

# Chapter 4

## Imprecise Discrete-Time Markov Chains



Gert de Cooman

**Abstract** I present a short and easy introduction to a number of basic definitions and important results from the theory of imprecise Markov chains in discrete time, with a finite state space. The approach is intuitive and graphical.

### 4.1 Introduction

Although imprecision and robustness in discrete-time Markov chains were already studied in the 1990s [6–8], more significant progress [2, 3, 5, 11] could be made after the graphical structure of imprecise probability trees underlying them was uncovered in 2008 [4]. Research has now moved firmly into the continuous-time domain, for which [1, 9] are good starting points.

In this paper, I give a concise and elementary overview of a number of basic ideas and results in discrete-time imprecise Markov chains, with an emphasis on their graphical representation. We begin with the basics of precise and imprecise probability models in Sects. 4.2 and 4.3. When such models are used in a dynamical context, precise and imprecise probability trees arise naturally; they and the use of the fundamental Law of Iterated Expectations for making inferences about them constitute the subjects of Sects. 4.4 and 4.5. Imprecise Markov chains correspond to special imprecise probability trees, and they and their basic inferences are discussed in Sect. 4.6, followed by a number of examples in Sect. 4.7. These examples hint at stationary distributions and ergodicity. These notions are briefly discussed in Sect. 4.8, which concludes the paper. Throughout, I have included a number of simple exercises to illustrate the arguments in the main text.

---

G. de Cooman (✉)

Foundations Lab for imprecise probabilities, Ghent University, Ghent, Belgium

e-mail: [gert.decooman@ugent.be](mailto:gert.decooman@ugent.be)

© The Author(s) 2022

L. Aslett et al. (eds.), *Uncertainty in Engineering*,  
SpringerBriefs in Statistics,

[https://doi.org/10.1007/978-3-030-83640-5\\_4](https://doi.org/10.1007/978-3-030-83640-5_4)

### 4.2 Precise Probability Models

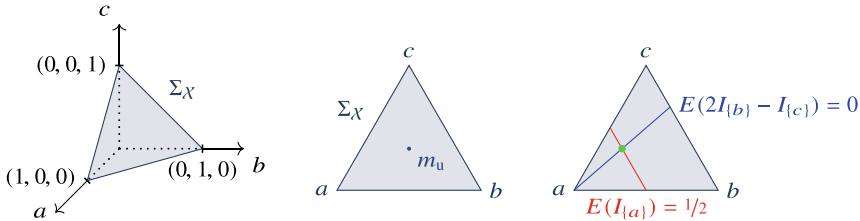
Assume we are uncertain about the value that a variable  $X$  assumes in some finite set of possible values  $\mathcal{X}$ . This is usually modelled by a *probability mass function*  $m$  on  $\mathcal{X}$ , satisfying  $(\forall x \in \mathcal{X})m(x) \geq 0$  and  $\sum_{x \in \mathcal{X}} m(x) = 1$ .

With  $m$  we can associate an *expectation operator*  $E_m$  as follows

$$E_m(f) := \sum_{x \in \mathcal{X}} m(x)f(x) \text{ where } f: \mathcal{X} \rightarrow \mathbb{R}.$$

If  $A \subseteq \mathcal{X}$  is an *event*, then its *probability* is given by  $P_m(A) = \sum_{x \in A} m(x) = E_m(I_A)$ , where  $I_A: \mathcal{X} \rightarrow \mathbb{R}$  is the *indicator* of  $A$  and assumes the value 1 on  $A$  and 0 elsewhere. This tells us that there are two equivalent mathematical languages for dealing with uncertainty: the language of probabilities and the language of expectations, and that we can go freely from one to the other.

All possible (precise) probability models are gathered in the *simplex*  $\Sigma_{\mathcal{X}}$  of all mass functions on  $\mathcal{X}$ :  $\Sigma_{\mathcal{X}} := \{m \in \mathbb{R}^{\mathcal{X}} : (\forall x \in \mathcal{X})m(x) \geq 0 \text{ and } \sum_{x \in \mathcal{X}} m(x) = 1\}$ . Any probability model for uncertainty about  $X$  is a point in that simplex, which indicated that mass functions have a geometrical interpretation. This is illustrated below for the case  $\mathcal{X} = \{a, b, c\}$  and the uniform mass function  $m_u$ .



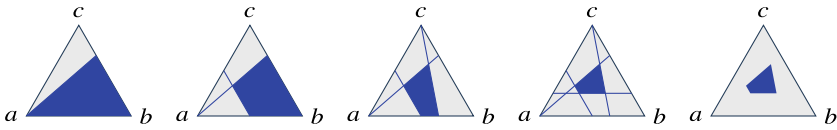
Expectation also has a geometrical interpretation: specifying a value  $E(f)$  for the expectation of a map  $f: \mathcal{X} \rightarrow \mathbb{R}$ , namely,  $\sum_{x \in \mathcal{X}} m(x)f(x) = E(f)$ , imposes a linear constraint on the possible values for  $m$  in  $\Sigma_{\mathcal{X}}$ . It corresponds to intersecting the simplex  $\Sigma_{\mathcal{X}}$  with a hyperplane, whose direction depends on  $f$ . This is also illustrated in the picture above; in this particular case two assessments turn out to completely determine a unique mass function.

