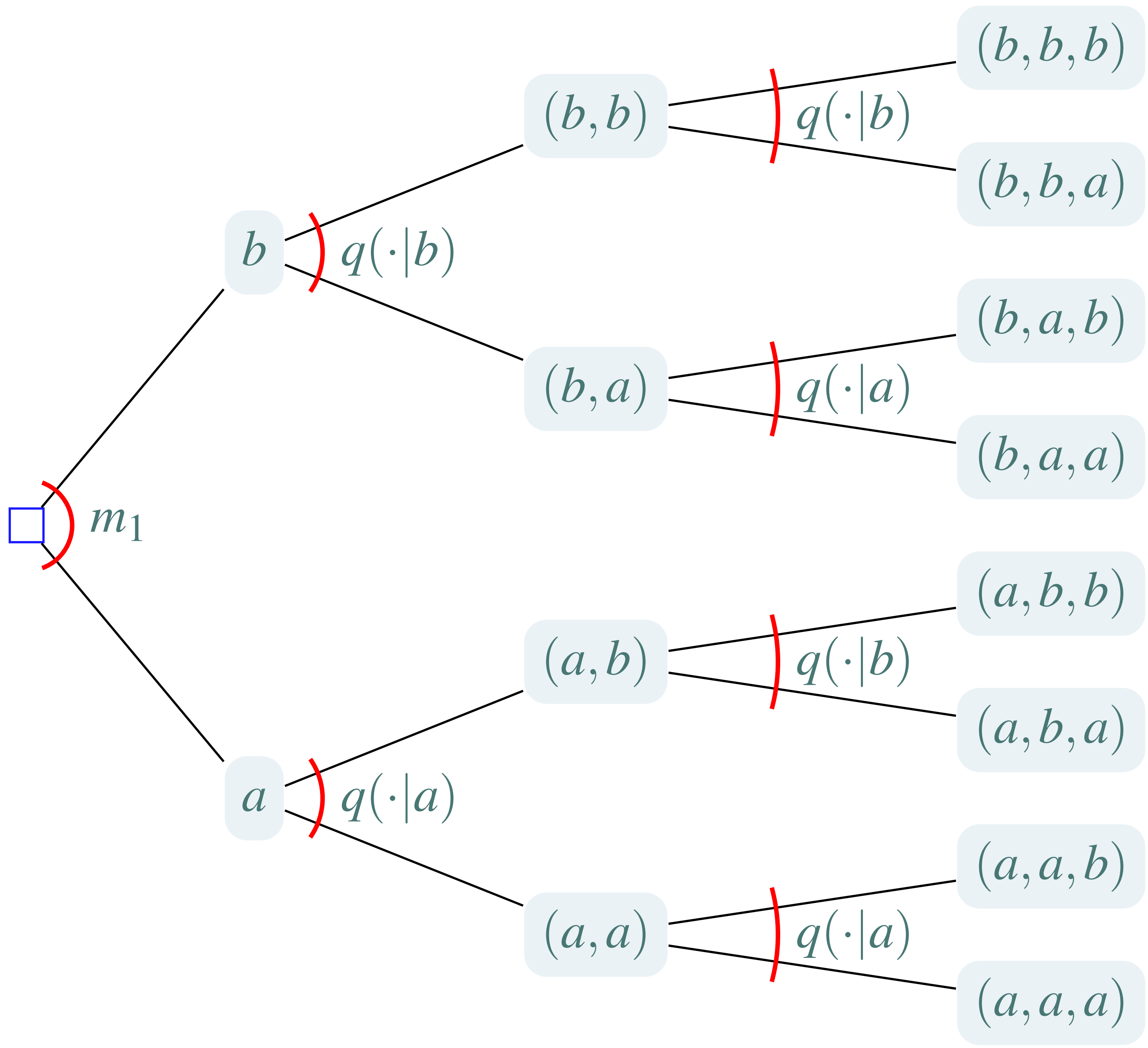
Definition

The uncertain process is a stationary precise Markov chain when all $\mathscr{M}_s$ are singletons (precise), and

1. $\mathscr{M}_\square = \{m_1\}$,
2. the Markov Condition is satisfied:

$$\boxed{\mathscr{M}_{(x_1,\ldots,x_n)} = \{q(\cdot|x_n)\}.}$$

Precise Markov chains: definition



Precise Markov chains: definition

Definition

The uncertain process is a stationary precise Markov chain when all $\mathscr{M}_s$ are singletons (precise), and

