Proceedings of Machine Learning Research 215:450–460, 2023 ISIPTA 2023

A Nonstandard Approach To Stochastic Processes Under Probability Bounding

Matthias C. M. Troffaes MATTHIAS.TROFFAES@DURHAM.AC.UK
Department of Mathematical Sciences, Durham University, UK

Abstract

This paper studies stochastic processes under probability bounding, using nonstandard conditional lower previsions within the framework of internal set theory. Following Nelson's approach to stochastic processes, we introduce elementary processes which are defined over a finite number of time points and that serve to approximate any standard process, including processes over continuous time. We show that every standard process can be represented by an elementary process, and that the shadow of every elementary process constitutes again a standard process. We then move to demonstrate how elementary processes can be used to define imprecise Markov chains both in discrete and continuous time. To demonstrate the benefits and downsides of this approach, we show how to recover some basic results for continuous time Markov chains through analysis of a nonstandard elementary process.

Keywords: imprecise Markov chain, stochastic process, lower prevision, internal set theory, nonstandard analysis

1. Introduction

The theory of imprecise Markov chains has been an active area of research and applications for nearly two decades [15, 2, 13, 3, 1, 6]. Moving from discrete time to continuous time however presents an analytical challenge, similar to the challenges encountered for precise stochastic processes. Nelson [9] developed an approach to stochastic processes based on internal set theory, which simplifies the analysis of continuous time, by introducing the notion of nearby stochastic processes, which comprise nonstandard discrete representations of continuous time processes that can be used to study continuous time processes in a far more algebraic manner, with fewer technicalities involved, at the expense of nonstandard logic.

This paper aims to explore Nelson's ideas as applied to stochastic processes described by probability bounding, or more precisely, by Williams coherent lower previsions. We introduce a Nelson-inspired approach to representing arbitrary imprecise stochastic processes by means of elementary (discrete) ones. As applications, we state a basic definition for imprecise Markov chains which covers both discrete and continuous time simultaneously, and we recover a standard result from imprecise continuous time Markov chains using simple algebraic means via Nelson's nonstandard logic. In doing so, we find (see Definition 17) that imprecise continuous time Markov chains can be defined as an envelope of precise processes that need not be 'well behaved' (with 'well behaved' in the sense of [6]), thereby making a modest generalisation to the existing theory.

Section 2 gives a brief introduction to internal set theory. Section 3 states a few useful results on convergence within internal set theory. Section 4 defines stochastic processes, and Section 5 studies discrete representations of such processes. Section 6 briefly touches upon how convergence within such discrete representations can be handled. Section 7 applies the theory to imprecise Markov chains, with Section 8 focusing on the continuous case, where we recover the well-known expression for the expectation in terms of the exponential of the lower rate operator. Section 9 concludes.

2. Internal Set Theory

In this section, we give a brief introduction to internal set theory. For a more in-depth treatment, we refer to for instance [8, 9, 4, 10].

2.1. Basic Concepts

In a nutshell, internal set theory:

- introduces a new predicate for sets, called 'standard' (so, each set can be either standard or not; one could imagine that each set now carries a colour, 'standard' or 'not standard'), and
- extends ZFC (the axioms of Zermelo-Fraenkel set theory plus the axiom of choice) by three new axiom schemata which govern the use of the predicate 'standard'.

Formulae that use 'standard' (such as "there is a standard natural number") are called *external*, and those that do not (such as "one is larger than zero") are called *internal*. All ZFC axioms are internal formulae, and all of mathematics that follows from ZFC is internal.

We note that, in ZFC, every object is a set, including numbers, functions, etc. For instance, a function is identified with its graph, i.e. in ZFC, a function $f\colon X \to Y$ is a subset of $X \times Y$ such that for each $x \in X$ there is a unique $y \in Y$ such that $(x, y) \in f$. So, since functions are sets, we can also talk about them being standard or not, and we will do so regularly in what follows.

Finally, we emphasize that the ZFC axiom of specification is not extended to external formulae, as doing so would lead to an inconsistent theory. This means that *we cannot form sets using external formulae*, and doing so is called illegal set formation. For example, we cannot form the set $\{n \in \mathbb{N}\colon n \text{ is standard}\}$, since '$n$ is standard' involves the predicate 'standard' and is therefore external. However, we can use nonstandard parameters to form sets. For example, we can form the set $\{n \in \mathbb{N}\colon n \leq N\}$ for some nonstandard $N \in \mathbb{N}$. This works because the formula $n \leq N$ is internal.

Next, we cover each of the three new axioms, interspersed with some important results that will be used in the remainder of the paper.

2.2. Transfer

We start with the *transfer principle*, which states that for every internal formula $A(x, t_1, \ldots, t_k)$ with free variables $x, t_1, \ldots, t_k$ (and no other free variables) we have that

$$\forall^{\mathrm{s}} t_1 \cdots \forall^{\mathrm{s}} t_k \big(\forall^{\mathrm{s}} x\colon A(x, t_1, \ldots, t_k) \implies \forall x\colon A(x, t_1, \ldots, t_k)\big) \quad (1)$$

Here, we used '$\forall^{\mathrm{s}} x \ldots$' as an abbreviation for '$\forall x (x \text{ standard} \implies \ldots)$'. We also have a dual version of transfer:

$$\forall^{\mathrm{s}} t_1 \cdots \forall^{\mathrm{s}} t_k \big(\exists x\colon A(x, t_1, \ldots, t_k) \implies \exists^{\mathrm{s}} x\colon A(x, t_1, \ldots, t_k)\big) \quad (2)$$

where we used '$\exists^{\mathrm{s}} x \ldots$' as an abbreviation for '$\exists x (x \text{ standard} \wedge \ldots)$'.

For example, for every standard function f and standard x in the domain of f, it follows that $f(x)$ is standard. Indeed,

$$\forall f\colon X \to Y \, \forall x \in X \, \exists y \in Y\colon f(x) = y \quad (3)$$

implies, by transfer (provided X and Y are standard),

$$\forall^{\mathrm{s}} f\colon X \to Y \, \forall^{\mathrm{s}} x \in X \, \exists^{\mathrm{s}} y \in Y\colon f(x) = y \quad (4)$$

As yet another important example, if there is a unique x such that some internal formula $A(x, t_1, \ldots, t_k)$ holds, then, provided $t_1, \ldots, t_k$ are standard, it follows that x must be standard. Consequently, all sets used in conventional mathematics, such as $\mathbb{N}$, $\mathbb{R}$, 0, 1, π, log, exp, sin, ... are all standard, because they can be uniquely characterized by internal formulae. For instance, one of the ZFC axioms states that there is a unique set $\emptyset$ such that "$\forall x\colon x \notin \emptyset$", so $\emptyset$ is standard because "$\forall x\colon x \notin \emptyset$" is internal. From the ZFC axioms, one can show that there is a unique set N satisfying

$$N = \cap\{X\colon (\emptyset \in X \wedge (\forall x \in X\colon x \cup \{x\} \in X))\} \quad (5)$$

This unique set N must be standard because this formula is internal and $\emptyset$ is a standard parameter; N is the set of natural numbers $\mathbb{N}$, where 0 is identified with $\emptyset$, 1 with $\emptyset \cup \{\emptyset\}$, and so on [11, §1.47, p. 31]. So we have shown that $\mathbb{N}$ is standard, and that 0 is standard. But we know that every natural number has a unique successor:

$$\forall n \in \mathbb{N}\colon \exists! m \in \mathbb{N}\colon m = n + 1 \quad (6)$$

so by transfer,

$$\forall^{\mathrm{s}} n \in \mathbb{N}\colon \exists!^{s} m \in \mathbb{N}\colon m = n + 1 \quad (7)$$

or in other words, if n is standard then its successor $n + 1$ is also standard.

