Efficient Computation of Updated Lower Expectations for Imprecise Continuous-Time Hidden Markov Chains

Thomas Krak   T.E.KRAK@UU.NL
[†]*Department of Information and Computing Sciences, Utrecht University (The Netherlands)*

Jasper De Bock   JASPER.DEBOCK@UGENT.BE
Department of Electronics and Information Systems, imec, IDLab, Ghent University (Belgium)

Arno Siebes[†]   A.P.J.M.SIEBES@UU.NL

Abstract

We consider the problem of performing inference with *imprecise continuous-time hidden Markov chains*, that is, *imprecise continuous-time Markov chains* that are augmented with random *output* variables whose distribution depends on the hidden state of the chain. The prefix 'imprecise' refers to the fact that we do not consider a classical continuous-time Markov chain, but replace it with a robust extension that allows us to represent various types of model uncertainty, using the theory of *imprecise probabilities*. The inference problem amounts to computing lower expectations of functions on the state-space of the chain, given observations of the output variables. We develop and investigate this problem with very few assumptions on the output variables; in particular, they can be chosen to be either discrete or continuous random variables. Our main result is a polynomial runtime algorithm to compute the lower expectation of functions on the state-space at any given time-point, given a collection of observations of the output variables.

Keywords: continuous-time hidden Markov chains; imprecise probabilities; updating.

1. Introduction

A continuous-time Markov chain (CTMC) is a stochastic model that describes the evolution of a dynamical system under uncertainty. Specifically, it provides a probabilistic description of how such a system might move through a finite state-space, as time elapses in a continuous fashion. There are various ways in which this model class can be extended.

One such extension are continuous-time *hidden* Markov chains (CTHMC's) (Wei et al., 2002). Such a CTHMC is a stochastic model that contains a continuous-time Markov chain as a latent variable—that is, the actual realised behaviour of the system cannot be directly observed. This model furthermore incorporates random *output* variables, which depend probabilistically on the current state of the system, and it is rather realisations of these variables that one observes. Through this stochastic dependency between the output variables and the states in which the system might be, one can perform inferences about quantities of interest that depend on these states—even though they have not been, or cannot be, observed directly.

Another extension of CTMC's, arising from the theory of *imprecise probabilities* (Walley, 1991), are *imprecise continuous-time Markov chains* (ICTMC's) (Škulj, 2015; Krak et al., 2016). This extension can be used to robustify against uncertain numerical parameter assessments, as well as the simplifying assumptions of time-homogeneity and that the model should satisfy the Markov property. Simply put, an ICTMC is a *set* of continuous-time stochastic processes, some of which are "traditional" time-homogeneous CTMC's. However, this set also contains more complicated processes, which are non-homogeneous and do not satisfy the Markov property.

In this current work, we combine these two extensions by considering *imprecise continuous-time hidden Markov chains*—a stochastic model analogous to a CTHMC, but where the latent CTMC is replaced by an ICTMC. We will focus in particular on practical aspects of the corresponding inference problem. That is, we provide results on how to efficiently compute lower expectations of functions on the state-space, given observed realisations of the output variables.

Throughout, all results are stated without proof. We have made available an extended version of this work (Krak et al., 2017), which includes an appendix containing the proofs of all our results.

1.1 Related Work

As should be clear from the description of CTHMC's in Section 1, this model class extends the well-known (discrete-time) *hidden Markov models* (HMM's) to a continuous-time setting. In the same sense, the present subject of ICTHMC's can be seen to extend previous work on *imprecise hidden Markov models* (iHMM's) (de Cooman et al., 2010) to a continuous-time setting. Hence, the model under consideration should hopefully be intuitively clear to readers familiar with (i)HMM's.

The main novelty of this present work is therefore not the (somewhat obvious) extension of iHMM's to ICTHMC's, but rather the application of recent results on ICTMC's (Krak et al., 2016) to derive an efficient solution to the continuous-time analogue of inference in iHMM's. The algorithm that we present is largely based on combining these results with the ideas behind the MePiCTIr algorithm (de Cooman et al., 2010) for inference in credal trees under epistemic irrelevance.

A second novelty of the present paper is that, contrary to most of the work in the literature on iHMM's, we allow the output variables of the ICTHMC to be either discrete or continuous. This allows the model to be applied to a much broader range of problems. At the same time, it turns out that this does not negatively influence the efficiency of the inference algorithm.

2. Preliminaries

We denote the reals as $\mathbb{R}$, the non-negative reals as $\mathbb{R}_{\geq 0}$, and the positive reals as $\mathbb{R}_{>0}$. The natural numbers are denoted by $\mathbb{N}$, and we define $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

Since we are working in a continuous-time setting, a *time-point* is an element of $\mathbb{R}_{\geq 0}$, and these are typically denoted by t or s. We also make extensive use of non-empty, finite sequences of time points $u \subset \mathbb{R}_{\geq 0}$. These are taken to be ordered, so that they may be written $u = t_0, \ldots, t_n$, for some $n \in \mathbb{N}_0$, and such that then $t_i < t_j$ for all $i, j \in \{0, \ldots, n\}$ for which $i < j$. Such sequences are usually denoted by u or v, and we let $\mathscr{U}$ be the entire set of them.

Throughout, we consider some fixed, finite state space $\mathscr{X}$. A generic element of $\mathscr{X}$ will be denoted by x. When considering the state-space at a specific time t, we write $\mathscr{X}_t := \mathscr{X}$, and x_t denotes a generic state-assignment at this time. When considering multiple time-points u simultaneously, we define the joint state-space as $\mathscr{X}_u := \prod_{t_i \in u} \mathscr{X}_{t_i}$, of which $x_u = (x_{t_0}, \ldots, x_{t_n})$ is a generic element.

For any $u \in \mathscr{U}$, we let $\mathscr{L}(\mathscr{X}_u)$ be the set of all real-valued functions on $\mathscr{X}_u$.

2.1 Imprecise Continuous-Time Markov Chains

We here briefly recall the most important properties of imprecise continuous-time Markov chains (ICTMC's), following the definitions and results of Krak et al. (2016). For reasons of brevity, we provide these definitions in a largely intuitive, non-rigorous manner, and refer the interested reader to this earlier work for an in-depth treatise on the subject.

An ICTMC will be defined below as a specific set of *continuous-time stochastic processes*. Simply put, a continuous-time stochastic process is a joint probability distribution over random variables X_t, for each time $t \in \mathbb{R}_{\geq 0}$, where each random variable X_t takes values in $\mathscr{X}$.