### 4.3 Imprecise Probability Models

We now turn to a generalisation of precise probability models, which we will call imprecise. To allow for more realistic and flexible assessments, we can envisage imposing linear inequality—rather than equality—constraints on the  $m$  in  $\Sigma_{\mathcal{X}}$ :

$$\underline{E}(f) \leq \sum_{x \in \mathcal{X}} m(x)f(x) \text{ or } \sum_{x \in \mathcal{X}} m(x)f(x) \leq \bar{E}(f).$$

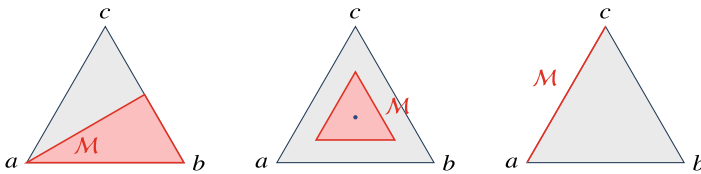
This corresponds to intersecting  $\Sigma_{\mathcal{X}}$  with affine semi-spaces:



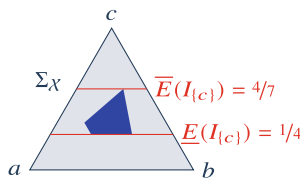
Any such number of assessments leads to a *credal set*  $\mathcal{M}$ , which is our first type of imprecise probability model.

**Definition 4.1** A credal set  $\mathcal{M}$  is a convex closed subset of  $\Sigma_{\mathcal{X}}$ .

Below, we show some more examples of such credal sets in the special case  $\mathcal{X} = \{a, b, c\}$ . The credal set on the left corresponds to the assessment: ‘ $b$  is at least as likely as  $c$ ’; the one in the middle is a convex mixture of the uniform mass function with the entire simplex; and the one on the right represents a statement in classical propositional logic: ‘ $X = a$  or  $X = c$ ’. This illustrates that the language of credal sets encompasses both precise probabilities and classical propositional logic.



Lower and upper expectations are our second type of imprecise probability model. To see how they come about, consider the credal set in the figure below on the right.



We can ask what we know about the probability of  $c$ , or the expectation of  $I_{\{c\}}$ , given this credal set: it is only known to belong to the closed interval  $[1/4, 4/7]$ . This can be generalised from events to arbitrary elements of the set  $\mathcal{L}(\mathcal{X}) = \mathbb{R}^{\mathcal{X}}$  of all real-valued maps  $f$  on  $\mathcal{X}$ : As  $m$  ranges over the credal set  $\mathcal{M}$ ,  $E_m(f)$  will similarly range over a closed interval that is completely determined by its lower and upper bounds.

This leads to the definition of the following two real functionals on  $\mathcal{L}(\mathcal{X})$ :

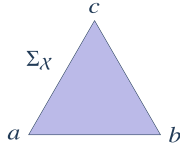
$$\begin{aligned} \underline{E}_{\mathcal{M}}(f) &= \min \{E_m(f) : m \in \mathcal{M}\} \text{ lower expectation} \\ \overline{E}_{\mathcal{M}}(f) &= \max \{E_m(f) : m \in \mathcal{M}\} \text{ upper expectation} \end{aligned} \quad \text{for all } f : \mathcal{X} \rightarrow \mathbb{R}.$$

Observe that these lower and upper expectations are mathematically equivalent models, because

$$\overline{E}_{\mathcal{M}}(f) = -\underline{E}_{\mathcal{M}}(-f) \text{ for all } f \in \mathcal{L}(\mathcal{X}).$$

We will in what follows focus on upper expectations.

**Exercise 4.1** What is the upper expectation  $\overline{E}_{\mathcal{M}}$  when  $\mathcal{M} = \Sigma_{\mathcal{X}}$ ?



Solution:  $\overline{E}_{\mathcal{M}}(f) = \max f$ .

This shows that we can go from the language of probabilities—and the use of  $\mathcal{M}$ —to the language of expectations—and the use of  $\overline{E}_{\mathcal{M}}$ . To see that we can also go the other way, we need the following definition:

**Definition 4.2** We call a real functional  $\overline{E}$  on  $\mathcal{L}(\mathcal{X})$  an *upper expectation* if it satisfies the following properties: for all  $f$  and  $g$  in  $\mathcal{L}(\mathcal{X})$  and all real  $\lambda \geq 0$ :

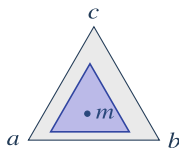
1.  $\overline{E}(f) \leq \max f$  [*boundedness*];
2.  $\overline{E}(f + g) \leq \overline{E}(f) + \overline{E}(g)$  [*sub-additivity*];
3.  $\overline{E}(\lambda f) = \lambda \overline{E}(f)$  [*non-negative homogeneity*].

Upper expectations are also called *coherent upper previsions* [10, 12]. They constitute a model that is mathematically equivalent to credal sets, in very much the same way as expectations are mathematically equivalent to probability mass functions:

**Theorem 4.1** A real functional  $\overline{E}$  is an upper expectation if and only if it is the upper envelope of some credal set  $\mathcal{M}$ .