Interestingly, these observations do not imply that all elements of $\mathbb{N}$ are standard, and therefore do not exclude the existence of nonstandard elements in $\mathbb{N}$. At first, this may seem counterintuitive. However, as we shall see next, we can do this because we do not extend the axiom of specification to external formulae.

2.3. Idealisation

The axiom of *idealisation* is somewhat involved, however for the purpose of this paper, we really only need a few consequences of this axiom [8, Theorems 1.1 & 1.2]. Before we state these consequences, we remind the reader that we say that a set A is *finite* whenever there is a natural number $n \in \mathbb{N}$ and a bijection between A and the set $\{m \in \mathbb{N}\colon m < n\}$ [11, §2.16]. This is the usual definition that most readers will be familiar with, and we want to emphasize here that nothing is changed about what it means for a set to be finite.

Theorem 1 *All elements of a set A are standard if and only if A is a finite standard set.*

Theorem 2 *For every set A, there is a finite subset B of A such that B contains all standard elements of A.*

Remember that every object in ZFC is a set (including numbers, functions, etc.), so every element of a set is also a set, thus it makes sense to talk about the standard elements of a set.

As an obvious consequence, there is a finite subset of $\mathbb{N}$ containing all standard elements of $\mathbb{N}$. Let N denote the maximum of this set. Then $N + 1$ must be nonstandard. In fact, N must also be nonstandard, since if it were standard, we would conclude that its successor $N+1$ would be standard too (by the proof given earlier).

So, if we depict $\mathbb{N}$, we have all standard naturals on the left (starting from 0), and all nonstandard naturals on the right, but we cannot say where the nonstandard naturals start as we are forbidden from forming the largest standard natural. Indeed, 'n is standard' is an external formula, and we are disallowed from forming the set of all standard naturals. At first this seems limiting. However, the idea that we cannot form certain sets turns out to be a very useful fact that can be used in proofs.

Let us now investigate the structure of $\mathbb{R}$ within internal set theory. We say that $x \in \mathbb{R}$ is *infinitesimal*, and write $x \simeq 0$, if for all standard $n \in \mathbb{N}$ we have that $|x| \leq \frac{1}{n}$. We write $x \simeq y$ if $x - y \simeq 0$. We say $x \in \mathbb{R}$ is *unlimited* if for all standard $n \in \mathbb{N}$ we have that $|x| \geq n$. We write $x \sim \infty$ if x is unlimited and positive. We saw that naturals are unlimited if and only if they are nonstandard. However, whilst every unlimited real is nonstandard, as we will see shortly, not all limited reals are standard (for instance, $x + \epsilon$ for standard x and infinitesimal $\epsilon \neq 0$).

Before we move on, now is perhaps a good time to reflect on Theorem 2. This theorem applies to *every* set, regardless of its cardinality: $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{R}^{\mathbb{R}}$, etc. To aid with intuition, 'standard' is sometimes understood to mean 'at any stage within the mathematical discourse, [...] uniquely defined' [4, §1.1.1, p. 2].[1] For example, take the collection of standard rationals to be the collection of all rationals that humanity will ever compute. If we assume humanity will perish one day, then this collection is finite, though unimaginably large. Similarly, the collection of reals that will ever be uniquely characterized by some explicitly written internal formula is finite, since humanity can only write finitely many internal formulae. Idealization abstracts this idea. In general, if a set A is infinite, then one can easily show that any finite set B containing all standard elements of A must necessarily have nonstandard cardinality, i.e. though finite, B will be unimaginably large, as desired.

2.4. Standardisation

Finally, we have the *standardisation principle*. As with idealisation, this axiom is quite involved, and here we will only state its main consequences that we need further. We already said that every unlimited real is nonstandard. To complete the picture of the reals, the following result follows from the standardisation principle [8, Theorem 1.4]:

Theorem 3 *For every limited real x there is a unique standard real y, called the* shadow *of x, such that $x \simeq y$.*

There is also a more general version of this theorem that applies to real-valued functions (the formulation here is a special case of [8, Theorem 1.3]):

Theorem 4 *Assume X is standard, and let $f\colon X \to \mathbb{R}$ be a function (standard or nonstandard) such that for every standard $x \in X$, we have that $f(x)$ is limited. Then there is a unique standard function $f_0\colon X \to \mathbb{R}$, called the* shadow *of f, such that for all standard $x \in X$ we have that $f(x) \simeq f_0(x)$.*

Together with Theorem 2, this theorem permits us to use nonstandard functions on finite domains to represent standard functions on arbitrary domains. We thus arrive at the following generalization of Theorem 4:

Theorem 5 *Assume $X \subseteq X_0$ where X_0 is standard and such that X and X_0 have the same standard elements. Let $f\colon X \to \mathbb{R}$ be a function such that for every standard $x \in X$, we have that $f(x)$ is limited. Then there is a unique standard function $f_0\colon X_0 \to \mathbb{R}$, called the* shadow *of f, such that for all standard $x \in X$ we have that $f(x) \simeq f_0(x)$.*

Proof Define $g\colon X_0 \to \mathbb{R}$ as follows: $g(x) := f(x)$ for all $x \in X$, and $g(x) := 0$ for all $x \in X_0 \setminus X$. Note that $g(x)$ is limited for all standard $x \in X_0$. By Theorem 4, g has a shadow, say f_0, and for this shadow we have that $f(x) = g(x) \simeq f_0(x)$ for all standard $x \in X$.

To prove uniqueness, let $f_0'\colon X_0 \to \mathbb{R}$ be any other standard function such that $\forall^{\mathrm{s}} x \in X\colon f(x) \simeq f_0'(x)$. By the previous part of the proof, we already know that $\forall^{\mathrm{s}} x \in X\colon f(x) \simeq f_0(x)$. So,

$$\forall^{\mathrm{s}} x \in X\colon f_0(x) \simeq f(x) \simeq f_0'(x) \tag{8}$$

but, since f_0 and f_0' are standard, it follows that $f_0(x)$ and $f_0'(x)$ are also standard for all standard $x \in X$, so

$$\forall^{\mathrm{s}} x \in X\colon f_0(x) = f_0'(x) \tag{9}$$

Now note that $\forall^{\mathrm{s}} x \in X$ is the same as $\forall^{\mathrm{s}} x \in X_0$. Since the formula is internal, we can apply transfer to conclude that $f_0(x) = f_0'(x)$ for all $x \in X_0$, establishing uniqueness. ∎

For example, let $X_0 = \mathbb{N}$ and $X = \{0, \dots, 2N\}$ for some nonstandard $N \in \mathbb{N}$. The shadow of the function $f(n) := 1/n$ for $n \in \{0, \dots, N\}$ and $f(n) := n$ for $n \in \{N+1, \dots, 2N\}$ is the function $f_0(n) = 1/n$ for $n \in \mathbb{N}$, because f_0 is standard (as it is defined using an internal formula) and $f(n) = f_0(n)$ for all standard $n \in X$. Note that the shadow of f does not depend on the choice of N: different functions can have the same shadow.

