# Kolmogorov Extension, Martingale Convergence, and Compositionality of Processes

Dexter Kozen
Cornell University
kozen@cs.cornell.edu

## Abstract

We show that the Kolmogorov extension theorem and the Doob martingale convergence theorem are two aspects of a common generalization, namely a colimit-like construction in a category of Radon spaces and reversible Markov kernels. The construction provides a compositional denotational semantics for lossless iteration in probabilistic programming languages, even in the absence of a natural partial order.

***Categories and Subject Descriptors*** G.3 [*Probability and Statistics*]: Markov processes, Stochastic processes; F.1.2 [*Modes of Computation*]: Probabilistic computation; F.3.2 [*Semantics of Programming Languages*]: Algebraic approaches to semantics, Denotational semantics, Operational semantics

***Keywords*** Markov processes, Kolmogorov extension, martingale convergence, compositionality, probabilistic computation

## 1. Introduction

*Compositionality* is a key desideratum in programming language semantics. The behavior of a large complex program depends on the behavior of its constituent parts, and it is essential for effective reasoning that this dependence be properly understood. This is no less true of probabilistic programs than deterministic ones.

Classical foundations of probability and measure theory provide little support for compositional reasoning. Standard formalizations of iterative processes prefer to construct a single monolithic sample space from which all random choices are made at once. The central result in this regard is the *Kolmogorov extension theorem* [16] (see [1] or [6, Theorem 3.3.6]), which identifies conditions under which a family of measures on finite subproducts of an infinite product space extend to a measure on the whole space. This theorem is typically used to construct a large sample space for an infinite iterative process when the behavior of each individual step of the process is known.

The Kolmogorov extension theorem is normally formulated in terms of measures alone, but for purposes of compositional reasoning, a more general formulation is needed. Probabilistic programs are commonly interpreted denotationally as *Markov kernels* [7, 19, 24, 25]. The chief benefit of this characterization is that it allows programs to be composed sequentially via Lebesgue integration. In [12], this formulation of the theorem was used to provide a compositional semantics for a lossless iteration operator * in a probabilistic language for packet switching networks.

Another important classical result that is relevant in this context is the *martingale convergence theorem* of Doob [8] (see [20] or [10, Theorem VII.9.2]). This theorem states that a martingale has a pointwise limit that is unique up to a nullset. Martingales are normally presented in introductory texts as a model of betting strategies, but in fact they are much more general and quite relevant in the semantics of probabilistic programming languages, as they characterize the evolution of conditional probabilities as an infinite process progresses.

Modern presentations of martingales go even further toward the removal of any compositional aspect. They are typically formulated in terms of a *filtration*, a sequence of ever finer $\sigma$-algebras on a common set of states $S$. Intuitively, the states of $S$ are infinite sequences of intermediate states in an iterative process, but the formalism does not reflect that intuition.[1]

Recently, martingales have emerged as a useful tool in the analysis of probabilistic algorithms. Dubhashi and Panconesi [9] give a generalization of *Chernoff-Hoeffding* bounds using martingales. These bounds exploit the tendency of certain smooth functions of random variables to concentrate asymptotically in a narrow range. Dubhashi and Panconesi's generalization applies to more general functions than just the sum and applies under weaker independence conditions. Chakarovand and Sankaranarayanan [3, 4] present general techniques based on martingales and supermartingales for the synthesis of expectation invariants for probabilistic loops and almost sure termination. Fioriti and Hermanns [11] also treat almost sure termination using supermartingales and provide soundness and completeness results.

In this paper we derive an unexpected connection between Kolmogorov extension and martingale convergence: they are different aspects of a common generalization, namely a colimit-like construction in a category of Radon spaces and reversible Markov kernels. A Markov kernel is *reversible* if it has a deterministic right[2] inverse; this is simply an abstract way of remembering history. We show that the limit object in the construction of [12], when applied to reversible kernels, is again reversible. Moreover, the martingale convergence theorem is essentially the statement of universality; intuitively, given an event in a continuation following an iteration, the finite approximants to the iteration comprise a martingale, which converges to the probability of the event conditioned on the final outcome of the iteration.



