Imprecise Continuous-Time Hidden Markov Chains and Efficiently Computing their Lower Expectations

Thomas Krak[a,*], Jasper De Bock[a], Arno Siebes[b]

[a] *Ghent University, Department of Electronics and Information Systems, Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium*
[b] *Utrecht University, Department of Information and Computing Sciences, Princetonplein 5, De Uithof, 3584 CC Utrecht, The Netherlands*

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 and/or on these 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 various such inference problems. In particular, we cover the computation of unconditional probabilities and density values of the output variables, as well as functions on the state-space given observations of the outputs.

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) [13]. 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* [12], are *imprecise continuous-time Markov chains* (ICTMC's) [10, 7]. 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.

*Corresponding author
Email addresses: `thomas.krak@ugent.be` (Thomas Krak), `jasper.debock@ugent.be` (Jasper De Bock), `a.p.j.m.siebes@uu.nl` (Arno Siebes)

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 and on the output variables. We also consider the computation of lower expectations with respect to the *updated* model; lower expectations of functions on the states-space given observed realisations of the output variables.

The proofs of all results are gathered in an appendix, where they are largely ordered by their chronological appearance in the main text.

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) [5] 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 [7] 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 [5] 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.

We would like to point out that our ideas and results on using continuous (output) variables are not, themselves, new or novel. Indeed, many or most of the ideas that we use were already discussed in [12, Section 6.10]. However, our results do provide the required rigour for using such variables in ICTHMC's. Furthermore, the work by Benavoli *et al.* [1] also considers continuous output variables in the related context of the (imprecise) *filtering problem*. However, in deriving their methods they assume that these continuous variables are idealisations that are only observed up to finite precision, whence they are effectively replaced by discrete and finite output variables.

Finally, we would like to point out that this present paper represents an extended version of previously published work [8]. The merit of this extended paper is on several points. First of all, the algorithm that we present is here framed much more generally than in this previous work. This widens the applicability of this result. To demonstrate this, we here also show that the algorithm can be used to efficiently solve other inference problems that were not covered in this previous work. Finally, all the proofs of our results have so far remained unpublished—they are now added to the appendix.

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 $\mathbb{N}_0 \coloneqq \mathbb{N} \cup \{0\}$.

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 $\mathcal{U}$ be the entire set of them. For any sequence of time points $u \in \mathcal{U}$ such that $u = t_0, \ldots, t_n$, we write for any $t \in \mathbb{R}_{\geq 0}$ that $u < t$ whenever $t_i < t$ for all $i \in \{0, \ldots, n\}$. Similarly, for any $u, v \in \mathcal{U}$, we write $u < v$ when all time-points in u are strictly less than all time-points in v.

Throughout, we consider some fixed, finite state space $\mathcal{X}$. A generic element of $\mathcal{X}$ will be denoted by x. When considering the state-space at a specific time t, we write $\mathcal{X}_t \coloneqq \mathcal{X}$, 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 $\mathcal{X}_u \coloneqq \prod_{t_i \in u} \mathcal{X}_{t_i}$, of which $x_u = (x_{t_0}, \dots, x_{t_n})$ is a generic element.

For any $u \in \mathcal{U}$, we let $\mathcal{L}(\mathcal{X}_u)$ be the set of all real-valued functions on $\mathcal{X}_u$. We endow these function spaces with the L^∞-norm, i.e. the norm $\|f\|$ of any $f \in \mathcal{L}(\mathcal{X}_u)$ is defined to be $\|f\| \coloneqq \|f\|_\infty = \max\{|f(x_u)| : x_u \in \mathcal{X}_u\}$. Limits of functions are to be interpreted under this norm.

Furthermore, we sometimes use the shorthand notation $\{a_i\}_{i\in\mathbb{N}} \to c$ for convergent sequences of quantities, which should be read as $\lim_{i\to+\infty} a_i = c$. If this limit is approached from above or below, we write $\{a_i\}_{i\in\mathbb{N}} \to c^+$ or $\{a_i\}_{i\in\mathbb{N}} \to c^-$, respectively.

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. [7]. 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 $\mathcal{X}$.

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 $\mathcal{X}_0$. 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 $|\mathcal{X}| \times |\mathcal{X}|$ 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

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

for some rate matrix Q_{t,x_u}, where I denotes the $|\mathcal{X}| \times |\mathcal{X}|$ 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 $\mathcal{Q}$, and a *set* of initial distributions $\mathcal{M}$. The corresponding ICTMC, denoted by $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$, is the set of all continuous-time stochastic processes whose dynamics can be described using the elements of $\mathcal{Q}$, and whose initial distributions are consistent with $\mathcal{M}$. That is, $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ is the set of stochastic processes P for which $P(X_0) \in \mathcal{M}$ and for which $Q_{t,x_u} \in \mathcal{Q}$ for every time t and history x_u.

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

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

where $\mathbb{E}_P[\cdot \,|\, \cdot]$ denotes the expectation with respect to the (precise) stochastic process P. The *upper expectation* $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}$ is defined similarly, and is derived through the well-known conjugacy property $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\cdot \,|\, \cdot] = -\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[-\cdot \,|\, \cdot]$. 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 $A \subseteq \mathcal{X}$, we have that $\underline{P}_{\mathcal{Q},\mathcal{M}}(X_s \in A \,|\, X_t) \coloneqq \inf\{P(X_s \in A | X_t) \,:\, P \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}\} = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\mathbb{I}_A(X_s) \,|\, X_t]$, where $\mathbb{I}_A$ is the indicator of A, defined for all $x \in \mathcal{X}$ 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 $\mathcal{M}$ is non-empty, and that $\mathcal{Q}$ is non-empty, bounded,[1] convex, and has *separately specified rows.* This latter property states that $\mathcal{Q}$ is closed under arbitrary recombination of rows from its elements; see [7, Definition 24] for a formal definition. Under these assumptions, $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ satisfies an *imprecise Markov property*, in the sense that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_s) \,|\, X_t, X_u = x_u] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_s) \,|\, X_t]$. This property explains why we call this model an imprecise continuous-time "Markov" chain.

Furthermore, it can be shown that under these assumptions, the corresponding lower expectation satisfies a very convenient property that is known as the law of *iterated lower expectation*:

Proposition 2.1 (Iterated Lower Expectation)**.** *[7, Theorem 6.5] Let $\mathcal{Q}$ be a non-empty, bounded and convex set of rate matrices, and let $\mathcal{M}$ be a non-empty set of probability mass functions on $\mathcal{X}_0$. Let $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ denote the corresponding ICTMC. Let $u \subset \mathbb{R}_{\geq 0}$ be a finite (possibly empty) sequence of time-points, and consider any $v, w \in \mathcal{U}$ such that $u < v < w$. Choose any $f \in \mathcal{L}(\mathcal{X}_{u\cup v\cup w})$. Then,*

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_u, X_v, X_w) \,|\, X_u] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_u, X_v, X_w) \,|\, X_u, X_v] \,|\, X_u\right].$$

This property is useful because it simplifies considerably the process of computing lower expectations of functions that depend on multiple time-points. Specifically, it allows us to repeatedly decompose such problems until we are left with a problem that requires only the computation on a single time-point. For details of this method, we refer to [7, Section 9.2] and Section 4 further on.

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. [7] 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 because, as mentioned in the previous section, the problem for multiple time-points can be reduced to this problem due to Proposition 2.1.

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

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

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 $f \in \mathcal{L}(\mathcal{X})$, it has been shown that

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_s) \,|\, X_t] = \underline{\mathbb{E}}_{\mathcal{Q}}[f(X_s) \,|\, X_t] \coloneqq \lim_{n\to+\infty}\left[I + \frac{(s-t)}{n}\underline{Q}\right]^n f\,, \tag{2}$$

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

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_s) \,|\, X_t] = \underline{\mathbb{E}}_{\mathcal{Q}}[f(X_s) \,|\, X_t] \approx \left[I + \Delta\underline{Q}\right]^n f\,. \tag{3}$$

Concretely, this means that if one is able to solve the minimisation problem in Equation (1)—which is relatively straightforward for "nice enough" $\mathcal{Q}$, e.g., convex hulls of finite sets of rate matrices—then one can also compute conditional lower expectations using the expression in Equation 3. In practice, we do this by first computing $f_1' \coloneqq \underline{Q}f$ using Equation (1), and then computing $f_1 \coloneqq f + \Delta f_1'$. Next, we compute $f_2' \coloneqq \underline{Q}f_1$, from which we obtain $f_2 \coloneqq f_1 + \Delta f_2'$. Proceeding in this fashion, after n steps we then finally

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

obtain $f_n \coloneqq [I + \Delta\underline{Q}]f_{n-1} = \big[I + \Delta\underline{Q}\big]^n f$, which is roughly the quantity of interest $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t]$ provided that n was taken large enough.[2][3]

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

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_0)] = \underline{\mathbb{E}}_{\mathcal{M}}[f(X_0)] \coloneqq \inf\left\{\sum_{x\in\mathcal{X}} p(x)f(x) \,:\, p \in \mathcal{M}\right\}. \tag{4}$$

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

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_t)] = \underline{\mathbb{E}}_{\mathcal{M}}\big[\underline{\mathbb{E}}_{\mathcal{Q}}[f(X_t)\,|\,X_0]\big], \tag{5}$$

which can therefore be computed by combining Equations (3) and (4). In particular, from a practical point of view, it suffices to first compute the conditional lower expectation $\underline{\mathbb{E}}_{\mathcal{Q}}[f(X_t)\,|\,X_0]$, using Equation (3). Once this quantity is obtained, it remains to compute the right-hand side of Equation (4), which again is relatively straightforward when $\mathcal{M}$ is "nice enough", e.g., the convex hull of some finite set of probability mass functions.

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 particular inference problems of interest, i.e., computing lower expectations of given functions. In particular, we will assume throughout that these functions only depend on a finite number of time-points. 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. This simplifies considerably the formalisation of the problem, because we can disregard the problem of ensuring "internal consistency" of the model at an infinite number of time-points simultaneously. In order to disambiguate the notation, we will henceforth denote stochastic processes as $P_{\mathcal{X}}$, 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 $\mathcal{Y}$ to be the outcome space of the random variable. We then let Σ be some algebra on $\mathcal{Y}$. Finally, for each $x \in \mathcal{X}$, we consider some finitely (and possibly σ-)additive probability measure $P_{\mathcal{Y}|\mathcal{X}}(\cdot|x)$ on $(\mathcal{Y}, \Sigma)$, with respect to which the random variable Y_t can be defined.

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

[2] We refer the reader to [7, 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 (3). We here briefly note that this bound scales polynomially in every relevant parameter. This means that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t]$ is numerically computable in polynomial time, provided that $\mathcal{Q}$ is such that Equation (1) can also be solved in the same time-complexity order.

[3] The approximation/computational method expressed in Equation (3) is *uniform*, in that it consists of a sequence of steps each with length $\Delta = (s-t)/n$. We refer to recent work by Erreygers and De Bock [6] for more efficient computational methods.

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

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 $x_u \in \mathcal{X}_u$, we also adopt the shorthand notation $P_{\mathcal{Y}|\mathcal{X}}(O_u|x_u) \coloneqq \prod_{t\in u} P_{\mathcal{Y}|\mathcal{X}}(O_t|x_t)$.

We will assume throughout that whenever we consider a function on the outputs, i.e. some $f : \mathcal{Y} \to \mathbb{R}$, say, that it is measurable, and in particular that we have some method of computing its expectation $\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[f(Y) \,|\, X = x]$ with respect to the measure $P_{\mathcal{Y}|\mathcal{X}}(\cdot|x)$, for all $x \in \mathcal{X}$.

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 $(\mathcal{Y}, \Sigma, P_{\mathcal{Y}|\mathcal{X}})$, some fixed continuous-time stochastic process $P_\mathcal{X}$ and some fixed, non-empty and finite sequence of time-points $u \in \mathcal{U}$ 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 $P_{\mathcal{Y}|\mathcal{X}}(\cdot \,|\, X_t)$ with any distribution $P_\mathcal{X}(X_t, \cdot)$ that includes X_t to obtain the joint distribution including Y_t: for any $t \in u$ and $v \in \mathcal{U}$ such that $t \notin v$, any $x_t \in \mathcal{X}_t$ and $x_v \in \mathcal{X}_v$, and any $O_t \in \Sigma$:

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

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

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

Other probabilities can be derived by appropriate marginalisation. We denote the resulting augmented stochastic process as $P = P_{\mathcal{Y}|\mathcal{X}} \otimes P_\mathcal{X}$, for the specific output model $P_{\mathcal{Y}|\mathcal{X}}$ and stochastic process $P_\mathcal{X}$ 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 $\mathcal{Z} \coloneqq \left\{P_{\mathcal{Y}|\mathcal{X}} \otimes P_\mathcal{X} \,:\, P_\mathcal{X} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}\right\}$. 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.

We conclude this section by introducing a simple example model, which we will also use in the sequel to illustrate some of our results.

Example 3.1. *For this example we will use an imprecise, hidden version of the well-known $M/M/1/k$ queue. In the precise and observable (non-hidden) case, this corresponds to a single-server queueing system that has Poisson arrivals at a rate of a per unit time, exponentially distributed departures at a rate of d*

per unit time, and a maximum queue length of k. The queue length X_t at a given time t is then a random variable of interest.

It is well-known that this queueing system can equivalently be described by means of a continuous-time Markov chain. The state-space $\mathcal{X} \coloneqq \{0,\dots,k\}$ then represents the possible lengths of the queue, obtained at time t by the state X_t of the CTMC, and the system's dynamics are described by the $(k+1)\times(k+1)$ transition rate matrix Q whose elements are defined, for all $i,j\in\{0,\dots,k\}$ such that $i\neq j$, by

$$Q(i,j) \coloneqq \begin{cases} a & \text{if } i<k \text{ and } j=i+1, \\ d & \text{if } i>0 \text{ and } j=i-1, \\ 0 & \text{otherwise,} \end{cases}$$

and the diagonal elements follow from the zero-row sum constraint, i.e., for all $i\in\{0,\dots,k\}$:

$$Q(i,i) = \begin{cases} -a & \text{if } i=0, \\ -d & \text{if } i=k, \\ -(a+d) & \text{otherwise.} \end{cases}$$

We can now make this system imprecise by specifying that the arrival rates a and departure rates d are not known precisely, but that we only know that they belong to the intervals $[\underline{a},\overline{a}]$ and $[\underline{d},\overline{d}]$ respectively. Next, we can relax the assumption that these rates are independent of the length of the queue; for instance, a longer queue might deter new arrivals due to higher expected waiting times. Hence, we then consider the set $\mathcal{Q}$ as the set of rate matrices $Q\in\mathcal{Q}$ for which, for all $i,j\in\{0,\dots,k\}$ with $i\neq j$,

$$Q(i,j) \in \begin{cases} [\,\underline{a},\overline{a}\,] & \text{if } i<k \text{ and } j=i+1, \\ [\,\underline{d},\overline{d}\,] & \text{if } i>0 \text{ and } j=i-1, \\ \{0\} & \text{otherwise,} \end{cases}$$

and for which $\sum_{j\in\{0,\dots,k\}} Q(i,j) = 0$ for all $i\in\{0,\dots,k\}$. Using this set $\mathcal{Q}$, we can now specify the corresponding ICTMC $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$. For simplicity, we assume that the queue always starts out empty, i.e., that $\mathcal{M}=\{p_0\}$, where p_0 is a probability mass function on $\{0,\dots,k\}$ that satisfies $p_0(0)=1$ (and hence $p_0(i)=0$ for all $i\in\{1,\dots,k\}$).

