Chapter 5
An Introduction to Imprecise Markov Chains

Thomas Krak

Abstract Stochastic processes in general provide a popular framework for modelling uncertainty about the evolution of dynamical systems. The theory of Markov chains uses a number of crucial assumptions about the (in)dependence of such a process on its history that make their analysis tractable. In practice however, the parameters of a Markov chain may not be known exactly, or there may exist doubt as to the applicability of these assumptions to the system under study. This chapter presents an introduction to imprecise Markov chains, which are a robust generalisation of these models that may be used when parameters are not known exactly or when such assumptions could be violated. Their treatment is grounded in the theory of imprecise probabilities. The generalised model can be interpreted as a set of (traditional) stochastic processes, which may or may not be Markovian and which may have different and varying parameter values. Inferences are then performed to ensure robustness with respect to variations within this set. This chapter assumes no advanced familiarity with Markov chains or imprecise probability theory. It aims to develop an intuitive and graphical understanding of (imprecise) Markov chains in discrete and in continuous time.

Keywords Imprecise probabilities · Model uncertainty · Stochastic processes · Imprecise Markov chains

5.1 Introduction

In many areas of science and engineering, we are interested in modelling uncertainty about the behaviour of dynamical systems, that is, systems whose state changes as time passes. For instance, we may want to model the evolution of the spatial trajectories of a system in motion; or the performance and reliability of a complex

T. Krak (✉)
IDLab, Ghent University, Ghent, Belgium
e-mail: thomas.krak@ugent.be

M. Vasile (ed.), *Optimization Under Uncertainty with Applications to Aerospace Engineering*, https://doi.org/10.1007/978-3-030-60166-9_5

composite system as its components wear out, break down and get replaced; or the spread of pathogens through a population; or the evolution of stock prices—and so on and so forth.

What all these systems have in common is that there is a *dynamic* component to their description—they change *over time*—and they are, in a sense, hard to describe *exactly*. For instance, this difficulty may arise because their behaviour depends on unknown external influences or because the system cannot reasonably be described at a sufficiently detailed level. Thus, there arises an uncertainty about how exactly the system will evolve over time, even if one can model how it will 'roughly' behave. Regardless of the interpretation that we want to assign to this uncertainty, such systems are modelled using *stochastic processes*. A stochastic process, then, is a probabilistic description of the system under study. In this sense, it provides a formal and integrated description of the system dynamics and the probabilistic uncertainty of its evolution.

On the other hand, we might also be uncertain about whether such a model is 'correct'. For instance, we might not know exactly the numerical values that the parameters of our model should take. Similarly, we might be aware that our modelling assumptions lead to simplifications that are not necessarily warranted, which introduces uncertainty about the accuracy or applicability of any assessments made on the basis of these models. It is therefore of interest to robustify our models also against these kinds of 'meta', or 'higher-order', uncertainties.

In this chapter, we consider stochastic processes for which this higher-order uncertainty is modelled using the theory of imprecise probabilities (IP). For an extended introduction to IP, we refer the reader back to Chap. 2. We constrain ourselves to briefly recalling that such imprecise probabilistic models can be interpreted as representing a *set* of traditional probabilistic models. So, in our current setting, we will be considering *sets* of stochastic processes. From an inference point of view, the aim is then to compute inferences which are robust with respect to variations within such a set. We recall from Chap. 2 that these robust inferences are captured in general by the *lower* and *upper* expectations with respect to the elements of the set that we are considering.

Our aim with the present chapter is to provide an extensive but intuitive introduction to the theory of imprecise stochastic processes and of imprecise Markov chains in particular. To this end, we will intentionally focus on the different representations of these processes. We will show how each of the different ways of looking at these models provides its own way of deriving useful properties and highlights different intuitive ways of reasoning about them. Important results and properties are stated, but we have made an effort to keep the discussion intuitive. We try to prevent technicalities and do not provide extended proofs; instead, we will provide pointers to the literature that the interested reader might pursue herself.

The remainder of this chapter is organised as follows. We start the discussion by giving a quick introduction to stochastic processes in Sect. 5.2. The first part basically uses the measure-theoretic approach (albeit in a rather simplified sense) to pin down some first concepts and notation. We then go on to present three different and graphical representations of stochastic processes, which can be used

when the time-dimension is discrete. Specifically, we cover the representation using probability trees, in Sect. 5.2.1; using Bayesian networks, in Sect. 5.2.2 and using transition graphs, in Sect. 5.2.3.

Once we have developed these different ways of reasoning about discrete-time processes, we generalise the discussion to *imprecise* discrete-time processes in Sect. 5.3. We use the previously developed graphical notions to provide intuition about how to reason and compute inferences using these models. The treatment of (imprecise) continuous-time processes is largely postponed until Sect. 5.4. Here the graphical and intuitive representations largely break down, but we can then use the previously developed understanding of the discrete-time case to reason about these models. To keep the main text as readable as possible, the discussion of the literature on which the material in this chapter is based is deferred to Sect. 5.5.

5.2 (Precise) Stochastic Processes

We will start the exposition around stochastic processes in a relatively general and abstract sense but will quickly make things more specific. Throughout the remainder of this chapter, we will consider some fixed abstract *state-space* $\mathscr{X}$. A *state* is an element $x \in \mathscr{X}$ and represents uniquely the relevant information about the underlying system that we are interested in modelling. So as not to complicate matters, we will assume throughout that $\mathscr{X}$ is finite, so that we can identify it without loss of generality as the set $\mathscr{X} = \{1, \ldots, k\} \subset \mathbb{N}$. Note that here and in what follows, we denote with $\mathbb{N}$ the natural numbers and will write $\mathbb{N}_0 \coloneqq \mathbb{N} \cup \{0\}$ when we include zero. Furthermore, the real numbers are written $\mathbb{R}$, the non-negative reals are $\mathbb{R}_{\geq 0}$ and the positive reals are $\mathbb{R}_{>0}$.

Because we are interested in modelling a system whose state $x \in \mathscr{X}$ changes over time, we next identify some *time-dimension* $\mathbb{T}$. A crucial choice to be made later on is whether we are considering processes in discrete-time, in which case we identify $\mathbb{T} = \mathbb{N}_0$, or processes in continuous-time, in which case $\mathbb{T} = \mathbb{R}_{\geq 0}$. For now we simply keep the discussion general without making this identification.

With the state-space and time-dimension in place, it now makes sense to talk about the *realisation* of some (yet to be identified) stochastic process. Such a realisation is also called a *sample path*, and it is a function $\omega : \mathbb{T} \to \mathscr{X}$. So, this ω describes for each point in time $t \in \mathbb{T}$ the state $\omega(t) \in \mathscr{X}$ that the system was in at that time. We collect in the set Ω all these sample paths. For technical reasons, it is sometimes required to restrict attention to paths that satisfy sufficient smoothness conditions; for instance, when $\mathbb{T} = \mathbb{R}_{\geq 0}$, it is common practice to let Ω only contain càdlàg functions, that is, paths $\omega(t)$ that are right-continuous and whose left-sided limits exist everywhere.

This set Ω thus contains all possible ways in which the system might behave over time; it can therefore be considered an *outcome space* of a stochastic model. Formally, we will consider some abstract underlying probability space $(\Omega, \mathscr{F}, P)$, where $\mathscr{F}$ is some appropriate σ-algebra on Ω and where P is a probability measure

on $(\Omega, \mathscr{F})$. Given this probability space, we can finally formalise the notion of a *stochastic process* as a collection $\{X_t\}_{t\in\mathbb{T}}$ of random variables associated to this probability space. We will here slightly restrict our definition to the following specific stochastic process:

Definition 5.1 (Stochastic process) Fix a time-dimension $\mathbb{T}$ and consider a probability space $(\Omega, \mathscr{F}, P)$. Then (the corresponding) stochastic process is the collection $\{X_t\}_{t\in\mathbb{T}}$ of random variables $X_t : \Omega \to \mathscr{X} : \omega \mapsto \omega(t)$, $t \in \mathbb{T}$, on this space.

Corollary 5.1 *Fix a time-dimension $\mathbb{T}$; consider a probability space $(\Omega, \mathscr{F}, P)$; and let $\{X_t\}_{t\in\mathbb{T}}$ be the corresponding stochastic process. Then for all $t \in \mathbb{T}$ and $x \in \mathscr{X}$, it holds that* $\Pr(X_t = x) = P\big(\{\omega \in \Omega : \omega(t) = x\}\big)$.

Proof Fix $t \in \mathbb{T}$, and recall the definition of a random variable: for all $x \in \mathscr{X}$, the probability $\Pr(X_t = x)$ of X_t taking the value x is equal to $P\big(X_t^{-1}(x)\big)$, the measure of its preimage in Ω. Since $X_t(\omega) = \omega(t)$, we have $X_t^{-1}(x) = \{\omega \in \Omega : \omega(t) = x\}$. □

The above is a formal way of saying that, and how, these random variables $\{X_t\}_{t\in\mathbb{T}}$ are associated to the given probability space. In words, for some fixed time $t \in \mathbb{T}$, X_t is a random variable that takes on a value $x \in \mathscr{X}$ with probability equal to the measure of the set of paths along which the state at time t is x. Conversely, if we fix the outcome $\omega \in \Omega$, then the collection $\{X_t\}_{t\in\mathbb{T}}$ can be considered a deterministic process, and $X_t(\omega) = \omega(t)$ for all $t \in \mathscr{X}$.

Note, therefore, that all the quantitative information about the probability of the process taking on certain values at given points in time are completely determined by the measure P. It is therefore also intuitive to instead consider this measure P to be 'the stochastic process', although this is technically an abuse of terminology. This is because, for a given probability space $(\Omega, \mathscr{F}, P)$, it is possible to define many different stochastic processes; *any* $\mathbb{T}$-indexed collection of random variables on this space satisfies the general definition. However, in a sense, the stochastic process in Definition 5.1 can be viewed as the 'canonical' stochastic process corresponding to the given probability space, since it specifically and exactly represents the uncertainty about which states might be obtained at different points in time. We will therefore, and for notational convenience, often refer to the measure P and its corresponding stochastic process $\{X_t\}_{t\in\mathbb{T}}$ interchangeably and without confusion.

Next, it will be convenient to have a standardised notation to index a subset of the random variables of a stochastic process. To this end, for any finite sequence of time points $\mathbf{t} = t_1, \ldots, t_n$ in $\mathbb{T}$, with $n \in \mathbb{N}$, we will write $X_{\mathbf{t}} = X_{t_1}, \ldots, X_{t_n}$. Typically, these sequences will be taken to be ordered, so that $t_1 < \cdots < t_n$. Note that each of the random variables X_{t_i}, $i = 1, \ldots, n$ takes values in $\mathscr{X}$. Hence, the sequence $X_{\mathbf{t}}$ takes values (jointly) in $\mathscr{X}^n = \times_{i=1}^n \mathscr{X}$. An element of this joint state-space is thus a vector $(x_1, \ldots, x_n) \in \mathscr{X}^n$. When we are explicitly talking about a sequence $\mathbf{t}$ of n time points, we will also write $x_{\mathbf{t}}$ to denote a generic element of $\mathscr{X}^n$.

In what follows, we will be interested in computing the expectation of some real-valued function, whose value depends on the specific realisation of the stochastic process. To prevent technical difficulties, we will assume that this function only depends on a finite number of time points; without loss of generality, we can then assume that it is a map $f : \mathscr{X}^n \to \mathbb{R}$, with $n \in \mathbb{N}$, whose value depends on the n random variables $X_{\mathbf{t}}$, with $\mathbf{t} = t_1, \ldots, t_n$ in $\mathbb{T}$. We collect in the set $\mathscr{L}(\mathscr{X}^n)$ all such real-valued functions on $\mathscr{X}^n$. The expected value of any such $f \in \mathscr{L}(\mathscr{X}^n)$ on the n time points $\mathbf{t}$ is defined as

$$\mathbb{E}_P\big[f(X_{\mathbf{t}})\big] \coloneqq \sum_{x_{\mathbf{t}} \in \mathscr{X}^n} f(x_{\mathbf{t}}) P(X_{\mathbf{t}} = x_{\mathbf{t}}) \,, \tag{5.1}$$

where we have implicitly introduced the intuitive notation for the set

$$(X_{\mathbf{t}} = x_{\mathbf{t}}) \coloneqq \Big\{\omega \in \Omega \,:\, \big(\forall i \in \{1, \ldots, n\} : \omega(t_i) = x_{t_i}\big)\Big\} \,.$$

In Eq. (5.1), we use the subscript P for the expectation operator $\mathbb{E}_P$ to make explicit that it is taken with respect to the measure P; this will be notationally convenient further on.

We finish this first introduction by recalling the notion of conditional probabilities and conditional expectations. For any two finite sequences of time points $\mathbf{t}$ and $\mathbf{s}$ in $\mathbb{T}$, the *conditional probability* of $X_{\mathbf{t}}$, given $X_{\mathbf{s}}$, is derived using *Bayes' rule*:

$$P(X_{\mathbf{t}} \,|\, X_{\mathbf{s}}) \coloneqq \frac{P(X_{\mathbf{s}}, X_{\mathbf{t}})}{P(X_{\mathbf{s}})} \,,$$

whenever $P(X_{\mathbf{s}})$ is strictly positive. The necessity of the final condition is obvious; it leads to a division by zero whenever it does not hold.

Using this notion of conditional probability, we can define conditional expectations analogously. Suppose the sequences $\mathbf{s}$ and $\mathbf{t}$ are of length $n, m \in \mathbb{N}$, respectively. Then for any $f \in \mathscr{L}(\mathscr{X}^{n+m})$ on $X_{\mathbf{s}}$, $X_{\mathbf{t}}$ we define, for all $x_{\mathbf{s}} \in \mathscr{X}^n$,

$$\mathbb{E}_P\big[f(X_{\mathbf{s}}, X_{\mathbf{t}}) \,\big|\, X_{\mathbf{s}} = x_{\mathbf{s}}\big] \coloneqq \sum_{x_{\mathbf{t}} \in \mathscr{X}^m} f(x_{\mathbf{s}}, x_{\mathbf{t}}) P(X_{\mathbf{t}} = x_{\mathbf{t}} \,|\, X_{\mathbf{s}} = x_{\mathbf{s}}) \,.$$

5.2.1 Probability Trees

The preceding discussion introduced stochastic processes in a very general, but rather abstract sense. We will build further intuition by next offering a different view and representation, by means of *probability trees*. In the remainder of this section, unless otherwise specified, we will focus on discrete-time stochastic processes, whence we identify $\mathbb{T} = \mathbb{N}_0$.

We next need some notation and definitions for 'partial paths', which in this setting are also called *situations*. As before, a (full) path is a map $\omega : \mathbb{N}_0 \to \mathscr{X}$. In contrast, a *situation* is defined as a (finite length) *prefix* of such a path. In other words, a situation is an element of a set $\mathscr{X}^n$, for some $n \in \mathbb{N}$. If $w \in \mathscr{X}^n$, $n \in \mathbb{N}$, is a situation, we write w_i for its $(i+1)$-th coordinate, $i \in \{0, \ldots, n-1\}$, and we say that its *length* is $|w| = n$. Note that the indexing over the coordinates is taken to start from zero rather than one—this is done for notational consistency with paths ω. Since we will need to refer to it so often, we introduce the shorthand notation $w_\top$ for the *last* element of w; so if w has length n, then $w_\top \coloneqq w_{n-1}$. The set of all non-empty situations is $\mathscr{X}^* \coloneqq \cup_{n\in\mathbb{N}} \mathscr{X}^n$, and we define $\mathscr{X}^*_\square \coloneqq \{\square\} \cup \mathscr{X}^*$, where we add the *empty situation* denoted by $\square$.