*LICS '16* July 05 - 08, 2016, New York, NY, USA

DOI: http://dx.doi.org/10.1145/2933575.2933610

[1] Dubhashi and Panconesi [9] remark that "the concept of martingales, as found in probability textbooks, poses quite a barrier to the computer scientist who is unfamiliar with the language of filters, partitions and measurable sets from measure theory."

[2] in diagrammatic order

We say the construction is colimit-*like* because it is not a colimit or even a weak colimit in the strict sense of the word. The exact nature of the discrepancy is technical, but it is essentially due to the fact that certain key properties hold only up to a nullset. Whether the construction can be characterized as a true colimit or weak colimit under stronger assumptions or using a point-free approach is a tantalizing topic for future investigation.

This construction can be used to give a compositional denotational semantics for iteration operators even when iteration is lossless. Normally, iteration is *lossy* in the sense that the probability of halting may be strictly less than unity [7, 18, 19, 22, 24, 25]. This is usually modeled by allowing the output of a program to be a subprobability distribution. However, in the system of [12], iteration is lossless: a program is a packet filter that consumes an input set of packets and produces an output set of packets according to some probability distribution. "Halting" is not a relevant concept; with probability one, the program "halts" and produces a set of packets (which may be the empty set). An important distinction is that lossy iteration aligns more closely with traditional domain-theoretic semantics, as there is a natural partial order of approximation by which the partial executions of a loop approximate the loop. With lossless iteration, there is no obvious partial order and no notion of least fixpoint. The system of [12] identifies an appropriate notion of convergence, but it is not order-theoretic.

Kolmogorov's original formulation of the extension theorem was in terms of finite subproducts of an infinite product space. Many authors [2, 5, 13, 21, 23, 26, 27] have observed that this is essentially a projective limit construction. Our development involves a projective limit as well, but it is important to note that, unlike projections, the morphisms of our category (reversible Markov kernels) go in the chronologically positive direction. Connections between Kolmogorov extension and martingale theory have also been previously drawn in earlier work [21, 23, 27], although the relationship presented here seems to have escaped notice.

## 2. Definitions and Notation

In this section we briefly review some basic definitions and notation. More comprehensive treatments can be found in [1, 6, 10, 14].

Let $(S, \mathcal{B})$ be a Borel space. A probability measure $\mu : \mathcal{B} \to [0, 1]$ is said to be *inner regular* if the measure of any $A \in \mathcal{B}$ can be approximated arbitrarily closely from below by compact sets. Formally, $\mu$ is inner regular if for all $A \in \mathcal{B}$ and $\varepsilon > 0$, there exists a compact set $C$ such that $C \subseteq A$ and $\mu(A - C) < \varepsilon$ (see e.g. [1]). A Borel space in which every probability measure is inner regular is called a *Radon space*.

We will restrict our attention to Radon spaces. This assumption is quite weak. Most natural spaces in real-life applications, including all Polish and Suslin spaces, are Radon. This assumption is needed for the Kolmogorov extension theorem.

***Markov Kernels*** Let $(S, \mathcal{B}_S)$ and $(T, \mathcal{B}_T)$ be Borel spaces. A *Markov kernel* (Markov transition, measurable kernel, stochastic kernel, stochastic relation) is a function $P : S \times \mathcal{B}_T \to [0, 1]$ such that

- for fixed $A \in \mathcal{B}_T$, the map $P(-, A) : S \to [0, 1]$ is a measurable function; and
- for fixed $s \in S$, the map $P(s, -) : \mathcal{B}_T \to [0, 1]$ is a probability measure.

Programs will be interpreted as Markov kernels. Some authors allow *subprobability measures* in which the universal event may occur with probability less than one; however, we will allow this only in the form of *guards* (see below).