Note that the ICTMC $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ further relaxes our assumptions about the model. For instance, not only can the arrival rate $a\in[\,\underline{a},\overline{a}\,]$ depend on the current state (queue length) X_t, it might also depend on the current time t or even on the entire history before time t. Note that, correspondingly, the arrival process is no longer assumed to be Poisson, but can be a much more general mechanism. In a similar vein, the distribution of the departures is allowed to be much more general than the exponential distribution that is assumed in the precise case. Observe that this model corresponds to the continuous-time analogue of the discrete-time imprecise Geo/Geo/1/k queue previously studied by Lopatatzidis et al. *[9].*

We will next turn this system into a hidden *model, and assume that the state X_t cannot be directly observed. To this end, we will first interpret the queue in a very literal sense as a physical system wherein items (or, say, people) arrive, arrange themselves in a linear fashion, and at some point leave the front of the queue. The system might then be (partially) observed by means of a sensor placed towards the back of the queue (using the assumption that the maximum length is finite). This sensor can then measure whether the queue is full at any given time. This system is schematically represented on the left of Figure 1.*

However, due to the physical interpretation of this system, the items (people) lining up in the queue might not do so perfectly, which introduces measurement errors to this sensor setup—see the right side of Figure 1. For the sake of this example, we will assume that this measurement error follows a simple distribution depending on the actual length of the queue. Specifically, we assume that

$$P_{\mathcal{Y}|\mathcal{X}}(\mathit{SensorSaysFull}{=}1\,|\,i) \coloneqq \frac{i}{k+1}\,, \quad \forall i\in\{0,\dots,k\}\,.$$

The combined system dynamics, described by $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$, and the sensor model, $P_{\mathcal{Y}|\mathcal{X}}$, can now be used to form an ICTHMC as in Definition 2. We will show in later examples how to perform inferences using this model. In these later examples, we will assume that the numerical parameters are given as follows: the maximum queue length $k=10$, the arrival rates $a\in[0.8,1.2]$, and the departure rates $d\in[0.9,1.3]$.

$\Diamond$

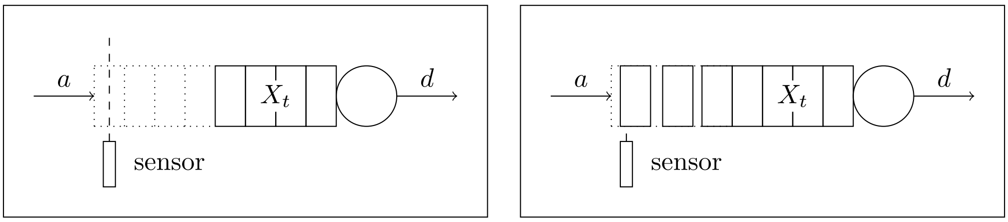

Figure 1: Left: Sketch of a finite-length, single server queue X_t with arrival rate a and departure rate d. A sensor is in place to measure whether the queue is full. Right: The same system as on the left, here with non-ideal filling of the queue. The sensor registers the queue as being full despite not all slots being filled.

4. Computing Lower Expectations for ICTHMC's

Having defined ICTHMC's in the previous section, we now consider how to compute the corresponding lower expectations of functions that are defined on the state-space $\mathcal{X}_u$ and the outputs $\mathcal{Y}_u$ at some given time-points u.[4] The following result will be crucial throughout.

Proposition 4.1. *Consider any ICTMC $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ and any output model $(\mathcal{Y}, \Sigma, P_{\mathcal{Y}|\mathcal{X}})$, and let $\mathcal{Z}$ be the corresponding ICTHMC. Fix any $u \in \mathcal{U}$, and consider any (measurable) function $f : \mathcal{X}_u \times \mathcal{Y}_u \to \mathbb{R}$. Then $\underline{\mathbb{E}}_{\mathcal{Z}}\left[f(X_u, Y_u)\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[f(X_u, Y_u) \,|\, X_u]\right]$.*

Observe that Proposition 4.1 implies a straightforward (if possibly computationally intensive) method for computing the lower expectation of arbitrary functions f on both the state space $\mathcal{X}_u$ and on the outputs $\mathcal{Y}_u$. Specifically, since by assumption we are able to compute the conditional expectation(s) for $P_{\mathcal{Y}|\mathcal{X}}$, we can simply construct a new function $g \in \mathcal{L}(\mathcal{X}_u)$ by computing

$$g(x_u) \coloneqq \mathbb{E}_{\mathcal{Y}|\mathcal{X}}\left[f(x_u, Y_u) \,|\, X_u = x_u\right],$$

for all $x_u \in \mathcal{X}_u$. Proposition 4.1 then implies that in order to compute $\underline{\mathbb{E}}_{\mathcal{Z}}[f(X_u, Y_u)]$, it suffices to instead compute $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)]$. Since this latter quantity is simply a lower expectation of a function g on the state space $\mathcal{X}_u$, that is taken with respect to an ICTMC $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$, we can now apply the known method from [7] to compute this quantity.

We here briefly recall that this method is based on the repeated application of the law of iterated lower expectation that was explained in Section 2.1. That is, writing $u = t_0, \ldots, t_n$, since

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[g(X_u)\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\Big[\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[g(X_u) \,|\, X_{u\setminus\{t_n\}}\right]\Big],$$

we can first compute a function $g' \in \mathcal{L}(\mathcal{X}_{u\setminus\{t_n\}})$, which is defined for all $x_{u\setminus\{t_n\}} \in \mathcal{X}_{u\setminus\{t_n\}}$, as

$$g'\left(x_{u\setminus\{t_n\}}\right) \coloneqq \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[g(X_u) \,|\, X_{u\setminus\{t_n\}} = x_{u\setminus\{t_n\}}\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[g(x_{u\setminus\{t_n\}}, X_{t_n}) \,|\, X_{u\setminus\{t_n\}} = x_{u\setminus\{t_n\}}\right],$$

which due to the imprecise Markov property and Equation (2), reduces to

$$g'\left(x_{u\setminus\{t_n\}}\right) = \underline{\mathbb{E}}_{\mathcal{Q}}\left[g(x_{u\setminus\{t_n\}}, X_{t_n}) \,|\, X_{t_{n-1}} = x_{t_{n-1}}\right]. \tag{6}$$

The right-hand side of this equality can be computed using Equation (3). Repeating this computation for all $x_{u\setminus\{t_n\}}$, we obtain the function g', and by substitution we find that

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[g(X_u)\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[g'(X_{u\setminus\{t_n\}})\right].$$

[4]More generally, we may be interested in functions $f : \mathcal{X}_u \times \mathcal{Y}_v \to \mathbb{R}$ with $u \neq v$. This case is handled by applying the methods that we consider to the trivial extension of f to $f' : \mathcal{X}_{u\cup v} \times \mathcal{Y}_{u\cup v} \to \mathbb{R} : (x_{u\cup v}, y_{u\cup v}) \mapsto f(x_u, y_v)$.

Note that this has reduced the problem to one of computing the lower expectation of a function g' on the time-points $u \setminus \{t_n\}$, which is one time-point fewer than in the original problem. Repeating the above procedure, we can iteratively remove the latest remaining time-points from u, until we are left with a function g^* on $\mathcal{X}_{t_0}$. It then remains to compute the right-hand side of

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g(X_u)\big] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g^*(X_{t_0})\big]\,,$$

which, since g^* is a function on a single time-point t_0, can be done using Equation (5) and the method explained in Section 2.2. For further details about the above method, we refer the interested reader to [7, Section 9.2].

4.1. An Algorithm for Factorising Functions

The method described in Section 4 above can be used to compute the lower expectation of arbitrary functions on $\mathcal{X}_u$, and therefore—due to Proposition 4.1—of arbitrary functions on $\mathcal{X}_u$ and $\mathcal{Y}_u$ jointly. However, it is computationally expensive, in that the runtime complexity scales exponentially with the number of time-points in u.

We here consider an adaptation of the above method—that is still based on the law of iterated lower expectation—which can be used for functions that factorise over the time-points in u, i.e. functions f that can be written as $f(X_u) = \prod_{t\in u} f_t(X_t)$, with $f_t \in \mathcal{L}(\mathcal{X}_t)$ for all $t \in u$. As we will see, this method achieves a runtime complexity that is linear in this number of time-points, and is therefore much more efficient than the general method described above. Later, we will show in Sections 4.2, 4.3, and 6, that a number of practically important inference problems for ICTHMC's can be reduced to computing lower expectations of functions that are exactly of this form. The method that we are about to describe can therefore be used to efficiently solve these inference problems.

So, consider any collection of functions $f_t \in \mathcal{L}(\mathcal{X}_t)$, $t \in u$, and define $f(x_u) \coloneqq \prod_{t\in u} f_t(x_t)$ for all $x_u \in \mathcal{X}_u$. The crucial observation is that for functions f of this form, the computation of the inner (conditional) lower expectation—see Equation (6) above—can be simplified considerably. In particular, due to the non-negative homogeneity of lower expectations and the conjugacy relation between lower and upper expectations, we have

$$\begin{aligned}
&\underline{\mathbb{E}}_{\mathcal{Q}}\left[f_{t_n}(X_{t_n}) \prod_{t\in u\setminus\{t_n\}} f_t(x_t) \,\middle|\, X_{t_{n-1}} = x_{t_{n-1}}\right] \\
&\quad = \begin{cases} \prod_{t\in u\setminus\{t_n\}} f_t(x_t)\underline{\mathbb{E}}_{\mathcal{Q}}\left[f_{t_n}(X_{t_n}) \,\middle|\, X_{t_{n-1}} = x_{t_{n-1}}\right] & \text{if } \prod_{t\in u\setminus\{t_n\}} f_t(x_t) \geq 0\,, \text{ and} \\ \prod_{t\in u\setminus\{t_n\}} f_t(x_t)\overline{\mathbb{E}}_{\mathcal{Q}}\left[f_{t_n}(X_{t_n}) \,\middle|\, X_{t_{n-1}} = x_{t_{n-1}}\right] & \text{otherwise.} \end{cases}
\end{aligned}$$

The next important fact is that the sign condition, $\prod_{t\in u\setminus\{t_n\}} f_t(x_t) \geq 0$, can itself be factored into sign conditions on $f_{t_{n-1}}(x_{t_{n-1}})$ and on $\prod_{t\in u\setminus\{t_{n-1},t_n\}} f_t(x_t)$. Roughly speaking, this implies that we can make a "tentative" choice based only on the sign of $f_{t_{n-1}}(x_{t_{n-1}})$, and—as a first step towards backward induction on n—only make a "definitive" choice once we are considering the earlier time-points.

Putting these observations together and applying backwards induction on n, leads to the following dynamic programming method. For all $t \in u$, we define auxiliary functions $f_t^+, f_t^- \in \mathcal{L}(\mathcal{X}_t)$, as follows. First, we define $f_{t_n}^+ \coloneqq f_{t_n}^- \coloneqq f_{t_n}$. Then, for all $i \in \{0, \ldots, n-1\}$ and all $x_{t_i} \in \mathcal{X}_{t_i}$, let

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

and,

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

These auxiliary functions can be given the following interpretation.

Lemma 4.2. *Let f_{t_i}, $f^+_{t_i}$ and $f^-_{t_i}$ be as defined above for all $i \in \{0,\dots,n\}$. Then for all $i \in \{0,\dots,n\}$:*

$$f^+_{t_i} = \underline{\mathbb{E}}_{\mathcal{Q}}\left[\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_i}\right], \qquad \textit{and,} \qquad f^-_{t_i} = \overline{\mathbb{E}}_{\mathcal{Q}}\left[\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_i}\right].$$

Combining the above with the law of iterated lower expectation one more time, we obtain the following result.

Proposition 4.3. *Let f_t, f^+_t and f^-_t be as defined above for all $i \in \{0,\dots,n\}$. Then it holds that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{i=0}^{n} f_{t_i}(X_{t_i})\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[f^+_{t_0}(X_{t_0})\right]$. Furthermore, $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{i=0}^{n} f_{t_i}(X_{t_i})\right] = \overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[f^-_{t_0}(X_{t_0})\right]$.*

These results lead us to consider Algorithm 1 for computing the lower expectation of such factorising functions. This algorithm can be read as follows.

First, on Lines 2 and 3, we initialise the base-case of the backward induction/dynamic programming. Lines 4,5, and 17 then perform the backwards iteration over the time-points in u. For each time-point, we start by computing the lower and upper expectations of the auxiliary functions $f^+_{t_{i+1}}$ and $f^-_{t_{i+1}}$, on Lines 6 and 7. Practically, this can be done using Equation (3); as mentioned in Section 2.2, this has polynomial time-complexity. At Lines 9-15, we compute the values of the auxiliary functions $f^+_{t_i}$ and $f^-_{t_i}$ at time-point t_i. Clearly, this has time-complexity linear in the number of states (see Line 8). In total, the time-complexity of this backwards induction is therefore n times the complexity of computing conditional lower expectations using Equation (3); this is clearly itself polynomial, and linear in n.

After all time-points in u have been resolved, it remains to reduce the problem to a conditional lower expectation on time zero (Line 19). Again, this is done using Equation (3) and in polynomial time. Finally, we compute the lower expectation over the probabilities on the initial state X_0, at Line 20, which can also be done in polynomial time (see Section 2.2 and the assumption that $\mathcal{M}$ is convex). Due to Equation (5) and Proposition 4.3, this yields the quantity of interest, which is finally returned by the algorithm.

4.2. Lower (Unconditional) Probabilities of Observations

As an immediate application of the method from the previous section, we consider the lower probability of output events $O_u \in \Sigma_u$ at time-points u. It is easy to see that for events of this type,

$$\begin{aligned}
\underline{P}_{\mathcal{Z}}(Y_u \in O_u) &= \underline{\mathbb{E}}_{\mathcal{Z}}\left[\mathbb{I}_{O_u}(Y_u)\right] \\
&= \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\mathbb{I}_{O_u}(Y_u) \,|\, X_u]\right] \\
&= \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[P_{\mathcal{Y}|\mathcal{X}}(O_u \,|\, X_u)\right] \\
&= \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{t\in u} P_{\mathcal{Y}|\mathcal{X}}(O_t \,|\, X_t)\right],
\end{aligned}$$

where we used Proposition 4.1 for the second equality, and the independence assumptions from Section 3.2 for the final equality. Clearly, the final right-hand side is a lower expectation of a function that factorises over the time-points in u. Hence, Algorithm 1 can be directly applied to compute this quantity of interest.

Example 4.1. *Consider again the imprecise, hidden queue model from Example 3.1. For the purposes of this example, suppose that we are interested in the lower probability of observing a negative sensor measurement at time 2 (indicating that the queue is not observed to be full), followed by a positive sensor management at time 10. That is, we are interested in computing*

$$\begin{aligned}
&\underline{P}_{\mathcal{Z}}\big(\mathit{SensorSaysFull}_2 = 0 \,\&\, \mathit{SensorSaysFull}_{10} = 1\big) = \\
&\qquad \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\Big[P_{\mathcal{Y}|\mathcal{X}}(\mathit{SensorSaysFull} = 0 \,|\, X_2) P_{\mathcal{Y}|\mathcal{X}}(\mathit{SensorSaysFull} = 1 \,|\, X_{10})\Big].
\end{aligned}$$

Following Algorithm 1, we therefore start by setting, for all $i \in \{0,\dots,k\}$,

$$f^+_{10}(i) \coloneqq f^-_{10}(i) \coloneqq P_{\mathcal{Y}|\mathcal{X}}(\mathit{SensorSaysFull} = 1 \,|\, X_{10} = i) = \frac{i}{k+1}.$$

Algorithm 1 Polynomial-time algorithm to compute the lower expectation of a factorising function $f(X_u) = \prod_{t\in u} f_t(X_t)$ with respect to an ICTMC $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$. It is assumed that $\mathcal{Q}$ and $\mathcal{M}$ satisfy the assumptions mentioned in Section 2.1, and additionally that $\mathcal{M}$ is convex.

Input: Set of rate matrices $\mathcal{Q}$, convex set of probability mass functions $\mathcal{M}$, sequence of time-points $u = t_0, \dots, t_n$, functions $f_t \in \mathcal{L}(\mathcal{X}_t)$ for all $t \in u$.