As a final point in this notational digression, for any $s, t \in \mathbb{N}_0$ such that $s \leq t$, we will introduce the shorthand notation $s : t$ to denote the sequence of time points $s, \ldots, t$. Using our previously introduced notation, we can then write $X_{s:t}$ for the random variables at these time points. Furthermore, for any $n \in \mathbb{N}_0$ and any situation $w \in \mathscr{X}^{n+1}$, we can then use the previously introduced notation to write $X_{0:n} = w$; this is understood to mean that the random variables at time points $0, \ldots, n$ obtained the states corresponding to the situation w.

We endow the set $\mathscr{X}^*_\square$ with the *prefix order*, denoted $\prec$, which is a partial order such that $\square \prec v$ for all $v \in \mathscr{X}^*$ and for all $v, w \in \mathscr{X}^*$ with lengths $n = |v|$ and $m = |w|$, it holds that $v \prec w$ if and only if $n < m$ and $v_i = w_i$ for all $i \in \{0, \ldots, n-1\}$. This is just a rigorous but somewhat obfuscated way of saying that $v \prec w$ if 'v is the beginning of w' or 'w is what you can get if v happens first, and then some other things happen' or, indeed, 'v is a prefix of w'.

The important thing to notice is that the ordered set $(\mathscr{X}^*_\square, \prec)$ induces a graphical tree structure, with all the situations as its vertices. This tree is what is known as the *event tree*. It has $\square$ as its root, and, for all $v, w \in \mathscr{X}^*_\square$, w is a descendant of v exactly if $v \prec w$. An example of such a tree is shown in Fig. 5.1, which (partially) shows the event tree corresponding to a binary state-space $\mathscr{X} = \{a, b\}$.

Such an event tree can be turned into an intuitive representation of a stochastic process by augmenting it into a *probability tree*. This is done by assigning to each situation $w \in \mathscr{X}^*_\square$ in the tree a *local model* p_w, which is a probability mass function on $\mathscr{X}$; that is, it is a map $p_w : \mathscr{X} \to \mathbb{R}_{\geq 0}$ such that $\sum_{x\in\mathscr{X}} p_w(x) = 1$. An example of this is again illustrated in Fig. 5.1.

Definition 5.2 (Probability tree) A probability tree is a tuple $(\mathscr{X}^*_\square, \prec, p_{(\cdot)})$, where $\mathscr{X}^*_\square$ is the set of all situations, $\prec$ is the prefix order on $\mathscr{X}^*_\square$ and $p_{(\cdot)} : \mathscr{X}^*_\square \times \mathscr{X} \to \mathbb{R}_{\geq 0}$ represents all local models, so that $\sum_{x\in\mathscr{X}} p_w(x) = 1$ for all $w \in \mathscr{X}^*_\square$.

The mechanism by which a stochastic process obtains a certain realisation $\omega \in \Omega$ can now be interpreted as performing a weighted, random walk along this probability tree, starting from $\square$. Following the tree in Fig. 5.1, this is done as follows: from $\square$, we transition either to a, with probability $p_\square(a)$, or to b, with probability $p_\square(b)$. Suppose we transition to a. From this new situation, the next step will take us either to aa, with probability $p_a(a)$, or to ab, with probability $p_a(b)$.

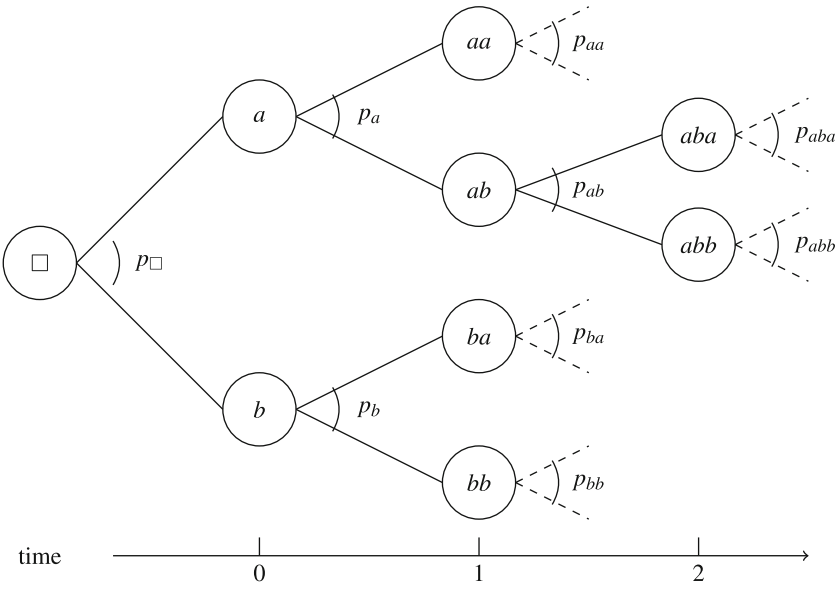

Fig. 5.1 A (partial) event tree for a binary state-space $\mathscr{X} = \{a, b\}$. The vertices are situations, i.e. elements of $\mathscr{X}_\square^*$, and the edges are induced by the prefix order $\prec$. Dashed lines represent branches that are not shown in the figure. The tree has been augmented to a probability tree, by assigning to each $w \in \mathscr{X}^*$ a local model p_w. A time axis represents at which point in time the situations can occur

Proceeding in this fashion, an infinite random walk along this tree generates a full path $\omega : \mathbb{N}_0 \to \mathscr{X}$, where, for all $t \in \mathbb{N}_0$, the state $\omega(t)$ represents the (randomly chosen) branch that we took along the tree at the $(t+1)$-th step.

This 'path construction' view allows us also to connect back to the measure-theoretic definition that we encountered earlier. To obtain this correspondence in one direction, fix a probability tree $(\mathscr{X}_\square^*, \prec, p_{(\cdot)})$ and let $(\Omega, \mathscr{F})$ be an appropriate measurable space of discrete-time sample paths, on which we will aim to construct the measure P quantifying, in the measure-theoretic sense, the uncertainty of the corresponding stochastic process $\{X_t\}_{t\in\mathbb{N}_0}$ on the resulting probability space.

We now reason intuitively by using the 'random walk' along the probability tree. Starting from $\square$, we transition to a first situation $x \in \mathscr{X}$ with probability $p_\square(x)$. From there, we could then perform the entire infinite random walk to generate the remainder of the path. So, a different way of saying this is that, of all the random paths $\omega \in \Omega$ that could be generated, a fraction of $p_\square(x)$ of them will start with $\omega(0) = x$. Using also the interpretation given by Corollary 5.1, it therefore makes sense to define the *first-step marginal measure* $P^*(X_0 = x) \coloneqq p_\square(x)$ for all $x \in \mathscr{X}$.

Let us now consider the next step, and assume the first step down the tree resulted in a situation $x \in \mathscr{X}$. Then, with probability $p_x(y)$, $y \in \mathscr{X}$, the next situation will be xy. In terms of paths that could be generated, a fraction of $p_x(y)$ of the paths that

satisfy $\omega(0) = x$ will furthermore satisfy $\omega(1) = y$. Therefore, we define for the *second*-step marginal measure $P^*(X_0 = x, X_1 = y) \coloneqq p_\Box(x)p_x(y)$.

Proceeding in this manner, for every situation $w \in \mathscr{X}^*$ with length $n+1$, $n \in \mathbb{N}_0$, we can compute the $(n+1)$-th step marginal measure as

$$P^*\big(X_{0:n} = w\big) \coloneqq p_\Box\big(w_0\big) \prod_{i=1}^{n} p_{w_0 \cdots w_{i-1}}\big(w_i\big)\,,$$

or in words, by multiplying all probabilities given by the local models of the situations encountered on the path from the root of the tree, down to the situation w.

A fundamental result in the measure-theoretic treatment of stochastic processes (known as the *Kolmogorov extension theorem*) states that the collection of all these n-th step marginal measures P^* induces ('coherently') a probability measure P on $(\Omega, \mathscr{F})$. Specifically, the finite n-th step marginals of P will correspond exactly to these n-th step marginal measures that we constructed from the probability tree. This establishes the connection between probability trees and discrete-time measure-theoretic stochastic processes, in that the latter can be constructed from the former.

For the other direction, so, to construct a probability tree from a given probability space $(\Omega, \mathscr{F}, P)$, we start with an event tree $(\mathscr{X}^*_\Box, \prec)$ and aim to construct the local models $p_{(\cdot)}$. Using the intuitive interpretation offered by Corollary 5.1, we start by setting $p_\Box(x) = P(X_0 = x)$ for all $x \in \mathscr{X}$. For all other situations $w \in \mathscr{X}^*$ with length $n+1$, $n \in \mathbb{N}_0$, the local model p_w is defined as the conditional measure constructed from Bayes' rule, i.e. for all $x \in \mathscr{X}$,

$$p_w(x) = P\big(X_{n+1} = x \,\big|\, X_{0:n} = w\big) = \frac{P\big(X_{0:n} = w, X_{n+1} = x\big)}{P\big(X_{0:n} = w\big)}\,. \tag{5.2}$$

This also establishes the connection in the other direction. It can be verified that, by now constructing from this probability tree a measure P^*, say, in the manner described above, we obtain again $P^* = P$; so, we conclude that this yields a one-to-one correspondence between probability trees and measure-theoretic stochastic processes.

It should be noted that the second direction in the preceding discussion has one (rather large) caveat: it does not work when there are partial paths that have zero probability to occur. This is because then Bayes' rule cannot define the conditional measure required to construct the local model for the situation corresponding to that partial path, since it would result in a division by zero.

To summarise, we can conclude that there is indeed a correspondence between the two representations that we have seen so far (up to some technical difficulties surrounding probabilities that are zero). We have seen that the graphical tree structure allows us to reason intuitively about how a stochastic process generates a sample path, by 'walking' from the root of the tree down its branches. As we will discuss next, we can also use this structure to 'reason backwards': from vertices

deep down in the tree back to the root. We will see that this allows one to intuitively derive *computational methods* for working with stochastic processes.

So, fix $n \in \mathbb{N}_0$, and let $f \in \mathscr{L}(\mathscr{X}^{n+1})$ be a real-valued function for which we aim to compute the expected value with respect to the random variables $X_{0:n}$ at the time points $0, \ldots, n \in \mathbb{N}_0$. Note that it suffices to consider this case, in the sense that any function defined on a subset of the variables $X_{0:n}$, can always be trivially extended to a function on all of them. Now first notice the following. For any situation $w \in \mathscr{X}^*$ with length $|w| = n + 1$, the value of f in w is easy to compute; it is simply $f(w)$. Hence in particular, the *expected value* of f, in w, is simply

$$\mathbb{E}\big[f(X_{0:n}) \,\big|\, X_{0:n} = w\big] = f(w)\,.$$

Recall that the situation w represents a node in the event tree. We will now 'pull back' the above expected value, to the time point $n - 1$. Consider therefore the parent situation of w in the probability tree; we will compute the expected value of f in this parent situation.

This parent is a situation v of length $|v| = |w| - 1 = n$, which entirely coincides with w: $v_i = w_i$ for all $i = 0, \ldots, n - 1$. Associated to v is the local probability model p_v which, as we have discussed above, represents the probability with which a random walk along the tree travels through the various children of v. In particular, such a random walk goes through the situation w, with probability $p_v(w_\top)$. Therefore, the contribution of the expected value in w, to the expected value in v, is the expected value in w weighted by $p_v(w_\top)$. Since this holds for all children of v, we can write

$$\mathbb{E}\big[f(X_{0:n}) \,\big|\, X_{0:(n-1)} = v\big] = \sum_{x \in \mathscr{X}} p_v(x)\mathbb{E}\big[f(X_{0:n}) \,\big|\, X_{0:(n-1)} = v, X_n = x\big]\,.$$

This 'pullback' operation is graphically illustrated in Fig. 5.2.

Now, observe that the above conditional expectation of f in v is itself a real-valued function in $\mathscr{L}(\mathscr{X}^n)$. Its value is determined by the states at times $0, \ldots, n - 1$. We can therefore repeat the above argument; we pull back to the parent of v, then to the parent of *that* situation and so on. Eventually, the parent that we are considering is the empty situation $\square$; we then finish by computing

$$\mathbb{E}\big[f(X_{0:n})\big] = \sum_{x \in \mathscr{X}} p_\square(x)\mathbb{E}\big[f(X_{0:n}) \,\big|\, X_0 = x\big]\,,$$

which is exactly the expected value of f that we started out wanting to compute.

This method to compute the expected value of a function by 'pulling back' the 'local', or conditional, expected values, uses the interpretation of a stochastic process as a probability tree. The method relies on a property that is called the *law of iterated expectation*, or alternatively the *law of total probability*. It can be

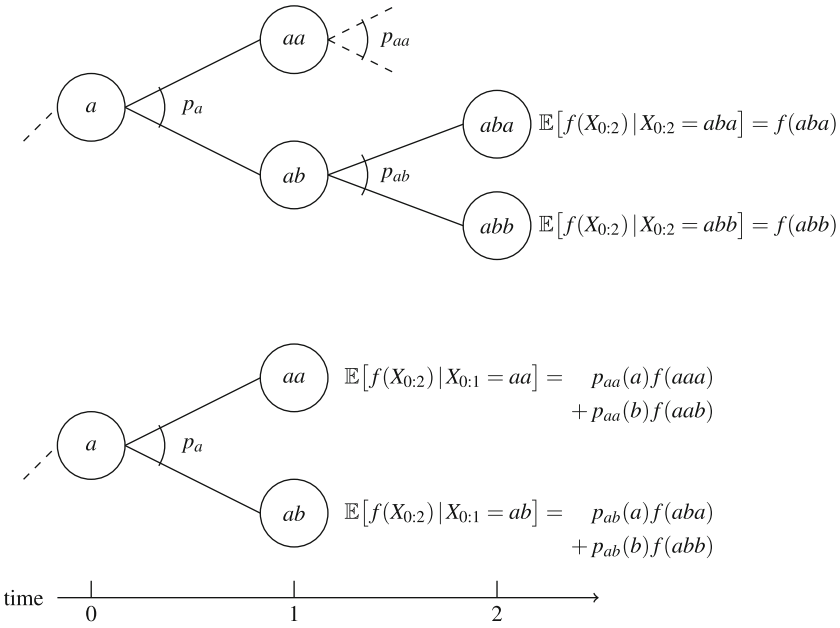

Fig. 5.2 Graphical illustration of 'pulling back' the expected value of a function f on $X_{0:2}$, in a probability tree on a binary state-space $\mathscr{X} = \{a, b\}$. Top: the function f is entirely determined by the situations of length 3, i.e. the expected value of the function in those situations is simply the value of the function evaluated in that situation. Bottom: the result after 'pulling back' the expectations by one step. The resulting conditional expectation is a function whose value is entirely determined by the situations of length 2. The values are the weighted average of the expectations in the child nodes, weighted by the local models $p_{(\cdot)}$

stated formally in the measure-theoretic context, where it is also easily stated for *continuous*-time stochastic processes.