1. $\mathscr{M}_\square = \{m_1\}$,
2. the Markov Condition is satisfied:

$$\boxed{\mathscr{M}_{(x_1,\dots,x_n)} = \{q(\cdot|x_n)\}.}$$

For each $x \in \mathscr{X}$, the transition mass function $q(\cdot|x)$ corresponds to an expectation operator:

$$E(f|x) = \sum_{z\in\mathscr{X}} q(z|x)f(z).$$

Precise Markov chains: transition operators

Definition

Consider the linear transformation T of $\mathscr{L}(\mathscr{X})$, called transition operator:

$$\mathrm{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \mathrm{T}f$$

where $\mathrm{T}f$ is the real map given by, for any $x \in \mathscr{X}$:

$$\boxed{\mathrm{T}f(x) := E(f|x) = \sum_{z \in \mathscr{X}} q(z|x)f(z)}$$

T is the dual of the linear transformation with Markov matrix M, with elements $M_{xy} := q(y|x)$.

Precise Markov chains: transition operators

Definition

Consider the linear transformation T of $\mathscr{L}(\mathscr{X})$, called transition operator:

$$\mathrm{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \mathrm{T}f$$

where $\mathrm{T}f$ is the real map given by, for any $x \in \mathscr{X}$:

$$\boxed{\mathrm{T}f(x) := E(f|x) = \sum_{z\in\mathscr{X}} q(z|x)f(z)}$$

T is the dual of the linear transformation with Markov matrix M, with elements $M_{xy} := q(y|x)$.

Then the Law of Iterated Expectation yields:

$$\boxed{E_n(f) = E_1(\mathrm{T}^{n-1}f), \text{ and dually, } m_n = M^{n-1}m_1.}$$

Complexity is linear in the number of time steps n.