Output: $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{t\in u} f_t(X_t)\right]$

function $\textsc{ComputeProductFunction}(\mathcal{Q}, \mathcal{M}, u, f_{t_0}, \dots, f_{t_n})$
  $f^+_{t_n} \leftarrow f_{t_n}$
  $f^-_{t_n} \leftarrow f_{t_n}$
  $i \leftarrow n-1$
  while $i \geq 0$ **do**
    $\hat{f}^+_{t_{i+1}} \leftarrow \underline{\mathbb{E}}_{\mathcal{Q}}\left[f^+_{t_{i+1}}(X_{t_{i+1}}) \,|\, X_{t_i}\right]$
    $\hat{f}^-_{t_{i+1}} \leftarrow \overline{\mathbb{E}}_{\mathcal{Q}}\left[f^-_{t_{i+1}}(X_{t_{i+1}}) \,|\, X_{t_i}\right]$
    for all $x_{t_i} \in \mathcal{X}_{t_i}$ **do**
      if $f_{t_i}(x_{t_i}) \geq 0$ **then**
        $f^+_{t_i}(x_{t_i}) \leftarrow f_{t_i}(x_{t_i})\hat{f}^+_{t_{i+1}}(x_{t_i})$
        $f^-_{t_i}(x_{t_i}) \leftarrow f_{t_i}(x_{t_i})\hat{f}^-_{t_{i+1}}(x_{t_i})$
      else
        $f^+_{t_i}(x_{t_i}) \leftarrow f_{t_i}(x_{t_i})\hat{f}^-_{t_{i+1}}(x_{t_i})$
        $f^-_{t_i}(x_{t_i}) \leftarrow f_{t_i}(x_{t_i})\hat{f}^+_{t_{i+1}}(x_{t_i})$
      end if
    end for
    $i \leftarrow i-1$
  end while
  $f_* \leftarrow \underline{\mathbb{E}}_{\mathcal{Q}}\left[f^+_{t_0}(X_{t_0}) \,|\, X_0\right]$
  return $\underline{\mathbb{E}}_{\mathcal{M}}\left[f_*(X_0)\right]$
end function

Then, using Equation (3), we compute for all $i \in \{0, \dots, k\}$:

$$\hat{f}^+_{10}(i) \coloneqq \underline{\mathbb{E}}_{\mathcal{Q}}\left[f^+_{10}(X_{10}) \,|\, X_2 = i\right], \quad \textit{and,} \quad \hat{f}^-_{10}(i) \coloneqq \overline{\mathbb{E}}_{\mathcal{Q}}\left[f^-_{10}(X_{10}) \,|\, X_2 = i\right] = -\underline{\mathbb{E}}_{\mathcal{Q}}\left[-f^-_{10}(X_{10}) \,|\, X_2 = i\right].$$

Next, we construct the functions f^+_2 and f^-_2. Because $P_{\mathcal{Y}|\mathcal{X}}(\mathit{SensorSaysFull} = 0 \,|\, X_2 = i)$ is never negative, we simply have for all $i \in \{0, \dots, k\}$ that

$$f^+_2(i) \coloneqq P_{\mathcal{Y}|\mathcal{X}}(\mathit{SensorSaysFull} = 0 \,|\, X_2 = i)\hat{f}^+_{10}(i) = \left(1 - i/k+1\right)\hat{f}^+_{10}(i), \quad \textit{and,} \quad f^-_2(i) \coloneqq \left(1 - i/k+1\right)\hat{f}^-_{10}(i).$$

We can then compute f_*, again using Equation (3), as

$$f_*(i) \coloneqq \underline{\mathbb{E}}_{\mathcal{Q}}\left[f^+_2(X_2) \,|\, X_0 = i\right].$$

From here it remains to compute the marginal lower expectation at time zero, i.e. with respect to $\mathcal{M}$. Since by assumption we always start with an empty queue, we have

$$\underline{\mathbb{E}}_{\mathcal{M}}[f_*(X_0)] = f_*(0).$$

Numerically, using the values from Example 3.1, we conclude that

$$\underline{P}_{\mathcal{Z}}(\mathit{SensorSaysFull}_2 = 0 \,\&\, \mathit{SensorySaysFull}_{10} = 1) \approx 0.112.$$

◊

4.3. Lower (Unconditional) Density Functions

We mentioned in Section 3.1 that we do not want to stipulate that the output variables Y_t be either discrete or continuous. However, by not making this choice we will at times want to consider an explicit special case: when the Y_t are continuous we may be interested in the corresponding density function. In this section we handle the case analogous to the lower probabilities from the previous section, i.e. lower unconditional density functions on the outputs. Later, in Section 5, we will also want to consider a special case of continuous outputs when considering the "updated" model.

The discussion that follows in this section might come across as somewhat abstract and didactic, explicitly building up the notion of density functions as limits of other functions. Clearly, for our present purposes we could also assume these concepts to be known and simply pass to density functions directly. However, we introduce the machinery here as a first step to the upcoming discussion in Section 5, where the required concepts are somewhat less straightforward.

So, suppose that $\mathcal{Y}$ is uncountable, and that we are interested in the lower unconditional density function at some point $y_u \in \mathcal{Y}_u$. Roughly speaking, this of course corresponds to the limiting value of a sequence of (lower) probabilities of shrinking regions around this point y_u, which are normalised by the "area" of these regions.

That is, consider any sequence $\{O_u^i\}_{i\in\mathbb{N}}$ of events in Σ_u that shrinks to $y_u \in \mathcal{Y}_u$—i.e. such that $O_u^i \supseteq O_u^{i+1}$ for all $i \in \mathbb{N}$, and $\cap_{i\in\mathbb{N}} O_u^i = \{y_u\}$—and suppose that there exists a sequence $\{\lambda_i\}_{i\in\mathbb{N}}$ in $\mathbb{R}_{>0}$ such that, for all $x_u \in \mathcal{X}_u$,

$$\phi_u(y_u \,|\, x_u) \coloneqq \lim_{i\to+\infty} \frac{P_{\mathcal{Y}|\mathcal{X}}(O_u^i \,|\, x_u)}{\lambda_i} \tag{7}$$

exists and is real-valued (in particular, finite). Because lower expectation operators are continuous—see e.g. [12, Proposition 2.6.1.ℓ] or Lemma A.2 in the appendix—this implies that also $\lim_{i\to+\infty} \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}} \left[{}^{P_{\mathcal{Y}|\mathcal{X}}(O_u^i \,|\, X_u)}/_{\lambda_i} \right]$ exists, is real-valued (in particular, finite), and, using Proposition 4.1,

$$\lim_{i\to+\infty} \frac{\underline{P}_{\mathcal{Z}}(Y_u \in O_u^i)}{\lambda_i} = \lim_{i\to+\infty} \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}} \left[\frac{P_{\mathcal{Y}|\mathcal{X}}(O_u^i \,|\, X_u)}{\lambda_i} \right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}} \left[\phi_u(y_u \,|\, X_u) \right] .$$

We call this quantity $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}} \left[\phi_u(y_u \,|\, X_u) \right]$ the lower unconditional density function evaluated at y_u.

Under some regularity conditions, the existence of the function $\phi_u(y_u \,|\, x_u)$ at *all* $y_u \in \mathcal{Y}_u$ is implied by the well-known Radon-Nikodym theorem. If this is the case, $\phi_u(\cdot \,|\, x_u)$ is called the Radon-Nikodym derivative of the measure $P_{\mathcal{Y}|\mathcal{X}}(\cdot \,|\, x_u)$ with respect to a reference measure λ on $(\mathcal{Y}_u, \Sigma_u^*)$ that (at least) satisfies $\lambda(O_u^i) = \lambda_i$ for all $i \in \mathbb{N}$, with respect to which $P_{\mathcal{Y}|\mathcal{X}}(\cdot \,|\, x_u)$ is absolutely continuous, and where Σ_u^* denotes the product (σ-)algebra of Σ over u (c.f. Σ_u). This function $\phi_u(\cdot \,|\, x_u)$ then satisfies $P_{\mathcal{Y}|\mathcal{X}}(O_u \,|\, x_u) = \int_{O_u} \phi_u(y_u \,|\, x_u) \,\mathrm{d}\lambda(y_u)$ for all $O_u \in \Sigma_u$, which motivates the terminology "density function". Here the integral is understood in the Lebesgue sense and with respect to the reference measure λ.

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 $x_t \in \mathcal{X}_t$,

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

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

Conversely, the above discussion can also be framed by starting directly from known density functions $\psi(\cdot \,|\, x) : \mathcal{Y} \to \mathbb{R}_{\geq 0}$. It is didactically convenient to then assume that $\mathcal{Y} = \mathbb{R}^d$ for some $d \in \mathbb{N}$ and that Σ is the Borel σ-algebra on $\mathcal{Y}$ under the usual topology. We can then make the preceding discussion less abstract by taking the reference measure λ to be the Lebesgue measure on $(\mathcal{Y}, \Sigma)$. The crucial assumption on these functions ψ is then that $\int_{\mathcal{Y}} \psi(y \,|\, x) \,\mathrm{d}y = 1$, for all $x \in \mathcal{X}$, and they are used to define the measures $P_{\mathcal{Y}|\mathcal{X}}(O \,|\, x) \coloneqq \int_O \psi(y \,|\, x) \,\mathrm{d}y$ for all $O \in \Sigma$ and all $x \in \mathcal{X}$.

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

$$\phi_t(y_t|x_t) = \lim_{i\to+\infty} \frac{P_{\mathcal{Y}|\mathcal{X}}(O_t^i|x_t)}{\lambda(O_t^i)} = \psi(y_t|x_t) \,, \tag{8}$$

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 $O_u^i \coloneqq \prod_{t\in u} O_t^i$, with each O_t^i chosen as above. If we then choose the sequence $\{\lambda_i\}_{i\in\mathbb{N}}$ as $\lambda_i \coloneqq \prod_{t\in u} \lambda(O_t^i)$ 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 (8).

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 $\psi(\cdot\,|x_t)$. This is generally possible because, arguably, most of the density functions that one encounters in practice will be continuous; this guarantees that the limit in Equation (8) exists.[5]

Finally, the above identification ensures that $\phi_u(y_u\,|\,x_u) = \prod_{t\in u} \psi(y_t\,|\,x_t)$; this is a function on $\mathcal{X}_u$ that clearly factorises over the time-points in u. Hence, the lower unconditional density function $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\phi_u(y_u\,|\,X_u)\right]$ can be efficiently computed using Algorithm 1.

5. Updating the Model

In the preceding section we have seen how to compute unconditional probabilities (and densities) of events of the outputs of a given ICTHMC. 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.

5.1. 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 5.1. *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*

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

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* [12]. 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.

Definition 3. *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*

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

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.

[5] Technically, there is still a uniqueness issue in that the limit might depend on the choice of events $\{O_u^i\}_{i\in\mathbb{N}}$—something that we circumvented in Equation (8) by restricting attention to sequences of open balls. In the more general case, e.g. of the limit in Equation (7), certain pathological sequences can be constructed however. To ensure uniqueness it suffices if, for all $t \in u$, there is a sequence of open balls $\{B_t^i\}_{i\in\mathbb{N}}$ in $\mathcal{Y}$ 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.

As is well known, the updated lower expectation that is obtained through regular extension satisfies Walley's *generalised Bayes' rule* [12]. 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 5.2. *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*

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

5.2. 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. The formalisation of this notion was already hinted at in our discussion of density functions in Section 4.3.

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

$$\mathbb{E}_P[f(X_v) \,|\, Y_u = y_u] \coloneqq \lim_{i\to+\infty} \mathbb{E}_P[f(X_v) \,|\, Y_u \in O_u^i] . \tag{9}$$

Perhaps unsurprisingly, it turns out that this limit is closely connected to the limit $\phi_u(y_u|x_u)$ from Equation (7). Specifically, 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 $\mathbb{E}_{P_{\mathcal{X}}}[\phi_u(y_u \,|\, X_u)] > 0$, then the following result holds:

Proposition 5.3. *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*

$$\mathbb{E}_P[f(X_v) \,|\, Y_u = y_u] \coloneqq \lim_{i\to+\infty} \mathbb{E}_P[f(X_v) \,|\, Y_u \in O_u^i] = \frac{\mathbb{E}_{P_{\mathcal{X}}}[f(X_v)\phi_u(y_u|X_u)]}{\mathbb{E}_{P_{\mathcal{X}}}[\phi_u(y_u \,|\, X_u)]} . \tag{10}$$

Note that, as was the case for the value of $\phi_u(y_u|x_u)$, the quantity $\mathbb{E}_P[f(X_v) \,|\, Y_u = y_u]$ 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. We do this because, as was noted in Section 4.3, in many (or, most) practical applications, the measures $P_{\mathcal{Y}|\mathcal{X}}(\cdot|x)$ will be constructed from some known density function ψ that tends to be continuous and relatively well-behaved. Recall that $\phi_u(y_u|x_u)$ represents a density function value, and can often be identified as $\phi_u(y_u|x_u) = \prod_{t\in u} \psi(y_t|x_t)$. Proposition 5.3 therefore implies uniqueness of the limit (9) under the same additional assumptions that were mentioned in Footnote 5 (and additionally under strict positivity at y_u). Finally, note that due to this interpretation, the right-hand side of Equation (10) is simply the well-known Bayes' rule for (finite) mixtures of densities.

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

Definition 4. *Let $\mathcal{Z}$ be an ICTHMC 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 and is real-valued, the updated lower expectation is defined by*

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

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

Similar to the results in Section 5.1, this updated lower expectation satisfies a "generalised Bayes' rule for mixtures of densities", in the following sense.

Proposition 5.4. *Let $\mathcal{Z}$ be an ICTHMC 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 $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] > 0$, then*

$$\underline{\mathbb{E}}_{\mathcal{Z}}\big[f(X_v)\,|\,Y_u = y_u\big] = \max\big\{\mu \in \mathbb{R} \,:\, \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[\phi_u(y_u|X_u)\big(f(X_v) - \mu\big)\big] \geq 0\big\}\,. \tag{11}$$

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

Proposition 5.5. *Let $\mathcal{Z}$ be an ICTHMC 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 $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] > 0$, then $\underline{\mathbb{E}}_{\mathcal{Z}}[f(X_v)|Y_u = y_u] = \lim_{i\to+\infty} \underline{\mathbb{E}}_{\mathcal{Z}}[f(X_v)|Y_u \in O_u^i]$.*

Now, recall that the requirement $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] > 0$ 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

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

whenever $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] > 0$; 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 (11) is equal to $\underline{\mathbb{E}}^{\mathrm{R}}_{\mathcal{Z}}[f(X_v)\,|\,Y_u = y_u]$ whenever $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] > 0$; see Proposition D.1 in the appendix for a proof. However, as the following example shows, the limit interpretation then fails to hold in the sense that there are cases where $\underline{\mathbb{E}}^{\mathrm{R}}_{\mathcal{Z}}[f(X_v)\,|\,Y_u = y_u] \neq \lim_{i\to+\infty} \underline{\mathbb{E}}_{\mathcal{Z}}[f(X_v)\,|\,Y_u \in O_u^i]$, with $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] > 0$ but $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] = 0$.

Example 5.1. *Let $\mathcal{X} \coloneqq \{x, \overline{x}\}$ be a binary state-space, and let $\mathcal{Y} \coloneqq [-1, 1]$, with Σ the Borel σ-algebra on $\mathcal{Y}$ under the usual topology. For the sequences of time-points, we choose $u = v = \{0\}$. Set $y_0 \coloneqq 0$. The trick will be to choose the measures $P_{\mathcal{Y}|\mathcal{X}}(\cdot|x)$ and $P_{\mathcal{Y}|\mathcal{X}}(\cdot|\overline{x})$ so that the first gives y_0 strictly positive support (i.e. density), while the second assigns zero support (i.e. density) to y_0 but positive support (i.e. density) to the region around y_0. To this end, let $P_{\mathcal{Y}|\mathcal{X}}(\cdot|x)$ be the uniform distribution on $[-1, 1]$.*

For the measure $P_{\mathcal{Y}|\mathcal{X}}(\cdot|\overline{x})$, we first define $\phi(y|\overline{x}) \coloneqq |y|$ for all $y \in \mathcal{Y}$. Then clearly, $\int_{\mathcal{Y}} \phi(y|\overline{x})\,\mathrm{d}y = 1$, and furthermore $\phi(y_0|\overline{x}) = 0$. For every $O \in \Sigma$, let now $P_{\mathcal{Y}|\mathcal{X}}(O|\overline{x}) \coloneqq \int_O \phi(y|\overline{x})\,\mathrm{d}y$.