Theorem 5.1 *Fix a time-dimension $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{R}_{\geq 0}\}$, and let $\{X_t\}_{t\in\mathbb{T}}$ be a stochastic process on $(\Omega, \mathscr{F}, P)$. Choose any three ordered sequences $\mathbf{s} = s_1, \ldots, s_n$; $\mathbf{t} = t_1, \ldots, t_m$ and $\mathbf{u} = u_1, \ldots, u_\ell$ in $\mathbb{T}$, with $n, m, \ell \in \mathbb{N}$ such that $s_n < t_1$ and $t_m < u_1$. Then for any real-valued function $f \in \mathscr{L}(\mathscr{X}^{n+m+\ell})$ on $X_\mathbf{s}$, $X_\mathbf{t}$, $X_\mathbf{u}$, it holds that*

$$\mathbb{E}\big[f(X_\mathbf{s}, X_\mathbf{t}, X_\mathbf{u})\,\big|\,X_\mathbf{s}\big] = \mathbb{E}\Big[\mathbb{E}\big[f(X_\mathbf{s}, X_\mathbf{t}, X_\mathbf{u})\,\big|\,X_\mathbf{s}, X_\mathbf{t}\big]\,\Big|\,X_\mathbf{s}\Big],$$

whenever $P(X_\mathbf{s})$ and $P(X_\mathbf{s}, X_\mathbf{t})$ are everywhere strictly positive.

In this result, the final constraint is required to ensure that the conditional expectations are all well-defined in the measure-theoretic sense. This point did not arise in the discussion using probability trees, because there the local (conditional) models are always properly defined by the model specification.

Having discussed how to interpret probability trees and how to use them to reason about the computation of expected values, we now move on to a discussion of their structural properties. Note that the specification of a probability tree is still relatively complicated. This is not really due to the structure of the tree; the situations $\mathscr{X}_\square^*$ and prefix order $\prec$ carry enough information to construct the tree up to any desired level, and their mathematical specification is straightforward. However, in order to specify all the local models $p_{(\cdot)}$, we need to provide an infinite number of probability mass functions on $\mathscr{X}$—one for each situation $w \in \mathscr{X}_\square^*$. This is why one often restricts attention to simpler models, where one needs fewer, and often only finitely many, local models.

These simplifications can be seen as a matter of degree. At the one extreme, we have the general definition that we used above, where each situation $w \in \mathscr{X}_\square^*$ has a local model p_w. This leads to a lot of possible structure but is hard to specify. At the other extreme is the *independent and identically distributed* (i.i.d.) process; this is when we only have a single probability mass function p, and we set $p_w \coloneqq p$ for all $w \in \mathscr{X}_\square^*$. For such a process, no matter what situation we are in, the next branch will always be chosen according to p. This process is easy to specify, but it does not yield a lot of structure that can capture the dynamics of the underlying system that we are trying to model.

A useful step up from the i.i.d. process is reached by the popular class of models known as *homogeneous Markov chains*. For a homogeneous Markov chain, the local model *only* depends on the last step of the corresponding situation, and not on what happened before that:

Definition 5.3 (Homogeneous Markov chain as probability tree) A probability tree $(\mathscr{X}_\square^*, \prec, p_{(\cdot)})$ is called a homogeneous Markov chain if $p_v = p_w$ for all situations $v, w \in \mathscr{X}^*$ such that $v_\top = w_\top$.

Corollary 5.2 *Let $(\mathscr{X}_\square^*, \prec, p_{(\cdot)})$ be a homogeneous Markov chain. Then $p_w = p_x$ for all $x \in \mathscr{X}$ and all $w \in \mathscr{X}^*$ such that $w_\top = x$.*

Proof Trivial from Definition 5.3 and the fact that all $x \in \mathscr{X}$ are also situations. □

An example for the binary state-space $\mathscr{X} = \{a, b\}$ is shown in Fig. 5.3. Additional degrees of freedom can be introduced back into this model by also letting the local models depend on the corresponding depth of the tree. The dynamics can then depend on the point in time, but not on the *specific* history up to that time. This yields the more general definition of a (non-homogeneous) *Markov chain*:

Definition 5.4 (Markov chain as probability tree) A probability tree $(\mathscr{X}_\square^*, \prec, p_{(\cdot)})$ is called a Markov chain if $p_v = p_w$ for all situations $v, w \in \mathscr{X}^*$ for which $|v| = |w|$ and $v_\top = w_\top$.

An example for the binary state-space $\mathscr{X} = \{a, b\}$ is shown in Fig. 5.4. It can be verified that a homogeneous Markov chain is a Markov chain, but not—in general—the other way around. Note that, in contrast to homogeneous Markov chains where we only needed to specify local models p_x for all $x \in \mathscr{X}$, we now need different

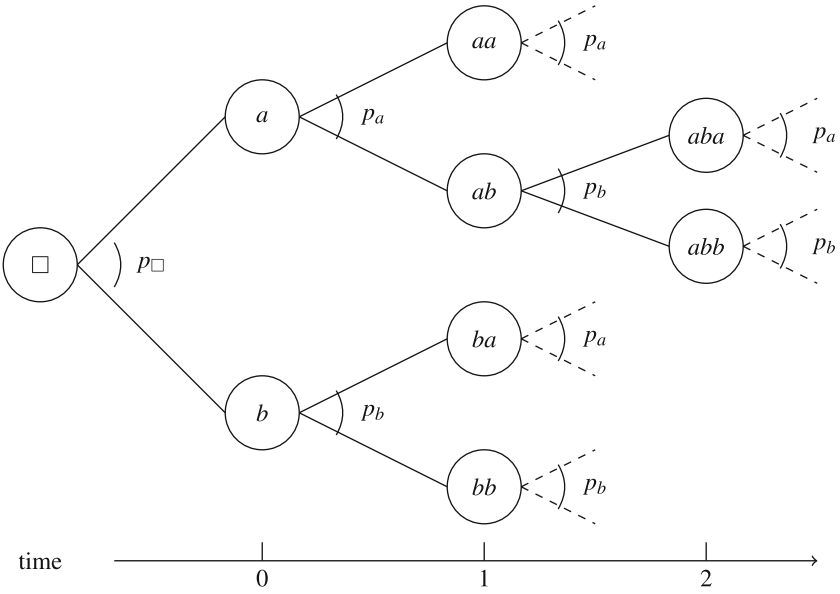

Fig. 5.3 A homogeneous Markov chain, represented as a probability tree

local models for each level of the tree. So, we are now back to needing an infinite number of local models in order to fully describe such a model.

These definitions of (homogeneous) Markov chains can also be conveniently translated back to the measure-theoretic context. We here give the general definition, for an arbitrary time-dimension (so, either $\mathbb{T} = \mathbb{N}_0$ or $\mathbb{T} = \mathbb{R}_{\geq 0}$) and multiple steps into the future:

Definition 5.5 (Markov chain as probability measure) A stochastic process $\{X_t\}_{t\in\mathbb{T}}$ on $(\Omega, \mathscr{F}, P)$ is called a Markov chain if for all $s_1, \ldots, s_n, t \in \mathbb{T}$, $n \in \mathbb{N}$, such that $s_1 < \cdots < s_n < t$, it holds that $P(X_t \mid X_{s_1}, \ldots, X_{s_n}) = P(X_t \mid X_{s_n})$. A stochastic process that is a Markov chain is said to have the Markov property.

Similarly, the notion of homogeneity can be defined measure-theoretically and for an arbitrary time-dimension:

Definition 5.6 (Homogeneous Markov chain as probability measure) A stochastic process $\{X_t\}_{t\in\mathbb{T}}$ on $(\Omega, \mathscr{F}, P)$ is called a homogeneous Markov chain if it is a Markov chain, and if additionally, for all $s, t \in \mathbb{T}$ such that $s < t$, it holds that $P(X_t \mid X_s) = P(X_{t-s} \mid X_0)$.

We leave it as an exercise to verify that, when $\mathbb{T} = \mathbb{N}_0$, Definitions 5.5 and 5.6 correspond to what we would expect from Definitions 5.4 and 5.3, respectively.

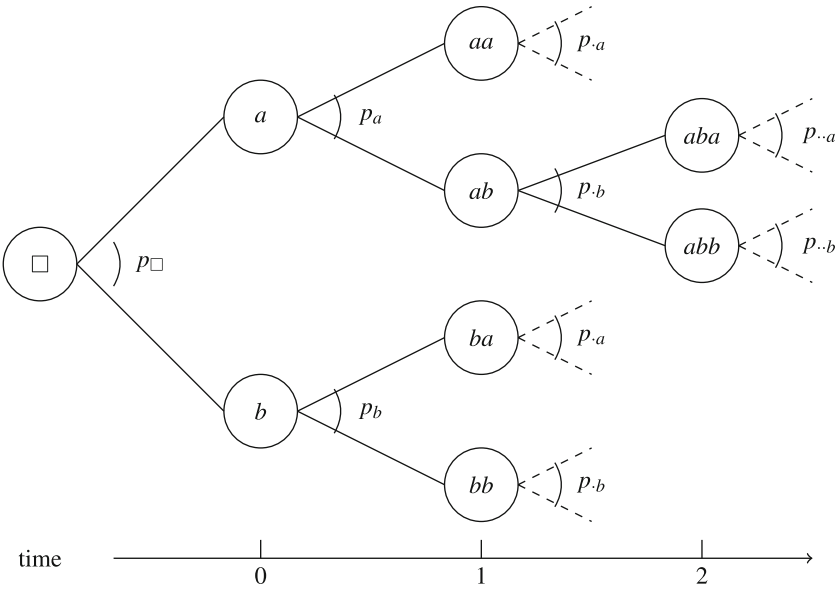

Fig. 5.4 A (non-homogeneous) Markov chain, represented as a probability tree as above

5.2.2 *Bayesian Networks*

We now move on to a different graphical representation of stochastic processes that is useful for Markov chains in particular: *Bayesian networks* (BNs), a specific type of probabilistic graphical model. While the graphical structure of probability trees in Sect. 5.2.1 emphasised the partial paths in the realisation of a stochastic process, the BN representation emphasises the individual random variables X_t.

The BN representation of a discrete-time Markov chain $\{X_t\}_{t\in\mathbb{N}_0}$ is given in Fig. 5.5. The structure is a directed acyclic graph, with one node associated to each random variable X_t and arcs representing the dependence of the receiving node's random variable's distribution, on the originating node's random variable's value. Due to the Markov property (c.f. Definition 5.5), each random variable X_n, $n \in \mathbb{N}$, is only ('directly') dependent on X_{n-1}, the value of the random variable immediately before it. The initial variable X_0 is somewhat of a special case, since it does not depend on any other variables; there are no time points preceding it. Due to these properties, the graphical structure is that of a chain; this may go some way in explaining the name 'Markov chain'. In the remainder of this section, we will refer to both a node in the BN and to its random variable, using the notation X_t.

It should be emphasised that the graphical structure is not saying that only nodes which are adjacent in the BN can influence each other. The formal interpretation is as follows: for any node X_n, $n \in \mathbb{N}$, *conditional on the value of the parent(s) of* X_n, the distribution of X_n is probabilistically independent of the non-parents,

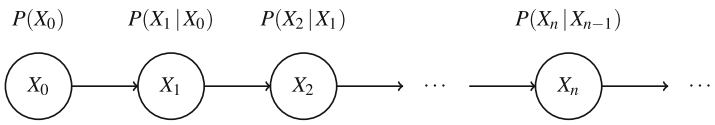

Fig. 5.5 Bayesian network representation of a discrete-time Markov chain $\{X_t\}_{t\in\mathbb{N}_0}$. Nodes represent random variables. An incoming arc on a node represents that the distribution of the corresponding random variable is influenced by the originating node of that arc. Correspondingly, each node associates a probability distribution to its random variable, conditional on the values of the random variables of the nodes on which it is dependent as before

non-descendants of X_n. This is the general interpretation of the independence properties of the arcs in a BN. In the special case of Markov chains that we are considering here, the interpretation vastly simplifies. Notably, the 'non-parents, non-descendants' of any node X_n are exactly its 'grandparents', 'great-grandparents' and so on; it is the set of nodes $\{X_m : m \in \mathbb{N}_0,\ m < n-1\}$.

Put differently, the value of X_n influences the distribution of *all* of its descendants (i.e. the nodes X_m, $m > n$), so long as we do not know the value of any of those descendants themselves. We will next consider how we can quantify this.

We start by observing that for each node X_n, $n \in \mathbb{N}$, we have the associated conditional probability $P(X_n \mid X_{n-1})$. Since the state-space $\mathscr{X}$ is taken to be finite, we can conveniently represent these conditional probabilities in a $|\mathscr{X}| \times |\mathscr{X}|$ matrix. For any $t \in \mathbb{N}_0$, this matrix T_t is defined, for all $x, y \in \mathscr{X}$, as

$$T_t(x, y) \coloneqq P(X_{t+1} = y \mid X_t = x)\,, \tag{5.3}$$

where the indexing is taken to be row-first. This matrix T_t is called the *transition matrix* of the Markov chain at time t. Its elements $T_t(x, y)$ are called the *transition probabilities from x to y*, and they are the probabilities that a system that is in state x at time t will be in state y at time $t+1$. This explains the subscript-indexing, whereby the matrix T_t contains the conditional probabilities associated to node X_{t+1}.