It will be convenient to have a way to numerically parameterise such a stochastic process P. For this, we require two different kinds of parameters. First, we need the specification of the initial distribution $P(X_0)$ over the state at time zero; this simply requires the specification of some probability mass function on $\mathscr{X}_0$. Second, we need to parameterise the dynamic behaviour of the model.

In order to describe this dynamic behaviour, we require the concept of a *rate matrix*. Such a rate matrix Q is a real-valued $|\mathscr{X}| \times |\mathscr{X}|$ matrix, whose off-diagonal elements are non-negative, and whose every row sums to zero—thus, the diagonal elements are non-positive. Such a rate matrix may be interpreted as describing the "rate of change" of the conditional probability $P(X_s | X_t, X_u = x_u)$, when s is close to t. In this conditional probability, it is assumed that $u < t$, whence the state assignment x_u is called the *history*. For small enough $\Delta \in \mathbb{R}_{>0}$, we may now write that

$$P(X_{t+\Delta} | X_t, X_u = x_u) \approx \left[I + \Delta Q_{t,x_u} \right] (X_t, X_{t+\Delta}),$$

for some rate matrix Q_{t,x_u}, where I denotes the $|\mathscr{X}| \times |\mathscr{X}|$ identity matrix, and where the quantity $[I + \Delta Q_{t,x_u}](X_t, X_{t+\Delta})$ denotes the element at the X_t-row and $X_{t+\Delta}$-column of the matrix $I + \Delta Q_{t,x_u}$. Note that in general, this rate matrix Q_{t,x_u} may depend on the specific time t and history x_u at which this relationship is stated.

If these rate matrices only depend on the time t and not on the history x_u, i.e. if $Q_{t,x_u} = Q_t$ for all t and all x_u, then it can be shown that P satisfies the *Markov property*: $P(X_s | X_t, X_u) = P(X_s | X_t)$. In this case, P is called a *continuous-time Markov chain*.

Using this method of parameterisation, an *imprecise continuous-time Markov chain* (ICTMC) is similarly parameterised using a *set* of rate matrices $\mathscr{Q}$, and a *set* of initial distributions $\mathscr{M}$. The corresponding ICTMC, denoted by $\mathbb{P}_{\mathscr{Q},\mathscr{M}}$, is the set of all continuous-time stochastic processes whose dynamics can be described using the elements of $\mathscr{Q}$, and whose initial distributions are consistent with $\mathscr{M}$. That is, $\mathbb{P}_{\mathscr{Q},\mathscr{M}}$ is the set of stochastic processes P for which $P(X_0) \in \mathscr{M}$ and for which $Q_{t,x_u} \in \mathscr{Q}$ for every time t and history x_u.

The *lower expectation* with respect to this set $\mathbb{P}_{\mathscr{Q},\mathscr{M}}$ is then defined as

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\cdot | \cdot] \coloneqq \inf \left\{ \mathbb{E}_P[\cdot | \cdot] : P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}} \right\},$$

where $\mathbb{E}_P[\cdot | \cdot]$ denotes the expectation with respect to the (precise) stochastic process P. The *upper expectation* $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}$ is defined similarly, and is derived through the well-known conjugacy property $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\cdot | \cdot] = -\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[- \cdot | \cdot]$. Note that it suffices to focus on lower (or upper) expectations, and that *lower* (and *upper*) *probabilities* can be regarded as a special case; for example, for any $A \subseteq \mathscr{X}$, we have that $\underline{P}_{\mathscr{Q},\mathscr{M}}(X_s \in A | X_t) \coloneqq \inf \{ P(X_s \in A | X_t) : P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}} \} = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\mathbb{I}_A(X_s) | X_t]$, where $\mathbb{I}_A$ is the indicator of A, defined for all $x \in \mathscr{X}$ by $\mathbb{I}_A(x) \coloneqq 1$ if $x \in A$ and $\mathbb{I}_A(x) \coloneqq 0$ otherwise.

In the sequel, we will assume that $\mathscr{M}$ is non-empty, and that $\mathscr{Q}$ is non-empty, bounded,[1] convex, and has *separately specified rows*. This latter property states that $\mathscr{Q}$ is closed under arbitrary recombination of rows from its elements; see (Krak et al., 2016, Definition 24) for a formal definition. Under these assumptions, $\mathbb{P}_{\mathscr{Q},\mathscr{M}}$ satisfies an *imprecise Markov property*, in the sense that $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_s) | X_t, X_u = x_u] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_s) | X_t]$. This property explains why we call this model an imprecise continuous-time "Markov" chain.

---

1. That is, that there exists a $c \in \mathbb{R}_{\geq 0}$ such that, for all $Q \in \mathscr{Q}$ and $x \in \mathscr{X}$, it holds that $|Q(x,x)| < c$.

2.2 Computing Lower Expectations for ICTMC's

Because we want to focus in this paper on providing efficient methods of computation, we here briefly recall some previous results from Krak et al. (2016) about how to compute lower expectations for ICTMC's. We focus in particular on how to do this for functions on a single time-point.

To this end, it is useful to introduce the *lower transition rate operator $\underline{Q}$* that corresponds to $\mathscr{Q}$. This operator is a map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$, defined for every $f \in \mathscr{L}(\mathscr{X})$ by

$$\left[\underline{Q}f \right](x) := \inf \left\{ \sum_{x' \in \mathscr{X}} Q(x,x')f(x') : Q \in \mathscr{Q} \right\} \quad \text{for all } x \in \mathscr{X}. \tag{1}$$

Using this lower transition rate operator $\underline{Q}$, we can compute conditional lower expectations in the following way. For any $t, s \in \mathbb{R}_{\geq 0}$, with $t \leq s$, and any $f \in \mathscr{L}(\mathscr{X})$, it has been shown that

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t] = \underline{\mathbb{E}}_{\mathscr{Q}}[f(X_s)\,|\,X_t] := \lim_{n \to +\infty} \left[I + \frac{(s-t)}{n}\underline{Q} \right]^n f,$$

where I is the identity operator on $\mathscr{L}(\mathscr{X})$, in the sense that $Ig = g$ for every $g \in \mathscr{L}(\mathscr{X})$. The notation $\underline{\mathbb{E}}_{\mathscr{Q}}$ is meant to indicate that this conditional lower expectation only depends on $\mathscr{Q}$, and not on $\mathscr{M}$. The above implies that for large enough $n \in \mathbb{N}$, and writing $\Delta := {(s-t)}/{n}$, we have

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t] = \underline{\mathbb{E}}_{\mathscr{Q}}[f(X_s)\,|\,X_t] \approx \left[I + \Delta\underline{Q} \right]^n f. \tag{2}$$