The Markov kernels are the morphisms of a category whose objects are measurable spaces, the *Kleisli category of the Giry monad*; see [7, 24, 25]. In this context, we write $P : (S, \mathcal{B}_S) \to (T, \mathcal{B}_T)$ or just $P : S \to T$. Sequential composition is given by Lebesgue integration: for $P : S \to T$ and $Q : T \to U$,

$$(P \,;\, Q)(s, A) = \int_{t \in T} P(s, dt) \cdot Q(t, A).$$

Associativity of composition is essentially Fubini's theorem (see [14, VII.36.C] or [6, p. 59]). The the identity kernels

$$1(s, A) = \chi_A(s) = \delta_s(A) = \begin{cases} 1, & s \in A, \\ 0, & s \notin A \end{cases}$$

are left and right identities for composition. Here $\chi_A$ is the characteristic function of $A \in \mathcal{B}_S$ and $\delta_s$ is the Dirac (point mass) measure on $s \in S$.

If $P : S \to T_1$ and $Q : S \to T_2$, the kernel $P \times Q : S \to T_1 \times T_2$ on a given $s \in S$ gives the product measure $P(s, -) \times Q(s, -)$ on $T_1 \times T_2$. Thus for $A \in \mathcal{B}_{T_1}$ and $B \in \mathcal{B}_{T_2}$,

$$(P \times Q)(s, A \times B) = P(s, A) \cdot Q(s, B).$$

***Convergence*** We will have the occasion to consider convergence of sequences of kernels. A common mode of convergence is *pointwise almost everywhere* (*a.e.*) *convergence* with respect to an ambient measure $\mu$. A sequence $P_n$ converges *pointwise a.e. to $Q$ with respect to $\mu$* if for all $B$, the measurable functions $P_n(-, B)$ converge to the measurable function $Q(-, B)$ pointwise outside a $\mu$-nullset. Note that this does not immediately imply that $Q$ is a kernel, as it may not be countably additive in its second argument; that has to be established separately. Note also that the $\mu$-nullsets on which the $P_n(-, B)$ fail to converge may differ for different $B$.

***Deterministic Kernels*** A Markov kernel $S \to T$ is *deterministic* iff there is a measurable function $f : S \to T$ such that the kernel's value on $s \in S$ and $A \in \mathcal{B}_T$ is

$$1_T(f(s), A) = 1_S(s, f^{-1}(A)).$$

Every measurable function $f : S \to T$ gives a deterministic kernel of this form, thus the deterministic kernels and the measurable functions are in one-to-one correspondence. We write $f$ for both the set function and its associated kernel, thus

$$f(s, A) = 1_T(f(s), A) = 1_S(s, f^{-1}(A)).$$

Deterministic kernels compose on the left and right with arbitrary kernels as follows:

$$(f \,;\, P)(s, A) = \int_t 1(f(s), dt) \cdot P(t, A) = P(f(s), A)$$

$$(P \,;\, f)(s, A) = \int_t P(s, dt) \cdot 1(t, f^{-1}(A)) = P(s, f^{-1}(A)).$$

***Guards*** Every measurable set $A \in \mathcal{B}_S$ gives rise to an associated *guard* $A : S \to S$ of the same name, a subprobability kernel

$$A(s, B) = 1_S(s, A \cap B).$$

Guards can be used in sequential composition expressions to limit integration:

$$(P \,;\, A \,;\, Q)(s, B) = \int_{t \in A} P(s, dt) \cdot Q(t, B).$$

***Reversible Kernels*** A Markov kernel $P : S \to T$ is *reversible* if it has a deterministic right inverse $f : T \to S$; thus $P \,;\, f = 1$. If $P : S \to T$ and $Q : T \to U$ are reversible with inverses $f : T \to S$ and $g : U \to T$ respectively, then $P \,;\, Q$ is reversible with inverse $g \,;\, f$. The measurable spaces and reversible kernels form a subcategory of the Kleisli category of the Giry monad.

Reversibility is simply an abstract way of saying that history is preserved. Normally this is done with projections as in the usual

formulation of the Kolmogorov extension theorem. Any Markov kernel $P : S \to T$ gives rise to a reversible kernel $P' : S \to S \times T$ by just remembering the first argument:

$$P'(s, A \times B) = \begin{cases} P(s, B), & s \in A, \\ 0, & s \notin A. \end{cases}$$

The inverse is then the projection onto the first component.