Observe that for any sequence $\{O_0^i\}_{i\in\mathbb{N}}$ of open intervals $O_0^i \coloneqq (y_0 - \delta_i, y_0 + \delta_i)$, with $\delta_i > 0$ such that $\{\delta_i\}_{i\in\mathbb{N}} \to 0^+$, we then clearly have

$$\phi(y_0|x) = \lim_{i\to+\infty} \frac{P_{\mathcal{Y}|\mathcal{X}}(O_0^i|x)}{\lambda(O_0^i)} = \lim_{i\to+\infty} \frac{\delta_i}{2\delta_i} = \frac{1}{2}\,,$$

and

$$\phi(y_0|\overline{x}) = \lim_{i\to+\infty} \frac{P_{\mathcal{Y}|\mathcal{X}}(O_0^i|\overline{x})}{\lambda(O_0^i)} = \lim_{i\to+\infty} \frac{\delta_i^2}{2\delta_i} = \lim_{i\to+\infty} \frac{\delta_i}{2} = 0\,,$$

where $\lambda(O_0^i)$ is the Lebesgue measure of O_0^i. Fix any such sequence $\{O_0^i\}_{i\in\mathbb{N}}$ and let $\{\lambda_i\}_{i\in\mathbb{N}} \coloneqq \{\lambda(O_0^i)\}_{i\in\mathbb{N}}$.

Let the set of initial distributions $\mathcal{M}$ be the entire set of probability mass functions on $\mathcal{X}_0$. Consider the two probability mass functions P_1, P_2 on $\mathcal{X}_0$ such that $P_1(x) = 1$, $P_1(\overline{x}) = 0$, and $P_2(x) = 0$, $P_2(\overline{x}) = 1$; clearly, $P_1, P_2 \in \mathcal{M}$. Let $\mathcal{Q}$ be any non-empty, bounded, and convex set of rate matrices with separately specified rows, and let $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ be the corresponding ICTMC. Construct $\mathcal{Z}$ from $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ and $(\mathcal{Y}, \Sigma, P_{\mathcal{Y}|\mathcal{X}})$ as in Definition 2. Let $f \in \mathcal{L}(\mathcal{X}_0)$ be defined by $f(x) \coloneqq 1$ and $f(\overline{x}) \coloneqq -1$.

This concludes the construction of the model and inference problem that we will use for this counter-example. It remains to show that it fails to satisfy the limit interpretation, as claimed above.

First, $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi(y_0|X_0)] = \overline{\mathbb{E}}_{\mathcal{M}}[\phi(y_0|X_0)] = P_1(x)\phi(y_0|x) + P_1(\overline{x})\phi(y_0|\overline{x}) = P_1(x)\phi(y_0|x) = \frac{1}{2} > 0$, and so the lower expectation $\underline{\mathbb{E}}^{\mathrm{R}}_{\mathcal{Z}}[f(X_v) \,|\, Y_u = y_u]$ is well-defined.

Now, for any $P \in \mathcal{Z}$ such that $\mathbb{E}_P[\phi(y_0|X_0)] > 0$, we have by Proposition 5.3 that

$$\mathbb{E}_P[f(X_0) \,|\, Y_0 = y_0] = \frac{\mathbb{E}_P[f(X_0)\phi(y_0|X_0)]}{\mathbb{E}_P[\phi(y_0|X_0)]} = \frac{P(x)f(x)\phi(y_0|x)}{P(x)\phi(y_0|x)} = f(x) = 1\,,$$

because $\phi(y_0|\overline{x}) = 0$. Hence,

$$\underline{\mathbb{E}}^{\mathrm{R}}_{\mathcal{Z}}[f(X_0) \,|\, Y_0 = y_0] = \inf\left\{\mathbb{E}_P[f(X_0) \,|\, Y_0 = y_0] \,:\, P \in \mathcal{Z}, \mathbb{E}_P[\phi(y_0|X_0)] > 0\right\} = 1\,.$$

However, for any $i \in \mathbb{N}$, since $P_2 \in \mathcal{M}$, there is some $P \in \mathcal{Z}$ such that

$$P(Y_0 \in O_0^i) = \mathbb{E}_{P_{\mathcal{X}}}[P_{\mathcal{Y}|\mathcal{X}}(O_0^i|X_0)] = \mathbb{E}_{P_2}[P_{\mathcal{Y}|\mathcal{X}}(O_0^i|X_0)] = P_{\mathcal{Y}|\mathcal{X}}(O_0^i|\overline{x}) > 0,$$

and so

$$\mathbb{E}_P[f(X_0) \,|\, Y_0 \in O_0^i] = \frac{\mathbb{E}_{P_{\mathcal{X}}}[f(X_0)P_{\mathcal{Y}|\mathcal{X}}(O_0^i|X_0)]}{\mathbb{E}_{P_{\mathcal{X}}}[P_{\mathcal{Y}|\mathcal{X}}(O_0^i|X_0)]} = \frac{f(\overline{x})P_{\mathcal{Y}|\mathcal{X}}(O_0^i|\overline{x})}{P_{\mathcal{Y}|\mathcal{X}}(O_0^i|\overline{x})} = f(\overline{x}) = -1.$$

Since we also know that $-1 = \min f \leq \underline{\mathbb{E}}_{\mathcal{Z}}[f(X_0)|Y_0 \in O_0^i] \leq \mathbb{E}_P[f(X_0) \,|\, Y_0 \in O_0^i]$, this allows us to infer that $\underline{\mathbb{E}}_{\mathcal{Z}}[f(X_0)|Y_0 \in O_0^i] = -1$. Because this holds for all $i \in \mathbb{N}$, we conclude from this that

$$-1 = \lim_{i \to +\infty} \underline{\mathbb{E}}_{\mathcal{Z}}[f(X_0)|Y_0 \in O_0^i] \neq \underline{\mathbb{E}}^{\mathrm{R}}_{\mathcal{Z}}[f(X_0) \,|\, Y_0 = y_0] = 1\,.$$

◊

We feel that the existence of this counterexample makes this more general updating scheme, using $\underline{\mathbb{E}}^{\mathrm{R}}_{\mathcal{Z}}$, somewhat troublesome from an interpretation (and hence philosophical) point of view.

Next, we recall that the existence of $\phi_u(y_u|X_u)$ and the positivity of $\mathbb{E}_{P_{\mathcal{X}}}[\phi_u(y_u|X_u)]$ are necessary and sufficient conditions for the limit in Equation (9) to exist and be computable using Equation (10). However, these conditions are sufficient but non-necessary for that limit to simply exist. Therefore, yet another way to generalise the imprecise updating method would be

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

whenever $\{P \in \mathcal{Z} \,:\, \mathbb{E}_P[f(X_v) \,|\, Y_u = y_u] \text{ exists}\} \neq \emptyset$. We conjecture that this updated model *does* satisfy the limit interpretation, but on the other hand, the following example shows that this model, in turn, no longer satisfies the above generalised Bayes' rule.

Example 5.2. *Let $\mathcal{Z}$, y_0, f and $\{O_0^i\}_{i \in \mathbb{N}}$ be the same as in Example 5.1.*

Now consider any $P \in \mathcal{Z}$, and note that clearly either $\mathbb{E}_{P_{\mathcal{X}}}[\phi(y_0|X_0)] > 0$, or $\mathbb{E}_{P_{\mathcal{X}}}[\phi(y_0|X_0)] = 0$. In the first case, as we already established in Example 5.1, we know that $\mathbb{E}_P[f(X_0)|Y_0 = y_0]$ exists and is equal to 1.

Suppose for the other case that $\mathbb{E}_{P_{\mathcal{X}}}[\phi(y_0|X_0)] = 0$. Because $\phi(y_0|x) > 0$ and $\phi(y_0|\overline{x}) = 0$, this clearly implies that $P_{\mathcal{X}}(X_0 = x) = 0$, or in other words, that $P_{\mathcal{X}}(X_0 = \overline{x}) = 1$. As we already established in the previous example, we then have for every $i \in \mathbb{N}$ that

$$\mathbb{E}_P[f(X_0)|Y_0 \in O_0^i] = \frac{f(\overline{x})P_{\mathcal{Y}|\mathcal{X}}(O_0^i|\overline{x})}{P_{\mathcal{Y}|\mathcal{X}}(O_0^i|\overline{x})} = -1\,.$$

Because this holds for all $i \in \mathbb{N}$, we clearly have that $\mathbb{E}_P[f(X_0)|Y_0 = y_0]$ exists, and

$$\mathbb{E}_P[f(X_0)|Y_0 = y_0] = \lim_{i \to +\infty} \mathbb{E}_P[f(X_0)|Y_0 \in O_0^i] = -1\,.$$

Since this exhaustively covers all cases, we conclude that $\mathbb{E}_P[f(X_0)|Y_0 = y_0]$ exists for all $P \in \mathcal{Z}$. Therefore,

$$\{P \in \mathcal{Z} : \mathbb{E}_P[f(X_0)\,|\,Y_0 = y_0] \text{ exists}\} = \mathcal{Z}\,,$$

which means that $\underline{\mathbb{E}}^{\mathrm{L}}_{\mathcal{Z}}[f(X_0)\,|\,Y_0 = y_0]$ is well-defined.

Furthermore, it follows from the above that, for all $P \in \mathcal{Z}$, we have either $\mathbb{E}_P[f(X_0)|Y_0 = y_0] = 1$, or $\mathbb{E}_P[f(X_0)|Y_0 = y_0] = -1$. Also, since $P_2 \in \mathcal{M}$, there is at least one $P \in \mathcal{Z}$ for which the second case applies. Hence, it follows that

$$\underline{\mathbb{E}}^{\mathrm{L}}_{\mathcal{Z}}[f(X_0)\,|\,Y_0 = y_0] \coloneqq \inf\bigl\{\mathbb{E}_P[f(X_0)\,|\,Y_0 = y_0] : P \in \mathcal{Z},\, \mathbb{E}_P[f(X_0)\,|\,Y_0 = y_0] \text{ exists}\bigr\} = -1.$$

However, we also know that, for any $P \in \mathcal{Z}$, if $\mathbb{E}_{P_\mathcal{X}}[\phi(y_0|X_0)] > 0$, then $\mathbb{E}_P[f(X_0)|Y_0 = y_0] = 1$. Since $P_1 \in \mathcal{M}$, there is at least one $P \in \mathcal{Z}$ for which this holds. Therefore,

$$\begin{aligned}1 &= \inf\{\mathbb{E}_P[f(X_0)\,|\,Y_0 = y_0] : P \in \mathcal{Z},\, \mathbb{E}_{P_\mathcal{X}}[\phi(y_0|X_0)] > 0\} \\ &= \underline{\mathbb{E}}^{\mathrm{R}}_{\mathcal{Z}}[f(X_0)\,|\,Y_0 = y_0] \\ &= \max\{\mu \in \mathbb{R} : \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\bigl[\phi(y_0|X_0)\bigl(f(X_0) - \mu\bigr)\bigr] \geq 0\}\,,\end{aligned}$$

where the third equality follows from Proposition D.1. We conclude that indeed

$$-1 = \underline{\mathbb{E}}^{\mathrm{L}}_{\mathcal{Z}}[f(X_0)\,|\,Y_0 = y_0] \neq \max\{\mu \in \mathbb{R} : \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\bigl[\phi(y_0|X_0)\bigl(f(X_0) - \mu\bigr)\bigr] \geq 0\} = 1\,.$$

◊

The above example makes this updating scheme, using $\underline{\mathbb{E}}^{\mathrm{L}}_{\mathcal{Z}}$, somewhat troublesome from a practical point of view because, as we discuss in the next section, the expression in Equation (11) is crucial for our method of efficient computation of the updated lower expectation.

6. Computing Updated Lower Expectations

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 3 or 4 directly, we can focus on evaluating the function $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\bigl[P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)\bigl(f(X_v) - \mu\bigr)\bigr]$, 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 6.1, 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. The important special case that we will consider here is that of functions f depending only on a single time-point; we will show that the required computations can then be performed using Algorithm 1.

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

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

In both cases, we can rewrite this expression as $\prod_{t \in u} g_t(X_t)$, where, for all $t \in u$, $g_t \in \mathcal{L}(\mathcal{X}_t)$ and $g_t \geq 0$. The function of interest is then $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\bigl[\bigl(\prod_{t \in u} g_t(X_t)\bigr)\bigl(f(X_v) - \mu\bigr)\bigr]$ and the sign conditions in Propositions 5.2 and 5.4 reduce to $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\prod_{t \in u} g_t(X_t)] > 0$ and $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\prod_{t \in u} g_t(X_t)] > 0$, respectively. Clearly, these sign conditions require the computation of an unconditional upper (respectively lower) expectation of a function that factorises over the time-points in u. Hence, we can directly use Algorithm 1 to compute these quantities.

For the particular case that we consider here, where the function f depends only on a single time-point—s, say—the function of interest can be written $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\left(\prod_{t\in u} g_t(X_t)\right)\left(f(X_s)-\mu\right)\right]$. For clarity of exposition, we can define an auxiliary function $f_\mu \coloneqq f-\mu$ on $\mathcal{X}_s$. The quantity that we are interested in is then $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[f_\mu(X_s)\prod_{t\in u} g_t(X_t)\right]$. This is again an unconditional lower expectation of a function that factorises over the time-points in $u\cup\{s\}$, whence Algorithm 1 provides an efficient method to compute its value for any given choice of μ.

6.1. Solving the Generalised Bayes' Rule

Finally, 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., [3, 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 6.1. *Let $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ be an ICTMC and consider any $u,v\in\mathcal{U}$, any $f\in\mathcal{L}(\mathcal{X}_v)$ and, for all $t\in u$, any $g_t\in\mathcal{L}(\mathcal{X}_t)$ 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 $G(\mu)\coloneqq\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\left(\prod_{t\in u} g_t(X_t)\right)\left(f(X_v)-\mu\right)\right]$. 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 $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{t\in u} g_t(X_t)\right]>0$, then G is (strictly) decreasing, and has a unique root.

G3: If $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{t\in u} g_t(X_t)\right]=0$ but $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{t\in u} g_t(X_t)\right]>0$, 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 $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{t\in u} g_t(X_t)\right]=0$, then G is identically zero, i.e. $\forall\mu\in\mathbb{R}:G(\mu)=0$.

Figure 6.1 provides an illustration of how well-behaved this function G is. This figure depicts the use of the generalised Bayes' rule for computing the updated lower expectation of the queue length of the model from Example 3.1, given a positive measurement of the sensor.

More generally, note that the function G in Proposition 6.1 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 [2].

A sketch of this algorithm can be given as follows. The algorithm starts by setting $\mu_-\coloneqq\min f$, and $\mu_+\coloneqq\max f$; 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 $\mu\coloneqq{}^1\!/_2(\mu_+-\mu_-)$; then, if $G(\mu)\geq 0$ set $\mu_-\coloneqq\mu$, otherwise set $\mu_+\coloneqq\mu$; then repeat. Clearly, the interval $[\mu_-,\mu_+]$ still contains the maximum root after each step. The procedure can be terminated whenever $(\mu_+-\mu_-)<\epsilon$, for some desired numerical precision $\epsilon>0$. Since the width of the interval is halved at each iteration, the runtime of this procedure is $O\big(\log\{(\max f-\min f)\epsilon^{-1}\}\big)$. Methods for improving the numerical stability of this procedure can be found in [3, Section 2.7.3].