These transition matrices make it easy to connect back to the probability tree representation of Markov chains that we encountered earlier:

Proposition 5.1 *Let $(\mathscr{X}^*_\square, \prec, p_{(\cdot)})$ be a probability tree that is a Markov chain, and let T_t denote the associated family of transition matrices, as defined above. Then for all $t \in \mathbb{N}$ and all $w \in \mathscr{X}^*$ such that $|w| = t$, it holds that $p_w(y) = T_t(w_\top, y)$ for all $y \in \mathscr{X}$.*

Proof Use Eq. (5.2), Definition 5.5 and Eq. (5.3). □

The reason that we represent these probabilities using matrices is that this opens up the entire toolbox of linear algebra. We will see that this allows us to very succinctly write down certain relations and properties. For instance, we can now write the influence of a node on its descendants, using a simple matrix product:

Proposition 5.2 *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time Markov chain, and let T_t be the associated family of transition matrices, as defined above. Then for all $s, t \in \mathbb{N}_0$ such that $s \leq t$, and all $x, y \in \mathscr{X}$, it holds that $P(X_{t+1} = y \,|\, X_s = x) = [T_s \cdots T_t](x, y)$.*

Proof We give a proof by induction. For $t = s$ the result is immediate from the definition of the transition matrix T_s. Now suppose the result is true for $t-1$; we show that it is also true for t:

$$\begin{aligned} P(X_{t+1} = y \,|\, X_s = x) &= \sum_{z\in\mathscr{X}} P(X_{t+1} = y, X_t = z \,|\, X_s = x) \\ &= \sum_{z\in\mathscr{X}} P(X_t = z \,|\, X_s = x) P(X_{t+1} = y \,|\, X_t = z,\ X_s = x) \\ &= \sum_{z\in\mathscr{X}} [T_s \cdots T_{t-1}](x, z) P(X_{t+1} = y \,|\, X_t = z) \\ &= \sum_{z\in\mathscr{X}} [T_s \cdots T_{t-1}](x, z) T_t(z, y) = \big[T_s \cdots T_{t-1} T_t\big](x, y)\,, \end{aligned}$$

where the first and second equalities are basic properties of probabilities, the third equality is due to the induction hypothesis and the Markov property (c.f. Definition 5.5), the fourth equality uses the definition of the transition matrix T_t and the final equality uses the definition of a matrix product. □

Another useful property of this representation is that it allows us to write conditional expectations of functions $f \in \mathscr{L}(\mathscr{X})$ using matrix-vector products. In particular, again because $\mathscr{X}$ is finite, any $f \in \mathscr{L}(\mathscr{X})$ can be interpreted as a vector in $\mathbb{R}^{|\mathscr{X}|}$; the coordinates are simply the values $f(x)$, $x \in \mathscr{X}$. Hence:

Proposition 5.3 *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time Markov chain, and let T_t be the associated family of transition matrices. Then, for all $f \in \mathscr{L}(\mathscr{X})$, all $t \in \mathbb{N}_0$ and all $x \in \mathscr{X}$, it holds that $\mathbb{E}\big[f(X_{t+1}) \,|\, X_t = x\big] = [T_t f](x)$.*

Proof Simply use the definition of the matrix-vector product:

$$[T_t f](x) = \sum_{y\in\mathscr{X}} T_t(x, y) f(y) = \sum_{y\in\mathscr{X}} P(X_{t+1} = y \,|\, X_t = x) f(y) = \mathbb{E}\big[f(X_{t+1}) \,|\, X_t = x\big].$$

□

The above properties can be combined to give a simplified version of the law of iterated expectation (Theorem 5.1) that we encountered in Sect. 5.2.1:

Corollary 5.3 *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time Markov chain, and let T_t be the associated family of transition matrices. Then, for all $f \in \mathscr{L}(\mathscr{X})$, all $s, t \in \mathbb{N}_0$ such that $s \leq t$ and all $x \in \mathscr{X}$, it holds that $\mathbb{E}\big[f(X_{t+1}) \,|\, X_s = x\big] = [T_s \cdots T_t f](x)$.*

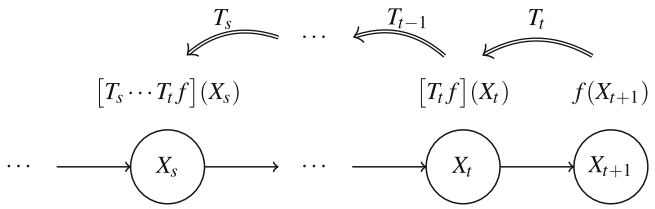

Fig. 5.6 Graphical representation of the 'pulling back' interpretation of the simplified version of the law of iterated expectation in Corollary 5.3. The function f, of which we want to compute the expectation on X_{t+1}, given X_s, starts at node X_{t+1}, where its value is trivial. The function is then 'pulled back' to the parent X_t of X_{t+1}, by taking the local expectation, by left-multiplying with T_t. This new function $T_t f$ on X_t is then 'pulled' back by multiplying with T_{t-1} and so forth. Eventually, the function $T_{s+1} \cdots T_t f$ is pulled into X_s, by left-multiplying with T_s. The resulting function on X_s is the conditional expectation of interest as before

Proof Immediate from Propositions 5.2 and 5.3. □

Note that, where the law of iterated expectation in Theorem 5.1 could be interpreted as 'pulling back' in the associated probability tree, the above simplified version can additionally be interpreted as 'pulling back' the conditional expectations in the associated BN, through the product of the transition matrices. This is graphically represented in Fig. 5.6.

5.2.3 Transition Graphs

We now move on to yet another graphical representation: the *transition graph* of a homogeneous (discrete-time) Markov chain. We start by noticing the following:

Proposition 5.4 *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time homogeneous Markov chain, and let T_t be the associated family of transition matrices. Then there is a unique matrix T such that $T_t = T$ for all $t \in \mathbb{N}_0$.*

Proof The matrix of interest can be identified as $T \coloneqq T_0$. Now, using the definition of a homogeneous Markov chain (Definition 5.6) and the transition matrix T_t for any $t \in \mathbb{N}_0$, it holds for all $x, y \in \mathscr{X}$ that

$$\begin{aligned} T(x,y) = T_0(x,y) &= P(X_1 = y \mid X_0 = x) \\ &= P(X_{(t+1)-t} = y \mid X_0 = x) = P(X_{t+1} = y \mid X_t = x) = T_t(x,y)\,, \end{aligned}$$

which concludes the proof; uniqueness is trivial. □

As an aside, note therefore that a discrete-time homogeneous Markov chain can be characterised (up to the initial distribution $P(X_0)$) by a single transition matrix T. In particular, this T can be seen as the canonical parameter of the Markov chain. This

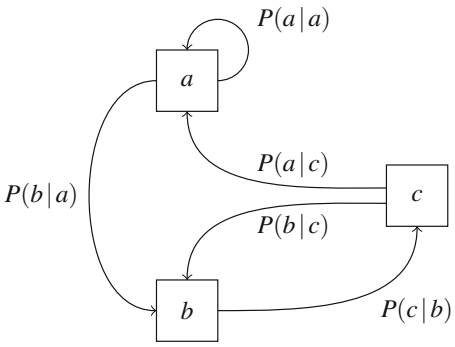

Fig. 5.7 Example transition graph for a discrete-time homogeneous Markov chain with a ternary state-space $\mathscr{X} = \{a, b, c\}$. The transition graph is a directed graph, with a vertex for each state and an arc from the vertex of x to that of y, with $x, y \in \mathscr{X}$, whenever $T(x, y) = P(X_1 = y \mid X_0 = x) > 0$. The arcs are labelled with the corresponding transition probabilities. The figure uses the shorthand notation $P(y|x)$ for the elements $T(x, y)$ of T as before

relative ease of parameterisation—compared to say an arbitrary stochastic process, which needs separate parameters for every possible history—is arguably one of the reasons that make homogeneous Markov chains such convenient and widely used models.

Moving on, the transition graph of a discrete-time homogeneous Markov chain is a graphical representation of its associated transition matrix T. In this way, this representation emphasises the interactions between the *states*, rather than the random variables. An example transition graph is shown in Fig. 5.7. The formal definition is as follows:

Definition 5.7 (Transition graph) Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time homogeneous Markov chain, and let T be its associated transition matrix. Then its associated *transition graph* is a directed graph (V, E) with one vertex for each state, $V = \mathscr{X}$, and, for all $x, y \in \mathscr{X}$, an arc $(x, y) \in E$ whenever $T(x, y) > 0$.

One of the reasons transition graphs are sometimes useful is that they allow one to study which parts of a system can be reached from other parts of the system. The simplest application is that of *communicating states*:

Definition 5.8 (Communicating states) Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time homogeneous Markov chain, and let T be its associated transition matrix. For any two states $x, y \in \mathscr{X}$, y is said to be *accessible* from x if there is some $n \in \mathbb{N}$ such that $T^n(x, y) > 0$. Furthermore, x and y are said to *communicate* if y is accessible from x, and x is accessible from y.

Note that in the above, the term T^n denotes the n-th matrix power of T (c.f. Proposition 5.2). This has an intuitive graphical interpretation:

Corollary 5.4 *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time homogeneous Markov chain. Then for any $x, y \in \mathscr{X}$, y is accessible from x if and only if there is a path from x to y*

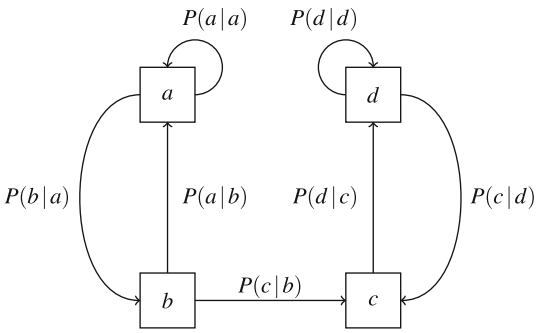

Fig. 5.8 Transition graph of a Markov chain that is *not* irreducible. It has two communication classes, $\{a, b\}$ and $\{c, d\}$. The set $\{c, d\}$ dominates $\{a, b\}$ and is the top (communication) class of the Markov chain. This Markov chain is top class regular

in the associated transition graph. Furthermore, x and y communicate if and only if there is a cycle in the associated transition graph that contains both x and y.

Proof Trivial from Definitions 5.7 and 5.8. □

Inspection of the transition graph in Fig. 5.7 shows that, in that example, all states communicate with each other. When this is the case, i.e. when all states communicate, the Markov chain is said to be *irreducible*. A maximal set of states that all communicate with each other is called a *communication class*. Hence, an irreducible Markov chain has only a single communication class, which is equal to $\mathscr{X}$.

Note that not every Markov chain is irreducible; in general there may be more than one communication class. An example is given in Fig. 5.8. When a communication class $\mathscr{A} \subset \mathscr{X}$ is accessible from a different communication class $\mathscr{B} \subset \mathscr{X}$, then $\mathscr{A}$ is said to dominate $\mathscr{B}$. A communication class which is not dominated is called *maximal*. When a Markov chain has only a single maximal communication class, this is called the *top (communication) class*.

Investigation of the communicating states in a Markov chain is often useful when one is interested in the long-term behaviour of the system. After all, while a system might begin in one state, it need not necessarily always eventually return to that state; this is the property that is illustrated in Fig. 5.8.

An important concept is that of the *regularity* of the communication classes of a Markov chain. A communication class is regular if there is a number $n \in \mathbb{N}$ such that it is possible to go from any state in the class to any other state in the class, in exactly n steps. Of particular importance is the notion of *top class regularity*:

Definition 5.9 (Top class regularity) Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time homogeneous Markov chain, and let T be its associated transition matrix. Then the Markov chain is said to be *top class regular* if

$$\left\{y \in \mathscr{X} \,:\, (\exists n \in \mathbb{N})(\forall x \in \mathscr{X})\, T^n(x, y) > 0\right\} \neq \emptyset\,,$$

and in that case the top class $\mathscr{X}_{\text{top}}$ of the Markov chain exists and is equal to this set. When furthermore $\mathscr{X}_{\text{top}} = \mathscr{X}$, the Markov chain itself is said to be *regular*.

The reason that this property is so important is that it provides a sufficient condition for the long-term behaviour of a Markov chain to converge to a stationary distribution, regardless of the state in which it started:

Theorem 5.2 *Let $\{X_t\}_{t\in\mathbb{N}_0}$ be a discrete-time homogeneous Markov chain, and let T be its associated transition matrix. Let this Markov chain be regular. Then there is a probability mass function $P_\infty : \mathscr{X} \to \mathbb{R}_{\geq 0}$ such that, for all $x, y \in \mathscr{X}$,*

$$P_\infty(y) = \lim_{n\to+\infty} T^n(x, y)\,.$$

5.3 Imprecise Discrete-Time Markov Chains

We will now move on to the discussion surrounding *imprecise (discrete-time) Markov chains* (IDTMCs). So, we still consider the time-dimension $\mathbb{T} = \mathbb{N}_0$. We will generalise each of the representations that we previously encountered to this new setting, where we roughly follow the same order as in Sect. 5.2.

So, let us start with the 'measure-theoretic' representation of imprecise stochastic processes. In this setting, we consider a set $\mathbb{P}$ of probability measures on the measurable space of paths $(\Omega, \mathscr{F})$. Then for each $P \in \mathbb{P}$, we have a probability space $(\Omega, \mathscr{F}, P)$, to which we can associate the precise stochastic process $\{X_t\}_{t\in\mathbb{N}_0}$ as in Definition 5.1. For any function $f \in \mathscr{L}(\mathscr{X}^n)$, $n \in \mathbb{N}$, we can express the expected value on the n time points $\mathbf{t} \subset \mathbb{N}_0$ as $\mathbb{E}_P[f(X_\mathbf{t})]$ as in Sect. 5.2. Recall from Chap. 2 that in this imprecise probabilistic context, we are more generally interested in the *lower* and *upper expectation* of f, which are defined, respectively, as

$$\underline{\mathbb{E}}_\mathbb{P}\big[f(X_\mathbf{t})\big] \coloneqq \inf_{P\in\mathbb{P}} \mathbb{E}_P\big[f(X_\mathbf{t})\big] \quad \text{and} \quad \overline{\mathbb{E}}_\mathbb{P}\big[f(X_\mathbf{t})\big] \coloneqq \sup_{P\in\mathbb{P}} \mathbb{E}_P\big[f(X_\mathbf{t})\big]\,.$$

We briefly recall the well-known conjugacy relation $\overline{\mathbb{E}}_\mathbb{P}\big[f(X_\mathbf{t})\big] = -\underline{\mathbb{E}}_\mathbb{P}\big[-f(X_\mathbf{t})\big]$, from which it follows that we can present the remainder of this discussion entirely in terms of lower expectations; any corresponding results on upper expectations follow directly through this relation.

Slightly more generally than the above, we will focus on *conditional* lower expectations. Similar to the precise case that we discussed before, these are defined for any $f \in \mathscr{L}(\mathscr{X}^{n+m})$, $n, m \in \mathbb{N}$, any $\mathbf{s}, \mathbf{t} \subset \mathbb{N}_0$ such that $\mathbf{s}$ and $\mathbf{t}$ are of length n and m, respectively, and any $x_\mathbf{s} \in \mathscr{X}^n$, as

$$\underline{\mathbb{E}}_\mathbb{P}\big[f(X_\mathbf{s}, X_\mathbf{t}) \,\big|\, X_\mathbf{s} = x_\mathbf{s}\big] \coloneqq \inf_{P\in\mathbb{P}} \mathbb{E}_P\big[f(X_\mathbf{s}, X_\mathbf{t}) \,\big|\, X_\mathbf{s} = x_\mathbf{s}\big]\,,$$

whenever $\underline{\mathbb{E}}_\mathbb{P}[\mathbb{I}_{x_\mathbf{s}}(X_\mathbf{s})] > 0$. In this last condition, $\mathbb{I}_{x_\mathbf{s}}$ is the indicator of $x_\mathbf{s}$; for all $y_\mathbf{s} \in \mathscr{X}^n$, $\mathbb{I}_{x_\mathbf{s}}(y_\mathbf{s}) \coloneqq 1$ if $x_\mathbf{s} = y_\mathbf{s}$ and $\mathbb{I}_{x_\mathbf{s}}(y_\mathbf{s}) \coloneqq 0$, otherwise. Note that then

$$0 < \underline{\mathbb{E}}_{\mathbb{P}}[\mathbb{I}_{x_{\mathbf{s}}}(X_{\mathbf{s}})] = \inf_{P \in \mathbb{P}} \mathbb{E}_P[\mathbb{I}_{x_{\mathbf{s}}}(X_{\mathbf{s}})] = \inf_{P \in \mathbb{P}} P(X_{\mathbf{s}} = x_{\mathbf{s}}) \,,$$

so this condition guarantees that the conditional expectations are well-defined for all the precise measures $P \in \mathbb{P}$. As before, there are formalisms where this condition is not strictly required—see, for example, the discussion around the local models of probability trees—or where it can be weakened. For simplicity, we keep the condition here to ensure that everything remains well-defined also under the measure-theoretic interpretation.