Concretely, this means that if one is able to solve the minimisation problem in Equation (1)—which is relatively straightforward for "nice enough" $\mathscr{Q}$, e.g., convex hulls of finite sets of rate matrices—then one can also compute conditional lower expectations using the expression in Equation 2. In practice, we do this by first computing $f_1' := \underline{Q}f$ using Equation (1), and then computing $f_1 := f + \Delta f_1'$. Next, we compute $f_2' := \underline{Q}f_1$, from which we obtain $f_2 := f_1 + \Delta f_2'$. Proceeding in this fashion, after n steps we then finally obtain $f_n := [I + \Delta\underline{Q}]f_{n-1} = \left[I + \Delta\underline{Q} \right]^n f$, which is roughly the quantity of interest $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t]$ provided that n was taken large enough.[2]

As noted above, the conditional lower expectation $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_s)|X_t]$ only depends on $\mathscr{Q}$. Similarly, and in contrast, the unconditional lower expectation at time zero only depends on $\mathscr{M}$. That is,

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{M}}[f(X_0)] := \inf \left\{ \sum_{x \in \mathscr{X}} p(x)f(x) : p \in \mathscr{M} \right\}. \tag{3}$$

Furthermore, the unconditional lower expectation at an arbitrary time $t \in \mathbb{R}_{\geq 0}$, is given by

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_t)] = \underline{\mathbb{E}}_{\mathscr{M}}\left[\underline{\mathbb{E}}_{\mathscr{Q}}[f(X_t)\,|\,X_0] \right], \tag{4}$$

which can therefore be computed by combining Equations (2) and (3). In particular, from a practical point of view, it suffices to first compute the conditional lower expectation $\underline{\mathbb{E}}_{\mathscr{Q}}[f(X_t)\,|\,X_0]$, using Equation (2). Once this quantity is obtained, it remains to compute the right-hand side of Equation (3), which again is relatively straightforward when $\mathscr{M}$ is "nice enough", e.g., the convex hull of some finite set of probability mass functions.

---

2. We refer the reader to (Krak et al., 2016, Proposition 8.5) for a theoretical bound on the minimum such n that is required to ensure a given maximum error on the approximation in Equation (2). We here briefly note that this bound scales polynomially in every relevant parameter. This means that $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[f(X_s)|X_t]$ is numerically computable in polynomial time, provided that $\mathscr{Q}$ is such that Equation (1) can also be solved in the same time-complexity order.

3. Imprecise Continuous-Time Hidden Markov Chains

In this section, we construct the *hidden* model that is the subject of this paper. Our aim is to augment the stochastic processes that were introduced in the previous section, by adding random *output* variables Y_t whose distribution depends on the state X_t at the same time point t.

We want to focus in this paper on the more practical aspect of solving the inference problem of interest, i.e., computing lower expectations on the state-space *given some observations*. Hence, we will assume that we are given some finite sequence of time points, and we then only consider these time points in augmenting the model. In order to disambiguate the notation, we will henceforth denote stochastic processes as $P_{\mathscr{X}}$, to emphasise that they are only concerned with the state-space.

3.1 Output Variables

We want to augment stochastic processes with random "output variables" Y_t, whose distribution depends on the state X_t. We here define the corresponding (conditional) distribution.

We want this definition to be fairly general, and in particular do not want to stipulate that Y_t should be either a discrete or a continuous random variable. To this end, we simply consider some set $\mathscr{Y}$ to be the outcome space of the random variable. We then let Σ be some algebra on $\mathscr{Y}$. Finally, for each $x \in \mathscr{X}$, we consider some finitely (and possibly σ-)additive probability measure $P_{\mathscr{Y}|\mathscr{X}}(\cdot|x)$ on $(\mathscr{Y}, \Sigma)$, with respect to which the random variable Y_t can be defined.

Definition 1 *An* output model *is a tuple* $(\mathscr{Y}, \Sigma, P_{\mathscr{Y}|\mathscr{X}})$*, where* $\mathscr{Y}$ *is an outcome space,* Σ *is an algebra on* $\mathscr{Y}$*, and, for all* $x \in \mathscr{X}$*,* $P_{\mathscr{Y}|\mathscr{X}}(\cdot|x)$ *is a finitely additive probability measure on* $(\mathscr{Y}, \Sigma)$*.*

When considering (multiple) explicit time points, we use notation analogous to that used for states; so, $\mathscr{Y}_t \coloneqq \mathscr{Y}$ for any time $t \in \mathbb{R}_{\geq 0}$, and for any $u \in \mathscr{U}$, we write $\mathscr{Y}_u \coloneqq \prod_{t \in u} \mathscr{Y}_t$.

We let Σ_u denote the set of all events of the type $O_u = \times_{t \in u} O_t$, where, for all $t \in u$, $O_t \in \Sigma$. This set Σ_u lets us describe observations using assessments of the form $(Y_t \in O_t$ for all $t \in u)$. For any $O_u \in \Sigma_u$ and $x_u \in \mathscr{X}_u$, we also adopt the shorthand notation $P_{\mathscr{Y}|\mathscr{X}}(O_u|x_u) \coloneqq \prod_{t \in u} P_{\mathscr{Y}|\mathscr{X}}(O_t|x_t)$.

3.2 Augmented Stochastic Processes

We now use this notion of an output model to define the stochastic model P that corresponds to a—precise—continuous-time *hidden* stochastic process. So, consider some fixed output model $(\mathscr{Y}, \Sigma, P_{\mathscr{Y}|\mathscr{X}})$, some fixed continuous-time stochastic process $P_{\mathscr{X}}$ and some fixed, non-empty and finite sequence of time-points $u \in \mathscr{U}$ on which observations of the outputs may take place.

We assume that Y_t is conditionally independent of *all* other variables, given the state X_t. This means that the construction of the augmented process P is relatively straightforward; we can simply multiply $P_{\mathscr{Y}|\mathscr{X}}(\cdot|X_t)$ with any distribution $P_{\mathscr{X}}(X_t, \cdot)$ that includes X_t to obtain the joint distribution including Y_t: for any $t \in u$ and $v \in \mathscr{U}$ such that $t \notin v$, any $x_t \in \mathscr{X}_t$ and $x_v \in \mathscr{X}_v$, and any $O_t \in \Sigma$:

$$P(Y_t \in O_t, X_t = x_t, X_v = x_v) \coloneqq P_{\mathscr{Y}|\mathscr{X}}(O_t|x_t) P_{\mathscr{X}}(X_t = x_t, X_v = x_v).$$

Similarly, when considering multiple output observations at once—say for the entire sequence u— then for any $v \in \mathscr{U}$ such that $u \cap v = \emptyset$, any $x_u \in \mathscr{X}_u$ and $x_v \in \mathscr{X}_v$, and any $O_u \in \Sigma_u$:

$$P(Y_u \in O_u, X_u = x_u, X_v = x_v) \coloneqq P_{\mathscr{Y}|\mathscr{X}}(O_u|x_u) P_{\mathscr{X}}(X_u = x_u, X_v = x_v).$$

Other probabilities can be derived by appropriate marginalisation. We denote the resulting augmented stochastic process as $P = P_{\mathscr{Y}|\mathscr{X}} \otimes P_{\mathscr{X}}$, for the specific output model $P_{\mathscr{Y}|\mathscr{X}}$ and stochastic process $P_{\mathscr{X}}$ that were taken to be fixed in this section.