In Section 4 and further, we will use Theorem 5 to represent continuous time processes through discrete time ones.

3. Convergence of Sequences

The next two results are canonical; for instance see [10, Theorems 3.4.1 & 3.4.9(b)].

[1] We might say more precisely, 'uniquely defined using an explicitly written internal formula'.

Lemma 6 *For every standard sequence* $(x_n)_{n\in\mathbb{N}}$,

$$\lim_{n\to\infty} x_n = 0 \iff (\forall n \sim \infty\colon x_n \simeq 0) \tag{10}$$

Lemma 7 (Robinson's Lemma) *For every sequence* $(x_n)_{n\in\mathbb{N}}$,

$$(\forall^{s} n\colon x_n \simeq 0) \iff (\exists N \sim \infty\colon \forall n \leq N\colon x_n \simeq 0) \tag{11}$$

We will also need the next result later, which allows us to study the limit of a standard sequence from a 'nearby' nonstandard finite sequence. It resembles Robinson's lemma (and indeed that lemma is used to prove both directions). It does not appear to be a canonical result from the literature, therefore a proof is provided.

Lemma 8 *Let* $(x_n)_{n\in\mathbb{N}}$ *be a standard sequence, and let* $(y_n)_{n=0}^{N}$ *be a finite sequence for some nonstandard* N *such that*

$$\forall^{s} n\colon x_n \simeq y_n \tag{12}$$

Then

$$\lim_{n\to\infty} x_n = 0 \iff \exists M \sim \infty, M \leq N\colon \forall n \sim \infty, n \leq M\colon y_n \simeq 0 \tag{13}$$

Proof First note that by Lemma 7 applied on the sequence $x_n - y_n$,

$$\exists M' \sim \infty\colon \forall n \leq M'\colon x_n - y_n \simeq 0 \tag{14}$$

'only if'. Assume $\lim_{n\to\infty} x_n = 0$. Then, by Lemma 6,

$$\forall n \sim \infty\colon x_n \simeq 0 \tag{15}$$

By Equation (14), with $M := \min\{M', N\}$, the right-hand side of Equation (13) must hold.

'if'. Assume the right-hand side of Equation (13) holds. By Equation (14), with $M'' := \min\{M', M\}$, we conclude that

$$\forall n \sim \infty, n \leq M''\colon x_n \simeq 0 \tag{16}$$

and consequently,

$$\forall n \leq M''\colon \forall^{s} \epsilon > 0\colon\ (|x_n| > \epsilon \implies n \ll \infty) \tag{17}$$

Fix any standard $\epsilon > 0$, and consider the set

$$\mathcal{N}_\epsilon := \{n \in \{0, \dots, M''\}\colon \forall m \in \{n, \dots, M''\}\colon |x_m| \leq \epsilon\} \tag{18}$$

This set is non-empty by Equation (16), so it has a minimum:

$$n_\epsilon := \min \mathcal{N}_\epsilon \leq M'' \tag{19}$$

and

$$n_\epsilon = 0 \text{ or } |x_{n_\epsilon - 1}| > \epsilon \tag{20}$$

Either way, by Equation (17), we must have that n_ϵ is standard too. So,

$$\forall^{s} \epsilon > 0\colon \exists^{s} n_\epsilon\colon \forall m \in \{n_\epsilon, \dots, M''\}\colon |x_m| \leq \epsilon \tag{21}$$

But $\{n_\epsilon, \dots, M''\}$ contains all standard $m \geq n_\epsilon$, so

$$\forall^{s} \epsilon > 0\colon \exists^{s} n_\epsilon\colon \forall^{s} m \geq n_\epsilon\colon |x_m| \leq \epsilon \tag{22}$$

Now use transfer to conclude that $\lim_n x_n = 0$. ■

In Lemma 8, though we only require that $x_n \simeq y_n$ for standard n, we can say something about the limit of x_n by studying y_n for nonstandard n. This is due to a phenomenon called *overspill* [9, pp. 18-19], and excellently demonstrates the power of nonstandard analysis.

Note that the limiting number 0 in the above results can be replaced by any standard real number z, since $\lim_n x_n = z$ if and only if $\lim_n (x_n - z) = 0$.

4. Stochastic Processes

We are now ready to turn to the main topic of the paper.

Consider a finite *state space* $\mathcal{X}$ (for instance, $\{1, \dots, n\}$ for some natural number n), representing the set of values that a process can take. The state space $\mathcal{X}$ is fixed throughout the paper, and is assumed to be a standard finite set; this implies that its cardinality is standard, and that all of its elements are standard too (see Theorem 1). We also need a totally ordered *index set* $T_0 \subseteq \mathbb{R}_+$, interpreted as a set of time points at which we can observe the state of the process. We assume T_0 to be standard.

Finally, we let Ω denote a *possibility space*. Throughout, we assume that Ω is a standard (though not necessarily finite) set. As with the state space, the possibility space is fixed throughout the paper.

Next, we fix a standard function $\xi\colon T_0 \times \Omega \to \mathcal{X}$. This function describes the time evolution of the process: given outcome ω and time t, the state of the process is $\xi(t, \omega)$.

We aim to study discrete versions of the process, that is, we want to study the process for (usually, finite) subsets T of T_0. Informally, a stochastic process on an index set T is a function $\xi'\colon T \times \Omega \to \mathcal{X}$ where $\xi' = \xi|_T$ along with some description of uncertainty on Ω pertaining to ξ'.

This paper is concerned with stochastic processes under severe uncertainty. Therefore, we will take uncertainty to be described by probability bounds rather than a probability measure. More specifically, we assume uncertainty on Ω to be described by a coherent conditional lower prevision, with coherence in the sense of Williams [16] [17, Section 3] (see Definition 9 further). Such approach includes the usual measure theoretic approaches, and more general approaches that allow finite additivity as well as probability bounding. Before we can formally define this, we need some notation to adequately describe the domain of the conditional lower prevision that we will be working with. For this paper, we keep our structures intentionally simple, as this will ease the nonstandard representation further.

First, let $\mathcal{A}(T)$ denote the algebra (or field) generated by events of the form $\{\xi(t) = x\}$ for $t \in T$ and $x \in \mathcal{X}$. For example, if B is a subset of $\mathcal{X}$ and $t \in T$ then

$$\{\xi(t) \in B\} \coloneqq \{\omega \in \Omega\colon \xi(t, \omega) \in B\} \tag{23}$$

is an element of $\mathcal{A}(T)$, representing the event that the state at time t is in B (note this works for arbitrary $B \subseteq \mathcal{X}$ because $\mathcal{X}$ is finite).

The indicator function of an event A is the function $I_A\colon \Omega \to \{0, 1\}$ which takes value 1 on A and 0 elsewhere. By $\mathcal{L}(T)$ we denote the linear span of the set of indicator functions induced by events in $\mathcal{A}(T)$. An element of $\mathcal{L}(T)$ is called a *gamble on* ξ, or simply a *gamble* if the process is clear from the context. For example, if $g \in \mathbb{R}^{\mathcal{X}}$ then $g(\xi(t))$ is the gamble

$$\sum_{x \in \mathcal{X}} g(x) I_{\{\xi(t)=x\}} \tag{24}$$

Finally, we define

$$\mathcal{K}(T) \coloneqq \mathcal{L}(T) \times (\mathcal{A}(T) \setminus \{\emptyset\}) \tag{25}$$

This domain is simpler but slightly larger than the one considered in [6, Definition 4.3].