**Lemma 1.**

(i) *For any $A \in \mathcal{B}_S$ and reversible $P : S \to T$ with right inverse $f : T \to S$, we have $A \mathbin{;} P = P \mathbin{;} f^{-1}(A)$.*
(ii) *For $A \in \mathcal{B}_S$ and deterministic $f : T \to S$, we have $f^{-1}(A) \mathbin{;} f = f \mathbin{;} A$.*
(iii) *The right inverse of any reversible kernel is surjective.*

*Proof.* (i) For any $s \in S$ and $B \in \mathcal{B}_T$,

$$\begin{aligned} (A \mathbin{;} P)(s, B) &= \int A(s, dt) \cdot P(t, B) \\ &= \begin{cases} P(s, B), & s \in A, \\ 0, & s \notin A, \end{cases} \\ (P \mathbin{;} f^{-1}(A))(s, B) &= \int P(s, dt) \cdot (f^{-1}(A))(t, B) \\ &= P(s, B \cap f^{-1}(A)). \end{aligned}$$

If $s \notin A$, then

$$\begin{aligned} P(s, B \cap f^{-1}(A)) &\le P(s, f^{-1}(A)) \\ &= (P \mathbin{;} f)(s, A) = 1(s, A) = 0, \end{aligned}$$

thus $A \mathbin{;} P$ and $P \mathbin{;} f^{-1}(A)$ agree in that case. If $s \in A$, then $s \notin {\sim}A$. By the above argument, $P(s, B \cap f^{-1}({\sim}A)) = 0$. Then

$$\begin{aligned} P(s, B \cap f^{-1}(A)) &= P(s, B \cap f^{-1}(A)) + P(s, B \cap f^{-1}({\sim}A)) \\ &= P(s, B), \end{aligned}$$

therefore $A \mathbin{;} P$ and $P \mathbin{;} f^{-1}(A)$ agree in that case as well.

(ii) For any $s \in S$ and $B \in \mathcal{B}_T$,

$$\begin{aligned} (f^{-1}(A) \mathbin{;} f)(s, B) &= f^{-1}(A)(s, f^{-1}(B)) \\ &= 1(s, f^{-1}(A \cap B)) = (1 \mathbin{;} f)(s, A \cap B) \\ &= f(s, A \cap B)) = (f \mathbin{;} A)(s, B). \end{aligned}$$

(iii) Suppose $P$ is reversible with right inverse $f$. By (i), for any $s \in S$,

$$\begin{aligned} (P \mathbin{;} f^{-1}(\{s\}) \mathbin{;} f)(s, S) &= (\{s\} \mathbin{;} P \mathbin{;} f)(s, S) \\ &= \{s\}(s, S) = 1(s, \{s\}) = 1, \end{aligned}$$

so it cannot be that $f^{-1}(\{s\}) = 0$. □

***Conditional Expectation*** Let $(S, \mathcal{B}, \mu)$ be a measure space with $\mu$ a probability measure. The *conditional expectation* $\mathcal{E}(X \mid \mathcal{F})$ of a $\mathcal{B}$-measurable function $X$ with respect to a $\sigma$-subalgebra $\mathcal{F}$ of $\mathcal{B}$ is any $\mathcal{F}$-measurable function such that for $A \in \mathcal{F}$,

$$\int_A \mathcal{E}(X \mid \mathcal{F})(s) \cdot \mu(ds) = \int_A X(s) \cdot \mu(ds).$$

The conditional expectation exists and is unique up to a $\mu$-nullset. It can be obtained as a Radon–Nikodým derivative, as the integral on the right-hand side, as a function of $A \in \mathcal{F}$, is absolutely continuous with respect to $\mu$; that is, the integral vanishes whenever $A \in \mathcal{F}$ and $\mu(A) = 0$.