**Proof** Use  $\mathcal{M} = \{m \in \Sigma_{\mathcal{X}} : (\forall f \in \mathcal{L}(\mathcal{X})) (E_m(f) \leq \overline{E}(f))\}$ . □

**Exercise 4.2** Consider any linear prevision  $E_m$  and any  $\epsilon \in [0, 1]$ . Verify that the so-called *linear-vacuous mixture*:



$$\overline{E}(f) = (1 - \epsilon)E_m(f) + \epsilon \max f$$

is an upper expectation.

**Solution:**  $E_m$  and  $\max$  are upper expectations by Theorem 4.1, because they are upper envelopes of the respective credal sets  $\{m\}$  and  $\Sigma_{\mathcal{X}}$ —see Exercise 4.1. Now verify that being an upper expectation is preserved by taking convex mixtures. The corresponding credal set  $(1 - \epsilon)\{m\} + \epsilon\Sigma_{\mathcal{X}} := \{(1 - \epsilon)m + \epsilon q : q \in \Sigma_{\mathcal{X}}\}$  is indicated in blue in the figure above. ◇

**Exercise 4.3** All upper expectations on a binary space  $\mathcal{X} = \{0, 1\}$  are such linear-vacuous mixtures, and the corresponding credal set can be depicted as

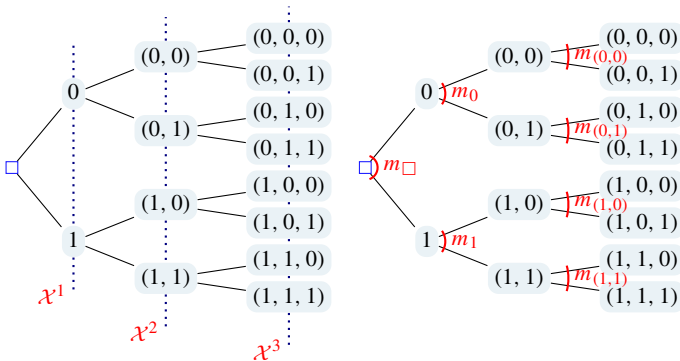


Let  $p := m(1)$  and  $q := 1 - p = m(0)$ . What is the relation between  $[\underline{p}, \bar{p}]$  and  $p, \epsilon$ ?

*Solution:*  $\underline{p} = \underline{E}(I_{\{1\}}) = (1 - \epsilon)p = p - \epsilon p$  and  $\bar{p} = \bar{E}(I_{\{1\}}) = (1 - \epsilon)p + \epsilon = p + \epsilon q$ . Hence,  $\bar{p} - \underline{p} = \epsilon$ .  $\diamond$

### 4.4 Discrete-Time Uncertain Processes

We now apply these ideas in a more dynamic context: the study of processes. We consider an uncertain process, which is a collection of uncertain variables  $X_1, X_2, \dots, X_n, \dots$  assuming values in some finite set of states  $\mathcal{X}$ . This can be represented graphically by a standard *event tree* with *nodes* (also called *situations*)  $s = (x_1, x_2, \dots, x_n)$  for  $x_k \in \mathcal{X}$  and  $n \geq 0$ . This is depicted below on the left for the special case that  $\mathcal{X} = \{0, 1\}$ , where we have limited ourselves to three variables  $X_1, X_2$ , and  $X_3$ ; but the idea should be clear. Observe that we use the symbol  $\square$  for the *initial situation*, or root node, of the event tree.



The event tree becomes a *probability tree* as soon as we attach to each node  $s = (x_1, x_2, \dots, x_n)$  a *local* probability mass function  $m_s$  on  $\mathcal{X}$  with associated expectation operator  $E_{m_s}$ , expressing the uncertainty about the next variable  $X_{n+1}$  after observing the earlier variables  $X_1 = x_1, \dots, X_n = x_n$ . This is depicted above on the right for the special case that  $\mathcal{X} = \{0, 1\}$ .

We now consider a very general inference problem in such a probability tree. Consider any function  $g : \mathcal{X}^n \rightarrow \mathbb{R}$  of the first  $n$  variables:  $g = g(X_1, X_2, \dots, X_n)$ . We want to calculate its expectation  $E(g|s)$  in the situation  $s = (x_1, \dots, x_k)$ , that is, after having observed the first  $k$  variables. Interestingly, this can be done efficiently using the following theorem, which is a reformulation of the Law of Total Probability:

**Theorem 4.2** (Law of Iterated Expectations) *If we know  $E(g|s, x)$  for all  $x \in \mathcal{X}$ , then we can calculate  $E(g|s)$  by backwards recursion using the local model  $m_s$ :*

$$E(g|s) = \underbrace{E_{m_s}}_{\text{local}}(E(g|s, \cdot)) = \sum_{x \in \mathcal{X}} m_s(x) E(g|s, x).$$

This shows that expectations can be calculated recursively using a very basic step, illustrated below for the case  $\mathcal{X} = \{0, 1\}$ :

$$E(g|s) = m_s(1)E(g|s, 1) + m_s(0)E(g|s, 0) \leftarrow s \begin{array}{l} \nearrow m_s \\ \searrow m_s \end{array} \begin{array}{l} (s, 1) \leftarrow E(g|s, 1) \\ (s, 0) \leftarrow E(g|s, 0) \end{array}$$