We are now ready to give the formal definition of an imprecise discrete-time Markov chain (IDTMC):

Definition 5.10 (IDTMC as set of processes) An *imprecise discrete-time Markov chain* is a set $\mathbb{P}$ of probability measures on the measurable space $(\Omega, \mathscr{F})$, with associated lower expectation operator $\underline{\mathbb{E}}_{\mathbb{P}}$ as defined above, such that, for all $f \in \mathscr{L}(\mathscr{X})$ and all $s_1, \ldots, s_n, t \in \mathbb{N}_0$ such that $s_1 < \cdots s_n < t$,

$$\underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_t) \,\big|\, X_{s_1}, \ldots, X_{s_n}\big] = \underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_t) \,\big|\, X_{s_n}\big] \,.$$

Furthermore, an imprecise discrete-time Markov chain is called *homogeneous* if, for all $s, t \in \mathbb{N}_0$, $s < t$, and all $f \in \mathscr{L}(\mathscr{X})$, it holds that $\underline{\mathbb{E}}_{\mathbb{P}}[f(X_t) \,|\, X_s] = \underline{\mathbb{E}}_{\mathbb{P}}[f(X_{t-s}) \,|\, X_0]$.

Let us compare this with Definition 5.5, the measure-theoretic definition of a precise Markov chain. The first difference is that the imprecise definition above is phrased in terms of (lower) expectations, whereas the precise definition used probabilities. We recall that this is because, in the framework of imprecise probability, it does not suffice to state results in terms of (lower) probabilities; instead the more general language of (lower) expectation operators is required.

Nevertheless, this definition implies that, in terms of lower probabilities,

$$\begin{aligned}\inf_{P \in \mathbb{P}} P(X_t = x \,|\, X_{s_1}, \ldots, X_{s_n}) &= \underline{\mathbb{E}}_{\mathbb{P}}\big[\mathbb{I}_x(X_t) \,\big|\, X_{s_1}, \ldots, X_{s_n}\big] \\ &= \underline{\mathbb{E}}_{\mathbb{P}}\big[\mathbb{I}_x(X_t) \,\big|\, X_{s_n}\big] = \inf_{P \in \mathbb{P}} P(X_t = x \,|\, X_{s_n}) \,,\end{aligned}$$

which displays this imprecise Markov condition in more familiar terms.

One may wonder at this point whether an imprecise Markov chain $\mathbb{P}$ is itself a set of Markov chains; the answer to this question is a resounding *no* (or at least, not necessarily). This point deserves the strongest possible emphasis:

> An element of an imprecise Markov chain $\mathbb{P}$ need **not** be a Markov chain! So, in general $P(X_t \,|\, X_{s_1}, \ldots, X_{s_n}) \neq P(X_t \,|\, X_{s_n})$ for $P \in \mathbb{P}$, with $s_1 < \cdots s_n < t$ in $\mathbb{N}_0$.

To clarify, the 'imprecise Markov condition' of an imprecise Markov chain is an 'independence' assessment about the *lower envelope* only. Formally, it is an assessment of *epistemic irrelevance*—a specific type of independence that arises in

imprecise probability theory—which is weaker than *strong independence* a different type of independence, and what would hold of all $P \in \mathbb{P}$ were Markov chains.

In a similar vein, the notion of homogeneity is here only enforced on the lower envelope. So, for an IDTMC $\mathbb{P}$ that is homogeneous, there may be processes $P \in \mathbb{P}$ that are neither Markov nor homogeneous.

The reason why we stress this so strongly is twofold. First of all, it implies that the structural assumptions of an imprecise Markov chain are in fact much weaker than those of a precise Markov chain—we no longer assume that future events are fully independent of the history, given the current state, or that their distribution is independent of the point in time. They might be, of course—there *are* elements $P \in \mathbb{P}$ that satisfy those properties—but it's not enforced as strictly. In other words, this model also represents 'higher-order' uncertainty about the *structural properties* of the system that we are trying to model.

The second reason is that this property is central to all the efficient computational methods that have been developed for working with imprecise Markov chains. We will next illustrate this point by moving the discussion to the representation of IDTMCs as *imprecise probability trees*.

5.3.1 Imprecise Probability Trees

Recall that for precise probability trees, we associate with each situation $w \in \mathscr{X}_\square^*$ a local model p_w, which is a probability mass function on $\mathscr{X}$. In contrast, in order to define *imprecise* probability trees, we will consider *imprecise local models*. Such an imprecise local model $\mathscr{P}_w$ is simply a *set* of probability mass functions on $\mathscr{X}$. This leads to the following definition:

Definition 5.11 (Imprecise probability tree) An imprecise probability tree is a tuple $(\mathscr{X}_\square^*, \prec, \mathscr{P}_{(\cdot)})$, where $(\mathscr{X}_\square^*, \prec)$ is an event tree and $\mathscr{P}_{(\cdot)}$ is a set-valued function such that, for all $w \in \mathscr{X}_\square^*$, $\mathscr{P}_w$ is a non-empty set of probability mass functions on $\mathscr{X}$.

An obvious question is how one should interpret such imprecise probability trees. As a first step, we consider the (precise) probability trees that are *compatible* with a given imprecise probability tree:

Definition 5.12 Let $(\mathscr{X}_\square^*, \prec, \mathscr{P}_{(\cdot)})$ be an imprecise probability tree. Then a (precise) probability tree $(\mathscr{X}_\square^*, \prec, p_{(\cdot)})$ is called *compatible* with this imprecise probability tree, if $p_w \in \mathscr{P}_w$ for all $w \in \mathscr{X}_\square^*$.

This immediately lets us connect back to the sets-of-measures that we discussed before. Specifically, consider an imprecise probability tree $(\mathscr{X}_\square^*, \prec, \mathscr{P}_{(\cdot)})$, and suppose the tree $(\mathscr{X}_\square^*, \prec, p_{(\cdot)})$ is compatible with it. Then, using the method outlined in Sect. 5.2.1, we can associate a (precise) measure P to this precise tree. Collecting in the set $\mathbb{P}$ all the associated measures of all precise trees that are compatible with the imprecise tree, we obtain a set representation as in Sect. 5.3.

The connection in the other direction is analogous but a bit more subtle. In particular, if we start from an IDTMC $\mathbb{P}$, then each $P \in \mathbb{P}$ induces a precise probability tree. Using the local models of this tree, we can construct set-valued local models by simply varying P over $\mathbb{P}$. These set-valued local models can then be used to construct an imprecise probability tree. Clearly, there are then precise trees that are compatible with this imprecise tree, and each such precise tree induces a precise measure P'. However, and this is the crucial observation, it is in general *not* guaranteed that such P' are included in $\mathbb{P}$!

As a simple example, suppose that $\mathscr{X} = \{a, b\}$ and we start with a set $\mathbb{P}$ containing only two i.i.d. processes, whose local models are given by p, h, respectively. Then, the induced imprecise probability tree has local models $\mathscr{P}_w = \{p, h\}$ for all $w \in \mathscr{X}_\square^*$. On the other hand, we can easily construct a non-i.i.d. process such that, for all $w \in \mathscr{X}_\square^*$, its local model is $p_w = p$ if $w_\top = a$ and $p_w = h$, otherwise. Then clearly this process was not in the original set $\mathbb{P}$, but it is compatible with the imprecise probability tree.

To prevent this from happening, we will require that the set representation $\mathbb{P}$ of the IDTMC is 'large enough'. Specifically, what we need is that it is already closed under such 'recombination' of local models at different points in time. Whenever this property holds, we will say that the IDTMC is *separately specified*. Clearly, when we start from an imprecise probability tree and construct its set of compatible processes, this IDTMC will then satisfy this property. In the remainder of this section, we will assume that a given set $\mathbb{P}$ is indeed separately specified. Further on, when we consider the parametrisation of an IDTMC, we will consider an easy condition that ensures this will hold.

With this connection between the two representations in place, we can again start to consider computational methods for lower expectations. Analogous to what we have seen before, in this context we have a *law of iterated lower expectation* that we can use as a computational tool. The imprecise probability tree representation again provides graphical intuition.

Similar to the exposition in Sect. 5.2.1, we start with a function $f \in \mathscr{L}(\mathscr{X}^{n+1})$ of which we want to compute the lower expectation with respect to the states at the time points $0, \ldots, n$. Then for any situation $w \in \mathscr{X}^*$ such that $|w| = n + 1$, the lower expectation is trivial:

$$\underline{\mathbb{E}}_{\mathbb{P}}\left[f(X_{0:n}) \,\middle|\, X_{0:n} = w\right] = f(w)\,.$$

We then again 'pull back' to the parent situation v of w; this is where the main difference with Sect. 5.2.1 occurs. Notably, we here have an *imprecise* local model $\mathscr{P}_v$ associated to this node v. The point to the law of iterated lower expectation is that it suffices to only compute the associated conditional lower expectation locally:

$$\underline{\mathbb{E}}_{\mathbb{P}}\left[f(X_{0:n}) \,\middle|\, X_{0:(n-1)}{=}v\right] = \inf_{p_v \in \mathscr{P}_v} \sum_{x \in \mathscr{X}} p_v(x) \underline{\mathbb{E}}_{\mathbb{P}}\left[f(X_{0:n}) \,\middle|\, X_{0:(n-1)}{=}v,\, X_n{=}x\right]\,.$$

Exactly analogous to the precise case, by repeatedly pulling back until we reach the root of the tree, we eventually compute

$$\underline{\mathbb{E}}_{\mathbb{P}}[f(X_{0:n})] = \inf_{p_{\Box} \in \mathscr{P}_{\Box}} \sum_{x \in \mathscr{X}} p_{\Box}(x) \underline{\mathbb{E}}_{\mathbb{P}}\left[f(X_{0:n}) \,\middle|\, X_0 = x\right],$$

which is the lower expectation of interest.

As before, the need to specify these (imprecise) local models $\mathscr{P}_w$ for all situations $w \in \mathscr{X}_{\Box}^*$ makes such a model difficult to work with. This is simplified for imprecise Markov chains; note that we here assume the analogue of homogeneity to hold implicitly:

Definition 5.13 (Homogeneous IDTMC as imprecise probability tree) An imprecise probability tree $(\mathscr{X}_{\Box}^*, \prec, \mathscr{P}_{(\cdot)})$ is called an imprecise homogeneous discrete-time Markov chain if $\mathscr{P}_v = \mathscr{P}_w$ for all $v, w \in \mathscr{X}^*$ for which $v_{\top} = w_{\top}$.

Corollary 5.5 *Let $(\mathscr{X}_{\Box}^*, \prec, \mathscr{P}_{(\cdot)})$ be a homogeneous IDTMC. Then $\mathscr{P}_w = \mathscr{P}_x$ for all $x \in \mathscr{X}$ and all $w \in \mathscr{X}^*$ such that $w_{\top} = x$.*

Proof Trivial from Definition 5.13 and the fact that all $x \in \mathscr{X}$ are also situations. □

As above, an IDTMC $(\mathscr{X}_{\Box}^*, \prec, \mathscr{P}_{(\cdot)})$ has a set of compatible precise probability trees, each of which induces a measure P, and these are collected in the set $\mathbb{P}$, which is the measure-theoretic IDTMC representation from Definition 5.10. Observe that a precise probability tree does *not* have to be a (homogeneous) Markov chain, for it to be compatible with a given IDTMC! That is, to be compatible, each local model p_w, $w \in \mathscr{X}_{\Box}^*$, should be in the set $\mathscr{P}_{w_{\top}}$, and this set depends only on the most recent state $w_{\top}$ of the situation w. But, while in a different situation v such that $v_{\top} = w_{\top}$, we do require that $p_v \in \mathscr{P}_{v_{\top}} = \mathscr{P}_{w_{\top}}$; we do *not* require that $p_v = p_w$!

We will next illustrate that the law of iterated lower expectation simplifies further for imprecise Markov chains. We do this again by considering the imprecise counterpart of Bayesian networks.

5.3.2 Credal Networks

We here consider the graphical representation of imprecise Markov chains as *credal networks*. This is the imprecise generalisation of the Bayesian network representation that we encountered in Sect. 5.2.2. The graphical structure is as before, with the notable differences being (i) the local models (which are here replaced with imprecise local models) and (ii) the interpretation of the independence properties induced by the arcs. Regarding the second point, it suffices for our present purpose to note that we interpret the structure as a credal network under epistemic irrelevance. This then has the same consequence as that stated in the beginning of Sect. 5.3: given the value of the parent of a node X_t, $t \in \mathbb{N}_0$, the lower expectation

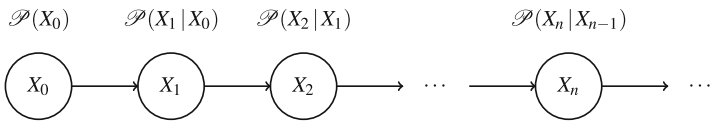

Fig. 5.9 Credal network representation of an imprecise discrete-time Markov chain. An incoming arc on a node represents that the local uncertainty model of the corresponding variable is influenced by the originating node of that arc. Correspondingly, each node associates an imprecise probability model to its variable, conditional on the values of the random variables of the nodes on which it is dependent

of any function dependent on X_t does not depend on the values of the non-parents, non-descendants (again, grandparents and so on) of X_t. For reference, the graphical representation is drawn in Fig. 5.9.

The interpretation in terms of sets of distributions is as would be expected; the model induces a set $\mathbb{P}$, each $P \in \mathbb{P}$ of which satisfies $P(X_n \mid X_{n-1}) \in \mathscr{P}(X_n \mid X_{n-1})$ for all $n \in \mathbb{N}$, and $P(X_0) \in \mathscr{P}(X_0)$. As before, the independence assumptions are not necessarily required to hold for these compatible precise models. Conversely, if we are given an IDTMC $\mathbb{P}$, then the local models $\mathscr{P}(X_n|X_{n-1})$ of the credal network are constructed by restricting attention to the conditional events $P(X_n \mid X_{n-1})$ and varying P over $\mathbb{P}$.

Similar to the discussion around the interpretation of imprecise probability trees, we here also need some 'closedness' assumptions to ensure this duality of representations holds. Specifically, we again require that $\mathbb{P}$ is separately specified. Furthermore, it is assumed that the local models $\mathscr{P}(X_n \mid X_{n-1})$ of the credal network have *separately specified rows*. This means that these local models are not arbitrary sets of conditional probabilities. If we let $\mathscr{P}(X_n \mid X_{n-1} = x) \coloneqq \left\{P(X_n \mid X_{n-1} = x) \in \mathscr{P}(X_n \mid X_{n-1})\right\}$ for all $x \in \mathscr{X}$, then what we require is that

$$\mathscr{P}(X_n \mid X_{n-1}) = \times_{x \in \mathscr{X}} \mathscr{P}(X_n \mid X_{n-1} = x)\,. \tag{5.4}$$

Under these conditions, we can straightforwardly switch between representations.