The general characterisation of Williams coherence of conditional lower previsions can be found in [17, Propositions 1 & 2]. For the domain considered here, the following properties (also due to [17, p. 370]) are necessary and sufficient; for a proof see for instance [14, Thm. 13.33].

Definition 9 (Coherence) $\underline{\mathbb{E}}\colon \mathcal{K}(T) \to \mathbb{R}$ *is said to be* coherent *if for all* f, $g \in \mathcal{L}(T)$, *all* A, $B \in \mathcal{A}(T)$ *such that* $A \cap B \neq \emptyset$, *and all* $\lambda \geq 0$, *we have that*

- $\underline{\mathbb{E}}(f \mid A) \geq \inf(f \mid A)$
- $\underline{\mathbb{E}}(f + g \mid A) \geq \underline{\mathbb{E}}(f \mid A) + \underline{\mathbb{E}}(g \mid A)$
- $\underline{\mathbb{E}}(\lambda f \mid A) = \lambda \underline{\mathbb{E}}(f \mid A)$
- $\underline{\mathbb{E}}(I_A(f - \underline{\mathbb{E}}(f \mid A \cap B)) \mid B) = 0$

Definition 10 (Self-Conjugacy) *We say that* $\underline{\mathbb{E}}$ *is* self-conjugate *when* $\underline{\mathbb{E}}(f \mid A) = -\underline{\mathbb{E}}(-f \mid A)$ *for all* $(f, A) \in \mathcal{K}(T)$.

It is known that a self-conjugate coherent lower prevision $\underline{\mathbb{E}}$ is an expectation operator induced by a conditional finitely additive probability measure in the sense of Dubins [5, Section 3], and is usually denoted by $\mathbb{E}$ (without the lower bar).

We can now formally define stochastic processes:

Definition 11 *A* (stochastic) process *on* T *is a coherent conditional lower prevision* $\underline{\mathbb{E}}$ *defined on* $\mathcal{K}(T)$. *If* T *is finite, then we say that the process is* elementary. *If* $\underline{\mathbb{E}}$ *is self-conjugate, then we say that the process is* precise.

We will show that, under certain conditions, every elementary process on T can be turned into a standard one on T_0. To do so, first, we prove a simple yet very useful lemma:

Lemma 12 *Assume* $T \subseteq T_0$ *and* T *contains all standard elements of* T_0. *Then,*

$$\forall^{\mathrm{s}}(f, A) \in \mathcal{K}(T_0)\colon (f, A) \in \mathcal{K}(T) \tag{26}$$

Proof Recall that $\mathcal{X}$ is standard and finite, so all elements x of $\mathcal{X}$ are standard too (see Theorem 1).

If $A \in \mathcal{A}(T_0)$ then A can be written through a finite number of unions and complements on sets of the form $\{\xi(t) = x\}$ with $t \in T_0$ and $x \in \mathcal{X}$. If, additionally, A is standard, then, by transfer, A can be expressed through a standard finite number of such unions and complements, where all t (and x) involved in these operations must be standard too. But that means that $t \in T$ for all these sets. Consequently, A is generated from sets of the form $\{\xi(t) = x\}$ with $t \in T$ and $x \in \mathcal{X}$. In other words, $A \in \mathcal{A}(T)$.

Similarly, for $f \in \mathcal{L}(T_0)$ to be standard, by transfer, it must be a standard linear combination of indicator functions of standard events in $\mathcal{A}(T_0)$. Since all those standard events are in $\mathcal{A}(T)$ by the previous part, it follows that $f \in \mathcal{L}(T)$. ■

To satisfy the conditions of Lemma 12, for example, if $T_0 = \mathbb{N}$, we can take $T \coloneqq \{0, 1, \dots, N\}$ for some nonstandard N.

Theorem 13 *Assume* $T \subseteq T_0$ *and* T *contains all standard elements of* T_0. *Let* $\underline{\mathbb{E}}\colon \mathcal{K}(T) \to \mathbb{R}$ *be any process. Then there is a unique standard process* $\underline{\mathbb{E}}_0\colon \mathcal{K}(T_0) \to \mathbb{R}$, *called the* shadow *of* $\underline{\mathbb{E}}$, *satisfying*

$$\forall^{\mathrm{s}}(f, A) \in \mathcal{K}(T_0)\colon \underline{\mathbb{E}}(f \mid A) \simeq \underline{\mathbb{E}}_0(f \mid A) \tag{27}$$

Note that the theorem needs that all standard elements in $\mathcal{K}(T_0)$ belong to $\mathcal{K}(T)$. This is guaranteed by Lemma 12.

Proof Since $|\underline{\mathbb{E}}(f \mid A)| \leq \sup(|f|)$, it follows that $\underline{\mathbb{E}}$ is limited for all standard $(f, A) \in \mathcal{K}(T)$ (because $\sup(|f|)$ is standard, and therefore limited, for standard f). Therefore, there is a unique function $\underline{\mathbb{E}}_0$ satisfying the conditions of the theorem by the standardisation principle (see Theorem 5).

Is the function $\underline{\mathbb{E}}_0$ coherent? Indeed, for all standard f, $g \in \mathcal{L}(T_0)$, all standard A, $B \in \mathcal{A}(T_0)$ such that $A \cap B \neq \emptyset$, and all standard $\lambda \geq 0$, we have[2]

[2] We rely on some properties of infinitesimal calculus; notably for standard a and b, one can show that if $a + \epsilon = b$ for some infinitesimal ϵ then $a = b$, and if $a + \epsilon \geq b$ for some infinitesimal ϵ then $a \geq b$.

- $\underline{\mathbb{E}}_0(f \mid A) \simeq \underline{\mathbb{E}}(f \mid A) \geq \inf(f \mid A)$ so $\underline{\mathbb{E}}_0(f \mid A) \geq \inf(f \mid A)$ since both sides are standard.
- $\underline{\mathbb{E}}_0(f + g \mid A) \simeq \underline{\mathbb{E}}(f + g \mid A) \geq \underline{\mathbb{E}}(f \mid A) + \underline{\mathbb{E}}(g \mid A) \simeq \underline{\mathbb{E}}_0(f \mid A) + \underline{\mathbb{E}}_0(g \mid A)$ so $\underline{\mathbb{E}}_0(f + g \mid A) \geq \underline{\mathbb{E}}_0(f \mid A) + \underline{\mathbb{E}}_0(g \mid A)$ since both sides are standard.
- $\underline{\mathbb{E}}_0(\lambda f \mid A) \simeq \underline{\mathbb{E}}(\lambda f \mid A) = \lambda\underline{\mathbb{E}}(f \mid A) \simeq \lambda\underline{\mathbb{E}}_0(f \mid A)$ so $\underline{\mathbb{E}}_0(\lambda f \mid A) = \lambda\underline{\mathbb{E}}_0(f \mid A)$ since both sides are standard.
- Note that

$$\begin{aligned}&\underline{\mathbb{E}}_0(I_A(f - \underline{\mathbb{E}}_0(f \mid A \cap B)) \mid B) \\ &\qquad \simeq \underline{\mathbb{E}}(I_A(f + \epsilon - \underline{\mathbb{E}}(f \mid A \cap B)) \mid B) \end{aligned} \tag{28}$$

  where $\epsilon := \underline{\mathbb{E}}(f \mid A \cap B) - \underline{\mathbb{E}}_0(f \mid A \cap B) \simeq 0$. It suffices to show that the right hand side of Equation (28) is infinitesimal. Indeed, by coherence

$$\begin{aligned}0 - |\epsilon| &= \underline{\mathbb{E}}(I_A(f - \underline{\mathbb{E}}(f \mid A \cap B)) \mid B) - |\epsilon| && (29)\\ &\leq \underline{\mathbb{E}}(I_A(f + \epsilon - \underline{\mathbb{E}}(f \mid A \cap B)) \mid B) && (30)\\ &\leq \underline{\mathbb{E}}(I_A(f - \underline{\mathbb{E}}(f \mid A \cap B)) \mid B) + |\epsilon| && (31)\\ &= 0 + |\epsilon| && (32)\end{aligned}$$

  and so, the conclusion follows since $\epsilon \simeq 0$.