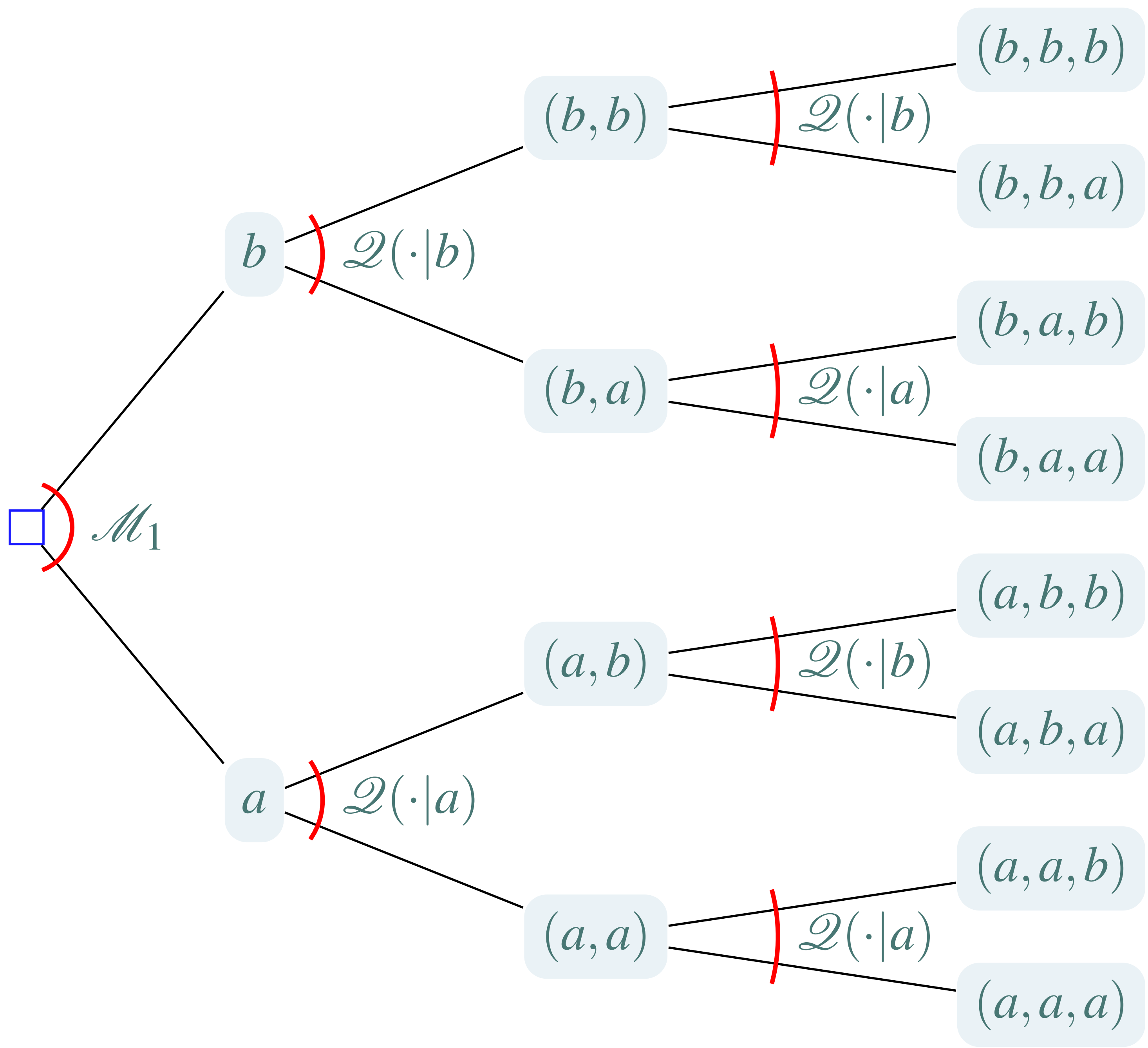
Imprecise Markov chains: definition

Definition

The uncertain process is a stationary imprecise Markov chain when the Markov Condition is satisfied:

$$\boxed{\mathscr{M}_{(x_1,\dots,x_n)} = \mathscr{Q}(\cdot|x_n).}$$

Imprecise Markov chains: definition



Imprecise Markov chains: definition

Definition

The uncertain process is a stationary imprecise Markov chain when the Markov Condition is satisfied:

$$\boxed{\mathscr{M}_{(x_1,\ldots,x_n)} = \mathscr{Q}(\cdot|x_n).}$$

An imprecise Markov chain can be seen as an infinity of probability trees.

For each $x \in \mathscr{X}$, the local transition model $\mathscr{Q}(\cdot|x)$ corresponds to lower and upper expectation operators:

$$\underline{E}(f|x) = \min\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\}$$
$$\overline{E}(f|x) = \max\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\}.$$

Imprecise Markov chains: transition operators

Definition

Consider the non-linear transformations $\underline{\mathrm{T}}$ and $\overline{\mathrm{T}}$ of $\mathscr{L}(\mathscr{X})$, called lower and upper transition operators:

$$\underline{\mathrm{T}} \colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \underline{\mathrm{T}}f$$
$$\overline{\mathrm{T}} \colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \overline{\mathrm{T}}f$$

where the real maps $\underline{\mathrm{T}}f$ and $\overline{\mathrm{T}}f$ are given by:

$$\boxed{\begin{aligned}\underline{\mathrm{T}}f(x) &:= \underline{E}(f|x) = \min\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\}\\ \overline{\mathrm{T}}f(x) &:= \overline{E}(f|x) = \max\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\}\end{aligned}}$$

Imprecise Markov chains: transition operators

Definition

Consider the non-linear transformations $\underline{\mathrm{T}}$ and $\overline{\mathrm{T}}$ of $\mathscr{L}(\mathscr{X})$, called lower and upper transition operators:

$$\underline{\mathrm{T}}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \underline{\mathrm{T}}f$$
$$\overline{\mathrm{T}}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \overline{\mathrm{T}}f$$

where the real maps $\underline{\mathrm{T}}f$ and $\overline{\mathrm{T}}f$ are given by:

$$\boxed{\begin{aligned} \underline{\mathrm{T}}f(x) &:= \underline{E}(f|x) = \min\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\} \\ \overline{\mathrm{T}}f(x) &:= \overline{E}(f|x) = \max\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\} \end{aligned}}$$

Then the Law of Iterated Expectation yields:

$$\boxed{\underline{E}_n(f) = \underline{E}_1(\underline{\mathrm{T}}^{n-1}f) \text{ and } \overline{E}_n(f) = \overline{E}_1(\overline{\mathrm{T}}^{n-1}f).}$$

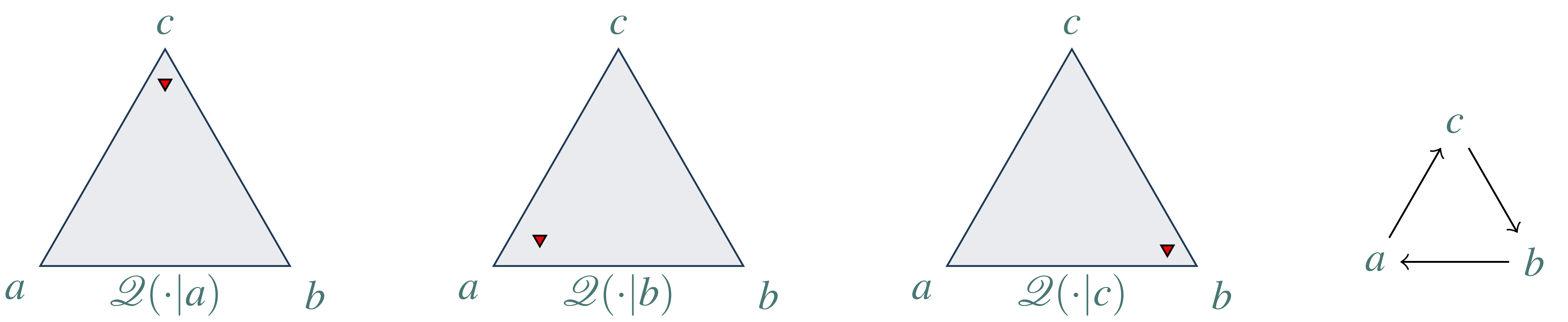
Complexity is still linear in the number of time steps n.