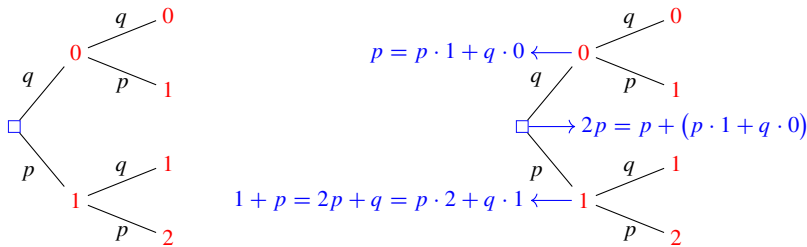
Hence, all expectations  $E(g|x_1, \dots, x_k)$  in the tree can be calculated from the local models  $m_s$  as follows:

1. start in the final cut  $\mathcal{X}^n$  and let  $E(g|x_1, x_2, \dots, x_n) = g(x_1, x_2, \dots, x_n)$ ;
2. do backwards recursion using the Law of Iterated Expectations:

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

3. go on until you get to the root node  $\square$ , where we can identify  $E(g|\square) = E(g)$ .

**Exercise 4.4** Consider flipping a coin twice independently, with probability  $p$  for heads—outcome 1—and  $q = 1 - p$  for tails—outcome 0. The corresponding probability tree for this experiment is given below on the left, with, in red, in the nodes, the corresponding number of heads. What is the expected number of heads?

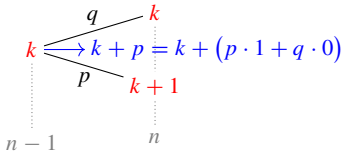


*Solution:* Above on the right, we apply the Law of Iterated Expectations recursively, from leaves to root; the solution is the expectation  $2p$  attached to the root.  $\diamond$

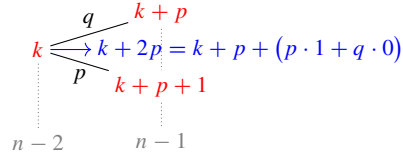
**Exercise 4.5** Extend the ideas in the solution to Exercise 4.4 to calculate the expected number of heads when the coin is flipped  $n$  times independently.

*Solution:* We apply the Law of Iterated Expectations recursively, from leaves to root. Below on the left, we consider starting from the leaves of the tree at depth  $n$ ; applying the Law reduces to adding  $p$  to the number of heads in each of their parent nodes at depth  $n - 1$ . On the right, we apply the Law to these nodes at depth  $n - 1$ , which reduces to adding  $2p$  to the number of heads in each of their parent nodes at depth  $n - 2$ .

at time  $n - 1$  in a situation with  $k$  heads



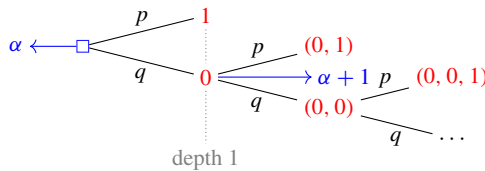
at time  $n - 2$  in a situation with  $k$  heads



Going on in this way, we see that the solution is the expectation  $np$  attached to the root at depth 0.  $\diamond$

**Exercise 4.6** We now flip the same coin time and time again, independently, until we reach heads for the first time. Calculate the expected number of coin flips.

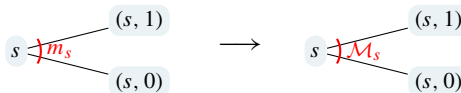
*Solution:* Below is the (unbounded) probability tree associated with this experiment.



Call the unknown expectation  $\alpha$ . We apply the Law of Iterated Expectations to the situations at depth 1. In the situation 1, the expected number of heads is 1, the actual number of heads there. In the situation 0, we see a copy of the original tree extending to the right, but since we have already flipped the coin once here, the expected number of heads in this situation is  $\alpha + 1$ . In the parent node, the expected number of heads  $\alpha$  is therefore also given by  $p \cdot 1 + q \cdot (\alpha + 1) = 1 + q\alpha$ , whence  $\alpha = 1/p$ .  $\diamond$

### 4.5 Imprecise Probability Trees

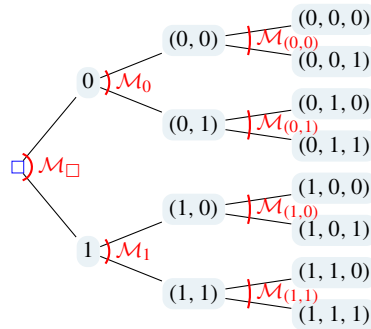
Until now, we have assumed that we have sufficient information in order to specify, in each node  $s$ , a local probability mass function  $m_s$  on the set  $\mathcal{X}$  of possible values for the next state.



We now let go of this major restrictive assumption by allowing for more general uncertainty models. We will consider credal sets as our more general local uncertainty models: closed convex subsets  $\mathcal{M}_s$  of  $\Sigma_{\mathcal{X}}$ . See the figure below for a special case when  $\mathcal{X} = \{0, 1\}$ .

**Definition 4.3** An *imprecise probability tree* is an event tree where in each node  $s$  the local uncertainty model is a credal set  $\mathcal{M}_s$ , or equivalently, its associated upper expectation  $\bar{E}_s$ , with  $\bar{E}_s(f) := \max \{E_m(f) : m \in \mathcal{M}_s\}$  for all  $f \in \mathcal{L}(\mathcal{X})$ .