7. 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 on the state-space of the chain that factorise

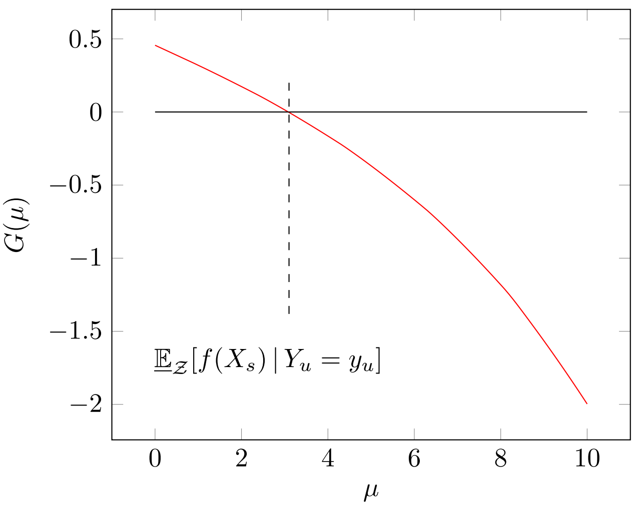

Figure 2: Plot of the function $G(\mu)$ used in the generalised Bayes rule. The dashed line marks the (in this case unique) root of $G(\mu)$, that is, the value of μ corresponding to the updated lower expectation of interest. The model considered here corresponds to the unobserved single-server queue from Example 3.1, with length X_t at most $k = 10$, with arrival rates $a \in [0.8, 1.2]$ and departure rates $d \in [0.9, 1.3]$, "full queue"-sensor output probabilities $P_{\mathcal{Y}|\mathcal{X}}(\text{SensorSaysFull=1}\,|\,i) = {}^{i}\!/_{k+1}$ for all $i \in \{0, \ldots, k\}$, and an empty initial queue $P(X_0 = 0) = 1$. The quantity of interest is the expected queue length, obtained by setting $f(x) \coloneqq x$, at time $s = 21$, given an observed positive sensor output at time $t = 20$. Then $\underline{\mathbb{E}}_{\mathcal{Z}}[f(X_{21})\,|\,\text{SensorSaysFull}_{20}\text{=1}] \approx 3.1$.

over a finite number of time-points. We have shown that we can use this algorithm to efficiently compute lower unconditional probabilities and densities with respect to the output variables. Previous work [11] has shown that inference problems of this kind have uses in robust classification using imprecise (discrete time) hidden Markov chains. These results might therefore be used for similar applications using continuous-time models.

Furthermore, we have shown that lower expectations of functions on the state-space of the updated model, i.e. given a collection of observations of the output variables, can be written as a maximisation problem over lower expectations of functions of this form. By combining our algorithm with a simple bisection algorithm, we have therefore provided an efficient method to compute such updated lower expectations of functions on the state-space.

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 $f \in \mathcal{L}(\mathcal{X}_v)$ that depend on more than one time-point. Another such problem is that of estimating state-sequences given observed output-sequences—as was previously done for (discrete-time) iHMM's [4].

Acknowledgements

The work in this paper was partially supported by the Research Foundation - Flanders (FWO) and the H2020-MSCA-ITN-2016 UTOPIAE, grant agreement 722734. The authors would like to thank Alexander Erreygers for stimulating discussions on the subject of imprecise continuous-time Markov chains.

[1] A. Benavoli, M. Zaffalon, and E. Miranda. Robust filtering through coherent lower previsions. *IEEE Transactions on Automatic Control* 56(7):1567-1581, 2011.

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

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

[4] 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.

[5] 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.

[6] A. Erreygers and J. De Bock. Imprecise Continuous-Time Markov Chains: Efficient Computational Methods with Guaranteed Error Bounds. *Proceedings of Machine Learning Research (Proceedings of ISIPTA 2017)*, 62:145-156, 2017.

[7] T. Krak, J. De Bock, and A. Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, 88:452-528, 2017.

[8] T. Krak, J. De Bock, and A. Siebes. Efficient computation of updated lower expectations for imprecise continuous-time hidden Markov chains. *Proceedings of Machine Learning Research (Proceedings of ISIPTA 2017)*, 62:193-204, 2017.

[9] S. Lopatatzidis, J. De Bock, G. de Cooman, S. De Vuyst, and J. Walraevens. Robust queueing theory: an initial study using imprecise probabilities. *Queueing Systems*, 82(1-2):75-101, 2016.

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

[11] Y. Soullard, A. Antonucci, and S. Destercke. Technical Gestures Recognition by Set-Valued Hidden Markov Models with Prior Knowledge. *Soft Methods for Data Science*, pp. 455-462, 2017.

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

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

Appendix A. Globally Required Proofs and Lemmas

The following property will be useful. The result is rather trivial, but we note it here explicitly to prevent confusion when we use it in our proofs.

Lemma A.1. *Let $P_{\mathcal{X}}$ be a stochastic process and consider any $u \in \mathcal{U}$ and any $\{f_i\}_{i\in\mathbb{N}} \to f$ in $\mathcal{L}(\mathcal{X}_u)$. Then $\lim_{i\to+\infty} \mathbb{E}_{P_{\mathcal{X}}}[f_i(X_u)] = \mathbb{E}_{P_{\mathcal{X}}}[f(X_u)]$.*

Proof. Trivial consequence of the definition of our norm $\|\cdot\|$ on $\mathcal{L}(\mathcal{X}_u)$. □

The following lemma states the imprecise analogue of the above result; this is essentially well-known, but we repeat it here for the sake of completeness.

Lemma A.2. *Let $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ be an ICTMC and consider any $u \in \mathcal{U}$ and any $\{f_i\}_{i\in\mathbb{N}} \to f$ in $\mathcal{L}(\mathcal{X}_u)$. Then $\lim_{i\to+\infty} \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f_i(X_u)] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[f(X_u)]$.*

Proof. Keeping u fixed, then since $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\,\cdot\,] : \mathcal{L}(\mathcal{X}_u) \to \mathbb{R}$ is an infimum over precise expectations $\mathbb{E}_{P_{\mathcal{X}}}[\,\cdot\,] : \mathcal{L}(\mathcal{X}_u) \to \mathbb{R}$, with $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$, we know from [12, Theorem 3.3.3] that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\,\cdot\,]$ is a coherent lower prevision on $\mathcal{L}(\mathcal{X}_u)$. Therefore, the statement follows directly from [12, Proposition 2.6.1.ℓ] and the definition of our norm $\|\cdot\|$ on $\mathcal{L}(\mathcal{X}_u)$. □

We provide the proof of Proposition 6.1 below; this is not in chronological order with respect to the main text, but it states a number of convenient properties that are required in the proofs of statements that appear before Proposition 6.1. We first need the following lemma.

Lemma A.3. *Let $P_{\mathcal{X}}$ be a stochastic process and consider any $u, v \in \mathcal{U}$, any $f \in \mathcal{L}(\mathcal{X}_v)$ and any $g \in \mathcal{L}(\mathcal{X}_u)$ such that $g \geq 0$. If $\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)] = 0$, then for all $\mu \in \mathbb{R}$ it holds that*

$$\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] = 0\,.$$

Proof. Because $\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)] = 0$, and since $g \geq 0$, we must clearly have that

$$P_{\mathcal{X}}(X_u = x_u) = 0\,,$$

for all $x_u \in \mathcal{X}_u$ for which $g(x_u) \neq 0$. Let $\mathcal{X}_u^0 \coloneqq \{x_u \in \mathcal{X}_u \,:\, g(x_u) \neq 0\}$. Then clearly for any $x_u \in \mathcal{X}_u^0$ it holds for all $x_{v\setminus u} \in \mathcal{X}_{v\setminus u}$ that also $P_{\mathcal{X}}(X_u = x_u, X_{v\setminus u} = x_{v\setminus u}) = 0$. Hence, for all $x_u \in \mathcal{X}_u$ and $x_{v\setminus u} \in \mathcal{X}_{v\setminus u}$, we find that $P_{\mathcal{X}}(X_u = x_u, X_{v\setminus u} = x_{v\setminus u})g(x_u) = 0$. Therefore,

$$\begin{aligned}\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] &= \sum_{x_{u\cup v}\in\mathcal{X}_{u\cup v}} P_{\mathcal{X}}(X_{u\cup v} = x_{u\cup v})g(x_u)\big(f(x_v) - \mu\big) \\ &= \sum_{x_u\in\mathcal{X}_u}\sum_{x_{v\setminus u}\in\mathcal{X}_{v\setminus u}} P_{\mathcal{X}}(X_u = x_u, X_{v\setminus u} = x_{v\setminus u})g(x_u)\big(f(x_v) - \mu\big) = 0.\end{aligned}$$

□

Proof. **of Proposition 6.1** For brevity, define $g \in \mathcal{L}(\mathcal{X}_u)$ as $g(x_u) \coloneqq \prod_{t\in u} g_t(x_t)$ for all $x_u \in \mathcal{X}_u$.

We start by proving Property G1. For continuity, consider any $\mu \in \mathbb{R}$. We will prove that G is continuous in μ, or in other words, that for every sequence $\{\mu_i\}_{i\in\mathbb{N}} \to \mu$ it holds that $\lim_{i\to\infty} G(\mu_i) = G(\mu)$. So, choose any sequence $\{\mu_i\}_{i\in\mathbb{N}} \to \mu$, and consider the induced sequence of functions $\big\{g(X_u)\big(f(X_v) - \mu_i\big)\big\}_{i\in\mathbb{N}}$ in $\mathcal{L}(\mathcal{X}_{u\cup v})$. Then, since $\{\mu_i\}_{i\in\mathbb{N}} \to \mu$, clearly also $\lim_{i\to+\infty} g(X_u)\big(f(X_v) - \mu_i\big) = g(X_u)\big(f(X_v) - \mu\big)$. Using Lemma A.2, we therefore find that

$$\lim_{i\to+\infty} \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)\big(f(X_v) - \mu_i\big)] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)\big(f(X_v) - \mu\big)]\,,$$

or in other words, that $\lim_{i\to+\infty} G(\mu_i) = G(\mu)$. Since the sequence $\{\mu_i\}_{i\in\mathbb{N}}$ was arbitrary, this concludes the proof.

We next prove that G is non-increasing. To this end, fix any $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$. Then, by the linearity of expectation operators,

$$\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] = \mathbb{E}_{P_{\mathcal{X}}}[g(X_u)f(X_v)] - \mu\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)]\,. \tag{A.1}$$

Since by assumption $g \geq 0$, it must hold that $\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)] \geq 0$, and so, that $\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big]$ is non-increasing in μ. Since this is true for all $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$, we find that $G(\mu)$ is a lower envelope of non-increasing functions, which must therefore be non-increasing itself. This concludes the proof.

For concavity, fix any $\mu, \nu \in \mathbb{R}$, and choose any $\lambda \in [0,1]$. Let $\mu' \coloneqq \lambda\mu + (1-\lambda)\nu$. We need to show that $\lambda G(\mu) + (1-\lambda)G(\nu) \leq G(\mu')$. To this end, fix any $\epsilon \in \mathbb{R}_{>0}$. Then, there is some $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$ such that

$$\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\big(f(X_v) - \mu'\big)\right] - \epsilon < G(\mu'). \tag{A.2}$$

By expanding the convex combination μ' using the linearity of expectation operators, we find that

$$\begin{aligned}
&\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\big(f(X_v) - \mu'\big)\right] \\
&= \mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)f(X_v)\right] - \mu'\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\right] \\
&= \mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)f(X_v)\right] - \lambda\mu\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\right] - (1-\lambda)\nu\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\right] \\
&= \big(\lambda + (1-\lambda)\big)\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)f(X_v)\right] - \lambda\mu\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\right] - (1-\lambda)\nu\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\right] \\
&= \lambda\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)f(X_v)\right] - \lambda\mu\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\right] + (1-\lambda)\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)f(X_v)\right] - (1-\lambda)\nu\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\right] \\
&= \lambda\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\big(f(X_v) - \mu\big)\right] + (1-\lambda)\mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\big(f(X_v) - \nu\big)\right] \\
&\geq \lambda\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[g(X_u)\big(f(X_v) - \mu\big)\right] + (1-\lambda)\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[g(X_u)\big(f(X_v) - \nu\big)\right] \\
&= \lambda G(\mu) + (1-\lambda)G(\nu)\,,
\end{aligned}$$

where the inequality follows from the fact that λ and $(1-\lambda)$ are non-negative, and $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$—hence in particular $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}} \leq \mathbb{E}_{P_{\mathcal{X}}}$. Combining with Equation (A.2), we find that

$$\lambda G(\mu) + (1-\lambda)G(\nu) - \epsilon \leq \mathbb{E}_{P_{\mathcal{X}}}\left[g(X_u)\big(f(X_v) - \mu'\big)\right] - \epsilon < G(\mu')\,.$$

Since the $\epsilon \in \mathbb{R}_{>0}$ was arbitrary, this concludes the proof.

To prove that the function has a root, first consider any $\mu < \min f$. Then, for every $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$, using the assumption that $g \geq 0$,

$$\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)f(X_v)] \geq \mathbb{E}_{P_{\mathcal{X}}}[g(X_u)\min f] = \mu\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)]\,,$$

and hence, using Equation (A.1), we find that $\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] \geq 0$. Since this is true for all $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$, we find that $G(\mu) \geq 0$ for this choice of μ. Next, consider any $\nu > \max f$. Then by a completely analogous argument—just reverse the inequalities—we find that $G(\nu) \leq 0$ for this choice of ν. Therefore, and since we already know that G is continuous, by the intermediate value theorem there must now be some $\mu_* \in [\mu, \nu]$ such that $G(\mu_*) = 0$. This concludes the proof.

We next prove G2, and start by showing that G is (strictly) decreasing if $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$. To this end, consider any $\mu \in \mathbb{R}$ and any $\Delta \in \mathbb{R}_{>0}$. We need to show that $G(\mu) > G(\mu + \Delta)$. Let $\epsilon \coloneqq \Delta\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)]$; clearly then $\epsilon > 0$. Therefore, there is some $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$ such that

$$\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] - \epsilon < G(\mu)\,. \tag{A.3}$$

Expanding the left hand side as in Equation (A.1), we find

$$\begin{aligned}
\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] - \epsilon &= \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)f(X_v)\big] - \mu\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big] - \epsilon \\
&= \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)f(X_v)\big] - \mu\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big] - \Delta\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] \\
&\geq \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)f(X_v)\big] - \mu\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big] - \Delta\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big] \\
&= \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)f(X_v)\big] - (\mu + \Delta)\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big] \\
&= \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - (\mu + \Delta)\big)\big] \\
&\geq G(\mu + \Delta)\,,
\end{aligned}$$

where the first inequality follows from the fact that Δ and $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)]$ are strictly positive, and $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] \leq \mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big]$ since $P_\mathcal{X} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$. Combining with Equation (A.3) shows that

$$G(\mu+\Delta) \leq \mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] - \epsilon < G(\mu)\,,$$

which concludes the proof.

To prove that G has a *unique* root under the assumption that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$, note that we already know that G has at least one root, i.e. $G(\mu) = 0$ for some $\mu \in \mathbb{R}$. By combining with the fact that G is strictly decreasing under the assumption that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$, the uniqueness of this root follows immediately.

We next prove G3. Lemma A.4 below states that G has a maximum root if $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$, so we do not need to prove this here. So, let $\mu_* \coloneqq \max\{\mu \in \mathbb{R} : G(\mu) \geq 0\}$ be this maximum root.

We will now prove that $G(\mu) = 0$ for all $\mu \leq \mu_*$ when $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] = 0$ but $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$. So, consider any $\mu < \mu_*$ (the case for $\mu = \mu_*$ is trivial). Since $\mu < \mu_*$, we already know that $G(\mu)$ is non-negative because G is non-increasing; so, it suffices to show that $G(\mu)$ is non-positive. Now, for every $P_\mathcal{X} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$ we have that

$$\begin{aligned}
\mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] &= \sum_{x_{u\cup v}} P_\mathcal{X}(X_{u\cup v} = x_{u\cup v})g(x_u)\big(f(x_v)-\mu\big) \\
&\leq \sum_{x_{u\cup v}} P_\mathcal{X}(X_{u\cup v} = x_{u\cup v})g(x_u)\,|f(x_v)-\mu| \\
&\leq \sum_{x_{u\cup v}} P_\mathcal{X}(X_{u\cup v} = x_{u\cup v})g(x_u)\,\|f-\mu\| \\
&= \sum_{x_u} P_\mathcal{X}(X_u = x_u)g(x_u)\,\|f-\mu\| = \|f-\mu\|\,\mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big],
\end{aligned} \tag{A.4}$$

where the first inequality follows from the fact that $g \geq 0$, the second inequality is due to the definition of the norm $\|\cdot\|$, and the second equality follows from the law of total probability (that is, $X_{v\setminus u}$ is marginalised out). Since $G(\mu) \leq \mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big]$, it follows from the above that if $\|f-\mu\| = 0$ then $G(\mu) \leq 0$, in which case we are done.