Lower and upper mass functions

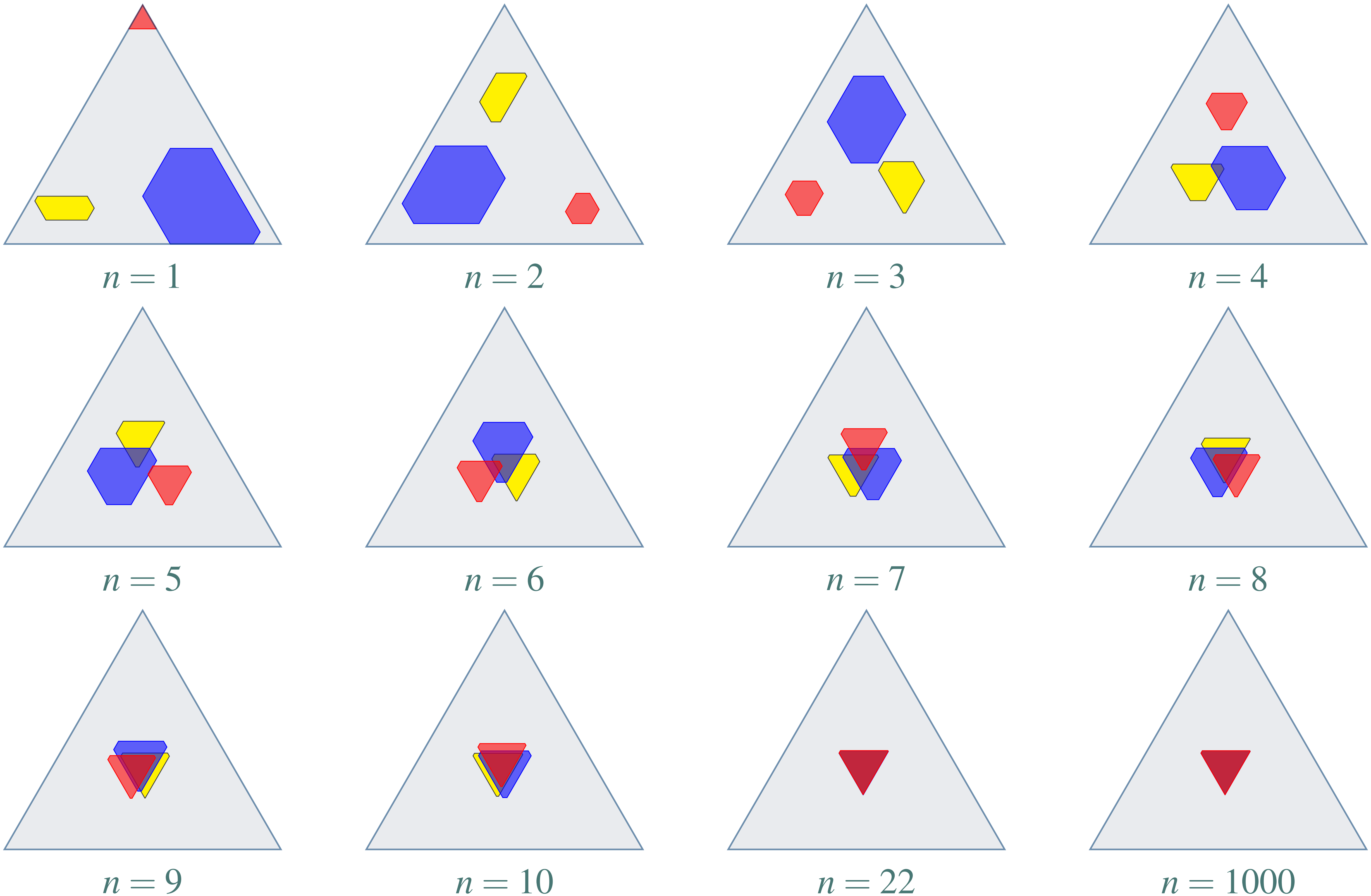
Another example with $\mathscr{X} = \{a, b, c\}$