We next generalise the exposition in Sect. 5.2.2 regarding the associated transition matrices. To this end, fix any $t \in \mathbb{N}_0$. Then, as in the precise case, each element $P(X_{t+1} \mid X_t) \in \mathscr{P}(X_{t+1} \mid X_t)$ induces a transition matrix T_t. So, let us now consider the *set* $\mathscr{T}_t$ of transition matrices that is induced by the imprecise local models:

$$\mathscr{T}_t \coloneqq \Big\{T_t : \big(\forall x, y \in \mathscr{X} : T_t(x, y) = P(X_{t+1} = y \mid X_t = x)\big),$$
$$P(X_{t+1} \mid X_t) \in \mathscr{P}(X_{t+1} \mid X_t)\Big\}\,.$$

A key insight is that we can use this set of transition matrices to define a convenient computational tool for lower expectations:

Definition 5.14 Let $\mathbb{P}$ be an IDTMC, and let $\mathscr{T}_t$ be the associated family of sets of transition matrices, as defined above. Then, for each $t \in \mathbb{N}_0$, the associated *lower transition operator* $\underline{T}_t : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$ is defined, for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{X}$, as

$$\big[\underline{T}_t f\big](x) \coloneqq \inf_{T_t \in \mathscr{T}_t} \big[T_t f\big](x)\,.$$

This lower transition operator essentially fulfils the same role as the transition matrices from which it is derived. In particular, we have the following:

Proposition 5.5 *Let $\mathbb{P}$ be an IDTMC, and let $\underline{T}_t$ be the associated family of lower transition operators. Then, for all $f \in \mathscr{L}(\mathscr{X})$, all $t \in \mathbb{N}_0$ and all $x \in \mathscr{X}$, it holds that*

$$\big[\underline{T}_t f\big](x) = \underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_{t+1}) \,|\, X_t = x\big]\,.$$

Proof Simply use the definitions together with Proposition 5.3:

$$\begin{aligned}
\big[\underline{T}_t f\big](x) &= \inf_{T_t \in \mathscr{T}_t} \big[T_t f\big](x) = \inf_{T_t \in \mathscr{T}_t} \sum_{y \in \mathscr{X}} f(y) T_t(x, y) \\
&= \inf_{P(X_{t+1}|X_t) \in \mathscr{P}(X_{t+1}|X_t)} \sum_{y \in \mathscr{X}} f(y) P(X_{t+1}{=}y | X_t = x) \\
&= \inf_{P \in \mathbb{P}} \sum_{y \in \mathscr{X}} f(y) P(X_{t+1}{=}y | X_t{=}x) \\
&= \inf_{P \in \mathbb{P}} \mathbb{E}_P\big[f(X_{t+1}) \,\big|\, X_t{=}x\big] = \underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_{t+1}) \,\big|\, X_t{=}x\big]\,,
\end{aligned}$$

where in the fourth equality, we used the definition of the compatible measures. □

As in Corollary 5.3, we can now state the simplified law of iterated lower expectation for imprecise Markov chains, using these lower transition operators:

Theorem 5.3 *Let $\mathbb{P}$ be an IDTMC that is separately specified, and let $\underline{T}_t$ be the associated family of lower transition operators. Then, for all $f \in \mathscr{L}(\mathscr{X})$, all $s, t \in \mathbb{N}_0$ such that $s \leq t$ and all $x \in \mathscr{X}$, it holds that*

$$\underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_t) \,\big|\, X_s = x\big] = \big[\underline{T}_s \cdots \underline{T}_t f\big](x)\,,$$

where the right-hand side represents an iterated operator product (composition).

We omit the full proof, but the interested reader can reconstruct the argument by using the general computational process of iterated lower expectation as explained in Sect. 5.3.1, the imprecise Markov property from Definition 5.10 and the interpretation of the lower transition operator from Proposition 5.5.

5.3.3 Limits of Homogeneous IDTMCs

We conclude the discussion of imprecise discrete-time Markov chains with some results about their limit behaviour, in analogy to the results in Sect. 5.2.3. We start again by restricting attention to homogeneous IDTMCs, and notice the following (we omit the proof, which is straightforward):

Proposition 5.6 *Let $\mathbb{P}$ be a homogeneous IDTMC, and let $\underline{T}_t$ be the associated family of lower transition operators. Then there is a unique lower transition operator $\underline{T} : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$, such that, for all $f \in \mathscr{L}(\mathscr{X})$, $\underline{T}_t f = \underline{T} f$ for all $t \in \mathbb{N}_0$.*

We take a moment here to remark on a property that was already encountered in Chap. 2: the duality between lower expectation operators and closed and convex sets of probability measures. Indeed, this correspondence was also used in Definition 5.14 above, where we used the sets $\mathscr{T}_t$ of transition matrices, to construct the lower transition operator $\underline{T}_t$. Since, as we have just seen, the dynamics of a *homogeneous* IDTMC can be completely described by a single $\underline{T}$, it now makes sense to think about the other direction.

Specifically, corresponding to $\underline{T}$, there exists a closed and convex set $\mathscr{T}$ of transition matrices, such that $\underline{T} f = \inf_{T \in \mathscr{T}} Tf$ for all $f \in \mathscr{L}(\mathscr{X})$. This implies that (up to the initial distribution at time zero) an IDTMC can also be characterised by such a set $\mathscr{T}$. So, whereas we noted in Sect. 5.2.2 that a (precise) discrete-time Markov chain's canonical parameter is a single transition matrix T, for a homogeneous IDTMC, the parameter can be understood as a single closed and convex set $\mathscr{T}$ of transition matrices. Moreover, if in this parametrisation we ensure that $\mathscr{T}$ has *separately specified rows*—essentially, satisfies a property exactly analogous to Eq. (5.4)—then the corresponding IDTMC will also be separately specified.

Furthermore, in Sect. 5.2.2 we used a property of the associated transition matrix T, to state a sufficient condition for the long-term behaviour of the Markov chain to converge to a distribution over the states, independently of the state in which it started. We here have a similar result, which starts by introducing the conjugate *upper* transition operator $\overline{T} : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto -\underline{T}(-f)$.

Now, recall that in the precise case, a homogeneous discrete-time Markov chain with transition matrix T was said to be *regular*, if there was some $n \in \mathbb{N}$ such that $T^n(x, y) > 0$ for all $x, y \in \mathscr{X}$. The interpretation is clear: the Markov chain is regular if and only if there is some finite number of steps n in which every state x can reach every state y. This is now generalised to the imprecise case:

Definition 5.15 (Regularity for homogeneous IDTMC) Let $\mathbb{P}$ be a homogeneous IDTMC with associated lower (and upper) transition operator $\underline{T}$ (and $\overline{T}$). Then the IDTMC is *regular* if there is some $n \in \mathbb{N}$ such that $\left[\overline{T}^n \mathbb{I}_y\right](x) > 0$ for all $x, y \in \mathscr{X}$.

Let us consider this definition. One difference with the precise case is the introduction of the indicator function $\mathbb{I}_y$ on the state $y \in \mathscr{X}$; this was introduced because,

in contrast to matrices, we cannot index the 'elements' of the transition operator. Specifically, using Theorem 5.3, we can interpret the condition as

$$0 < \left[\overline{T}^n \mathbb{I}_y\right](x) = \overline{\mathbb{E}}_{\mathbb{P}}\left[\mathbb{I}_y(X_n) \,\middle|\, X_0 = x\right] = \sup_{P \in \mathbb{P}} P(X_n = y \mid X_0 = x)\,,$$

for all $x, y \in \mathscr{X}$ and some $n \in \mathbb{N}$. What regularity asks for, then, is for there to be some $n \in \mathbb{N}$ such that is possible for all $x, y \in \mathscr{X}$ to move from x to y in exactly n steps, *according to some* $P \in \mathbb{P}$. In particular, the (precise) measure P for which this needs to be possible can be different for every pair $x, y \in \mathscr{X}$. Regularity for IDTMCs then is in a sense a much weaker—easier to satisfy—condition than that for precise Markov chains. Nevertheless, the condition is sufficient for the following:

Theorem 5.4 *Let $\mathbb{P}$ be a homogeneous IDTMC that is separately specified and regular, with associated lower transition operator $\underline{T}$. Then, there is a unique lower expectation operator $\underline{\mathbb{E}}_{\mathbb{P}}[\cdot(X_{+\infty})] : \mathscr{L}(\mathscr{X}) \to \mathbb{R}$ such that, for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{X}$,*

$$\underline{\mathbb{E}}_{\mathbb{P}}\left[f(X_{+\infty})\right] = \lim_{n \to +\infty} \underline{\mathbb{E}}_{\mathbb{P}}\left[f(X_n) \mid X_0 = x\right] = \lim_{n \to +\infty} \left[\underline{T}^n f\right](x)\,.$$

Furthermore, this is the unique $\underline{T}$-invariant lower expectation on $\mathscr{L}(\mathscr{X})$, meaning that $\underline{\mathbb{E}}_{\mathbb{P}}[f(X_{+\infty})] = \underline{\mathbb{E}}_{\mathbb{P}}\left[[\underline{T} f](X_{+\infty})\right]$ for all $f \in \mathscr{L}(\mathscr{X})$.

5.4 Imprecise Continuous-Time Markov Chains

We now move on to the discussion about (imprecise) *continuous*-time Markov chains. We have already encountered this setting several times in the preceding discussions but have generally skipped over any details. Let us recall from Sect. 5.2 that continuous-time stochastic processes are identified with a time-dimension $\mathbb{T} = \mathbb{R}_{\geq 0}$ and that the elements ω of the outcome space of paths Ω are maps $\omega : \mathbb{R}_{\geq 0} \to \mathscr{X}$. The measure-theoretic definition is then as before, where we consider the abstract probability space $(\Omega, \mathscr{F}, P)$, and the stochastic process $\{X_t\}_{t \in \mathbb{R}_{\geq 0}}$ is a family of random variables on this space. Furthermore, measure-theoretic definitions of (homogeneous) continuous-time Markov chains (CTMCs) have already been encountered in Definitions 5.5 and 5.6.

How, then, can these models be interpreted? Let us start by considering the simplest case, *viz.*, a precise and homogeneous Markov chain in continuous-time. According to the previous definitions, this is a stochastic process such that

1. $P(X_t \mid X_{s_1}, \ldots, X_{s_n}) = P(X_t \mid X_{s_n})$ for all $s_1 < \cdots < s_n < t$ in $\mathbb{R}_{\geq 0}$, and
2. $P(X_t \mid X_s) = P(X_{t-s} \mid X_0)$ for all $s < t$ in $\mathbb{R}_{\geq 0}$.

The immediate difficulty of moving on from this abstract representation is that the time-dimension is now, in a sense, too big to use any of the previous representations.

For instance, we could try to draw a 'continuous-time' probability tree, where the local model of a situation with terminal state $w_\top$ is given by a probability mass function $P(X_t \,|\, X_0 = w_\top)$. But what is the time t that we should use? When we were working in discrete-time, the approach was to use the *next* time point, as viewed from the current situation. But of course, there is no 'next' time t when working in continuous-time! This difficulty of using graphical representations is the main reason that we have postponed the treatment of continuous-time processes until now, thereby hopefully allowing the reader to first develop some graphical intuition for the discrete-time case.

Nevertheless, all is not lost; the first interpretation that we will consider is to view continuous-time processes as limits of discrete-time ones. To this end, it will be convenient to consider the transition-matrix T associated with a homogeneous DTMC. Let us recall from Sects. 5.2.2 and 5.2.3 that the elements of such a matrix represent the 'transition probabilities' of the system, that is, the probability of moving from a state x to a state y, in one time step:

$$T(x, y) = P(X_1 = y \,|\, X_0 = x)\,.$$

We can use this formalism to interpret the continuous-time case, by simply 'fixing the length of the step'. That is, consider some 'step size' $\Delta > 0$. Then, for a homogeneous CTMC, we known that

$$P(X_{t+\Delta} \,|\, X_t) = P(X_\Delta \,|\, X_0)\,,$$

for all $t \in \mathbb{R}_{\geq 0}$, so we can collect these 'transition probabilities' in a matrix T_Δ:

$$T_\Delta(x, y) \coloneqq P(X_\Delta = y \,|\, X_0 = x) \qquad \text{for all } x, y \in \mathscr{X}\,.$$

Clearly, the elements of T_Δ are the probabilities for the system to end up in a state y, if it is currently in a state x, after a time duration of Δ has elapsed. Provided, then, that we are not interested in a granularity of the time-dimension that is finer than Δ, this representation suffices. The matrix T_Δ can be associated with a DTMC, and all the previous results can be used. For instance, for any multiple $n \in \mathbb{N}$ of Δ, we use Proposition 5.2 to find that

$$P(X_{n\Delta} = y \,|\, X_0 = x) = T_\Delta^n(x, y)\,.$$

But, of course, the point of using the continuous-time representation is that we *are* interested in an arbitrarily fine granularity of the time-dimension. In particular, the measure-theoretic definition encodes this arbitrary granularity, and it seems a waste to only focus on the restriction to a single step size Δ. The 'trick', then, is to take the limit as Δ goes to zero, and somehow usefully represent this limit. It is hopefully clear from the above discussion that, as we decrease Δ further and further, the associated transition matrix T_Δ covers increasingly smaller steps along the time-

dimension. And, for each such positive Δ, we can associate a discrete-time Markov chain and use all the previous interpretations that we developed.

We first remark that the naive limit does not encode a lot of information; ignoring possible issues of continuity, it trivially holds that

$$\lim_{\Delta \to 0^+} P(X_\Delta = y \mid X_0 = x) = P(X_0 = y \mid X_0 = x) = \begin{cases} 1 & \text{if } y = x, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{5.5}$$

In matrix notation this reads as $\lim_{\Delta \to 0^+} T_\Delta = I$, where I denotes the $|\mathscr{X}| \times |\mathscr{X}|$ identity matrix. Colloquially, we might understand this as saying that 'if time does not evolve, the system does not change'. This is clearly an almost tautological statement to make of what may be interpreted as a dynamical system. So let us consider how the system *does* change as time evolves. The natural representation for this is obviously the *derivative* of the transition matrix T_Δ; this is the limit interpretation that we shall use. Ignoring technical issues of differentiability, we have

$$\left.\frac{\mathrm{d}\, T_\Delta}{\mathrm{d}\, \Delta}\right|_{\Delta=0} = \lim_{\Delta \to 0^+} \frac{T_\Delta - I}{\Delta} =: Q\,, \tag{5.6}$$

where we have used the previous observation that $T_0 = I$. On the right-hand side, the term Q is called the *transition rate matrix* of the homogeneous CTMC (or sometimes simply the *rate matrix*). It is clear from the above definition that it encodes the *rate of change* of the transition probabilities around time zero. It satisfies the following properties:

Definition 5.16 (Transition Rate Matrix) A real-valued $|\mathscr{X}| \times |\mathscr{X}|$ matrix Q is called a *transition rate matrix* if, for all $x \in \mathscr{X}$, it holds that

1. $Q(x, y) \geq 0$ for all $y \in \mathscr{X}$ such that $x \neq y$ and
2. $\sum_{y \in \mathscr{X}} Q(x, y) = 0$.

The elements $Q(x, y)$ of a rate matrix can be interpreted as the 'speed' with which the process moves from the state x to the state y. In the above definition, the two conditions imply that the diagonal elements $Q(x, x)$ are always non-positive. On the other hand, the first condition states that the off-diagonal elements are non-negative. Combined this can be understood as saying that the system will move 'out' of the current state (the non-positivity of the diagonal elements) and 'into' some other states (the non-negativity of the off-diagonals).