Hence, we can assume without loss of generality that $\|f-\mu\| > 0$. Choose any $\epsilon \in \mathbb{R}_{>0}$. Then, because $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] = 0$, there is some $P_\mathcal{X} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$ such that $\mathbb{E}_{P_\mathcal{X}}[g(X_u)] < \epsilon/\|f-\mu\|$, which implies that

$$\epsilon > \|f-\mu\|\,\mathbb{E}_{P_\mathcal{X}}[g(X_u)] \geq \mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] \geq G(\mu)\,,$$

using Equation (A.4) for the second inequality. Since the $\epsilon \in \mathbb{R}_{>0}$ was arbitrary, this implies that $G(\mu)$ is non-positive, which concludes the proof.

We next show that G is strictly decreasing for $\mu > \mu_*$. So consider any $\mu \in \mathbb{R}$ such that $\mu > \mu_*$, and any $\Delta \in \mathbb{R}_{>0}$; we need to show that $G(\mu) > G(\mu+\Delta)$. Because $\mu > \mu_*$, and since $\mu_* = \max\{\mu \in \mathbb{R} : G(\mu) \geq 0\}$, we know that $G(\mu) < 0$.

First note that for any $\epsilon \in \mathbb{R}_{>0}$, there is some $P_\mathcal{X} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$ for which

$$G(\mu) > \mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] - \epsilon\,,$$

and clearly also

$$\mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu_*\big)\big] \geq G(\mu_*) = 0\,.$$

Therefore,

$$\begin{aligned}
G(\mu)+\epsilon > \mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] &\geq \mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] - \mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu_*\big)\big] \\
&= -\mu\mathbb{E}_{P_\mathcal{X}}[g(X_u)] + \mu_*\mathbb{E}_{P_\mathcal{X}}[g(X_u)] = (\mu_*-\mu)\mathbb{E}_{P_\mathcal{X}}[g(X_u)]
\end{aligned}$$

and so, negating both sides and noting again that $G(\mu) < 0$,

$$(\mu-\mu_*)\mathbb{E}_{P_\mathcal{X}}[g(X_u)] > |G(\mu)| - \epsilon\,,$$

and dividing through,

$$\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)] > \frac{|G(\mu)| - \epsilon}{\mu - \mu_*}\,, \tag{A.5}$$

for any $P_{\mathcal{X}}$ that ϵ-approaches $G(\mu)$.

The idea is now the following: Equation (A.5) provides a lower bound on the (absolute magnitude of the) slope of the function $\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big]$ for any $P_{\mathcal{X}}$ that ϵ-approaches $G(\mu)$. Since this function can still be a distance of ϵ above $G(\mu)$, we now need to make sure that the slope is such that this function will decrease by more than ϵ after increasing μ by Δ; this will guarantee that this function becomes strictly lower than $G(\mu)$ when evaluated at $\mu + \Delta$, and since it is an upper bound on $G(\mu + \Delta)$, this will complete the proof.

Note that this lower bound increases as we decrease ϵ. In order to ensure that we get a large enough slope, we now solve the following for ϵ:

$$\begin{aligned}
\Delta \frac{|G(\mu)| - \epsilon}{\mu - \mu_*} > \epsilon &\Leftrightarrow \frac{\Delta\,|G(\mu)|}{\mu - \mu_*} - \frac{\Delta\epsilon}{\mu - \mu_*} > \epsilon \\
&\Leftrightarrow \frac{\Delta\,|G(\mu)|}{\mu - \mu_*} > \epsilon + \frac{\Delta\epsilon}{\mu - \mu_*} \\
&\Leftrightarrow \Delta\,|G(\mu)| > (\mu - \mu_*)\epsilon + \Delta\epsilon \\
&\Leftrightarrow \Delta\,|G(\mu)| > \epsilon\big((\mu - \mu_*) + \Delta\big) \Leftrightarrow \frac{\Delta\,|G(\mu)|}{\Delta + (\mu - \mu_*)} > \epsilon.
\end{aligned} \tag{A.6}$$

So, choose any $\epsilon \in \mathbb{R}_{>0}$ such that $\epsilon < {}^{\Delta|G(\mu)|}/_{(\Delta+\mu-\mu_*)}$; since the right-hand side is clearly strictly positive, this is always possible. Then there is some $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$ such that

$$\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] - \epsilon < G(\mu)\,.$$

For this same $P_{\mathcal{X}}$, we then have

$$\begin{aligned}
G(\mu + \Delta) &\leq \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - (\mu + \Delta)\big)\big] \\
&= \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] - \Delta\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big] \\
&< \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] - \Delta\frac{|G(\mu)| - \epsilon}{\mu - \mu_*} \\
&< \mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] - \epsilon \\
&< G(\mu)\,,
\end{aligned}$$

where the second inequality is by Equation (A.5), the third inequality is by Equation (A.6), and the final inequality is by the choice of $P_{\mathcal{X}}$. So, we have found that indeed $G(\mu) > G(\mu + \Delta)$, which concludes the proof.

We finally prove G4, i.e. that $G(\mu)$ is identically zero whenever $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] = 0$. Clearly, this assumption implies that $\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)] = 0$ for all $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$. Therefore, and since $g \geq 0$, it follows from Lemma A.3 that for any $\mu \in \mathbb{R}$, it holds that $\mathbb{E}_{P_{\mathcal{X}}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] = 0$ for all $P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$. It immediately follows that indeed $G(\mu) = 0$ for all $\mu \in \mathbb{R}$. □

The following lemma states a more general version of the result that the generalised Bayes' rule computes the updated lower expectation of a model under regular extension. We state it here because we use the result for various proofs throughout this appendix.

Lemma A.4. *Let $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ be an ICTMC, and consider any $u, v \in \mathcal{U}$, $f \in \mathcal{L}(\mathcal{X}_v)$ and $g \in \mathcal{L}(\mathcal{X}_u)$ such that $g \geq 0$. Then, if $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$, it holds that*

$$\begin{aligned}
&\max\big\{\mu \in \mathbb{R} \,:\, \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g(X_u)\big(f(X_v) - \mu\big)\big] \geq 0\big\} \\
&\qquad = \inf\left\{\frac{\mathbb{E}_{P_{\mathcal{X}}}[f(X_v)g(X_u)]}{\mathbb{E}_{P_{\mathcal{X}}}[g(X_u)]} \,:\, P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}, \mathbb{E}_{P_{\mathcal{X}}}[g(X_u)] > 0\right\}.
\end{aligned}$$

Proof. Let $\mathcal{P} \coloneqq \{P_\mathcal{X} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}} : \mathbb{E}_{P_\mathcal{X}}[g(X_u)] > 0\}$, and note that $\mathcal{P}$ is non-empty due to the assumption that $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$. For all $P_\mathcal{X} \in \mathcal{P}$, define

$$\mu_{P_\mathcal{X}} \coloneqq \frac{\mathbb{E}_{P_\mathcal{X}}[f(X_v)g(X_u)]}{\mathbb{E}_{P_\mathcal{X}}[g(X_u)]}\,,$$

and let μ_* be defined by

$$\mu_* \coloneqq \inf\left\{\mu_{P_\mathcal{X}} : P_\mathcal{X} \in \mathcal{P}\right\}.$$

Now define the following function, parameterised in $\mu \in \mathbb{R}$,

$$\underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu)] \coloneqq \inf\{\mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - \mu)] : P_\mathcal{X} \in \mathcal{P}\}\,,$$

and consider $\underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)]$. We start by showing that this quantity is non-negative. To this end, fix any $\epsilon > 0$. Then, there is some $P_\mathcal{X} \in \mathcal{P}$ such that

$$\mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - \mu_*)] - \epsilon < \underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)]\,.$$

Using Equation (A.1), the function $\mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - \mu)]$ is non-increasing in μ. Therefore, and since $\mu_* \leq \mu_{P_\mathcal{X}}$, we have

$$\mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - \mu_{P_\mathcal{X}})] - \epsilon \leq \mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - \mu_*)] - \epsilon < \underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)]\,.$$

Due to the choice of $\mu_{P_\mathcal{X}}$, we have $\mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - \mu_{P_\mathcal{X}})] = 0$, and so we find that

$$-\epsilon < \underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)]\,.$$

Since this is true for every $\epsilon > 0$, we conclude that $0 \leq \underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)]$. Next, we show that this quantity is also non-positive, or in other words, that μ_* is a root of this function.

To this end, fix any $\epsilon > 0$, and define $\epsilon' \coloneqq \epsilon/\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)]$; since by assumption $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$, we have $\epsilon' > 0$. Now consider $P_\mathcal{X} \in \mathcal{P}$ such that

$$\mu_{P_\mathcal{X}} - \epsilon' < \mu_*\,.$$

Then, since $\underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu)]$ is non-increasing in μ—because it is a lower envelope of non-increasing functions—we have that

$$\underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)] \leq \underline{\mathbb{E}}[g(X_u)(f(X_v) - (\mu_{P_\mathcal{X}} - \epsilon'))] \leq \mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - (\mu_{P_\mathcal{X}} - \epsilon'))]\,.$$

Expanding the r.h.s. using the linearity of expectation operators, and by the definition of $\mu_{P_\mathcal{X}}$, we then have

$$\begin{aligned}&\mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - (\mu_{P_\mathcal{X}} - \epsilon'))]\\ &\qquad = \mathbb{E}_{P_\mathcal{X}}[g(X_u)f(X_v)] - \mu_{P_\mathcal{X}}\mathbb{E}_{P_\mathcal{X}}[g(X_u)] + \epsilon'\mathbb{E}_{P_\mathcal{X}}[g(X_u)] = \epsilon'\mathbb{E}_{P_\mathcal{X}}[g(X_u)]\,,\end{aligned}$$

and since $P_\mathcal{X} \in \mathcal{P} \subseteq \mathbb{P}_{\mathcal{Q},\mathcal{M}}$,

$$\epsilon'\mathbb{E}_{P_\mathcal{X}}[g(X_u)] \leq \epsilon'\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] = \epsilon\,,$$

and so we find that

$$\underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)] \leq \epsilon\,.$$

Since this is true for every $\epsilon > 0$, and since we already know that $\underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)]$ is non-negative, we conclude that

$$\underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu_*)] = 0\,.$$

Now consider any $\mu' > \mu_*$. There must then be some $P_\mathcal{X} \in \mathcal{P}$ such that $\mu_* \leq \mu_{P_\mathcal{X}} < \mu'$, and furthermore,

$$\underline{\mathbb{E}}[g(X_u)(f(X_v) - \mu')] \leq \mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - \mu')] < \mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v) - \mu_{P_\mathcal{X}})] = 0\,,$$

where the strict inequality follows from the fact that $\mathbb{E}_{P_\mathcal{X}}[g(X_u)(f(X_v)-\mu)]$ is strictly decreasing, and $\mu_{P_\mathcal{X}} < \mu'$. Since this is true for every $\mu' > \mu_*$, we conclude that

$$\mu_* = \max\left\{\mu \in \mathbb{R} \,:\, \underline{\mathbb{E}}[g(X_u)(f(X_v)-\mu)] \geq 0\right\}.$$

Hence, because of our definition for μ_*, we are left to prove that

$$\max\left\{\mu \in \mathbb{R} \,:\, \underline{\mathbb{E}}[g(X_u)(f(X_v)-\mu)] \geq 0\right\} = \max\left\{\mu \in \mathbb{R} \,:\, \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)(f(X_v)-\mu)] \geq 0\right\}. \tag{A.7}$$

If $\mathcal{P} = \mathbb{P}_{\mathcal{Q},\mathcal{M}}$, this is trivially true. Therefore, we can assume without loss of generality that $\mathcal{P} \neq \mathbb{P}_{\mathcal{Q},\mathcal{M}}$. Let $\mathcal{P}_0 \coloneqq \mathbb{P}_{\mathcal{Q},\mathcal{M}} \setminus \mathcal{P}$. Due to the definition of $\mathcal{P}$, we then have for every $P_\mathcal{X} \in \mathcal{P}_0$ that $\mathbb{E}_{P_\mathcal{X}}[g(X_u)] = 0$. It follows from Lemma A.3 that $\mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] = 0$ for all $\mu \in \mathbb{R}$ and all $P_\mathcal{X} \in \mathcal{P}_0$. Hence,

$$\inf\left\{\mathbb{E}_{P_\mathcal{X}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] \,:\, P_\mathcal{X} \in \mathcal{P}_0\right\} = 0 \text{ for all } \mu \in \mathbb{R}.$$

Because $\mathcal{P} \cup \mathcal{P}_0 = \mathbb{P}_{\mathcal{Q},\mathcal{M}}$, we therefore have

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] = \min\left\{\underline{\mathbb{E}}[g(X_u)(f(X_v)-\mu)],\, 0\right\},$$

and so we conclude that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g(X_u)\big(f(X_v)-\mu\big)\big]$ is negative if and only if $\underline{\mathbb{E}}[g(X_u)(f(X_v)-\mu)]$ is negative. This clearly implies Equation (A.7). □

Corollary A.5. *Let $\mathcal{Z}$ be an ICTHMC, and consider any $u, v \in \mathcal{U}$, $f \in \mathcal{L}(\mathcal{X}_v)$ and $g \in \mathcal{L}(\mathcal{X}_u)$ such that $g \geq 0$. Then, if $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$, it holds that*

$$\max\left\{\mu \in \mathbb{R} \,:\, \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\big[g(X_u)\big(f(X_v)-\mu\big)\big] \geq 0\right\} = \inf\left\{\frac{\mathbb{E}_{P_\mathcal{X}}[f(X_v)g(X_u)]}{\mathbb{E}_{P_\mathcal{X}}[g(X_u)]} \,:\, P \in \mathcal{Z}, \mathbb{E}_{P_\mathcal{X}}[g(X_u)] > 0\right\}.$$

Proof. Trivial consequence of Lemma A.4 and Definition 2. □

Due to the independence assumptions on augmented stochastic processes and the fact that we use a fixed output distribution $P_{\mathcal{Y}|\mathcal{X}}$, (lower) probabilities of outputs can be conveniently rewritten as follows.

Lemma A.6. *Let $P = P_{\mathcal{Y}|\mathcal{X}} \otimes P_\mathcal{X}$ be an augmented stochastic process and consider any $u \in \mathcal{U}$ and any $O_u \in \Sigma_u$. Then $P(Y_u \in O_u) = \mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)]$.*

Proof. Due to the construction of augmented stochastic processes in Section 3.2, we have that

$$P(Y_u \in O_u) = \sum_{x_u \in \mathcal{X}_u} P(Y_u \in O_u, X_u = x_u) = \sum_{x_u \in \mathcal{X}_u} P_{\mathcal{Y}|\mathcal{X}}(O_u|x_u) P_\mathcal{X}(X_u = x_u) = \mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)].$$

□

Lemma A.7. *Let $\mathcal{Z}$ be an ICTHMC with corresponding ICTMC $\mathbb{P}_{\mathcal{Q},\mathcal{M}}$ and consider any $u \in \mathcal{U}$ and $O_u \in \Sigma_u$. Then $\overline{P}_\mathcal{Z}(Y_u \in O_u) = \overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)]$.*

Proof. Trivial consequence of Definition 2 and Lemma A.6. □

Appendix B. Proofs of the Results in Section 4

Proof. **of Proposition 4.1** Fix $P \in \mathcal{Z}$, and write $P = P_{\mathcal{Y}|\mathcal{X}} \otimes P_\mathcal{X}$. The corresponding expectation of f is then

$$\mathbb{E}_P[f(X_u, Y_u)] = \mathbb{E}_P\big[\mathbb{E}_P[f(X_u, Y_u) \,|\, X_u]\big] = \mathbb{E}_{P_\mathcal{X}}\big[\mathbb{E}_P[f(X_u, Y_u) \,|\, X_u]\big], \tag{B.1}$$

where the first equality follows from the law of iterated expectation, and the second equality follows from the fact that the inner (conditional) expectation is itself a function on $\mathcal{X}_u$; whence the outer expectation is only taken with respect to the random variable X_u, whose distribution is given by $P_{\mathcal{X}}$, since $P = P_{\mathcal{Y}|\mathcal{X}} \otimes P_{\mathcal{X}}$.