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



For each real map  $g = g(X_1, \dots, X_n)$ , each node  $s = (x_1, \dots, x_k)$ , and each such *compatible precise probability tree*, we can calculate the expectation  $E(g|x_1, \dots, x_k)$  using the backwards recursion method described before. By varying over each compatible probability tree, we get a closed real interval, completely characterised by lower and upper expectations  $\underline{E}(g|x_1, \dots, x_k)$  and  $\overline{E}(g|x_1, \dots, x_k)$ :  $[\underline{E}(g|x_1, \dots, x_k), \overline{E}(g|x_1, \dots, x_k)]$ . The complexity of calculating these bounds in this way is clearly exponential in the number of time steps  $n$ . But, there is a more efficient method to calculate them:

**Theorem 4.3** (Law of Iterated Upper Expectations [4, 5]) *If we know  $\overline{E}(g|s, x)$  for all  $x \in \mathcal{X}$ , then we can calculate  $\overline{E}(g|s)$  by backwards recursion using the local model  $\overline{E}_s$ :*

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

This shows that expectations can be calculated recursively using a very basic step, illustrated below for the case  $\mathcal{X} = \{0, 1\}$ :

$$\overline{E}(g|s) = \overline{E}_s(\overline{E}(g|s, \cdot)) \leftarrow s \begin{array}{l} (s, 1) \leftarrow \overline{E}(g|s, 1) \\ \overline{\mathcal{M}}_s \\ (s, 0) \leftarrow \overline{E}(g|s, 0) \end{array}$$

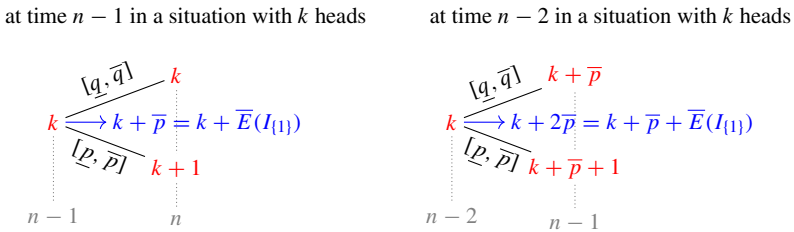
The method for, and the complexity of, calculating the  $\overline{E}(g|s)$ , as a function of  $n$ , is therefore essentially the same as in the precise case!

**Exercise 4.7** Extend the ideas in the solution to Exercise 4.5 to calculate the upper expected number of heads when the coin is flipped  $n$  times independently, but where now we have an imprecise probability model for a coin flip, with a probability interval  $[p, \overline{p}]$  for heads, and a corresponding interval  $[q, \overline{q}] = [1 - \overline{p}, 1 - p]$  for tails.

*Solution:* We apply the Law of Iterated Upper Expectations recursively, from leaves to root. Below on the left, we consider starting from the leaves of the tree at depth  $n$ ;



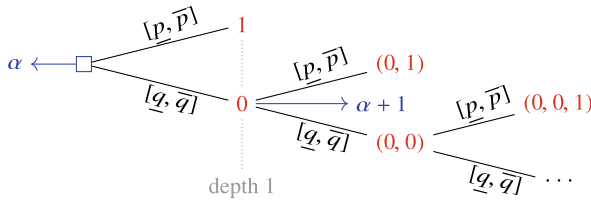
applying the Law reduces to adding  $\bar{p}$  to the number of heads in each of their parent nodes at depth  $n - 1$ .



On the right, we apply the Law to these nodes at depth  $n - 1$ , which reduces to adding  $2\bar{p}$  to the number of heads in each of their parent nodes at depth  $n - 2$ . Going on in this way, we see that the solution is the expectation  $n\bar{p}$  attached to the root at depth 0. A similar result holds for the lower expectation.  $\diamond$

**Exercise 4.8** We now flip the same coin with the imprecise probability model time and time again, independently, until we reach heads for the first time. Calculate the upper expected number of coin flips.

*Solution:* Below is the (unbounded) probability tree associated with this experiment.



Call the unknown upper expectation  $\alpha$ . We apply the Law of Iterated Upper Expectations to the situations at depth 1. In the situation 1, the upper expected number of heads is 1, the actual number of heads there. In the situation 0, we see a copy of the original tree extending to the right, but since we have already flipped the coin once here, the upper expected number of heads in this situation is  $\alpha + 1$ . In the parent node, the upper expected number of heads  $\alpha$  is therefore also given by  $1 + \bar{E}(\alpha I_{(0)}) = 1 + \alpha\bar{q}$ , whence  $\alpha = 1/\bar{p}$ . A similar result holds for the lower expectation.  $\diamond$

The attentive reader will have observed that in all these simple exercises, we can also obtain the ‘imprecise’ result from the ‘precise’ one by optimising over the single parameter  $p$ . We have to warn against too much optimism: in more involved examples, this will no longer be the case.

### 4.6 Imprecise Markov Chains

We now look at a special instance of a probability tree, corresponding to a stationary (precise) Markov chain. This happens when the precise local models  $m_{(x_1, \dots, x_n)}$  only

depend on the last observed state  $x_n$ —this is the *Markov Condition*—and also do not depend explicitly on the time step  $n$ :

$$m_{(x_1, \dots, x_n)} = q(\cdot | x_n)$$

for some family of *transition mass functions*  $q(\cdot | x)$ ,  $x \in \mathcal{X}$ .