Applied to characteristic functions of measurable sets $B \in \mathcal{B}$, conditional expectations are also measures as functions of $B$. As such, they are Markov kernels $\mathcal{E}_{\mathcal{F}} : (S, \mathcal{F}) \to (S, \mathcal{B})$ with $\mathcal{E}_{\mathcal{F}}(s, B) = \mathcal{E}(\chi_B \mid \mathcal{F})(s)$. This representation affords some notational advantages:

(2.1) A well-known property is that for $\mathcal{F} \subseteq \mathcal{G} \subseteq \mathcal{B}$,

$$\mathcal{E}(\mathcal{E}(X \mid \mathcal{G}) \mid \mathcal{F}) = \mathcal{E}(X \mid \mathcal{F}).$$

In our notation, this translates to

$$\mathcal{E}_{\mathcal{F}} \mathbin{;} \mathcal{E}_{\mathcal{G}} = \mathcal{E}_{\mathcal{F}}.$$

(2.2) Suppose that $\mathcal{F} \subseteq \mathcal{G}$ and we are given two kernels $P : (S, \mathcal{F}) \to T$ and $Q : (S, \mathcal{G}) \to T$, and we wish to show that $P(-, B) = \mathcal{E}(Q(-, B) \mid \mathcal{F})$ with respect to an ambient measure $\mu$. We would need to show that for all $A \in \mathcal{F}$,

$$\int_A P(s, B) \cdot \mu(ds) = \int_A Q(s, B) \cdot \mu(ds).$$

Regarding $\mu$ as a kernel $\mu : S_0 \to S$ on a one-point space $S_0$, it suffices to show

$$\mu \mathbin{;} A \mathbin{;} P = \mu \mathbin{;} A \mathbin{;} Q$$

for all $A \in \mathcal{F}$. This gives the desired equation for all $B \in \mathcal{B}_T$ uniformly. In particular, if $Q(-, B) = \chi_B$, so $Q = 1$, then it suffices to show

$$\mu \mathbin{;} A \mathbin{;} P = \mu \mathbin{;} A.$$

(2.3) A special case of the martingale convergence theorem is the *Lévy zero-one law*, which states that if $\mathcal{F}_n$ is a sequence of $\sigma$-algebras on $S$ such that $\mathcal{F}_m \subseteq \mathcal{F}_n$ for $m \le n$, and if $\mathcal{F}_\omega$ is the smallest $\sigma$-algebra containing $\bigcup_n \mathcal{F}_n$, then for any $X : S \to \mathbb{R}$ measurable with respect to $\mathcal{F}_\omega$, $\mathcal{E}(X \mid \mathcal{F}_n)$ converges to $\mathcal{E}(X \mid \mathcal{F}_\omega)$ pointwise outside of a $\mu$-nullset. In our notation, this becomes

$$\mathcal{E}_{\mathcal{F}_n} \to \mathcal{E}_{\mathcal{F}_\omega} \text{ pointwise a.e.}$$

***Martingales*** Let $(S, \mathcal{F}_\omega, \mu)$ be a measure space. A sequence of random variables and $\sigma$-algebras $(X_n, \mathcal{F}_n)$ on $S$, $n \ge 0$, is called a *martingale* if

(i) $\mathcal{F}_m \subseteq \mathcal{F}_n$ for $m \le n$,
(ii) $\mathcal{F}_\omega$ is the $\sigma$-algebra generated by $\bigcup_n \mathcal{F}_n$,
(iii) $X_n$ is $\mathcal{F}_n$-measurable,
(iv) $X_m = \mathcal{E}(X_n \mid \mathcal{F}_m)$ for all $m \le n$.

The martingale convergence theorem of Doob [8] (see [20] or [10, Theorem VII.9.2]) states that the sequence $X_n$ has a pointwise limit $X_\omega$ outside a $\mu$-nullset.