Furthermore, for every $x_u \in \mathcal{X}_u$, we have

$$\mathbb{E}_P[f(X_u, Y_u) \,|\, X_u = x_u] = \mathbb{E}_P[f(x_u, Y_u) \,|\, X_u = x_u] = \mathbb{E}_{P_{\mathcal{Y}|\mathcal{X}}}[f(x_u, Y_u) \,|\, X_u = x_u] \,,$$

where the first equality follows from the basic rules of probability. The second equality follows from the facts that the function for which the expectation is computed is $f(x_u, Y_u)$—the restriction of f to $\mathcal{Y}_u$—and that therefore the remaining expectation is taken only with respect to the conditional random variable $Y_u | X_u = x_u$, whose distribution is given by $P_{\mathcal{Y}|\mathcal{X}}$, since $P = P_{\mathcal{Y}|\mathcal{X}} \otimes P_{\mathcal{X}}$. Because this holds for all $x_u \in \mathcal{X}_u$, substitution into Equation (B.1) yields

$$\mathbb{E}_P[f(X_u, Y_u)] = \mathbb{E}_{P_{\mathcal{X}}}\left[\mathbb{E}_{P_{\mathcal{Y}|\mathcal{X}}}[f(X_u, Y_u) \,|\, X_u]\right].$$

Because this holds for all $P \in \mathcal{Z}$, due to Definition 2 we now find that

$$\begin{aligned}
\underline{\mathbb{E}}_{\mathcal{Z}}[f(X_u, Y_u)] &= \inf\left\{\mathbb{E}_P[f(X_u, Y_u)] \,:\, P \in \mathcal{Z}\right\} \\
&= \inf\left\{\mathbb{E}_{P_{\mathcal{Y}|\mathcal{X}} \otimes P_{\mathcal{X}}}[f(X_u, Y_u)] \,:\, P_{\mathcal{Y}|\mathcal{X}} \otimes P_{\mathcal{X}} \in \mathcal{Z}\right\} \\
&= \inf\left\{\mathbb{E}_{P_{\mathcal{X}}}\left[\mathbb{E}_{P_{\mathcal{Y}|\mathcal{X}}}[f(X_u, Y_u) \,|\, X_u]\right] \,:\, P_{\mathcal{Y}|\mathcal{X}} \otimes P_{\mathcal{X}} \in \mathcal{Z}\right\} \\
&= \inf\left\{\mathbb{E}_{P_{\mathcal{X}}}\left[\mathbb{E}_{P_{\mathcal{Y}|\mathcal{X}}}[f(X_u, Y_u) \,|\, X_u]\right] \,:\, P_{\mathcal{X}} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}\right\} \\
&= \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\mathbb{E}_{P_{\mathcal{Y}|\mathcal{X}}}[f(X_u, Y_u) \,|\, X_u]\right].
\end{aligned}$$

□

Proof. **of Lemma 4.2** We provide a proof by induction. Clearly, the result is trivial for $i = n$. So, assume that it is true for i. We show that it is then also true for $i - 1$ (with $i > 0$).

We focus on $f^+_{t_{i-1}}$, and consider the two cases in its definition separately. So, consider any $x_{t_{i-1}} \in \mathcal{X}_{t_{i-1}}$. Then, if $f_{t_{i-1}}(x_{t_{i-1}}) \geq 0$, we have

$$\begin{aligned}
f^+_{t_{i-1}}(x_{t_{i-1}}) &= f_{t_{i-1}}(x_{t_{i-1}})\underline{\mathbb{E}}_{\mathcal{Q}}\left[f^+_{t_i}(X_{t_i}) \,|\, X_{t_{i-1}} = x_{t_{i-1}}\right] \\
&= f_{t_{i-1}}(x_{t_{i-1}})\underline{\mathbb{E}}_{\mathcal{Q}}\left[\underline{\mathbb{E}}_{\mathcal{Q}}\left[\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_i}\right] \,\middle|\, X_{t_{i-1}} = x_{t_{i-1}}\right] \\
&= f_{t_{i-1}}(x_{t_{i-1}})\underline{\mathbb{E}}_{\mathcal{Q}}\left[\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_{i-1}} = x_{t_{i-1}}\right] \\
&= \underline{\mathbb{E}}_{\mathcal{Q}}\left[f_{t_{i-1}}(x_{t_{i-1}})\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_{i-1}} = x_{t_{i-1}}\right],
\end{aligned}$$

where the first equality is by definition, the second is by the induction hypothesis, the third by iterated lower expectation (Proposition 2.1), and the final by the non-negative homogeneity of lower expectations and the assumption that $f_{t_{i-1}}(x_{t_{i-1}}) \geq 0$.

For the other case, assume that $f_{t_{i-1}}(x_{t_{i-1}}) < 0$. Then,

$$
\begin{aligned}
f^+_{t_{i-1}}(x_{t_{i-1}}) &= f_{t_{i-1}}(x_{t_{i-1}})\overline{\mathbb{E}}_{\mathcal{Q}}\left[f^-_{t_i}(X_{t_i}) \,|\, X_{t_{i-1}} = x_{t_{i-1}}\right] \\
&= f_{t_{i-1}}(x_{t_{i-1}})\overline{\mathbb{E}}_{\mathcal{Q}}\left[\overline{\mathbb{E}}_{\mathcal{Q}}\left[\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_i}\right] \,\middle|\, X_{t_{i-1}} = x_{t_{i-1}}\right] \\
&= f_{t_{i-1}}(x_{t_{i-1}})\overline{\mathbb{E}}_{\mathcal{Q}}\left[\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_{i-1}} = x_{t_{i-1}}\right] \\
&= -f_{t_{i-1}}(x_{t_{i-1}})\underline{\mathbb{E}}_{\mathcal{Q}}\left[-\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_{i-1}} = x_{t_{i-1}}\right] \\
&= \underline{\mathbb{E}}_{\mathcal{Q}}\left[f_{t_{i-1}}(x_{t_{i-1}})\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_{i-1}} = x_{t_{i-1}}\right],
\end{aligned}
$$

where the first equality is by definition, the second equality by the induction hypothesis, the third by iterated upper expectation (Proposition 2.1 combined with conjugacy), the fourth by conjugacy of upper- and lower expectation, and the final by the non-negative homogeneity of lower expectations and the assumption that $f_{t_{i-1}}(x_{t_{i-1}}) < 0$.

Since this covers both cases in the definition of $f^+_{t_{i-1}}(x_{t_{i-1}})$, we find that

$$
f^+_{t_{i-1}}(X_{t_{i-1}}) = \underline{\mathbb{E}}_{\mathcal{Q}}\left[f_{t_{i-1}}(X_{t_{i-1}})\prod_{j=i}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_{i-1}}\right] = \underline{\mathbb{E}}_{\mathcal{Q}}\left[\prod_{j=i-1}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_{i-1}}\right],
$$

which concludes the proof for $f^+_{t_{i-1}}$. The proof for $f^-_{t_{i-1}}$ is completely analogous. □

Proof. **of Proposition 4.3** By combining Lemma 4.2 with iterated lower expectation (Proposition 2.1), we find that

$$
\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[f^+_{t_0}(X_{t_0})\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\underline{\mathbb{E}}_{\mathcal{Q}}\left[\prod_{j=0}^{n} f_{t_j}(X_{t_j}) \,\middle|\, X_{t_0}\right]\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\prod_{j=0}^{n} f_{t_j}(X_{t_j})\right],
$$

and similarly for the upper expectation. □

The below proves some of the properties that are claimed in the main text of Section 4.3. We first need the following result, which is essentially well-known, but which we repeat here for the sake of completeness.

Lemma B.1. *Fix $d \in \mathbb{N}$ and consider any absolutely integrable function $\psi : \mathbb{R}^d \to \mathbb{R}$. Then, for any $y \in \mathbb{R}^d$ and any sequence $\{B_i\}_{i\in\mathbb{N}}$ of open balls that are centred on, and shrink to, y, if ψ is continuous at y it holds that*

$$
\psi(y) = \lim_{i\to+\infty} \frac{1}{\lambda(B_i)} \int_{B_i} \psi(\gamma)\,\mathrm{d}\gamma\,,
$$

where the integral is understood in the Lebesgue sense, and where $\lambda(B_i)$ denotes the Lebesgue measure of B_i.

Proof. Fix any $\epsilon > 0$. We need to show that there is some $n \in \mathbb{N}$ such that, for all $i > n$, it holds that

$$
\left|\psi(y) - \frac{1}{\lambda(B_i)} \int_{B_i} \psi(\gamma)\,\mathrm{d}\gamma\right| < \epsilon\,.
$$

Now, because ψ is continuous at y, there is some open ball B that is centred on y, such that for all $\gamma \in B$, it holds that

$$
|\psi(y) - \psi(\gamma)| \le \epsilon\,.
$$

Furthermore, because the sequence $\{B_i\}_{i\in\mathbb{N}}$ is centred on, and shrinks to, y, there must be some $n \in \mathbb{N}$ such that for all $i > n$, it holds that $B_i \subset B$. Fix any such $i > n$. Then,

$$\begin{aligned}\left|\frac{1}{\lambda(B_i)}\int_{B_i}\psi(\gamma)\,\mathrm{d}\gamma-\psi(y)\right| &= \left|\frac{1}{\lambda(B_i)}\int_{B_i}\psi(\gamma)-\psi(y)\,\mathrm{d}\gamma\right| \\ &\leq \frac{1}{\lambda(B_i)}\int_{B_i}\left|\psi(\gamma)-\psi(y)\right|\,\mathrm{d}\gamma \\ &\leq \frac{1}{\lambda(B_i)}\int_{B_i}\epsilon\,\mathrm{d}\gamma = \frac{\epsilon}{\lambda(B_i)}\int_{B_i}1\,\mathrm{d}\gamma = \frac{\epsilon}{\lambda(B_i)}\lambda(B_i)=\epsilon.\end{aligned}$$

□

Proof. **of claims in Section 4.3** We start by proving the claim that $\phi_u(y_u|x_u)$ exists and is real-valued if it can be constructed "piecewise", as explained in the main text. So, fix any $y_u \in \mathcal{Y}_u$, choose any sequence $\{O_u^i\}_{i\in\mathbb{N}}$ in Σ_u that shrinks to y_u, and suppose that for every $t \in u$, there is a sequence $\{\lambda_{t,i}\}_{i\in\mathbb{N}}$ in $\mathbb{R}_{>0}$ such that, for all $x_t \in \mathcal{X}_t$,

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

exists and is real-valued. Recall that, for every $x_u \in \mathcal{X}_u$ and every $i \in \mathbb{N}$, we have $P_{\mathcal{Y}|\mathcal{X}}(O_u^i|x_u) = \prod_{t\in u} P_{\mathcal{Y}|\mathcal{X}}(O_t^i|x_t)$. So, by choosing $\{\lambda_i\}_{i\in\mathbb{N}}$ as $\lambda_i \coloneqq \prod_{t\in u}\lambda_{t,i}$, it follows that for every $x_u \in \mathcal{X}_u$,

$$\phi_u(y_u|x_u) = \lim_{i\to+\infty}\frac{P_{\mathcal{Y}|\mathcal{X}}(O_u^i|x_u)}{\lambda_i} = \lim_{i\to+\infty}\prod_{t\in u}\frac{P_{\mathcal{Y}|\mathcal{X}}(O_t^i|x_t)}{\lambda_{t,i}} = \prod_{t\in u}\lim_{i\to+\infty}\frac{P_{\mathcal{Y}|\mathcal{X}}(O_t^i|x_t)}{\lambda_{t,i}} = \prod_{t\in u}\phi_t(y_t|x_t)\,,$$

using the existence of the $\phi_t(y_t|x_t)$ for the third equality. Hence, $\phi_u(y_u|x_u)$ exists and, since each $\phi_t(y_t|x_t)$, $t \in u$, is real-valued, so is $\phi_u(y_u|x_u)$. This concludes the proof of this statement.

Next, we prove that the limit expression, and in particular the second equality, in Equation (8) are true when $\psi(\cdot|x_t)$ is continuous (at y_t). To this end, note that $P_{\mathcal{Y}|\mathcal{X}}(\cdot\,|x_t)$ was defined by

$$P_{\mathcal{Y}|\mathcal{X}}(O\,|x_t) \coloneqq \int_O \psi(y|x_t)\,\mathrm{d}y\,,$$

for all $O \in \Sigma$. For a sequence of open balls $\{O_t^i\}_{i\in\mathbb{N}}$ that are centred on, and shrink to, y_t, we therefore need to prove that

$$\psi(y_t|x_t) = \lim_{i\to+\infty}\frac{P_{\mathcal{Y}|\mathcal{X}}(O_t^i|x_t)}{\lambda(O_t^i)} = \lim_{i\to+\infty}\frac{1}{\lambda(O_t^i)}\int_{O_t^i}\psi(y|x_t)\,\mathrm{d}y\,.$$

Because $\psi(y_t|x_t)$ is by assumption continuous at y_t, this result follows immediately from Lemma B.1.

We finally prove the claim that $\phi_u(y_u|x_u)$ is the same for almost every sequence $\{O_u^i\}_{i\in\mathbb{N}}$ that shrinks to y_u, provided that for all $x_t \in \mathcal{X}$, $P_{\mathcal{Y}|\mathcal{X}}(\cdot|x_t)$ is constructed from a density $\psi(\cdot\,|x_t)$ that is continuous at y_t. To this end, assume that $\{O_u^i\}_{i\in\mathbb{N}}$ satisfies the assumptions in Footnote 5; i.e. that for all $t \in u$, there is a sequence of open balls $\{B_t^i\}_{i\in\mathbb{N}}$ in $\mathcal{Y}$ that shrinks to y_t such that, for all $i \in \mathbb{N}$, $\lambda(O_t^i) > 0$ and $O_t^i \subseteq B_t^i$.

We now aim to show that

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

is independent of the sequence $\{O_u^i\}_{i\in\mathbb{N}}$. We will prove this "piecewise". In particular, we will show that for all $t \in u$,

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

where we use the assumption $\lambda(O_t^i) > 0$ to ensure that each element of the sequence is well-defined.

So, consider any $t \in u$. Note that we have $P_{\mathcal{Y}|\mathcal{X}}(O_t^i|x_t) \coloneqq \int_{O_t^i} \psi(y|x_t)\,\mathrm{d}y$ by definition, and choose any $\epsilon \in \mathbb{R}_{>0}$. Because $\psi(\cdot|x_t)$ is continuous at y_t, there is some open ball $B_* \in \mathcal{Y}$ that is centred on y_t, and such that for all $y \in B_*$,

$$|\psi(y_t|x_t) - \psi(y|x_t)| \leq \epsilon\,.$$

Furthermore, because the sequence $\{B_t^i\}_{i\in\mathbb{N}}$ shrinks to y_t, there is some $n \in \mathbb{N}$ such that, for all $i > n$, it holds that $B_t^i \subseteq B_*$. Furthermore, because each $O_t^i \subseteq B_t^i$, also clearly $O_t^i \subseteq B_*$ for all $i > n$. Now consider any $i > n$. Then,

$$\begin{aligned}
\left|\psi(y_t|x_t) - \frac{P_{\mathcal{Y}|\mathcal{X}}(O_t^i|x_t)}{\lambda(O_t^i)}\right| &= \left|\psi(y_t|x_t) - \frac{1}{\lambda(O_t^i)}\int_{O_t^i}\psi(y|x_t)\,\mathrm{d}y\right| \\
&\leq \frac{1}{\lambda(O_t^i)}\int_{O_t^i}|\psi(y|x_t) - \psi(y_t|x_t)|\,\mathrm{d}y \\
&\leq \frac{1}{\lambda(O_t^i)}\int_{O_t^i}\epsilon\,\mathrm{d}y = \frac{\epsilon}{\lambda(O_t^i)}\int_{O_t^i}1\,\mathrm{d}y = \frac{\epsilon}{\lambda(O_t^i)}\lambda(O_t^i) = \epsilon\,.
\end{aligned}$$

So, we conclude that indeed

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

as claimed. Because this holds for all $t \in u$, it follows from what we discussed above that $\phi_u(y_u|x_u)$ can be constructed "piecewise", that is,

$$\phi_u(y_u|x_u) = \lim_{i\to+\infty}\frac{P_{\mathcal{Y}|\mathcal{X}}(O_u^i|x_u)}{\lambda(O_u^i)} = \lim_{i\to+\infty}\prod_{t\in u}\frac{P_{\mathcal{Y}|\mathcal{X}}(O_t^i|x_t)}{\lambda(O_t^i)} = \prod_{t\in u}\psi(y_t|x_t)\,,$$

which implies that $\phi_u(y_u|x_u)$ exists and is the same for every sequence $\{O_u^i\}_{i\in\mathbb{N}}$ for which the assumed properties hold.