3.3 Imprecise Continuous-Time Hidden Markov Chains

An *imprecise continuous-time hidden Markov chain* (ICTHMC) is a set of augmented stochastic processes, obtained by augmenting all processes in an ICTMC with some given output model.

Definition 2 *Consider any ICTMC $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$, and any output model $(\mathcal{Y},\Sigma,P_{\mathcal{Y}|\mathcal{X}})$. Then, the corresponding* imprecise continuous-time hidden Markov chain *(ICTHMC) $\mathcal{Z}$ is the set of augmented stochastic processes that is defined by $\mathcal{Z} := \left\{ P_{\mathcal{Y}|\mathcal{X}} \otimes P_{\mathcal{X}} : P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}} \right\}$. The lower expectation with respect to $\mathcal{Z}$ will be denoted by $\underline{\mathbb{E}}_{\mathcal{Z}}$.*

Note that we leave the parameters $\mathcal{M}$, $\mathcal{Q}$ and $P_{\mathcal{Y}|\mathcal{X}}$ implicit in the notation of the ICTHMC $\mathcal{Z}$—we will henceforth take these parameters to be fixed.

Also, the output model is taken to be precise, and shared by all processes in the set. One further generalisation that we aim to make in the future is to allow for an imprecise specification of this output model. However, this would force us into choosing an appropriate notion of independence; e.g., whether to enforce the independence assumptions made in Section 3.2, leading to strong or complete independence, or to only enforce the lower envelopes to have these independence properties, leading to epistemic irrelevance. It is currently unclear which choice should be preferred, e.g. with regard to computability, so at present we prefer to focus on this simpler model.

4. Updating the Model

Suppose now that we have observed that some event $(Y_u \in O_u)$ has taken place, with $O_u \in \Sigma_u$. We here use the terminology that we *update* our model with these observations, after which the updated model reflects our revised beliefs about some quantity of interest. These updated beliefs, about some function $f \in \mathcal{L}(\mathcal{X}_v)$, say, are then denoted by $\mathbb{E}_P[f(X_v)|Y_u \in O_u]$ or $\underline{\mathbb{E}}_{\mathcal{Z}}[f(X_v)|Y_u \in O_u]$, depending on whether we are considering a precise or an imprecise model. In this section, we provide definitions and alternative expressions for such updated (lower) expectations.

4.1 Observations with Positive (Upper) Probability

When our assertion $(Y_u \in O_u)$ about an observation at time points u has positive probability, we can—in the precise case—update our model by application of Bayes' rule. The following gives a convenient expression for the updated expectation $\mathbb{E}_P[f(X_v)|Y_u \in O_u]$, which makes use of the independence assumptions in Section 3.2 for augmented stochastic processes.

Proposition 3 *Let P be an augmented stochastic process and consider any $u,v \in \mathcal{U}$, $O_u \in \Sigma_u$ and $f \in \mathcal{L}(\mathcal{X}_v)$. Then the updated expectation is given by*

$$\mathbb{E}_P[f(X_v)|Y_u \in O_u] := \sum_{x_v \in \mathcal{X}_v} f(x_v) \frac{P(X_v = x_v, Y_u \in O_u)}{P(Y_u \in O_u)} = \frac{\mathbb{E}_{P_{\mathcal{X}}}[f(X_v)P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)]}{\mathbb{E}_{P_{\mathcal{X}}}[P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)]},$$

whenever $P(Y_u \in O_u) = \mathbb{E}_{P_{\mathcal{X}}}[P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)] > 0$, and is left undefined, otherwise.

Having defined above how to update all the precise models $P \in \mathcal{Z}$, we will now update the imprecise model through *regular extension* (Walley, 1991). This corresponds to simply discarding from $\mathcal{Z}$ those precise models that assign zero probability to $(Y_u \in O_u)$, updating the remaining models, and then computing their lower envelope.

198

Definition 4 *Let $\mathcal{Z}$ be an ICTHMC and consider any $u, v \in \mathcal{U}$, $O_u \in \Sigma_u$ and $f \in \mathcal{L}(\mathcal{X}_v)$. Then the updated lower expectation is defined by*

$$\underline{\mathbb{E}}_{\mathcal{Z}}\big[f(X_v)\,|\,Y_u \in O_u\big] \coloneqq \inf\big\{\mathbb{E}_P[f(X_v)\,|\,Y_u \in O_u] : P \in \mathcal{Z}, P(Y_u \in O_u) > 0\big\},$$

whenever $\overline{P}_{\mathcal{Z}}(Y_u \in O_u) = \overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[P_{\mathcal{Y}|\mathcal{X}}(O_u\,|\,X_u)] > 0$, and is left undefined, otherwise.

As is well known, the updated lower expectation that is obtained through regular extension satisfies Walley's *generalised Bayes' rule* (Walley, 1991). The following proposition gives an expression for this generalised Bayes' rule, rewritten using some of the independence properties of the model. We will shortly see why this expression is useful from a computational perspective.

Proposition 5 *Let $\mathcal{Z}$ be an ICTHMC and consider any $u, v \in \mathcal{U}$, $O_u \in \Sigma_u$ and $f \in \mathcal{L}(\mathcal{X}_v)$. Then, if $\overline{P}_{\mathcal{Z}}(Y_u \in O_u) = \overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[P_{\mathcal{Y}|\mathcal{X}}(O_u\,|\,X_u)] > 0$, the quantity $\underline{\mathbb{E}}_{\mathcal{Z}}\big[f(X_v)\,|\,Y_u \in O_u\big]$ satisfies*

$$\underline{\mathbb{E}}_{\mathcal{Z}}\big[f(X_v)\,|\,Y_u \in O_u\big] = \max\big\{\mu \in \mathbb{R} : \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[P_{\mathcal{Y}|\mathcal{X}}(O_u\,|\,X_u)\big(f(X_v) - \mu\big)\big] \geq 0\big\}.$$

4.2 Uncountable Outcome Spaces, Point Observations, and Probability Zero

An important special case where observations have probability zero for all precise models, but where we can still make informative inferences, is when we have an uncountable outcome space $\mathcal{Y}$ and the observations are points $y_u \in \mathcal{Y}_u$—i.e., when Y_u is continuous. In this case, it is common practice to define the updated expectation $\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u]$ as a limit of *conditional* expectations, where each conditioning event is an increasingly smaller region around this point y_u. We will start by formalising this idea in a relatively abstract way, but will shortly make this practicable. For the sake of intuition, note that we are working towards the introduction of probability density functions.