Note that we used Lemma 12. Apply transfer to conclude that $\underline{\mathbb{E}}_0$ is coherent on all of $\mathcal{K}(T_0)$. ■

The above proof, though short and simple, is one of the main contributions of the paper, as without it we could not construct nonstandard representations of standard processes. Indeed, the shadow of a function generally does not inherit all properties of that function. For example, the shadow of a countably additive probability measure may not retain countable additivity, and only retains finite additivity. In this sense, Nelson's radically elementary probability theory [9] is a theory of finite additivity (even though this fact is not mentioned in [9]). For this reason, it appears that internal set theory is generally not used in nonstandard probability, in favour of other approaches that stay within the realm of countable additivity such as those based on Loeb spaces [7]. However, coherent lower previsions are not bound by countable additivity, and that gives us the flexibility needed to stay within internal set theory.

5. Nearby Elementary Processes

This paper aims to study standard processes by means of 'nearby' (usually, nonstandard) elementary processes. Note that, by idealisation, for any standard T_0, we know there always is a finite (and usually nonstandard) subset T of T_0 which contains all standard elements of T_0, so we can always find a set $T \subseteq T_0$ such that T contains all standard elements of T_0 (see Theorem 2).

There are various ways in which we can define a notion of 'nearby'. Here is one, using the shadow introduced in Theorem 13:

Definition 14 *A standard process $\underline{\mathbb{E}}_0$ on T_0 is said to be* nearby *an elementary process $\underline{\mathbb{E}}$ on T if $T \subseteq T_0$, T contains all standard elements of T_0, and $\underline{\mathbb{E}}_0$ is the shadow of $\underline{\mathbb{E}}$.*

Definition 14 differs from Nelson's definition [9, p. 81] in several ways. Besides the obvious difference that ours applies to coherent conditional lower previsions and not just to probability measures, ours is more restrictive in that we do not allow perturbations of the process ξ itself (for standard indices t). This would only make sense if we studied real-valued processes like Nelson, but we do not in this paper: our processes only take values in a standard finite set. Secondly, we allow perturbations of our uncertainty structure, whereas Nelson does not: Nelson's elementary process has the same probability measure as the standard process (though restricted to a finite algebra). Our definition permits a nonstandard elementary definition of continuous-time imprecise Markov chains further in the paper: in this application, the lower previsions of the induced standard process are only approximately equal to the lower previsions of the elementary process. Despite these differences, we have opted for the same terminology, as the overall intention is identical: study of processes through elementary approximation.

We already know that every elementary process has a unique nearby standard process (namely, its shadow). Conversely, every standard process also has a nearby elementary process:

Theorem 15 *If $\underline{\mathbb{E}}_0$ is a standard process on T_0, T is finite with $T \subseteq T_0$, and T contains all standard elements of T_0, then $\underline{\mathbb{E}}_0|_{\mathcal{K}(T)}$ is an elementary process nearby $\underline{\mathbb{E}}_0$.*

Proof Using Lemma 12, it is immediately verified that $\underline{\mathbb{E}} := \underline{\mathbb{E}}_0|_{\mathcal{K}(T)}$ satisfies the condition of Theorem 13. ■

By transfer, every internal statement that holds for all standard processes extends to all processes. Therefore, it suffices to study standard processes only. But every standard process has nearby elementary processes, and every elementary process determines a standard process. *So, it suffices to study elementary processes only.*

As these processes are defined on a finite index set T, their study is much easier. For instance, if for any $x \in \mathcal{X}^T$ and $X \subseteq \mathcal{X}^T$, we use the notation

$$\{\xi = x\} := \{\omega \in \Omega\colon (\forall t \in T)(\xi(t)(\omega) = x(t))\} \tag{33}$$

$$\{\xi \in X\} := \bigcup_{x \in X} \{\xi = x\} \tag{34}$$

then, since T is finite,

$$\mathcal{A}(T) = \{\{\xi \in X\} : X \subseteq \mathcal{X}^T\} \tag{35}$$

$$\mathcal{L}(T) = \left\{ \sum_{x \in \mathcal{X}^T} g(x) I_{\xi = x} : g : \mathcal{X}^T \to \mathbb{R} \right\} \tag{36}$$

So, for an elementary process ξ, events correspond to arbitrary sets of paths (of which there are only finitely many), and gambles correspond to arbitrary real-valued functions of paths. Because $\mathcal{A}(T)$ forms a finite algebra, $\mathcal{L}(T)$ forms a finite dimensional vector space. Obviously, this comes at the expense of the more complicated nonstandard logic.

> **For the remainder of this paper, $\underline{\mathbb{E}}_0$ will denote a standard process and $\underline{\mathbb{E}}$ will denote a nearby elementary process.**

6. Convergence of Processes

As a simple example, let us show how convergence of a standard process with index set $T_0 = \mathbb{N}$ translates to a nearby elementary process. All the hard work for this has been done in Lemma 8.

Theorem 16 *Let $T_0 = \mathbb{N}$ and $T = \{0, 1, \dots, N\}$ for some nonstandard $N \in \mathbb{N}$. Let $\underline{\mathbb{E}}_*$ be a standard coherent lower prevision on $\mathbb{R}^{\mathcal{X}}$. Then for all $f \in \mathbb{R}^{\mathcal{X}}$ and all $x \in \mathcal{X}$ we have that*

$$\lim_{n \to \infty} \underline{\mathbb{E}}_0(f(\xi(n)) \mid \xi(0) = x) = \underline{\mathbb{E}}_*(f) \tag{37}$$

if and only if for all standard $f \in \mathbb{R}^{\mathcal{X}}$ and all $x \in \mathcal{X}$,

$$\exists M \in T, M \sim \infty \colon \forall n \in T, n \sim \infty, n \leq M \colon \underline{\mathbb{E}}(f(\xi(n)) \mid \xi(0) = x) \simeq \underline{\mathbb{E}}_*(f) \tag{38}$$

Proof Immediate from Lemma 8, since for all standard $n \in \mathbb{N}$ we have that

$$\underline{\mathbb{E}}(f(\xi(n)) \mid \xi(0) = x) \simeq \underline{\mathbb{E}}_0(f(\xi(n)) \mid \xi(0) = x) \tag{39}$$

■

To make a statement for more general T_0, for instance, one can use a standard sequence $(t_n)_{n \in \mathbb{N}}$ in T_0 and replace $\xi(n)$ with $\xi(t_n)$ in the above theorem.