□

Appendix C. Proofs of the Results in Section 5.1

Proof. **of Proposition 5.1** The result follows from some simple manipulations, but we need to keep track of the overlap $u \cap v$ between the time-points of interest to prevent double-counting those time-points. Let $w \coloneqq u \setminus v$; then clearly $v \cup w = v \cup (u \setminus v) = u \cup (v \setminus u) = u \cup v$.

Starting from the definition of $\mathbb{E}_P[f(X_v)\,|\,Y_u \in O_u]$ using Bayes' rule, we find

$$\begin{aligned}
\mathbb{E}_P[f(X_v)\,|\,Y_u \in O_u] &\coloneqq \sum_{x_v\in\mathcal{X}_v} f(x_v)\frac{P(X_v = x_v, Y_u \in O_u)}{P(Y_u \in O_u)} \\
&= \sum_{x_v\in\mathcal{X}_v}\sum_{x_w\in\mathcal{X}_w} f(x_v)\frac{P(X_v = x_v, X_w = x_w, Y_u \in O_u)}{P(Y_u \in O_u)} \\
&= \sum_{x_u\in\mathcal{X}_u}\sum_{x_{v\setminus u}\in\mathcal{X}_{v\setminus u}} f(x_v)\frac{P(X_u = x_u, X_{v\setminus u} = x_{v\setminus u}, Y_u \in O_u)}{P(Y_u \in O_u)} \\
&= \sum_{x_u\in\mathcal{X}_u}\sum_{x_{v\setminus u}\in\mathcal{X}_{v\setminus u}} f(x_v)\frac{P_{\mathcal{Y}|\mathcal{X}}(O_u|x_u)P_{\mathcal{X}}(X_u = x_u, X_{v\setminus u} = x_{v\setminus u})}{P(Y_u \in O_u)}\,,
\end{aligned}$$

where the first equality is by definition, the second is by the basic rules of probability, the third is by changing the indexing using $v \cup w = u \cup (v \setminus u)$, and the fourth is by the definition of augmented stochastic processes in Section 3.2. Joining the indexes using $u \cup (v \setminus u) = u \cup v$, we find that

$$
\begin{aligned}
\mathbb{E}_P[f(X_v) \,|\, Y_u \in O_u] &= \sum_{x_u \in \mathcal{X}_u} \sum_{x_{v\setminus u} \in \mathcal{X}_{v\setminus u}} f(x_v) \frac{P_{\mathcal{Y}|\mathcal{X}}(O_u|x_u) P_{\mathcal{X}}(X_u = x_u, X_{v\setminus u} = x_{v\setminus u})}{P(Y_u \in O_u)} \\
&= \sum_{x_{u\cup v} \in \mathcal{X}_{u\cup v}} f(x_v) \frac{P_{\mathcal{Y}|\mathcal{X}}(O_u|x_u) P_{\mathcal{X}}(X_{u\cup v} = x_{u\cup v})}{P(Y_u \in O_u)} \\
&= \frac{\mathbb{E}_{P_{\mathcal{X}}}[f(X_v) P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)]}{P(Y_u \in O_u)} ,
\end{aligned}
$$

where the third equality is by the definition of expectation. Applying Lemma A.6 to the denominator, we finally obtain

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

□

Proof. **of Proposition 5.2** This is a special case of Corollary A.5. To see this, set $g(X_u) \coloneqq P_{\mathcal{Y}|\mathcal{X}}(O_u|X_u)$ in that corollary's statement, apply Proposition 5.1 to the quantities $\mathbb{E}_P[f(X_v) \,|\, Y_u \in O_u]$ and apply Lemma A.6 to the quantities $P(Y_u \in O_u)$.

The hypothesis $\overline{P}_{\mathcal{Z}}(Y_u \in O_u) > 0$ then implies the hypothesis $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[g(X_u)] > 0$ of Corollary A.5, due to Lemma A.7. □

Appendix D. Proofs of the Results in Section 5.2

Proof. **of Proposition 5.3** Assume that there is a sequence $\{\lambda_i\}_{i\in\mathbb{N}}$ in $\mathbb{R}_{>0}$ such that, for all $x_u \in \mathcal{X}_u$,

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

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

Then, the existence of $\phi_u(y_u|X_u)$ clearly implies that

$$
\lim_{i\to+\infty} \frac{f(X_v) P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)}{\lambda_i} = f(X_v)\phi_u(y_u|X_u) ,
$$

and so, using Lemma A.1, we find that

$$
\lim_{i\to+\infty} \frac{\mathbb{E}_{P_{\mathcal{X}}}[f(X_v) P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]}{\lambda_i} = \mathbb{E}_{P_{\mathcal{X}}}[f(X_v)\phi_u(y_u|X_u)]
$$

exists, and similarly that

$$
\lim_{i\to+\infty} \frac{\mathbb{E}_{P_{\mathcal{X}}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]}{\lambda_i} = \mathbb{E}_{P_{\mathcal{X}}}[\phi_u(y_u|X_u)] > 0 .
$$

Furthermore, this latter inequality implies that there is some $n \in \mathbb{N}$ such that for all $i > n$,

$$
\frac{\mathbb{E}_{P_{\mathcal{X}}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]}{\lambda_i} > 0 ,
$$

and hence in particular $\mathbb{E}_{P_{\mathcal{X}}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)] > 0$ for all $i > n$. Furthermore, since by construction $O_u^j \supseteq O_u^{j+1}$ for all $j \in \mathbb{N}$, monotonicity of the measure $P_{\mathcal{Y}|\mathcal{X}}$ implies that $P_{\mathcal{Y}|\mathcal{X}}(O_u^j|x_u) \geq P_{\mathcal{Y}|\mathcal{X}}(O_u^{j+1}|x_u)$ for all $j \in \mathbb{N}$ and all $x_u \in \mathcal{X}_u$. So, for all $j \leq n$ we also have $\mathbb{E}_{P_{\mathcal{X}}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^j|X_u)] > 0$, and so we have found that

$\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)] > 0$ for all $i \in \mathbb{N}$. Due to Proposition 5.1, this implies that each $\mathbb{E}_P[f(X_v)\,|\,Y_u \in O_u^i]$ is well-defined.

By the limit definition of $\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u]$, and applying Proposition 5.1 to each step,

$$\begin{aligned}
\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u] &= \lim_{i\to+\infty} \mathbb{E}_P[f(X_v)\,|\,Y_u \in O_u^i] \\
&= \lim_{i\to+\infty} \frac{\mathbb{E}_{P_\mathcal{X}}[f(X_v)P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]}{\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]} \\
&= \lim_{i\to+\infty} \frac{\lambda_i}{\lambda_i} \frac{\mathbb{E}_{P_\mathcal{X}}[f(X_v)P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]}{\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]} \\
&= \lim_{i\to+\infty} \frac{\mathbb{E}_{P_\mathcal{X}}[f(X_v)P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]/\lambda_i}{\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]/\lambda_i} \\
&= \frac{\lim_{i\to+\infty} \mathbb{E}_{P_\mathcal{X}}[f(X_v)P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]/\lambda_i}{\lim_{i\to+\infty} \mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]/\lambda_i} = \frac{\mathbb{E}_{P_\mathcal{X}}[f(X_v)\phi_u(y_u|X_u)]}{\mathbb{E}_{P_\mathcal{X}}[\phi_u(y_u|X_u)]},
\end{aligned}$$

using the above established existence and properties of the individual limits for the penultimate step. □

Proof. **of Proposition 5.4** This is a special case of Corollary A.5, obtained by setting $g(X_u) \coloneqq \phi_u(y_u|X_u)$ in that corollary's statement, and applying Proposition 5.3 to the quantities $\mathbb{E}_P[f(X_v)\,|\,Y_u = y_u]$. □

Proof. **of Proposition 5.5** Assume that there is a sequence $\{\lambda_i\}_{i\in\mathbb{N}}$ such that $\phi_u(y_u|X_u)$ exists, is real-valued, and satisfies $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] > 0$.

Note that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)] > 0$ implies that $\mathbb{E}_{P_\mathcal{X}}[\phi_u(y_u|X_u)] > 0$ for all $P_\mathcal{X} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$. Furthermore, for any $P_\mathcal{X} \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$, because by assumption $\lim_{i\to+\infty} P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)/\lambda_i = \phi_u(y_u|X_u)$, it follows from Lemma A.1 that

$$\lim_{i\to+\infty} \frac{\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)]}{\lambda_i} = \mathbb{E}_{P_\mathcal{X}}[\phi_u(y_u|X_u)] > 0\,.$$

This implies that there is some $n \in \mathbb{N}$ such that for all $j > n$,

$$\frac{\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^j|X_u)]}{\lambda_j} > 0\,,$$

which implies that also $\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^j|X_u)] > 0$. Furthermore, since $O_u^i \supseteq O_u^{i+1}$, we have that $P_{\mathcal{Y}|\mathcal{X}}(O_u^i|x_u) \geq P_{\mathcal{Y}|\mathcal{X}}(O_u^{i+1}|x_u)$ for all $x_u \in \mathcal{X}_u$, by monotonicity of measure. It follows that also for all $k \leq n$

$$\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^k|X_u)] \geq \mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^j|X_u)] > 0\,.$$

Hence, we have found that $\mathbb{E}_{P_\mathcal{X}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)] > 0$ for all $i \in \mathbb{N}$. Since $P \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}$, it now follows that $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)] = \overline{P}_\mathcal{Z}(Y_u \in O_u^i) > 0$ for all $i \in \mathbb{N}$ by Lemma A.7.

Now define the sequence $\{\mu_i\}_{i\in\mathbb{N}}$ as $\mu_i \coloneqq \underline{\mathbb{E}}_\mathcal{Z}[f(X_v)\,|\,Y_u \in O_u^i]$, for all $i \in \mathbb{N}$. Fix any $i \in \mathbb{N}$. Then, because $\overline{P}_\mathcal{Z}(Y_u \in O_u^i) > 0$, it follows from Definition 3 that $\mu_i \in [\min f, \max f]$ and furthermore, by Proposition 5.2, that

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)\big(f(X_v) - \mu_i\big)] = 0\,.$$

Therefore, and by the non-negative homogeneity of lower expectations, it also holds that

$$\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\frac{P_{\mathcal{Y}|\mathcal{X}}(O_u^i|X_u)}{\lambda_i}\big(f(X_v) - \mu_i\big)\right] = 0\,. \tag{D.1}$$

Let now $\{\mu_{i_k}\}_{k\in\mathbb{N}}$ be any convergent subsequence; since the sequence $\{\mu_i\}_{i\in\mathbb{N}}$ is in the compact interval $[\min f, \max f]$, the Bolzano-Weierstrass theorem implies that at least one such subsequence exists. Let $\mu_* \coloneqq \lim_{k\to+\infty} \mu_{i_k}$.

We now clearly have that

$$\lim_{k\to+\infty} \frac{P_{\mathcal{Y}|\mathcal{X}}(O_u^{i_k}|X_u)}{\lambda_{i_k}}\big(f(X_v)-\mu_{i_k}\big) = \phi_u(y_u|X_u)\big(f(X_v)-\mu_*\big)\,,$$

and therefore, by Lemma A.2, that

$$\lim_{k\to+\infty} \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\frac{P_{\mathcal{Y}|\mathcal{X}}(O_u^{i_k}|X_u)}{\lambda_{i_k}}\big(f(X_v)-\mu_{i_k}\big)\right] = \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\phi_u(y_u|X_u)\big(f(X_v)-\mu_*\big)\right]\,.$$

Furthermore, using Equation (D.1), we find that $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\phi_u(y_u|X_u)\big(f(X_v)-\mu_*\big)\right]=0$.

Due to Proposition 6.1, and because $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)]>0$, we conclude that μ_* corresponds to the unique root of the function $G(\mu) \coloneqq \underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\phi_u(y_u|X_u)\big(f(X_v)-\mu\big)\right]$. Furthermore, since the convergent subsequence $\{\mu_{i_k}\}_{k\in\mathbb{N}}$ was arbitrary, we find that μ_* is the limit of *every* convergent subsequence of $\{\mu_i\}_{i\in\mathbb{N}}$.

We next show that $\{\mu_i\}_{i\in\mathbb{N}}$ itself also converges to μ_*. Assume *ex absurdo* that this is false. Then, there is some $\epsilon>0$ such that, for all $n\in\mathbb{N}$, there is some $k>n$ such that $|\mu_k-\mu_*|\geq\epsilon$. This implies that we can construct a subsequence $\{\mu_{i_k}\}_{k\in\mathbb{N}}$ such that $|\mu_{i_k}-\mu_*|\geq\epsilon$ for all $k\in\mathbb{N}$. This subsequence is again in the compact interval $[\min f,\max f]$, which implies that it has a convergent subsequence $\{\mu_{i_{k_\ell}}\}_{\ell\in\mathbb{N}}$, and clearly $\lim_{\ell\to+\infty}\mu_{i_{k_\ell}}\neq\mu_*$ because $\left|\mu_{i_{k_\ell}}-\mu_*\right|\geq\epsilon$ for all $\ell\in\mathbb{N}$. However, since $\{\mu_{i_{k_\ell}}\}_{\ell\in\mathbb{N}}$ is a convergent subsequence of the original sequence $\{\mu_i\}_{i\in\mathbb{N}}$, this contradicts our above conclusions. Hence, we must have that indeed $\lim_{i\to\infty}\mu_i=\mu_*$.

Since we already know that μ_* is the unique root of the function $G(\mu)$ defined above, and since this function is strictly decreasing due to Proposition 6.1, we must have that

$$\mu_* = \max\{\mu\in\mathbb{R}\,:\,G(\mu)\geq 0\}\,.$$

Therefore, due to Proposition 5.4, we conclude that

$$\begin{aligned}\underline{\mathbb{E}}_{\mathcal{Z}}[f(X_v)\,|\,Y_u=y_u] &= \max\left\{\mu\in\mathbb{R}\,:\,\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}\left[\phi_u(y_u|X_u)\big(f(X_v)-\mu\big)\right]\geq 0\right\}\\ &= \mu_* = \lim_{i\to+\infty}\mu_i = \lim_{i\to+\infty}\underline{\mathbb{E}}_{\mathcal{Z}}[f(X_v)\,|\,Y_u\in O_u^i]\,.\end{aligned}$$

□

The following proposition proves, as claimed in the main text, that the first of the two alternative imprecise updating methods that were suggested in Section 5.2 also satisfies the generalised Bayes' rule.

Proposition D.1. *Let $\mathcal{Z}$ be an ICTHMC 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 $\overline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}[\phi_u(y_u|X_u)]>0$, then the model defined by*

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

satisfies

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

Proof. This is a special case of Corollary A.5, obtained by setting $g(X_u)\coloneqq\phi_u(y_u|X_u)$ in that corollary's statement, and applying Proposition 5.3 to the quantities $\mathbb{E}_P[f(X_v)\,|\,Y_u=y_u]$. □