**Definition 4.4** The uncertain process is a stationary *precise Markov chain* when all  $\mathcal{M}_s$  are singletons  $\{m_s\}$  and  $\mathcal{M}_{(x_1, \dots, x_n)} = \{q(\cdot | x_n)\}$ , for some family of *transition mass functions*  $q(\cdot | x)$ ,  $x \in \mathcal{X}$ .

For each  $x \in \mathcal{X}$ , the transition mass function  $q(\cdot | x)$  corresponds to an expectation operator, given by  $E(f|x) = \sum_{z \in \mathcal{X}} q(z|x) f(z)$  for all  $f \in \mathcal{L}(\mathcal{X})$ .

**Definition 4.5** Consider the linear transformation  $T$  of  $\mathcal{L}(\mathcal{X})$ , called *transition operator*:  $T: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X})$ :  $f \mapsto Tf$ , where  $Tf$  is the real map defined by:

$$Tf(x) := E(f|x) = \sum_{z \in \mathcal{X}} q(z|x) f(z) \text{ for all } x \in \mathcal{X}.$$

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

Up to now, we have mainly been concerned with conditional expectations of the type  $E(\cdot | s)$ . We will now look at particular *unconditional* expectations, where  $s = \square$ . For any  $n \geq 0$ , we define the expectation for the (single) state  $X_n$  at time  $n$  by

$$E_n(f) = E(f(X_n)) = E(f(X_n) | \square) \text{ for all } f: \mathcal{X} \rightarrow \mathbb{R}$$

and we denote the corresponding mass function by  $m_n$ . Applying the Law of Iterated Expectations in Theorem 4.2 now yields, with also  $E_1 = E_{m_\square}$  and  $m_1 = m_\square$ :

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

so the complexity of calculating  $E_n(f)$  is linear in the number of time steps  $n$ .

**Exercise 4.9** Consider the stochastic process where we first flip a fair coin. From then on, on heads, we select a biased coin with probability  $p$  for heads for the next coin flip, and on tails, a biased coin with probability  $q = 1 - p$  for heads, and keep on flipping one of the two biased coins, selected on the basis of the outcome of the previous coin flip. This produces a Markov chain. Find  $Tf$ ,  $T^2f$ , and  $E_1(f)$ ,  $E_2(f)$  and  $E_3(f)$  for  $f \in \mathcal{L}(\{0, 1\})$ .

*Solution:* Clearly,  $E_1(f) = \frac{1}{2}f(1) + \frac{1}{2}f(0)$ ,  $Tf(0) = E(f|0) = qf(1) + pf(0)$  and  $Tf(1) = E(f|1) = pf(1) + qf(0)$ , whence

$$E_2(f) = E_1(Tf) = \frac{p+q}{2}f(1) + \frac{p+q}{2}f(0) = \frac{1}{2}f(1) + \frac{1}{2}f(0).$$

Similarly,

$$\begin{aligned} T^2 f(0) &= qTf(1) + pTf(0) = q[pf(1) + qf(0)] + p[qf(1) + pf(0)] \\ &= (p^2 + q^2)f(0) + 2pqf(1) \\ T^2 f(1) &= (p^2 + q^2)f(1) + 2pqf(0), \end{aligned}$$

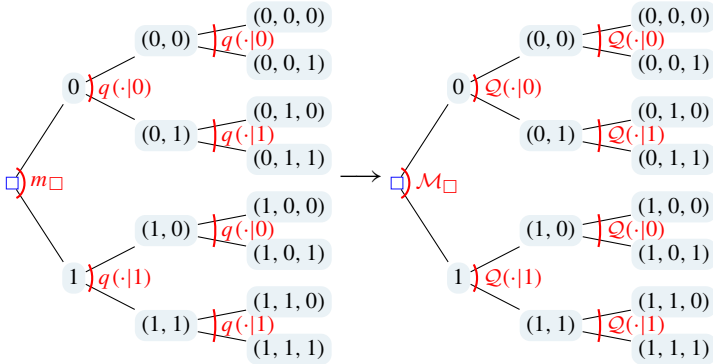
whence

$$E_3(f) = E_1(T^2 f) = \frac{p^2 + q^2 + 2pq}{2} f(1) + \frac{2pq + p^2 + q^2}{2} f(0) = \frac{1}{2} f(1) + \frac{1}{2} f(0),$$

and so on. We see that at the level of expectations of single state variables, the process cannot be distinguished from flipping a fair coin.  $\diamond$

The generalisation from precise to imprecise Markov chains goes as follows:

**Definition 4.6** The uncertain process is a stationary *imprecise Markov chain* when the Markov Condition is satisfied with stationarity:  $\mathcal{M}_{(x_1, \dots, x_n)} = \mathcal{Q}(\cdot|x_n)$  for some family of *transition credal sets*  $\mathcal{Q}(\cdot|x)$ ,  $x \in \mathcal{X}$ .



An imprecise Markov chain can be seen as an infinity of (precise) probability trees: choose a precise mass function from  $\mathcal{M}_s$  in each situation  $s$ . It should be clear that not all of these satisfy the Markov property or stationarity. This implies that solving the optimisation problem in order to find the tight upper bounds  $\bar{E}(g|s)$ , as discussed in Sect. 4.5, is not (necessary always) simply an optimisation over a parametrised collection of stationary (or even non-stationary) Markov chains, although it can turn out to be so simple in a number of special cases.