A more concrete way to interpret the rate-matrix is through a linearised approximation of the transition probabilities over a small enough time step. That is, it follows from Eq. (5.6) that, for 'small enough' $\Delta > 0$, it holds that $Q \approx (T_\Delta - I)^{1}/_{\Delta}$; hence also

$$T_\Delta \approx I + \Delta Q\,. \tag{5.7}$$

We therefore see that the matrix Q can be used to approximately compute the transition probabilities over a small enough time step.

An obvious next question is if we can extrapolate this to compute the matrix T_t that contains the transition probabilities over an arbitrary duration t. Indeed we can, although it requires a bit of setup. For any $t \in \mathbb{R}_{\geq 0}$, first define the transition matrix of the CTMC after time t:

$$T_t(x, y) \coloneqq P(X_t = y \mid X_0 = x) \qquad \text{for all } x, y \in \mathscr{X} .$$

Then we differentiate in t; to this end, first fix $\Delta > 0$, and use the Markov property and homogeneity to derive that $T_{t+\Delta} = T_t T_\Delta = T_\Delta T_t$ (c.f. Proposition 5.2). Then we proceed by using Eq. (5.6):

$$\frac{\mathrm{d}\, T_t}{\mathrm{d}\, t} = \lim_{\Delta \to 0^+} \frac{T_{t+\Delta} - T_t}{\Delta} = \lim_{\Delta \to 0^+} \frac{T_\Delta T_t - T_t}{\Delta} = \left(\lim_{\Delta \to 0^+} \frac{T_\Delta - I}{\Delta} \right) T_t = Q T_t .$$

Using also Eq. (5.5), we can now write the matrix differential equation

$$\frac{\mathrm{d}\, T_t}{\mathrm{d}\, t} = Q T_t , \qquad T_0 = I ,$$

whose solution is the *matrix exponential* of Qt:

$$T_t = e^{Qt} .$$

We recall from Proposition 5.4 that the dynamic behaviour of a homogeneous discrete-time Markov chain can be characterised by a single transition matrix T and that therefore this matrix constitutes the canonical parameter of the process. Because the matrix Q can be used to (re-)construct the transition matrices of a homogeneous CTMC over any time duration, it plays the same role here.

Proposition 5.7 *Let $\{X_t\}_{t \in \mathbb{R}_{\geq 0}}$ be a continuous-time homogeneous Markov chain, with transition rate matrix Q as defined above. Then for all $t \in \mathbb{R}_{\geq 0}$, the transition probabilities $P(X_t = y \mid X_0 = x)$, $x, y \in \mathscr{X}$ after time t are given by the elements $T_t(x, y)$ of the transition matrix $T_t = e^{Qt}$.*

While we do not aim to give a complete treatment on the interpretation of the matrix exponential, some properties are worth pointing out. First of all, it can be defined analogously to the exponential function of real numbers, that is, through a Taylor expansion around zero. Specifically, it holds that

$$T_t = e^{Qt} \coloneqq \sum_{k=0}^{+\infty} \frac{t^k Q^k}{k!} .$$

Thus, the approximation in Eq. (5.7) can be seen as a first-order truncation of the series above.

As a second important point, we can consider the entire family of transition matrices T_t for all $t \in \mathbb{R}_{\geq 0}$. Then this family constitutes a *semi-group* of transition matrices, and Q is the *generator* of this semi-group. Specifically, it holds that $T_{t+s} = T_t T_s$ for all $t, s \in \mathbb{R}_{\geq 0}$—this is called the *semi-group* property. Observe that it is analogous to the result in Proposition 5.2 and that we already used this property for the matrix $T_{t+\Delta}$ when constructing the derivative.

These properties immediately yield a different representation for the matrix exponential, which will be convenient further on. We omit the proof.

Proposition 5.8 *Let $\{X_t\}_{t \in \mathbb{R}_{\geq 0}}$ be a continuous-time homogeneous Markov chain, with transition rate matrix Q, and let T_t be the associated family of transition matrices. Then, for all $t \in \mathbb{R}_{\geq 0}$, it holds that*

$$T_t = \lim_{n \to +\infty} \left(I + \frac{t}{n} Q \right)^n .$$

One way to think about this is that, for some fixed (but large enough) $n \in \mathbb{N}$, each factor $(I + {}^t/_n Q)$ is, due to Eq. (5.7), roughly the 'small step' transition matrix $T_{t/n}$. The multiplication of these n terms $(I + {}^t/_n Q)^n$ is then analogous to the composition in Proposition 5.2, whereby we cover the duration t in steps of size ${}^t/_n$. It should be noted that this only becomes exact in the limit (as the result states), but the intuition behind it is the same regardless.

Furthermore, let us again remark that the transition-matrix representation is also convenient in that it offers an alternative representation of the conditional expectation operator:

Proposition 5.9 *Let $\{X_t\}_{t \in \mathbb{R}_{\geq 0}}$ be a continuous-time homogeneous Markov chain, with transition rate matrix Q, and let T_t be the associated family of transition matrices. Then, for all $f \in \mathscr{L}(\mathscr{X})$, all $t \in \mathbb{R}_{\geq 0}$ and all $x \in \mathscr{X}$, it holds that $\mathbb{E}[f(X_t) \mid X_0 = x] = \left[T_t f\right](x)$.*

Proof Analogous to the proof of Proposition 5.3. □

Let us consider the importance of the homogeneity assumption in the preceding exposition. Indeed, it is this property that crucially allows the parametrisation to only require a single rate matrix Q. More generally, we may consider a non-homogeneous CTMC and consider the derivatives at each time point; first write the transition matrix for the interval $[s, t]$ as

$$T_s^t(x, y) \coloneqq P(X_t = y \mid X_s = x) ,$$

and differentiate to obtain

$$\left. \frac{\mathrm{d}\, T_s^t}{\mathrm{d}\, t} \right|_{t=s} = \lim_{t \to s^+} \frac{T_s^t - I}{t - s} =: Q_s ,$$

whence the parametrisation now requires an entire family Q_s of rate matrices—one for each point in time. Note, though, that these matrices are still transition rate matrices, in that they satisfy the properties in Definition 5.16. However, the corresponding matrix differential equation is no longer solved by a simple matrix exponential.

More generally still, for arbitrary continuous-time stochastic processes (that are neither homogeneous nor Markov) we may consider the transition rates (derivatives) not only for specific points in time but also for specific histories leading up to that time. For instance, with $\mathbf{s} = s_1, \ldots, s_n$ and t in $\mathbb{R}_{\geq 0}$ and $x_{\mathbf{s}} \in \mathscr{X}^n$, we may write

$$\frac{\mathrm{d}}{\mathrm{d}\,u} P\big(X_u = y \mid X_{\mathbf{s}} = x_{\mathbf{s}}, X_t = x\big)\Big|_{u=t} =: Q_{x_{\mathbf{s}},t}(x, y)\,. \tag{5.8}$$

Thus, the parametrisation requires the specification of a transition rate matrix for each point in time and for each possible history before that time. It should be clear that this leads to a rather unwieldy process specification, which again goes some way in illustrating why homogeneity and Markovianity are such popular simplifying assumptions.

5.4.1 Imprecise Continuous-Time Markov Chains

With the notation and concepts for precise continuous-time stochastic processes in place, let us now turn to the imprecise generalisation. In what follows, we will consider imprecise, homogeneous continuous-time Markov chains (ICTMC). As before, we start by considering the abstract sets-of-measures definition:

Definition 5.17 (ICTMC as set of processes) An *imprecise continuous-time Markov chain* is a set $\mathbb{P}$ of probability measures on the measurable space $(\Omega, \mathscr{F})$ of (continuous-time) paths, with associated lower expectation operator $\underline{\mathbb{E}}_{\mathbb{P}}$ such that, for all $f \in \mathscr{L}(\mathscr{X})$ and all $s_1, \ldots, s_n, t \in \mathbb{R}_{\geq 0}$ such that $s_1 < \cdots < s_n < t$, it holds that

$$\underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_t) \,\big|\, X_{s_1}, \ldots, X_{s_n}\big] = \underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_t) \,\big|\, X_{s_n}\big]\,.$$

Furthermore, an imprecise continuous-time Markov chain is called *homogeneous* if, for all $s, t \in \mathbb{R}_{\geq 0}$, $s < t$, and all $f \in \mathscr{L}(\mathscr{X})$, it holds that $\underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_t) \,\big|\, X_s\big] = \underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_{t-s}) \,\big|\, X_0\big]$.

As in the discussion about imprecise discrete-time Markov chains, we distinguish between the definition by epistemic irrelevance—which is what is used above—and the definition by strong independence, which would imply that all $P \in \mathbb{P}$ are precise (homogeneous) Markov chains, and which we are explicitly not using.

Let us now consider the parametrisation of such an ICTMC. We recall that in the precise case, the canonical parameter is a single transition rate matrix Q. In

contrast, for the imprecise case, the 'parameter' of interest is a *set* $\mathscr{Q}$ of transition rate matrices. Because a precise homogeneous CTMC is identified with a rate matrix Q, it is clear that such a set $\mathscr{Q}$ induces a set of precise processes: simply consider all processes for which the associated rate matrix is included in $\mathscr{Q}$. However, this induced set then only includes homogeneous Markov processes, and, as remarked above, we aim to relax these independence assumptions. Using the parametrisation of more general precise processes, we introduce the notion of compatibility with a given set of rate matrices:

Definition 5.18 Let $\mathscr{Q}$ be a set of transition rate matrices. Then a continuous-time stochastic process P is called *compatible* with $\mathscr{Q}$ if, for all $\mathbf{s} = s_1, \ldots, s_n$ and $t \in \mathbb{R}_{\geq 0}$ such that $s_1 < \cdots < s_n < t$, and all $x_{\mathbf{s}} \in \mathscr{X}^n$, it holds that $Q_{x_{\mathbf{s}},t} \in \mathscr{Q}$, where $Q_{x_{\mathbf{s}},t}$ is the time- and history-dependent rate matrix associated with P, as in Eq. (5.8).

It can be verified that this definition includes, as a special case, the compatibility of homogeneous CTMCs with rate matrix Q, with a given set $\mathscr{Q}$, if $Q \in \mathscr{Q}$. Similarly, a non-homogeneous CTMC that is parametrised by a family Q_t is compatible with such a set if $Q_t \in \mathscr{Q}$ for all $t \in \mathbb{R}_{\geq 0}$. The ICTMC $\mathbb{P}$ corresponding to a given set $\mathscr{Q}$, then, is taken to be the largest set of continuous-time stochastic processes that are compatible with this $\mathscr{Q}$. While perhaps not obvious, it can be proven that this set $\mathbb{P}$ is then indeed a homogeneous ICTMC, in the sense that its corresponding lower expectations satisfy the properties of Definition 5.17.

With this ICTMC in place, let us now again consider the main inferential challenge: how to compute the corresponding lower expectation. A first attempt could be to use Propositions 5.7 and 5.9 and optimise over $\mathscr{Q}$; for some fixed $f \in \mathscr{L}(\mathscr{X})$, this would give

$$\inf_{Q \in \mathscr{Q}} e^{Qt} f \,.$$

If we think about what this computes, we come to the conclusion that for each $Q \in \mathscr{Q}$, there is a homogeneous CTMC for which the conditional expectation of f at time $t \in \mathbb{R}_{\geq 0}$ is indeed $e^{Qt} f$. We therefore conclude that this computes the lower expectation with respect to all *homogeneous* CTMCs that are compatible with $\mathscr{Q}$. But what about the non-homogeneous and/or non-Markovian stochastic processes that we know are also included in $\mathbb{P}$? It turns out that the above expression ignores their corresponding expectations and hence only yields an *upper bound* on the actual *lower* expectation. In other words, we cannot use this expression to compute the lower expectation for $\mathbb{P}$.

The way to proceed is analogous to the approach in Sect. 5.3.2; we first define a *local* 'lower' operator and then find the global lower expectation using repeated compositions of this operator through the law of iterated lower expectation. To this end, we associate with the set $\mathscr{Q}$ the corresponding *lower transition rate operator* $\underline{Q} : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$, which is defined for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{X}$ as

$$\big[\underline{Q}f\big](x) \coloneqq \inf_{Q\in\mathscr{Q}}\big[Qf\big](x)\,. \tag{5.9}$$

Intuitively, for small $\Delta > 0$, we can then approximate the lower expectation as

$$\underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_\Delta)\,|\,X_0\big] \approx \inf_{Q\in\mathscr{Q}}(I+\Delta Q)f = (I+\Delta\underline{Q})f\,,$$

where the approximation is again due to Eq. (5.7). It turns out that we can make this exact and extend the result to any time t, analogously to Proposition 5.8:

Theorem 5.5 *Let $\mathscr{Q}$ be a non-empty set of transition rate matrices, and let $\underline{Q}$ be the corresponding lower transition rate operator, as in Eq. (5.9). Then, for all $t \in \mathbb{R}_{\geq 0}$, there is an operator $\underline{T}_t : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$, such that*

$$\underline{T}_t = \lim_{n\to+\infty}\left(I+\frac{t}{n}\underline{Q}\right)^n.$$

These operators satisfy $\underline{T}_0 = I$, $\underline{T}_{t+s} = \underline{T}_t\underline{T}_s$ for all $t, s \in \mathbb{R}_{\geq 0}$ and $\mathrm{d}/\mathrm{d}t\,\underline{T}_t = \underline{Q}\underline{T}_t$.

Observe that this family of operators $\underline{T}_t$ satisfies in large part the same properties as the matrix exponentials of Qt—c.f. the discussion after Proposition 5.7—with the main difference being that they are *non-linear* operators. We can now finally present the result that allows the computation of lower expectations for ICTMCs.

Theorem 5.6 *Let $\mathscr{Q}$ be a non-empty set of transition rate matrices, with corresponding lower transition rate operator $\underline{Q}$, and let $\mathbb{P}$ be the corresponding ICTMC. Suppose that $\mathscr{Q}$ is closed, convex and has separately specified rows (i.e. is closed under recombination of the rows of its elements). Then, for all $f \in \mathscr{L}(\mathscr{X})$, all $t \in \mathbb{R}_{\geq 0}$ and all $x \in \mathscr{X}$, it holds that*

$$\underline{\mathbb{E}}_{\mathbb{P}}\big[f(X_t)\,|\,X_0 = x\big] = \big[\underline{T}_t f\big](x)\,. \tag{5.10}$$

Observe that this result needs some constraints on the rate matrix set $\mathscr{Q}$. This can be explained in the sense that the right-hand side of Eq. (5.10) depends, through Theorem 5.5, on the lower transition rate operator $\underline{Q}$. In turn, $\underline{Q}$ depends on $\mathscr{Q}$ through Eq. (5.9). Conversely, the left-hand side (the lower expectation) depends on the set $\mathbb{P}$, which in turn depends on $\mathscr{Q}$ through the compatibility as in Definition 5.18. It turns out that for these different dependencies on $\mathscr{Q}$ to be equivalent, we need some regularity conditions on this latter set—these are the constraints mentioned in the theorem above.

5.4.2 Limits of ICTMCs

Let us finally consider the long-term behaviour of a given homogeneous ICTMC $\mathbb{P}$ with transition rate matrix set $\mathscr{Q}$ and associated lower transition rate operator $\underline{Q}$; we assume these to be fixed in the remainder of this section. What, then, can we say about the lower expectation of a function as time goes to infinity?