By property (iv) of martingales and the definition of conditional expectation, for all $A_m \in \mathcal{F}_m$ and $n \ge m$,

$$\begin{aligned} \int_{A_m} X_m(s) \cdot \mu(ds) &= \int_{A_m} \mathcal{E}(X_n \mid \mathcal{F}_m)(s) \cdot \mu(ds) \\ &= \int_{A_m} X_n(s) \cdot \mu(ds), \end{aligned}$$

thus by the martingale convergence theorem,

$$\begin{aligned} \int_{A_m} X_m(s) \cdot \mu(ds) &= \int_{A_m} X_\omega(s) \cdot \mu(ds) \\ &= \int_{A_m} \mathcal{E}(X_\omega \mid \mathcal{F}_m) \cdot \mu(ds). \end{aligned}$$

Again by the definition of conditional expectation, we have that

$$X_m = \mathcal{E}(X_\omega \mid \mathcal{F}_m). \tag{2.4}$$

## 3. Main Results

Suppose we have a chain of Radon spaces $(S_n, \mathcal{B}_n)$, $n \in \omega$ along with reversible Markov kernels $P_{mn} : S_m \to S_n$ for each $m \leq n$ such that

$$P_{kn} = P_{km} \; ; P_{mn}, \quad k \leq m \leq n \qquad P_{nn} = 1_{S_n}. \tag{3.1}$$

Since the $P_{mn}$ are reversible, their deterministic inverses $f_{nm} : S_n \to S_m$ for $m \leq n$ satisfy

$$f_{nk} = f_{nm} \; ; f_{mk}, \quad k \leq m \leq n \qquad f_{nn} = 1_{S_n}. \tag{3.2}$$

The chain $S_n$ has a projective limit $S_\omega$, where

$$S_\omega = \{(s_n \mid n \in \omega) \in \prod_{n \in \omega} S_n \mid \forall m \leq n \; f_{nm}(s_n) = s_m\}.$$

Let $\mathcal{B}_\omega$ be the weakest $\sigma$-algebra on $S_\omega$ such that all projections $\pi_m : S_\omega \to S_m$ are measurable. Then $(S_\omega, \mathcal{B}_\omega)$ is the limit of the spaces $(S_n, \mathcal{B}_n)$ in the category of Radon spaces and measurable functions.

The following local consistency condition corresponds to the premise needed to apply the Kolmogorov extension theorem (see [6, Theorem 3.3.6]).

**Lemma 2.** *For all $k \leq m \leq n$,*

$$P_{km}(s, A) = P_{kn}(s, f_{nm}^{-1}(A)).$$

*Proof.* This is equivalent to the assertion that $P_{km} = P_{kn} \; ; f_{nm}$. But

$$P_{km} = P_{km} \; ; 1_{S_m} = P_{km} \; ; P_{mn} \; ; f_{nm} = P_{kn} \; ; f_{nm}. \qquad \square$$

Let $\mathcal{F}_n = \{\pi_n^{-1}(A_n) \mid A_n \in \mathcal{B}_n\} \subseteq \mathcal{B}_\omega$. Then $\bigcup_n \mathcal{F}_n$ is a Boolean subalgebra of $\mathcal{B}_\omega$. By the monotone class theorem ([14, Theorem I.6.B] or [6, Theorem 2.1.1]), $\mathcal{B}_\omega$ is the smallest class containing $\bigcup_n \mathcal{F}_n$ and closed under unions of countable ascending chains and intersections of countable descending chains.

**Lemma 3.**

(i) *If $m \leq n$, $A_m \in \mathcal{F}_m$, and $A_n \in \mathcal{F}_n$, then $\pi_n^{-1}(A_n) = \pi_m^{-1}(A_m)$ if and only if $A_n = f_{nm}^{-1}(A_m)$.*
(ii) *If $m \leq n$, then $\mathcal{F}_m \subseteq \mathcal{F}_n$.*

*Proof.* Clause (ii) and the reverse implication of clause (i) follow from the fact that $\pi_m = \pi_n \; ; f_{nm}$. Thus

$$\pi_m^{-1}(A_m) = \pi_n^{-1}(f_{nm}^{-1}(A_m)) \in \mathcal{F}_n.$$

For the forward implication of clause (i), assume that $\pi_n(s) \in A_n$ iff $\pi_m(s) \in A_m$. Since $\pi_m = \pi_n \; ; f_{nm}$, we have that

$$\pi_n(s) \in A_n \Leftrightarrow f_{nm}(\pi_n(s)) \in A_m \Leftrightarrow \pi_n(s) \in f_{nm}^{-1}(A_m).$$

It follows from Lemma 1(ii) that the $\pi_n : S_\omega \to S_n$ are surjective, thus $A_n = f_{nm}^{-1}(A_m)$. $\square$

We now wish to show that $(S_\omega, \mathcal{B}_\omega)$ is the colimiting object of the chain. We need to define reversible Markov kernels $P_{m\omega} : S_m \to S_\omega$ that commute with the $P_{mn}$. As with the Kolmogorov extension theorem, inner regularity is needed for this part of the argument; once this is done, the assumption of inner regularity is no longer needed.