As another quick example, consider a (precise) discrete time Markov chain with transition matrix $P \simeq P_0$ for some standard stochastic matrix P_0. Then, $\lim_{n \to \infty} P_0^n$ converges to a standard stochastic matrix Π if and only if there is an $M \sim \infty$ such that $P^n \simeq \Pi$ (element-wise) for all $n \sim \infty$, $n \leq M$. So, we can use nonstandard powers of P to study the convergence properties of P_0.

7. Imprecise Markov Chains

We now assume that:

- $T_0 \subseteq \mathbb{R}$, with $\min T_0 = 0$, and
- $\Omega = \mathcal{X}^{T_0}$ with $\xi(t, \omega) := \omega(t)$ (i.e. we work in the so-called canonical representation).

As before, T is a finite subset of T_0 containing all standard elements of T_0. In particular, $0 \in T$. For any function ϕ on T and $t \in T$, by $\phi(0 : t)$ we denote the restriction of ϕ to $[0, t] \cap T$. Define $T' := T \setminus \{\max T\}$. If $t \in T'$ then $t + dt$ denotes the successor of t in T, i.e.

$$dt := \min\{t' \in T : t' > t\} - t \tag{40}$$

Let $\underline{\mathbb{I}}$ be a coherent lower prevision on $\mathbb{R}^{\mathcal{X}}$. For each x in $\mathcal{X}$, and $t \in T'$, let $\underline{\mathbb{T}}_t(\cdot)(x)$ be a coherent lower prevision on $\mathbb{R}^{\mathcal{X}}$, or in other words, for each $t \in T'$, let $\underline{\mathbb{T}}_t$ be a *lower transition operator*. So, $\underline{\mathbb{I}} \colon \mathbb{R}^{\mathcal{X}} \to \mathbb{R}$ and $\underline{\mathbb{T}} \colon T' \to (\mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}})$.

Definition 17 *We say that a precise elementary process $\mathbb{E}$ on T is* compatible *with $(\underline{\mathbb{I}}, \underline{\mathbb{T}})$ if for all paths $x \colon T' \to \mathcal{X}$, all $t \in T'$, and all $f \in \mathbb{R}^{\mathcal{X}}$,*

$$\mathbb{E}(f(\xi(0)) \geq \underline{\mathbb{I}}(f) \tag{41}$$

$$\mathbb{E}(f(\xi(t + dt)) \mid \xi(0 : t) = x(0 : t)) \geq \underline{\mathbb{T}}_t(f)(x(t)) \tag{42}$$

If $\underline{\mathbb{E}}$ denotes the lower envelope of all these compatible precise elementary processes, then $\underline{\mathbb{E}}$ is called the elementary imprecise Markov chain *induced by $(\underline{\mathbb{I}}, \underline{\mathbb{T}})$. Its shadow is called the* imprecise Markov chain *induced by $(\underline{\mathbb{I}}, \underline{\mathbb{T}})$.*

The definition of compatibility is identical to the one given for imprecise discrete time Markov chains in [2, p. 605] (here formulated in terms of expectations instead of probability mass functions). The only new element in the above definition is that we also consider the shadow of elementary imprecise Markov chains to be an imprecise Markov chain. Besides the shadow, we emphasize that all other parts of the definition are completely internal as T can be any arbitrary finite set. In particular, the formula for the set of compatible processes is internal, so we are not committing illegal set formation, and we can take the lower envelope.

This single definition encompasses both continuous and discrete time imprecise Markov chains. For discrete time, take the above definition with $T = \{0, 1, \dots, N\}$ for some nonstandard natural number N, whereas for continuous time, take T to be a finite subset of $\mathbb{R}_+$ containing all standard $t \geq 0$.

In continuous time, the definition allows for processes whose shadow depends on the details of how T approximates T_0, as in the following example.

Example 1 *Let $T := \{t_0, t_1, \dots, t_N\}$ contain all standard non-negative reals, and let $\mathcal{X} := \{0,1\}$. Set*

$$\mathbb{E}(f(\xi(0)) := f(0) \tag{43}$$
$$\mathbb{E}(f(\xi(t+dt)) \mid \xi(0:t) = x(0:t)) := f(1 - x(t)) \tag{44}$$

This models a precise (even deterministic) process which starts at state 0 and then changes state at every time point in T. The state of the system at time 1 therefore depends on the details of T. However, it is compatible with $(\underline{\mathbb{I}}, \underline{\mathbb{T}})$ where:

$$\underline{\mathbb{I}}(f) := \inf f \tag{45}$$
$$\underline{\mathbb{T}}_t(f)(x) := \inf f \tag{46}$$

Is this problematic? Whilst many of the compatible precise processes $\mathbb{E}$ may depend on unspecified details of T, since our inferences about the lower envelope $\underline{\mathbb{E}}$ will be expressed in terms of $\underline{\mathbb{I}}$ and $\underline{\mathbb{T}}$, there is no issue for us to allow such precise processes, as long as $\underline{\mathbb{I}}$ and $\underline{\mathbb{T}}$ are sufficiently regular so that the shadow of $\underline{\mathbb{E}}$ does not depend on how T was constructed from T_0. One way of doing so will be described in Section 8.

In the literature, imprecise continuous time Markov chains have been defined as the envelope of 'well behaved' processes [6, Definition 6.4]. Could we, in principle, restrict the envelope to processes whose shadow is 'well behaved' and thereby, hopefully, remove the issue seen in the above example? Whilst this would allow us to study the connection with standard definitions of continuous time Markov chains, the nonstandard definition here leads to the same inferences found in the literature, at least for the events and gambles that we are studying here. So, although there might be a way to make a strong connection, it complicates the analysis for no immediate benefit, and we will not further explore this connection here. One of the contributions of this paper is that there is no need to restrict the precise processes in the envelope to well-behaved processes when dealing with imprecise continuous time Markov chains.

Because an elementary imprecise Markov chain is simply a non-homogeneous finite horizon discrete time imprecise Markov chain indexed by T, it follows from the usual discrete time theory (in the canonical representation) that the bounds are coherent, in the sense that [2, Section 3]:

$$\underline{\mathbb{E}}(f(\xi(0)) = \underline{\mathbb{I}}(f) \tag{47}$$
$$\underline{\mathbb{E}}(f(\xi(t+dt)) \mid \xi(0:t) = x(0:t)) = \underline{\mathbb{T}}_t(f)(x(t)) \tag{48}$$

For an arbitrary gamble $g \colon \mathcal{X}^T \to \mathbb{R}$, we can recursively calculate the following for all paths $(x_0, \dots, x_{\max T})$ and all $t \in T'$ [2, Section 3]:

$$g_{\max T}(x_0, \dots, x_{\max T}) := g(x_0, \dots, x_{\max T}) \tag{49}$$
$$g_t(x_0, \dots, x_t) := \underline{\mathbb{T}}_t(g_{t+dt}(x_0, \dots, x_t, X_{t+dt}))(x_t) \tag{50}$$

where $g_{t+dt}(x_0, \dots, x_t, X_{t+dt})$ is considered as a gamble on $x_{t+dt} \in \mathcal{X}$ for fixed $x_0, \dots, x_t$. One can show that

$$\underline{\mathbb{E}}(g(\xi) \mid \xi(0) = x) = g_0(x) \tag{51}$$

and consequently [2, Eq. (21)],

$$\underline{\mathbb{E}}(f(\xi(t)) \mid \xi(0) = x) = \left(\prod_{s<t} \underline{\mathbb{T}}_s\right)(f)(x) \tag{52}$$

where the product considers $s \in T'$ in increasing order (and $s < t$).