Fix any $P \in \mathcal{Z}$, consider any $y_u \in \mathcal{Y}_u$ and choose a sequence $\{O_u^i\}_{i \in \mathbb{N}}$ of events in Σ_u which shrink to y_u—i.e., such that $O_u^i \supseteq O_u^{i+1}$ for all $i \in \mathbb{N}$, and such that $\cap_{i \in \mathbb{N}} O_u^i = \{y_u\}$. We then define

$$\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u] \coloneqq \lim_{i \to +\infty} \mathbb{E}_P[f(X_v)\,|\,Y_u \in O_u^i]. \tag{5}$$

This limit exists if there is a sequence $\{\lambda_i\}_{i \in \mathbb{N}}$ in $\mathbb{R}_{>0}$ such that, for every $x_u \in \mathcal{X}_u$, the limit

$$\phi_u(y_u\,|\,x_u) \coloneqq \lim_{i \to +\infty} \frac{P_{\mathcal{Y}|\mathcal{X}}(O_u^i\,|\,x_u)}{\lambda_i}$$

exists, is real-valued—in particular, finite—and satisfies $\mathbb{E}_{P_{\mathcal{X}}}[\phi_u(y_u\,|\,X_u)] > 0$:

Proposition 6 *Let P be an augmented stochastic process and consider any $u, v \in \mathcal{U}$, $y_u \in \mathcal{Y}_u$ and $f \in \mathcal{L}(\mathcal{X}_v)$. For any $\{O_u^i\}_{i \in \mathbb{N}}$ in Σ_u that shrinks to y_u, if for some $\{\lambda_i\}_{i \in \mathbb{N}}$ in $\mathbb{R}_{>0}$ the quantity $\phi_u(y_u\,|\,X_u)$ exists, is real-valued, and satisfies $\mathbb{E}_{P_{\mathcal{X}}}[\phi_u(y_u\,|\,X_u)] > 0$, then*

$$\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u] \coloneqq \lim_{i \to +\infty} \mathbb{E}_P[f(X_v)\,|\,Y_u \in O_u^i] = \frac{\mathbb{E}_{P_{\mathcal{X}}}[f(X_v)\phi_u(y_u\,|\,X_u)]}{\mathbb{E}_{P_{\mathcal{X}}}[\phi_u(y_u\,|\,X_u)]}. \tag{6}$$

Note that $\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u]$ is clearly dependent on the exact sequence $\{O_u^i\}_{i \in \mathbb{N}}$. Unfortunately, this is the best we can hope for at the level of generality that we are currently dealing with. For brevity, we nevertheless omit from the notation the updated expectation's dependency on this sequence. However, as we will explain below, this should not be problematic for most practical applications.

It is also useful to note that $\phi_u(y_u|x_u)$ can often be constructed "piecewise". That is, if for every $t \in u$ there is a sequence $\{\lambda_{t,i}\}_{i \in \mathbb{N}}$ in $\mathbb{R}_{>0}$ such that, for all $x_t \in \mathscr{X}_t$,

$$\phi_t(y_t|x_t) := \lim_{i \to +\infty} \frac{P_{\mathscr{Y}|\mathscr{X}}(O_t^i|x_t)}{\lambda_{t,i}}$$

exists and is real-valued, then choosing $\{\lambda_i\}_{i \in \mathbb{N}}$ as $\lambda_i := \prod_{t \in u} \lambda_{t,i}$ yields $\phi_u(y_u|x_u) = \prod_{t \in u} \phi_t(y_t|x_t)$.

Now, to make the above practicable, we can for example assume that if $\mathscr{Y}$ is uncountable, then it is the set $\mathscr{Y} = \mathbb{R}^d$, for some $d \in \mathbb{N}$, and that Σ is the Borel σ-algebra on $\mathbb{R}^d$. For each $x \in \mathscr{X}$, we then assume that the measure $P_{\mathscr{Y}|\mathscr{X}}(\cdot|x)$ is induced by some given *probability density function*: a measurable function $\psi(\cdot|x) : \mathscr{Y} \to \mathbb{R}_{\geq 0}$ such that $\int_{\mathscr{Y}} \psi(y|x)\,\mathrm{d}y = 1$ and, for every $O \in \Sigma$,

$$P_{\mathscr{Y}|\mathscr{X}}(O|x) := \int_O \psi(y|x)\,\mathrm{d}y,$$

where the integrals are understood in the Lebesgue sense.

Then choose any $y_u \in \mathscr{Y}_u$, any $t \in u$, any sequence $\{O_t^i\}_{i \in \mathbb{N}}$ of open balls in $\mathscr{Y}_t$ that are centred on, and shrink to, y_t, and fix any $x_u \in \mathscr{X}_u$. If $\psi(\cdot|x_t)$ is continuous at y_t, it can be shown that

$$\phi_t(y_t|x_t) = \lim_{i \to +\infty} \frac{P_{\mathscr{Y}|\mathscr{X}}(O_t^i|x_t)}{\lambda(O_t^i)} = \psi(y_t|x_t), \tag{7}$$

where $\lambda(O_t^i)$ denotes the Lebesgue measure of O_t^i. So, we can construct the sequence $\{O_u^i\}_{i \in \mathbb{N}}$ such that every $O_u^i := \prod_{t \in u} O_t^i$, with each O_t^i chosen as above. If we then choose the sequence $\{\lambda_i\}_{i \in \mathbb{N}}$ as $\lambda_i := \prod_{t \in u} \lambda(O_t^i)$ for each $i \in \mathbb{N}$, we find $\phi_u(y_u|x_u) = \prod_{t \in u} \phi_t(y_t|x_t) = \prod_{t \in u} \psi(y_t|x_t)$, provided that each $\phi_t(y_t|x_t)$ satisfies Equation (7). It can now be seen that, under these assumptions, the right-hand side of Equation (6) is simply the well-known Bayes' rule for (finite) mixtures of densities.