Recall that, in the discrete-time case, Theorem 5.4 established a sufficient condition for such a lower expectation to converge. This condition was *regularity* of the IDTMC. Essentially, this meant that it was possible for the IDTMC to move from any state to any other state, in exactly n steps, for some $n \in \mathbb{N}$. In the continuous-time case that we consider here, there is a similar condition: *upper reachability* between all pairs of states.

We first remark that this condition is defined using the conjugate upper transition rate operator defined as $\overline{Q} f \coloneqq -\underline{Q}(-f)$ for all $f \in \mathscr{L}(\mathscr{X})$. The definition of upper reachability is then analogous to that of accessibility in discrete-time but is instead defined using the transition *rates*, rather than probabilities:

Definition 5.19 Let $\mathbb{P}$ be an ICTMC with associated upper transition rate operator $\overline{Q}$, as defined above. For any two states $x, y \in \mathscr{X}$, y is said to be *upper reachable* from x, if there is a sequence $x_0, \ldots, x_n \in \mathscr{X}$, $n \in \mathbb{N}$, such that $x_0 = x$, $x_n = y$ and, for all $i \in \{1, \ldots, n\}$, it holds that $x_i \neq x_{i-1}$ and $\left[\overline{Q}\, \mathbb{I}_{x_i}\right](x_{i-1}) > 0$.

Let us in particular consider the final condition in this definition. From the conjugacy between the lower and upper transition rate operators, and the definition of the former, we can rewrite this requirement as saying that

$$0 < \left[\overline{Q}\, \mathbb{I}_{x_i}\right](x_{i-1}) = \sup_{Q \in \mathscr{Q}} \left[Q\, \mathbb{I}_{x_i}\right](x_{i-1}) = \sup_{Q \in \mathscr{Q}} Q(x_{i-1}, x_i)\,.$$

Thus, upper reachability of y, from x, requires that there exists a sequence of states from x to y such that, at each step in this sequence, there is *some* transition rate matrix $Q \in \mathscr{Q}$ which assigns strictly positive 'speed' of moving from the current state in this sequence, to the next one. In other words, it should be possible for these transitions to happen according to some of the models in our set $\mathbb{P}$, but not necessarily all, and there can be a different model allowing for this possibility at each step. This can now be used to state the following result:

Theorem 5.7 *Let $\mathbb{P}$ be an ICTMC and suppose that, for all $x, y \in \mathscr{X}$, y is upper reachable from x. Then, there is a unique lower expectation operator $\underline{\mathbb{E}}_{\mathbb{P}}\left[\cdot(X_{+\infty})\right] : \mathscr{L}(\mathscr{X}) \to \mathbb{R}$ such that, for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{X}$,*

$$\underline{\mathbb{E}}_{\mathbb{P}}\left[f(X_{+\infty})\right] = \lim_{t \to +\infty} \underline{\mathbb{E}}_{\mathbb{P}}\left[f(X_t) \,\middle|\, X_0 = x\right] = \lim_{t \to +\infty} \left[\underline{T}_t f\right](x)\,.$$

5.5 Literature and Further Reading

Let us conclude this chapter by providing pointers to the literature on which the material in this chapter is based. We will also briefly discuss some parts of the literature that are related but not quite the same as what we covered here.

First of all, there exists an extensive body of literature on (precise) Markov chains, both in discrete- and in continuous times. It would be nigh impossible to give a complete overview here, but we think that [1, 38] make excellent introductory reads. For a broad and general introduction to the theory for imprecise probability, which lies at the heart of the models that we discussed here, we refer the reader to [3, 50]. The difference between the notions of strong independence and epistemic irrelevance—which we have stressed repeatedly and which is a crucial property of imprecise Markov chains as we treated them here—is discussed, e.g. in [5, 37].

For the interpretation of Markov chains using probability trees, see, for example, [13, 18, 32]. This interpretation is also closely related to the *game-theoretic* formalisation of probabilities using the theory of martingales (which we did not cover here). The interested reader may want to pursue [13, 32, 49].

For an account of the general theory of Bayesian networks, see [39]. For their imprecise generalisation—credal networks—references [2, 6, 7, 9, 11] discuss a lot of the general theory.

Imprecise *discrete*-time Markov chains are discussed, e.g. in [15, 17, 27]. For imprecise *continuous*-time Markov chains, see [30, 44]. A treatment of the matrix exponential, which is crucial to computational methods for CTMCs, is given in [48]. Reference [19] discusses the current state-of-the-art to efficiently compute the imprecise generalisation of the matrix exponential, which we have seen is crucial for computing inferences in ICTMCs.

Detailed treatments on the long-term (limit) behaviour in IDTMCs can be found in [14, 16, 26, 45]. Reference [10] provides the necessary and sufficient conditions for the limit behaviour of ICTMCs, and [19] also discusses computational methods to numerically approximate this limit. We remark that Theorems 5.4 and 5.7 in this chapter are stated in a simplified form compared to their statement in the literature. In particular, the results in [10, 14] are stronger; for instance, [10] in fact provides *necessary* and sufficient conditions for the convergence of an ICTMC, whereas Theorem 5.7 only states a sufficient condition.

Some examples of the merits of imprecise Markov chains in applications are provided by [40, 46, 47]. A domain for which the applicability of (imprecise) Markov chains has been extensively studied, is queueing theory [8, 33–36].

A generalisation of Markov chains that we have not discussed, but which is nevertheless important in many practical applications, is *hidden* Markov chains. There, the stochastic process cannot be observed directly, but only through a noisy measurement model. Their imprecise treatment is discussed, e.g. in [4, 12, 31].

Fields that are closely related to the theory of imprecise Markov chains are controlled Markov processes [21] and Markov decision processes [22, 28, 41, 51]. There also, the process under study has its parameters changed over time. However,

the goal there is not to represent uncertainty and change these parameters to compute robust bounds on quantities of interest. Rather, their aim is to optimise the process evolution towards some operational target.

Finally, we again emphasise that our treatment uses epistemic irrelevance, which we distinguish from using strong independence. There is, however, an extended body of literature also on the latter. These alternative models are known as Markov chains under strong independence, e.g. in [27], as interval Markov chains [20, 29, 42, 43] or as Markov set chains [23–25].

Acknowledgments The author wishes to express his sincere gratitude to Gert de Cooman and Jasper De Bock, for their helpful comments and suggestions during the writing of this chapter. He also wants to thank the reviewer, whose comments and suggestions further helped to improve this work.

References

1. W.J. Anderson, *Continuous-Time Markov Chains, An Applications-Oriented Approach*. Springer Series in Statistics (Springer, New York, 1991)
2. A. Antonucci, C.P. de Campos, M. Zaffalon, Probabilistic graphical models, in *Introduction to Imprecise Probabilities*, ed. by T. Augustin, F.P.A. Coolen, G. De Cooman, M.C.M. Troffaes (Wiley, New York, 2014)
3. T. Augustin, F.P.A. Coolen, G. De Cooman, M.C.M. Troffaes, *Introduction to Imprecise Probabilities* (Wiley, New York, 2014)
4. A. Benavoli, M. Zaffalon, E. Miranda, Robust filtering through coherent lower previsions. IEEE Trans. Autom. Control **56**(7), 1567–1581 (2011)
5. I. Couso, S. Moral, P. Walley, A survey of concepts of independence for imprecise probabilities. Risk Decis. Policy **5**(2), 165–181 (2000)
6. F.G. Cozman, Credal networks. Artif. Intell. **120**, 199–233 (2000)
7. F.G. Cozman, Graphical models for imprecise probabilities. Int. J. Approx. Reason. **39**(2–3), 167–184 (2005)
8. R.J. Crossman, P. Coolen-Schrijner, F.P. Coolen, Time-homogeneous birth-death processes with probability intervals and absorbing state. J. Stat. Theory Pract. **3**(1), 103–118 (2009)
9. J. De Bock, Credal networks under epistemic irrelevance: theory and algorithms. Ph.D. thesis, Ghent University (2015)
10. J. De Bock, The limit behaviour of imprecise continuous-time Markov chains. J. Nonlinear Sci. **27**(1), 159–196 (2017)
11. J. De Bock, Credal networks under epistemic irrelevance. Int. J. Approx. Reason. **85**, 107–138 (2017)
12. J. De Bock, G. De Cooman, An efficient algorithm for estimating state sequences in imprecise hidden Markov models. J. Artif. Intell. Res. **50**, 189–233 (2014)
13. G. De Cooman, F. Hermans, Imprecise probability trees: Bridging two theories of imprecise probability. Artificial Intelligence **172**(11), 1400–1427 (2008)
14. G. De Cooman, F. Hermans, E. Quaeghebeur, Imprecise Markov chains and their limit behavior. Probab. Eng. Inf. Sci. **23**(4), 597–635 (2009)
15. G. De Cooman, F. Hermans, A. Antonucci, M. Zaffalon, Epistemic irrelevance in credal nets: the case of imprecise Markov trees. Int. J. Approx. Reason. **51**(9), 1029–1052 (2010)
16. G. De Cooman, J. De Bock, S. Lopatatzidis, A pointwise ergodic theorem for imprecise Markov chains, in *Proceedings of ISIPTA 2015*, pp. 107–115 (2015)

17. G. De Cooman, J. De Bock, S. Lopatatzidis, Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems. Int. J. Approx. Reason. **76**, 18–46 (2016)
18. S. Destercke, G. De Cooman, Relating epistemic irrelevance to event trees. *Soft Methods for Handling Variability and Imprecision* (Springer, New York, 2008), pp. 66–73
19. A. Erreygers, J. De Bock, Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds, in *Proceedings of ISIPTA 2017*, pp. 145–156 (2017)
20. S. Galdino, Interval continuous-time Markov chains simulation, in *Proceedings of the 2013 International Conference on Fuzzy Theory and Its Applications*, pp. 273–278 (2013)
21. X. Guo, O. Hernández-Lerma, Continuous-time controlled Markov chains. Ann. Appl. Probab. **13**(1), 363–388 (2003)
22. X. Guo, O. Hernández-Lerma, *Continuous-Time Markov Decision Processes* (Springer, New York, 2009)
23. D.J. Hartfiel, Sequential limits in Markov set-chains. J. Appl. Probab. **28**(4), 910–913 (1991)
24. D.J. Hartfiel, *Markov Set-Chains*. Lecture Notes in Mathematics, vol. 1695 (Springer, New York, 1998)
25. D.J. Hartfiel, E. Seneta, On the theory of Markov set-chains. Adv. Appl. Probab. **26**, 947–964 (1994)
26. F. Hermans, G. De Cooman, Characterisation of ergodic upper transition operators. Int. J. Approx. Reason. **53**(4), 573–583 (2012)
27. F. Hermans, D. Škulj, Stochastic processes, in *Introduction to Imprecise Probabilities*, ed. by T. Augustin, F.P.A. Coolen, G. De Cooman, M.C.M. Troffaes (Wiley, New York, 2014)
28. H. Itoh, K. Nakamura, Partially observable Markov decision processes with imprecise parameters. Artificial Intelligence **171**, 453–490 (2007)
29. I.O. Kozine, L.V. Utkin, Interval-valued finite Markov chains. Reliable Computing **8**(2), 97–113 (2002)
30. T. Krak, J. De Bock, A. Siebes, Imprecise continuous-time Markov chains. Int. J. Approx. Reason. **88**, 452–528 (2017)
31. T. Krak, J. De Bock, A. Siebes, Efficient computation of updated lower expectations for imprecise continuous-time hidden Markov chains, in *Proceedings of ISIPTA 2017*, pp. 193–204 (2017)
32. S. Lopatatzidis, Robust modelling and optimisation in stochastic processes using imprecise probabilities, with an application to queueing theory. Ph.D. thesis, Ghent University (2017)
33. S. Lopatatzidis, J. De Bock, G. De Cooman, Calculating bounds on expected return and first passage times in finite-state imprecise birth-death chains, in *Proceedings of ISIPTA 2015*, pp. 177–186 (2015)
34. S. Lopatatzidis, J. De Bock, G. De Cooman, Computational methods for imprecise continuous-time birth-death processes: a preliminary study of flipping times, in *Proceedings of ISIPTA 2015*, p. 344 (2015)
35. S. Lopatatzidis, J. De Bock, G. De Cooman, S. De Vuyst, J. Walraevens, Robust queueing theory: an initial study using imprecise probabilities. Queueing Systems **82**(1–2), 75–101 (2016)
36. S. Lopatatzidis, J. De Bock, G. De Cooman, Computing lower and upper expected first passage and return times in imprecise birth-death chains. Int. J. Approx. Reason. **80**, 137–173 (2017)
37. E. Miranda, G. De Cooman, Structural judgements, in *Introduction to Imprecise Probabilities*, ed. by T. Augustin, F.P.A. Coolen, G. De Cooman, M.C.M. Troffaes (Wiley, New York, 2014)
38. J.R. Norris, *Markov Chains* (Cambridge University Press, Cambridge, 1998)
39. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, 1988)
40. C. Rottondi, A. Erreygers, G. Verticale, J. De Bock, Modelling spectrum assignment in a two-service flexi-grid optical link with imprecise continuous-time Markov chains, in *Proceedings of DRCN 2017*, pp. 39–46 (2017)
41. J.K. Satia, R.E. Lave, Markovian decision processes with uncertain transition probabilities. Operations Research **21**, 728–740 (1973)

42. D. Škulj, Finite discrete time Markov chains with interval probabilities, in *Soft Methods for Integrated Uncertainty Modelling*, ed. by J. Lawry, E. Miranda, A. Bugarin, S. Li, M.A. Gil, P. Grzegorzewski, O. Hryniewicz (Springer, New York, 2006), pp. 299–306
43. D. Škulj, Regular finite Markov chains with interval probabilities, in *Proceedings of ISIPTA 2007*, pp. 405–413 (2007)
44. D. Škulj, Efficient computation of the bounds of continuous time imprecise Markov chains. Appl. Math. Comput. **250**(C), 165–180 (2015)
45. D. Škulj, R. Hable, Coefficients of ergodicity for Markov chains with uncertain parameters. Metrika **76**(1), 107–133 (2013)
46. Y. Soullard, A. Antonucci, S. Destercke, Technical gestures recognition by set-valued hidden Markov models with prior knowledge, in *Soft Methods for Data Science*, pp. 455–462 (2017)
47. M. Troffaes, J. Gledhill, D. Škulj, S. Blake, Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair, in *Proceedings of ISIPTA 2015*, pp. 287–294 (2015)
48. C.F. Van Loan, A Study of the Matrix Exponential, Numerical Analysis Report No. 10, University of Manchester, Manchester, UK, August 1975, Reissued as MIMS EPrint 2006.397, Manchester Institute for Mathematical Sciences, The University of Manchester, UK
49. V. Vovk, G. Shafer, Game-theoretic probability, in *Introduction to Imprecise Probabilities*, ed. by T. Augustin, F.P.A. Coolen, G. De Cooman, M.C.M. Troffaes, (Wiley, New York, 2014)
50. P. Walley, *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London, 1991)
51. C.C. White, H.K. Eldeib, Markov decision-processes with imprecise transition-probabilities. Operations Research **42**, 739–749 (1994)