$$\begin{bmatrix}\underline{\mathrm{T}}I_{\{a\}} & \underline{\mathrm{T}}I_{\{b\}} & \underline{\mathrm{T}}I_{\{c\}}\end{bmatrix} = \begin{bmatrix}\underline{q}(a|a) & \underline{q}(b|a) & \underline{q}(c|a)\\ \underline{q}(a|b) & \underline{q}(b|b) & \underline{q}(c|b)\\ \underline{q}(a|c) & \underline{q}(b|c) & \underline{q}(c|c)\end{bmatrix} = {}^{1}\!/_{200}\begin{bmatrix}9 & 9 & 162\\ 144 & 18 & 18\\ 9 & 162 & 9\end{bmatrix}$$

$$\begin{bmatrix}\overline{\mathrm{T}}I_{\{a\}} & \overline{\mathrm{T}}I_{\{b\}} & \overline{\mathrm{T}}I_{\{c\}}\end{bmatrix} = \begin{bmatrix}\overline{q}(a|a) & \overline{q}(b|a) & \overline{q}(c|a)\\ \overline{q}(a|b) & \overline{q}(b|b) & \overline{q}(c|b)\\ \overline{q}(a|c) & \overline{q}(b|c) & \overline{q}(c|c)\end{bmatrix} = {}^{1}\!/_{200}\begin{bmatrix}19 & 19 & 172\\ 154 & 28 & 28\\ 19 & 172 & 19\end{bmatrix}$$



Another example with $\mathscr{X} = \{a, b, c\}$



A non-linear Perron–Frobenius Theorem

Generalising the linear case

Theorem (De Cooman, Hermans and Quaeghebeur, 2008)

Consider a stationary imprecise Markov chain with finite state set $\mathscr{X}$ *and an upper transition operator* $\overline{\mathrm{T}}$*. Suppose that* $\overline{\mathrm{T}}$ *is* regular*, meaning that there is some* $n > 0$ *such that* $\min \overline{\mathrm{T}}^n I_{\{x\}} > 0$ *for all* $x \in \mathscr{X}$*. Then for every initial upper expectation* $\overline{E}_1$*, the upper expectation* $\overline{E}_n = \overline{E}_1 \circ \overline{\mathrm{T}}^{n-1}$ *for the state at time* n *converges point-wise to the same upper expectation* $\overline{E}_\infty$*:*

$$\boxed{\lim_{n\to\infty} \overline{E}_n(h) = \lim_{n\to\infty} \overline{E}_1(\overline{\mathrm{T}}^{n-1} h) := \overline{E}_\infty(h)}$$

for all h *in* $\mathscr{L}(\mathscr{X})$*. Moreover, the corresponding limit upper expectation* $\overline{E}_\infty$ *is the* only $\overline{\mathrm{T}}$-invariant *upper expectation on* $\mathscr{L}(\mathscr{X})$*, meaning that* $\overline{E}_\infty = \overline{E}_\infty \circ \overline{\mathrm{T}}$*.*

Ergodicity

And in that case also

Theorem (De Cooman, De Bock and Lopatatzidis, 2016)

Consider a stationary imprecise Markov chain with finite state set $\mathscr{X}$ *and an upper transition operator* $\overline{\mathrm{T}}$*. Suppose that* $\overline{\mathrm{T}}$ *is Perron–Frobenius-like with stationary upper expectation* $\overline{E}_\infty$*. Then*

$$\boxed{\underline{E}_\infty(h) \leq \liminf_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} h(X_k) \leq \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} h(X_k) \leq \overline{E}_\infty(h)}$$

almost surely for all h *in* $\mathscr{L}(\mathscr{X})$*.*

Literature

Literature

Gert de Cooman and Filip Hermans.
Imprecise probability trees: bridging two theories of imprecise probability.
Artificial Intelligence, 172:1400–1427, 2008.

Gert de Cooman, Filip Hermans, and Erik Quaeghebeur.
Imprecise Markov chains and their limit behaviour.
Probability in the Engineering and Informational Sciences, 23:597–635, 2009.

Gert de Cooman, Jasper De Bock, and Stavros Lopatatzidis.
Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems.
International Journal of Approximate Reasoning, 76:18–46, 2016.

Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors.
Introduction to Imprecise Probabilities.
John Wiley & Sons, 2014.