As an important special case, if $T = \{0, 1, \dots, N\}$ and $\underline{\mathbb{T}}_n$ does not depend on n, then for all $f \in \mathcal{X}$ and $n \in T$,

$$\underline{\mathbb{E}}(f(\xi(n)) \mid \xi(0) = x) = \underline{\mathbb{T}}_0^n(x)(f) \tag{53}$$

where $\underline{\mathbb{T}}_0^n$ denotes the n-th power of $\underline{\mathbb{T}}_0$.

8. Imprecise Continuous Time Markov Chains

Let now $T_0 = \mathbb{R}_+$. As before, let T be any finite subset of T_0 containing all standard elements of T_0. We must have that $dt \simeq 0$ for all limited $t \in T'$. Indeed, if dt were non-infinitesimal, then the open interval $(t, t+dt)$ would necessarily contain the shadow of $t + \max\{dt/2, 1\}$, contradicting our assumption that T contains all standard elements of T_0. Without loss of generality, we can assume that $dt \simeq 0$ for all $t \in T'$ (i.e. even unlimited t). Indeed, if T does not satisfy this requirement, simply pick any natural number $N \geq \max T$ (N is necessarily nonstandard since T contains all standard reals), and add all rationals of the form n/N for $0 \leq n \leq N^2$. This set is still finite and leaves no non-infinitesimal gaps between points.

Let I denote the identity operator on $\mathbb{R}^{\mathcal{X}}$.

Definition 18 *The* lower rate operator $\underline{\mathbb{Q}} \colon T' \to (\mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}})$ *induced by $\underline{\mathbb{T}} \colon T' \to (\mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}})$ is defined as:*

$$\underline{\mathbb{Q}}_t := \frac{\underline{\mathbb{T}}_t - I}{dt} \tag{54}$$

We are now interested in the case where $\underline{\mathbb{Q}}_t$ is standard and does not depend on t; we will simply write $\underline{\mathbb{Q}}$ and omit the time index. Note that $\underline{\mathbb{T}}_t$ may still depend on t, as the points in T cannot be chosen in an equidistant manner (doing so would prevent T from covering all standard elements of $\mathbb{R}_+$). Another way of saying this is that we fix a standard function $\underline{\mathbb{Q}} \colon \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ and define

$$\underline{\mathbb{T}}_t := I + dt\underline{\mathbb{Q}} \tag{55}$$

Because $dt \simeq 0$ for all $t \in T'$, and $\underline{\mathbb{Q}}$ is standard, it follows that $\underline{\mathbb{T}}_t(\cdot)(x)$ is a coherent lower prevision for every $t \in T'$ and $x \in \mathcal{X}$, provided $\underline{\mathbb{Q}}$ satisfies the properties listed in the following definition:

Definition 19 *A function* $\underline{\mathbb{Q}}\colon \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ *is called a* lower rate operator *whenever*

$$\forall \mu \in \mathbb{R}\colon \underline{\mathbb{Q}}(\mu) = 0 \tag{56}$$
$$\forall x, y \in \mathcal{X}, x \neq y\colon \underline{\mathbb{Q}}(I_y) \geq 0 \tag{57}$$
$$\forall f, g \in \mathbb{R}^{\mathcal{X}}\colon \underline{\mathbb{Q}}(f+g) \geq \underline{\mathbb{Q}}(f) + \underline{\mathbb{Q}}(g) \tag{58}$$
$$\forall f \in \mathbb{R}^{\mathcal{X}}\colon \forall \lambda \geq 0\colon \underline{\mathbb{Q}}(\lambda f) = \lambda \underline{\mathbb{Q}}(f) \tag{59}$$

The proof of the following result is an easy exercise (for instance, see [1, Propositions 5 & 6] and note that the norm of a standard $\underline{\mathbb{Q}}$ is standard).

Theorem 20 *A standard function* $\underline{\mathbb{Q}}\colon \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ *is a lower rate operator if for at least one infinitesimal* $\delta > 0$ *we have that* $I + \delta\underline{\mathbb{Q}}$ *is a lower transition operator. In such case,* $I + \delta\underline{\mathbb{Q}}$ *is a lower transition operator for all infinitesimal* $\delta > 0$.

With this notation, for all $t \in T'$ we have

$$\underline{\mathbb{E}}(f(\xi(t+dt)) \mid \xi(0:t) = x(0:t)) = (I + dt\underline{\mathbb{Q}})(f)(x(t)) \tag{60}$$

and

$$\underline{\mathbb{E}}(f(\xi(t)) \mid \xi(0) = x) = \left(\prod_{s<t}(I + ds\underline{\mathbb{Q}})\right)(x)(f) \tag{61}$$

Definition 21 *For any non-negative* $t \in \mathbb{R}$ *and lower rate operator* $\underline{\mathbb{Q}}\colon \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$*, the* exponential *of* $t\underline{\mathbb{Q}}\colon \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ *is:*

$$e^{t\underline{\mathbb{Q}}}(f)(x) := \lim_{n\to\infty}\left(\left(I + \frac{1}{n}t\underline{\mathbb{Q}}\right)^n (f)(x)\right) \tag{62}$$

It can be shown that the above limit always exists, and that $e^{t\underline{\mathbb{Q}}}$ is a lower transition operator; see for instance [1, Section 6].

We now need just three simple algebraic results to establish our main result for this section. First,

Lemma 22 *For all non-negative* $t_1, \dots, t_n$,

$$\prod_{i=1}^n e^{t_i\underline{\mathbb{Q}}} = e^{\sum_{i=1}^n t_i\underline{\mathbb{Q}}} \tag{63}$$

Proof This is the semi-group property; see for instance [1, Eq. (10)] or [6, Proposition 7.13]. For fun, we give a quick sketch of a nonstandard proof, for the case where all t_i are rational. We may assume that all $t_i > 0$. By transfer, we may assume that n and $t_1, \dots, t_n$ are standard. Pick any non-standard natural number N such that all Nt_i are natural. Define $k_i := Nt_i$. Note that all $k_i \sim \infty$. Then, since n is standard, and by Lemma 6,

$$\prod_{i=1}^n e^{t_i\underline{\mathbb{Q}}} \simeq \prod_{i=1}^n \left(I + \frac{1}{k_i}t_i\underline{\mathbb{Q}}\right)^{k_i} = \prod_{i=1}^n \left(I + \frac{1}{N}\underline{\mathbb{Q}}\right)^{Nt_i} \tag{64}$$

$$= \left(I + \frac{1}{N}\underline{\mathbb{Q}}\right)^{N\sum_{i=1}^n t_i} = \left(I + \frac{\sum_{i=1}^n t_i}{M}\underline{\mathbb{Q}}\right)^M \tag{65}$$

$$\simeq e^{\sum_{i=1}^n t_i\underline{\mathbb{Q}}} \tag{66}$$

where $M = N\sum_{i=1}^n t_i \sim \infty$. The case for non-rational t_i follows from a continuity argument (omitted here, but this can also be done using nonstandard arguments). ■

For any (non-negatively homogeneous) operator $\Phi\colon \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ (such as a lower transition or lower rate operator), let $\|\Phi\|$ denote its operator norm:

$$\|\Phi\| := \sup\{|\Phi(f)|\colon f \in [-1,1]^{\mathcal{X}}\} \tag{67}$$

Since this definition is internal, $\|\Phi\|$ is standard whenever Φ is. Note that $\|\underline{\mathbb{T}}\| = 1$ for lower transition operators $\underline{\mathbb{T}}$ [6, LT4], and $\|\underline{\mathbb{Q}}\| < +\infty$ for lower rate operators $\underline{\mathbb{Q}}$ [6, LR5].