In most practical applications, therefore, the function $\phi_u(\cdot|x_u)$ is known explicitly; one may assume, for example, that Y_t follows a Normal distribution with parameters depending on X_t, and the functions $\phi_t(\cdot|x_t)$—and by extension, $\phi_u(\cdot|x_u)$—then follow directly by identification with $\psi(\cdot|x_t)$. Furthermore, arguably, most of the density functions that one encounters in practice will be continuous and strictly positive at y_t. This guarantees that the limit in Equation (7) exists, and largely solves the interpretation issue mentioned above: when $\phi_u(y_u|X_u) = \prod_{t \in u} \psi(y_t|X_t)$ is continuous and positive at y_u, $\mathbb{E}_P[f(X_v)|Y_u = y_u]$ exists and is the same for almost[3] all sequences $\{O_u^i\}_{i \in \mathbb{N}}$.

Moving on, note that if $\phi_u(y_u|X_u)$ exists and satisfies $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] > 0$, then the updated expectation $\mathbb{E}_P[f(X_v)|Y_u = y_u]$ is well-defined for every $P \in \mathscr{X}$. Hence, we can then update the imprecise model by updating each of the precise models that it consists of.

Definition 7 *Let $\mathscr{X}$ be an ICTHMC and consider any $u, v \in \mathscr{U}$, $y_u \in \mathscr{Y}_u$, and $f \in \mathscr{L}(\mathscr{X}_v)$. For any $\{O_u^i\}_{i \in \mathbb{N}}$ in Σ_u that shrinks to y_u, if for some $\{\lambda_i\}_{i \in \mathbb{N}}$ in $\mathbb{R}_{>0}$ the quantity $\phi_u(y_u|X_u)$ exists and is real-valued, the updated lower expectation is defined by*

$$\underline{\mathbb{E}}_{\mathscr{X}}\left[f(X_v)|Y_u = y_u\right] := \inf\{\mathbb{E}_P[f(X_v)|Y_u = y_u] : P \in \mathscr{X}\},$$

whenever $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] > 0$, and is left undefined, otherwise.

---

3. It suffices if, for all $t \in u$, there is a sequence of open balls $\{B_t^i\}_{i \in \mathbb{N}}$ in $\mathscr{Y}$ that shrinks to y_t such that, for all $i \in \mathbb{N}$, O_t^i has positive Lebesgue measure and is contained in B_t^i.

Similar to the results in Section 4.1, this updated lower expectation satisfies a "generalised Bayes' rule for mixtures of densities", in the following sense.

Proposition 8 *Let $\mathscr{X}$ be an ICTHMC and consider any $u, v \in \mathscr{U}$, $y_u \in \mathscr{Y}_u$ and $f \in \mathscr{L}(\mathscr{X}_v)$. For any $\{O_u^i\}_{i \in \mathbb{N}}$ in Σ_u that shrinks to y_u, if for some $\{\lambda_i\}_{i \in \mathbb{N}}$ in $\mathbb{R}_{>0}$ the quantity $\phi_u(y_u|X_u)$ exists, is real-valued, and satisfies $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] > 0$, then*

$$\underline{\mathbb{E}}_{\mathscr{X}}\big[f(X_v)\,|\,Y_u = y_u\big] = \max\big\{\mu \in \mathbb{R} : \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}\big[\phi_u(y_u|X_u)\big(f(X_v) - \mu\big)\big] \geq 0\big\}. \tag{8}$$

Furthermore, this updated imprecise model can be given an intuitive limit interpretation.

Proposition 9 *Let $\mathscr{X}$ be an ICTHMC and consider any $u, v \in \mathscr{U}$, $y_u \in \mathscr{Y}_u$ and $f \in \mathscr{L}(\mathscr{X}_v)$. For any $\{O_u^i\}_{i \in \mathbb{N}}$ in Σ_u that shrinks to y_u, if for some $\{\lambda_i\}_{i \in \mathbb{N}}$ in $\mathbb{R}_{>0}$ the quantity $\phi_u(y_u|X_u)$ exists, is real-valued, and satisfies $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] > 0$, then $\underline{\mathbb{E}}_{\mathscr{X}}[f(X_v)|Y_u = y_u] = \lim_{i \to +\infty} \underline{\mathbb{E}}_{\mathscr{X}}[f(X_v)|Y_u \in O_u^i]$.*

Now, recall that the requirement $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] > 0$ for updating the imprecise model is a sufficient condition to guarantee that *all* the precise updated models are well-defined. However, one may wonder whether it is also possible to update the imprecise model under weaker conditions. Indeed, one obvious idea would be to define the updated model more generally as

$$\underline{\mathbb{E}}_{\mathscr{X}}^{\mathrm{R}}[f(X_v)\,|\,Y_u = y_u] := \inf\{\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u] : P \in \mathscr{X}, \mathbb{E}_{P_{\mathscr{X}}}[\phi_u(y_u|X_u)] > 0\},$$

whenever $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] > 0$; this guarantees that *some* of the precise updated models are well-defined. This updated lower expectation satisfies the same generalised Bayes' rule as above, i.e. the right-hand side of Equation (8) is equal to $\underline{\mathbb{E}}_{\mathscr{X}}^{\mathrm{R}}[f(X_v)\,|\,Y_u = y_u]$ whenever $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] > 0$. However, the limit interpretation then fails to hold, in the sense that it is possible to construct an example where $\underline{\mathbb{E}}_{\mathscr{X}}^{\mathrm{R}}[f(X_v)\,|\,Y_u = y_u] \neq \lim_{i \to +\infty} \underline{\mathbb{E}}_{\mathscr{X}}[f(X_v)\,|\,Y_u \in O_u^i]$, with $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] > 0$ but $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\phi_u(y_u|X_u)] = 0$. We feel that this makes this more general updating scheme somewhat troublesome from an interpretation (and hence philosophical) point of view.

On the other hand, we recall that the existence of $\phi_u(y_u|X_u)$ and the positivity of $\mathbb{E}_{P_{\mathscr{X}}}[\phi_u(y_u|X_u)]$ are necessary and sufficient conditions for the limit in Equation (5) to exist and be computable using Equation (6). However, these conditions are sufficient but non-necessary for that limit to simply exist. Therefore, a different way to generalise the imprecise updating method would be

$$\underline{\mathbb{E}}_{\mathscr{X}}^{\mathrm{L}}[f(X_v)\,|\,Y_u = y_u] := \inf\{\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u] : P \in \mathscr{X}, \mathbb{E}_P[f(X_v)\,|\,Y_u = y_u] \text{ exists}\},$$

whenever $\{P \in \mathscr{X} : \mathbb{E}_P[f(X_v)\,|\,Y_u = y_u] \text{ exists}\} \neq \emptyset$. We conjecture that this updated model *does* satisfy the limit interpretation, but on the other hand, it is possible to show that this, in turn, no longer satisfies the above generalised Bayes' rule. That makes this updating scheme somewhat troublesome from a practical point of view because, as we discuss below, the expression in Equation (8) is crucial for our method of efficient computation of the updated lower expectation.