For each  $x \in \mathcal{X}$ , the local transition model  $\mathcal{Q}(\cdot|x)$  corresponds to an upper expectation operator  $\bar{E}(\cdot|x)$ , with  $\bar{E}(f|x) = \max \{E_p(f) : p \in \mathcal{Q}(\cdot|x)\}$  for all  $f \in \mathcal{L}(\mathcal{X})$ . This leads to the following definition, which generalises the definition of transition operators for precise Markov chains:

**Definition 4.7** Consider the non-linear transformation  $\bar{T}$  of  $\mathcal{L}(\mathcal{X})$ , called the *upper transition operator*:  $\bar{T}: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X}) : f \mapsto \bar{T}f$  where the real map  $\bar{T}f$  is defined by  $\bar{T}f(x) := \bar{E}(f|x) = \max \{E_p(f) : p \in \mathcal{Q}(\cdot|x)\}$  for all  $x \in \mathcal{X}$ .

For any  $n \geq 0$ , we define the upper expectation for the (single) state  $X_n$  at time  $n$  by

$$\bar{E}_n(f) = \bar{E}(f(X_n)) = \bar{E}(f(X_n)|\square) \text{ for all } f: \mathcal{X} \rightarrow \mathbb{R}.$$

Then the Law of Iterated Upper Expectations of Theorem 4.3 yields, with also  $\bar{E}_1 = \bar{E}_{\mathcal{M}_\square}$ :

$$\bar{E}_n(f) = \bar{E}_1(\bar{T}^{n-1} f) \text{ for all } n \geq 1 \text{ and all } f \in \mathcal{L}(\mathcal{X}),$$

so the complexity of calculating  $\bar{E}_n(f)$  is still linear in the number of time steps  $n$ .

### 4.7 Examples

Consider a two-element state space  $\mathcal{X} = \{1, 0\}$ , with upper expectation  $\bar{E}_1 = \bar{E}_{\mathcal{M}_\square}$  for the first variable, and for each  $(x_1, \dots, x_n) \in \{1, 0\}^n$ , with  $0 < \epsilon \leq 1$ ,  $\mathcal{M}_{(x_1, \dots, x_n)} = \mathcal{M}_{x_n} = (1 - \epsilon)\{q(\cdot|x_n)\} + \epsilon\Sigma_{\{1,0\}}$ , or equivalently, for the upper transition operator  $\bar{T} = (1 - \epsilon)T + \epsilon \max$ . In other words, each transition credal set  $\mathcal{Q}(\cdot|x)$  is a linear-vacuous mixture (see Exercise 4.2, also for the notations used) centred on the transition mass function  $q(\cdot|x)$ , where the mixture coefficient  $\epsilon$  is the same in each state  $x$ .

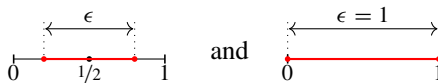
It is a matter of simple and direct verification that for  $n \geq 1$  and  $f \in \mathcal{L}(\mathcal{X})$ :  $\bar{T}^n f = (1 - \epsilon)^n T^n f + \epsilon \sum_{k=0}^{n-1} (1 - \epsilon)^k \max T^k f$ , and therefore, using the Law of Iterated Expectations,  $\bar{E}_{n+1}(f) = \bar{E}_1(\bar{T}^n f) = (1 - \epsilon)^n \bar{E}_1(T^n f) + \epsilon \sum_{k=0}^{n-1} (1 - \epsilon)^k \max T^k f$ . If we now let  $n \rightarrow \infty$ , it is not too hard to see that the limit exists and is independent of the initial upper expectation  $\bar{E}_1$ :

$$\lim_{n \rightarrow \infty} \bar{E}_n(f) = \epsilon \sum_{k=0}^{\infty} (1 - \epsilon)^k \max T^k f \text{ for all } f \in \mathcal{L}(\mathcal{X}).$$

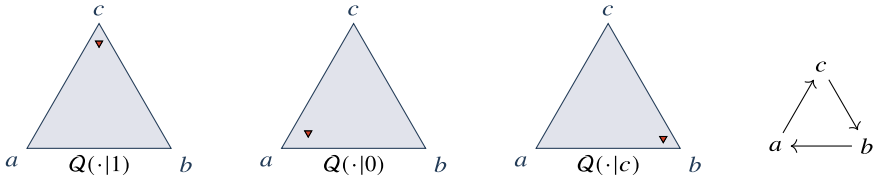
We consider two special cases:

1. Contaminated random walk: when  $Tf(1) = Tf(0) = 1/2[f(1) + f(0)]$ , the underlying precise Markov chain is actually like flipping a fair coin. We then find that  $\bar{E}_\infty(f) = (1 - \epsilon)1/2[f(1) + f(0)] + \epsilon \max f$  for all  $f \in \mathcal{L}(\mathcal{X})$ .
2. Contaminated cycle: when  $Tf(1) = f(0)$  and  $Tf(0) = f(1)$ , the underlying precise Markov chain is actually like deterministic cycle between the states 0 and 1. We then find that  $\bar{E}_\infty(f) = \max f$  for all  $f \in \mathcal{L}(\mathcal{X})$ .