Lemma 23 *For all* $n \in \mathbb{N}$ *and all infinitesimal* $\delta > 0$,

$$\left\| I + \delta\underline{\mathbb{Q}} - \left(I + \frac{1}{n}\delta\underline{\mathbb{Q}}\right)^n \right\| \leq \delta^2\|\underline{\mathbb{Q}}\|^2 \tag{68}$$

Proof This is a special case of [6, Lemma E.5]. ■

Lemma 24 *For any two finite sequences of lower transition operators* $T_1, \dots, T_n$ *and* $S_1, \dots, S_n$,

$$\left\| \prod_{i=1}^n T_i - \prod_{i=1}^n S_i \right\| \leq \sum_{i=1}^n \|T_i - S_i\| \tag{69}$$

Proof See [6, Lemma E.4]. ■

We can now formulate and prove the following key result:

Theorem 25 *Assume* $\underline{\mathbb{Q}}$ *is a standard lower rate matrix. Then for all* $x \in \mathcal{X}$*,* $f \in \mathbb{R}^{\mathcal{X}}$*, and* $t \geq 0$,

$$\underline{\mathbb{E}}_0(f(\xi(t)) \mid \xi(0) = x) = e^{t\underline{\mathbb{Q}}}(f)(x) \tag{70}$$

Proof The statement is trivial for $t = 0$. Assume $t > 0$. By transfer, we only need to establish the equality for standard t. By Equation (61), it suffices to show that

$$\left\| \prod_{s<t}(I + ds\underline{\mathbb{Q}}) - e^{t\underline{\mathbb{Q}}} \right\| \simeq 0 \tag{71}$$

Indeed, by Lemma 22,

$$\left\| \prod_{s<t}(I + ds\underline{\mathbb{Q}}) - e^{t\underline{\mathbb{Q}}} \right\| = \left\| \prod_{s<t}(I + ds\underline{\mathbb{Q}}) - \prod_{s<t} e^{ds\underline{\mathbb{Q}}} \right\| \tag{72}$$

and so by Lemma 24,

$$\leq \sum_{s<t} \left\| I + ds\underline{\mathbb{Q}} - e^{ds\underline{\mathbb{Q}}} \right\| \tag{73}$$

and so by Lemma 23 with $n \sim \infty$, together with Definition 21 and Lemma 6,

$$\lesssim \sum_{s<t} (ds)^2 \|\underline{\mathbb{Q}}\|^2 \tag{74}$$

and with $\delta := \max_{s<t} ds$,

$$\leq \sum_{s<t} (ds)\delta \|\underline{\mathbb{Q}}\|^2 = t\delta \|\underline{\mathbb{Q}}\|^2 \simeq 0 \tag{75}$$

since t and $\|\underline{\mathbb{Q}}\|$ are standard and therefore limited, and $\delta \simeq 0$. ∎

Note the final bound in the above proof closely resembles the right hand side of [6, Lemma E.6], to no surprise.

So, we have shown that an elementary stochastic process with infinitesimal time step δ and standard lower rate operator $\underline{\mathbb{Q}}$ can capture key behaviour of continuous-time Markov chains, using mostly simple algebraic means.

9. Conclusion

We have explored an application of internal set theory to imprecise stochastic processes. It is known that every internal result that can be proved within internal set theory can also be proved within ZFC, therefore there are no fundamentally new mathematical insights to be gained [8]. Yet, whilst the axioms of internal set theory are unusual and take some time to digest, the theory has allowed us to simplify definitions and proofs concerning imprecise stochastic processes. We coincidentally (and, modestly) generalized the theory of continuous time imprecise Markov chains: we found that the compatible precise processes need not be well behaved.

Continuous time processes are often explained in terms of small time increments. Internal set theory allows us to fully formalize that intuition, and allows the mathematical work to focus on the key algebraic elements, which are subjectively more pleasant than the analytic details. This was shown in the unified definition between discrete and continuous time imprecise Markov chains, as well as in some proofs at the end of the paper. Readers familiar with the literature may appreciate these simplifications.

There are some limitations. First, the notion of a 'nearby elementary process' takes quite a bit of effort to set up. It would be interesting to see if Nelson's definition could be incorporated with probability bounds. We did not do so here to keep things simpler. Another obstacle is that Nelson's approach is inherently one of finite additivity. The theory of lower previsions, and especially, Williams's approach (as opposed to Walley's), fully embraces finite additivity, giving the flexibility needed, and it worked for the results needed in this paper. But without resorting to external sets (such as in the approach of Loeb [7]) it is unclear how the theory could deal with conglomerability, or with the Shafer-Vovk-Ville formulae for stochastic processes under probability bounding [12, 3].

Acknowledgments

I am indebted to Teddy Seidenfeld for introducing me to Nelson's internal set theory. I also thank the reviewers for their comments which helped improving the manuscript.

References

[1] Jasper De Bock. The limit behaviour of imprecise continuous-time markov chains. *Journal of Nonlinear Science*, 27:159–196, 2017.

[2] Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, October 2009. doi:10.1017/S0269964809990039.

[3] Gert De Cooman, Jasper De Bock, and Stavros Lopatatzidis. Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems. *International Journal of Approximate Reasoning*, 76:18–46, 2016. doi:10.1016/j.ijar.2016.04.009.

[4] Francine Diener and Marc Diener, editors. *Nonstandard Analysis in Practice*. Springer, 1995.

[5] Lester E. Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability*, 3(1):89–99, 1975. doi:10.1214/aop/1176996451.

[6] Thomas Krak, Jasper De Bock, and Arno Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, 88:452–528, September 2017. doi:10.1016/j.ijar.2017.06.012.

[7] Peter A. Loeb. Conversion from nonstandard to standard measure spaces and applications in probability theory. *Transactions of the American Mathematical Society*, 211:113–122, 1975. doi:10.2307/1997222.

[8] Edward Nelson. Internal set theory: A new approach to nonstandard analysis. *Bulletin of the American Mathematical Society*, 83(6):1165–1198, 1977.

[9] Edward Nelson. *Radically Elementary Probability Theory*. Annals of Mathematical Studies. Princeton University Press, New Jersey, 1987. URL `https://web.math.princeton.edu/~nelson/books/rept.pdf`.

[10] Alain Robert. *Nonstandard Analysis*. Dover Publications, Inc., 2003.

[11] Eric Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, San Diego, 1997.

[12] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, 2001.

[13] Damjan Škulj. Efficient computation of the bounds of continuous time imprecise Markov chains. *Applied Mathematics and Computation*, 250:165–180, 2015. doi:10.1016/j.amc.2014.10.092.

[14] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley Series in Probability and Statistics. Wiley, 2014. ISBN 978-0-470-72377-7.

[15] Damjan Škulj. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, 50(8):1314–1329, 2009. doi:10.1016/j.ijar.2009.06.007.

[16] Peter M. Williams. Notes on conditional previsions. Technical report, School of Math. and Phys. Sci., Univ. of Sussex, 1975.

[17] Peter M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44 (3):366–383, 2007. doi:10.1016/j.ijar.2006.07.019.