5. Inference Algorithms

In the previous section, we have seen that we can use the generalised Bayes' rule for updating our ICTHMC with some given observations. From a computational point of view, this is particularly useful because, rather than having to solve the non-linear optimisation problems in Definitions 4 or 7 directly, we can focus on evaluating the function $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}\big[P_{\mathscr{Y}|\mathscr{X}}(O_u|X_u)\big(f(X_v) - \mu\big)\big]$, or its

density-analogue, for some fixed value of μ. Finding the updated lower expectation is then a matter of finding the maximum value of μ for which this quantity is non-negative. As we will discuss in Section 5.1, this is a relatively straightforward problem to solve numerically.

Therefore, in order for this approach to be computationally tractable, we require efficient algorithms that can evaluate this quantity for a given value of μ. In Section 5.2, we provide such an algorithm for the important case where the function f depends on a single time-point.

We first generalise the problem so that these results are applicable both for observations of the form $(Y_u \in O_u)$, and for point-observations $(Y_u = y_u)$ in an uncountable outcome space. Recall that

$$P_{\mathscr{Y}|\mathscr{X}}(O_u|X_u) = \prod_{t \in u} P_{\mathscr{Y}|\mathscr{X}}(O_t|X_t) \qquad \text{and} \qquad \phi_u(y_u|X_u) = \prod_{t \in u} \phi_t(y_t|X_t).$$

In both cases, we can rewrite this expression as $\prod_{t \in u} g_t(X_t)$, where, for all $t \in u$, $g_t \in \mathscr{L}(\mathscr{X}_t)$ and $g_t \geq 0$. The function of interest is then $\mathbb{E}_{\mathscr{Q},\mathscr{M}}\left[\left(\prod_{t \in u} g_t(X_t)\right)\left(f(X_v) - \mu\right)\right]$ and the sign conditions in Propositions 5 and 8 reduce to $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\prod_{t \in u} g_t(X_t)] > 0$ and $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\prod_{t \in u} g_t(X_t)] > 0$, respectively.

5.1 Solving the Generalised Bayes' Rule

Finding the maximum value of μ for which the function of interest in the generalised Bayes' rule is non-negative, is relatively straightforward numerically. This is because this function, parameterised in μ, is very well-behaved. The proposition below explicitly states some of its properties. These are essentially well-known, and can also be found in other work; see, e.g., (De Bock, 2015, Section 2.7.3). The statement below is therefore intended to briefly recall these properties, and is stated in a general form where we can also use it when working with densities.

Proposition 10 *Let $\mathbb{P}_{\mathscr{Q},\mathscr{M}}$ be an ICTMC and consider any $u,v \in \mathscr{U}$, any $f \in \mathscr{L}(\mathscr{X}_v)$ and, for all $t \in u$, any $g_t \in \mathscr{L}(\mathscr{X}_t)$ such that $g_t \geq 0$. Consider the function $G : \mathbb{R} \to \mathbb{R}$ that is given, for all $\mu \in \mathbb{R}$, by $G(\mu) \coloneqq \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}\left[\left(\prod_{t \in u} g_t(X_t)\right)\left(f(X_v) - \mu\right)\right]$. Then the following properties hold:*

G1: G is continuous, non-increasing, concave, and has a root, i.e. $\exists \mu \in \mathbb{R} : G(\mu) = 0$.

G2: If $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}\left[\prod_{t \in u} g_t(X_t)\right] > 0$, then G is (strictly) decreasing, and has a unique root.

G3: If $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}\left[\prod_{t \in u} g_t(X_t)\right] = 0$ but $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}\left[\prod_{t \in u} g_t(X_t)\right] > 0$, then G has a maximum root μ_*, satisfies $G(\mu) = 0$ for all $\mu \leq \mu_*$, and is (strictly) decreasing for $\mu > \mu_*$.

G4: If $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}\left[\prod_{t \in u} g_t(X_t)\right] = 0$, then G is identically zero, i.e. $\forall \mu \in \mathbb{R} : G(\mu) = 0$.

Note that the function G in Proposition 10 can behave in three essentially different ways. These correspond to the cases where the observed event has strictly positive probability(/density) for *all* processes in the set; to where it only has positive probability(/density) for *some* processes; and to where it has *zero* probability(/density) for *all* processes. In the first two cases—which are the important ones to apply the generalised Bayes' rule—the function is "well-behaved" enough to make finding its maximum root a fairly simple task. For instance, a standard bisection/bracketing algorithm can be applied here, known in this context as Lavine's algorithm (Cozman, 1997).

We sketch this method below. First, note that due to Propositions 5 and 8, the maximum root will always be found in the interval $[\min f, \max f]$. The properties above therefore provide us with a way to check the sign conditions for updating. That is, for any $\mu > \max f$, we see that $G(\mu) < 0$ if and only if $\overline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}[\prod_{t \in u} g_t(X_t)] > 0$; similarly, for any $\mu < \min f$, we see that $G(\mu) > 0$ if and only if

if $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\prod_{t \in u} g_t(X_t)] > 0$. Evaluating G at such values of μ is therefore sufficient to check the sign conditions in Propositions 5 and 8.

The algorithm now starts by setting $\mu_- := \min f$, and $\mu_+ := \max f$; if $G(\mu_+) = 0$, we know that μ_+ is the quantity of interest. Otherwise, proceed iteratively in the following way. Compute the half-way point $\mu := 1/2(\mu_+ - \mu_-)$; then, if $G(\mu) \geq 0$ set $\mu_- := \mu$, otherwise set $\mu_+ := \mu$; then repeat. Clearly, the interval $[\mu_-, \mu_+]$ still contains the maximum root after each step. The procedure can be terminated whenever $(\mu_+ - \mu_-) < \varepsilon$, for some desired numerical precision $\varepsilon > 0$. Since the width of the interval is halved at each iteration, the runtime of this procedure is $O\big(\log\{(\max f - \min f)\varepsilon^{-1}\}\big)$. Methods for improving the numerical stability of this procedure can be found in (De Bock, 2015, Section 2.7.3).