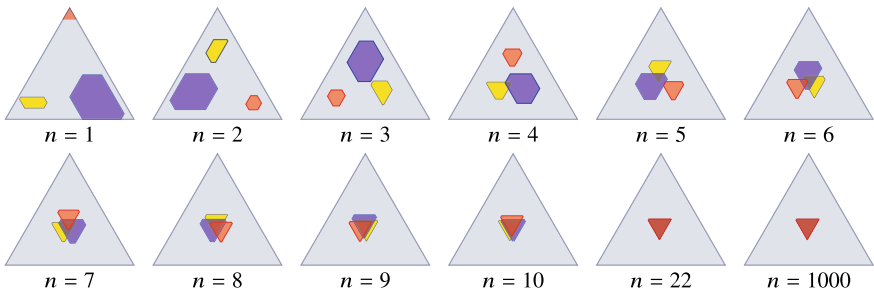
The probability intervals for 1 corresponding to these two limit models are given by



As another example, we consider  $\mathcal{X} = \{a, b, c\}$  and the transition models depicted below, which are imprecise models not very far from a simple cycle:



Below, we depict the time evolution of the  $\bar{E}_n$  (as credal sets) for three cases (red, yellow and blue). We see that, here too, regardless of the initial distribution  $\bar{E}_1$ , the distribution  $\bar{E}_n$  seems to converge to the same distribution.



### 4.8 A Non-linear Perron–Frobenius Theorem, and Ergodicity

The convergence behaviour in the previous examples can also be observed in general imprecise Markov chains under fairly weak conditions. The following theorems can be derived from the more general discussions and results in [3, 5].

**Theorem 4.4** Consider a stationary imprecise Markov chain with finite state set  $\mathcal{X}$  and upper transition operator  $\bar{T}$ . Suppose that  $\bar{T}$  is regular, meaning that there is some  $n > 0$  such that  $\min \bar{T}^n I_{\{x\}} > 0$  for all  $x \in \mathcal{X}$ . Then for every initial upper expectation  $\bar{E}_1$ , the upper expectation  $\bar{E}_n = \bar{E}_1 \circ \bar{T}^{n-1}$  for the state at time  $n$  converges point-wise to the same stationary upper expectation  $\bar{E}_\infty$ :  $\lim_{n \rightarrow \infty} \bar{E}_n(h) = \lim_{n \rightarrow \infty} \bar{E}_1(\bar{T}^{n-1}h) := \bar{E}_\infty(h)$  for all  $h$  in  $\mathcal{L}(\mathcal{X})$ . The limit upper expectation  $\bar{E}_\infty$  is the only  $\bar{T}$ -invariant upper expectation on  $\mathcal{L}(\mathcal{X})$ , meaning that  $\bar{E}_\infty = \bar{E}_\infty \circ \bar{T}$ .

In that case we also have an interesting *ergodicity* result. For a detailed description of the notion of ‘almost surely’, we refer to [3], but it roughly means ‘with upper probability one’.

**Theorem 4.5** Consider a stationary imprecise Markov chain with finite state set  $\mathcal{X}$  and upper transition operator  $\bar{T}$ . Suppose that  $\bar{T}$  is regular with stationary upper expectation  $\bar{E}_\infty$ . Then, almost surely, for all  $h$  in  $\mathcal{L}(\mathcal{X})$ :

$$\underline{E}_\infty(h) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n h(X_k) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n h(X_k) \leq \bar{E}_\infty(h).$$

## 4.9 Conclusion

The discussion in this paper lays bare a few interesting but quite basic aspects of inference in imprecise probability trees and Markov chains in discrete time. A more general and deeper treatment of these matters can be found in [3–5]. For recent work on imprecise Markov chains in continuous time, I refer the interested reader to [1, 9].

## References

1. Jasper De Bock. The Limit Behaviour of Imprecise Continuous-Time Markov Chains. *Journal of Nonlinear Science*, 27(1):159–196, 2017.
2. Jasper De Bock and Gert de Cooman. An efficient algorithm for estimating state sequences in imprecise hidden Markov models. *Journal of Artificial Intelligence Research*, 50:189–233, 2014.
3. Gert de Cooman, Jasper De Bock, and Stavros Lopatzizidis. Imprecise stochastic processes in discrete time: global models, imprecise markov chains, and ergodic theorems. *International Journal Of Approximate Reasoning*, 76(C):18–46, 2016.
4. Gert de Cooman and Filip Hermans. Imprecise probability trees: Bridging two theories of imprecise probability. *Artificial Intelligence*, 172(11):1400–1427, 2008.
5. Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behaviour. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, 2009. [arXiv:0801.0980](https://arxiv.org/abs/0801.0980).
6. D. J. Hartfiel and E. Seneta. On the theory of Markov set-chains. *Advances in Applied Probability*, 26:947–964, 1994.
7. Darald J. Hartfiel. Sequential limits in Markov set-chains. *Journal of Applied Probability*, 28(4):910–913, 1991.
8. Darald J. Hartfiel. *Markov Set-Chains*. Number 1695 in Lecture Notes in Mathematics. Springer, Berlin, 1998.
9. Thomas Krak, Jasper De Bock, and Arno Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*. 88:452–528, 2017.
10. Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley, 2014.
11. Damjan Škulj and Robert Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76(1):107–133, 2013.
12. Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