5.2 Functions on a Single Time Point

Having discussed an efficient method to find the maximum root of the function $G(\mu)$ in Section 5.1, it now remains to provide an efficient method to numerically *evaluate* this function for a given value of μ. Clearly, any such method will depend on the choice of f.

We focus on a particularly useful special case, which can be used to compute the updated lower expectation of a function $f \in \mathcal{L}(\mathcal{X}_s)$ on a single time point s, given observations at time points u. If $s \notin u$, then it will be notationally convenient to define $g_s := f - \mu$, and to let $u' := u \cup \{s\}$. We can then simply focus on computing

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\left(\prod_{t \in u} g_t(X_t)\right)(f(X_s) - \mu)\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{t \in u'} g_t(X_t)\right].$$

On the other hand, if $s = t$ for some $t \in u$, we let $u' := u$ and replace g_t by $(f - \mu)g_t$. Clearly, the above equality then also holds; the point is simply to establish a uniform indexing notation over all time-points and functions. The right hand side of the above equality can now be computed using the following dynamic programming technique.

For all $t \in u'$, we define auxiliary functions $g_t^+, g_t^- \in \mathcal{L}(\mathcal{X}_t)$, as follows. Writing $u' = t_0, \ldots, t_n$, let $g_{t_n}^+ := g_{t_n}^- := g_{t_n}$. Next, for all $i \in \{0, \ldots, n-1\}$ and all $x_{t_i} \in \mathcal{X}_{t_i}$, let

$$g_{t_i}^+(x_{t_i}) := \begin{cases} g_{t_i}(x_{t_i})\underline{\mathbb{E}}_{\mathcal{Q}}[g_{t_{i+1}}^+(X_{t_{i+1}})\,|\,X_{t_i} = x_{t_i}] & \text{if } g_{t_i}(x_{t_i}) \geq 0, \\ g_{t_i}(x_{t_i})\overline{\mathbb{E}}_{\mathcal{Q}}[g_{t_{i+1}}^-(X_{t_{i+1}})\,|\,X_{t_i} = x_{t_i}] & \text{if } g_{t_i}(x_{t_i}) < 0 \end{cases}$$

and

$$g_{t_i}^-(x_{t_i}) := \begin{cases} g_{t_i}(x_{t_i})\overline{\mathbb{E}}_{\mathcal{Q}}[g_{t_{i+1}}^-(X_{t_{i+1}})\,|\,X_{t_i} = x_{t_i}] & \text{if } g_{t_i}(x_{t_i}) \geq 0, \\ g_{t_i}(x_{t_i})\underline{\mathbb{E}}_{\mathcal{Q}}[g_{t_{i+1}}^+(X_{t_{i+1}})\,|\,X_{t_i} = x_{t_i}] & \text{if } g_{t_i}(x_{t_i}) < 0. \end{cases}$$

Clearly, backward recursion allows us to compute all these functions in a time-complexity order that is linear in the number of time points in u'. Practically, at each step, computing the quantities $\underline{\mathbb{E}}_{\mathcal{Q}}[g_{t_{i+1}}^+(X_{t_{i+1}})\,|\,X_{t_i} = x_{t_i}]$ and $\overline{\mathbb{E}}_{\mathcal{Q}}[g_{t_{i+1}}^-(X_{t_{i+1}})\,|\,X_{t_i} = x_{t_i}]$ can be done using Equation (2) and the method described in Section 2.2. Due to the results in (Krak et al., 2016), each of these quantities is computable in polynomial time. So, the total complexity of computing all these functions is clearly also polynomial. We now have the following result.

Proposition 11 *For all $t \in u'$, let g_t, g_t^+ and g_t^- be as defined above. Then the function of interest is given by $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\prod_{t \in u'} g_t(X_t)] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g_{t_0}^+(X_{t_0})\big]$. Also, $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\prod_{t \in u'} g_t(X_t)] = \overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g_{t_0}^-(X_{t_0})\big]$.*

So, in order to evaluate the function of interest, it remains to compute $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g_{t_0}^+(X_{t_0})\big]$. Since $g_{t_0}^+$ is a function on a single time point t_0, this can again be done in polynomial time, using Equation (4).

6. Conclusions and Future Work

We considered the problem of performing inference with *imprecise continuous-time hidden Markov chains*; an extension of *imprecise continuous-time Markov chains* obtained by augmenting them with random *output* variables, which may be either discrete or continuous. Our main result is an efficient, polynomial runtime, algorithm to compute lower expectations of functions that depend on the state-space at any given time-point, given a collection of observations of the output variables.

In future work, we intend to further generalise this model, by also allowing for imprecise output variables. Furthermore, we also aim to develop algorithms for other inference problems, such as the problem of computing updated lower expectations of functions $f \in \mathscr{L}(\mathscr{X}_v)$ that depend on more than one time-point. Similarly, we aim to investigate predictive output inferences, i.e., the lower probability/density of observations, which has uses in classification problems. Another such problem is that of estimating state-sequences given observed output-sequences—as was previously done for (discrete-time) iHMM's (De Bock and de Cooman, 2014).

Acknowledgments

References

F. Cozman. *Alternatives to Lavine's algorithm for calculation of posterior bounds given convex sets of distributions*. Carnegie Mellon University, The Robotics Institute, 1997.

J. De Bock. *Credal networks under epistemic irrelevance: theory and algorithms*. PhD thesis, Ghent University, 2015.

J. De Bock and G. de Cooman. An efficient algorithm for estimating state sequences in imprecise hidden Markov models. *Journal of Artificial Intelligence Research*, 50:189–233, 2014.

G. de Cooman, F. Hermans, A. Antonucci, and M. Zaffalon. Epistemic irrelevance in credal nets: the case of imprecise Markov trees. *International Journal of Approximate Reasoning*, 51(9): 1029–1052, 2010.

T. Krak, J. De Bock, and A. Siebes. Imprecise continuous-time Markov chains. *Under Review. Pre-print:* `https://arxiv.org/abs/1611.05796`, 2016.

T. Krak, J. De Bock, and A. Siebes. Efficient computation of updated lower expectations for imprecise continuous-time hidden Markov chains. *Extended ArXiv pre-print:* `https://arxiv.org/abs/1702.06791`, 2017.

D. Škulj. Efficient computation of the bounds of continuous time imprecise Markov chains. *Applied Mathematics and Computation*, 250(C):165–180, Jan. 2015.

P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.

W. Wei, B. Wang, and D. Towsley. Continuous-time hidden Markov models for network performance evaluation. *Performance Evaluation*, 49(1):129–146, 2002.