For each $\pi_n^{-1}(A_n) \in \mathcal{F}_n$ with $m \leq n$, define

$$P_{m\omega}(s, \pi_n^{-1}(A_n)) = P_{mn}(s, A_n). \tag{3.3}$$

We must argue that $P_{m\omega}$ is well defined. If $k \leq m \leq n$ with $\pi_m^{-1}(A_m) = \pi_n^{-1}(A_n)$, we have by Lemma 3(i) that $A_n = f_{nm}^{-1}(A_m)$. Then

$$\begin{aligned} P_{kn}(s, A_n) &= P_{kn}(s, f_{nm}^{-1}(A_m)) \\ &= (P_{kn} \; ; f_{nm})(s, A_m) = P_{km}(s, A_m). \end{aligned}$$

**Theorem 1.** *The map $P_{n\omega} : S_n \times \bigcup_n \mathcal{F}_n \to [0,1]$ extends to a reversible Markov kernel $P_{n\omega} : S_n \to S_\omega$ with right inverse $\pi_n$.*

*Proof.* We must show:

(i) For fixed $s \in S_n$, the map $P_{n\omega}(s, -) : \bigcup_n \mathcal{F}_n \to [0,1]$ extends to a measure $P_{n\omega}(s, -) : \mathcal{B}_\omega \to [0,1]$.
(ii) For fixed $A \in \mathcal{B}_\omega$, the map $P_{n\omega}(-, A) : S_n \to [0,1]$ is a measurable function.

For (i), using inner regularity one can show that for fixed $s \in S_n$, the map $P_{n\omega}(s, -) : \bigcup_n \mathcal{F}_n \to [0,1]$ is countably additive on $\bigcup_n \mathcal{F}_n$, therefore by the Carathéodory extension theorem (see [14, Theorem 13.A] or [17, Theorem 7.27.7]) extends to a measure $P_{n\omega}(s, -) : S_\omega \to [0,1]$. This is essentially the Kolmogorov extension theorem in this setting.

For (ii), the proof is by induction on the stage at which $A$ becomes an element of $\mathcal{B}_\omega$ via the monotone class theorem. The basis is (3.3). For the induction step, we use the fact that the pointwise supremum of a countable ascending chain of uniformly bounded measurable functions is measurable. If $A = \bigcup_n A_n$ for a chain $A_0 \subseteq A_1 \subseteq \cdots$, we have that the functions $P_{n\omega}(-, A_i)$ are measurable by the inductive hypothesis, and $P_{n\omega}(-, \bigcup_i A_i)$ is the pointwise supremum of the $P_{n\omega}(-, A_i)$, therefore measurable. The argument for intersections of countable descending chains is similar.

That $\pi_n$ is the right inverse of $P_{n\omega}$, that is, $P_{n\omega} \; ; \pi_n = 1_{S_n}$, is just (3.3) with $m = n$. $\square$

The next theorem establishes a universality property of the space $(S_\omega, \mathcal{B}_\omega)$ as a form of colimit of the $(S_n, \mathcal{B}_n)$ with coprojections $P_{n\omega} : S_n \to S_\omega$ in the category of Radon spaces and reversible Markov kernels. As mentioned, it is not a true colimit or even a weak colimit; nevertheless, the space $(S_\omega, \mathcal{B}_\omega)$ is universal in a sense to be made precise by part (ii) of the theorem.

**Theorem 2.**

(i) *The kernels $P_{n\omega}$ commute with the kernels $P_{mn}$ in the sense that for all $m \leq n$, $P_{m\omega} = P_{mn} \; ; P_{n\omega}$.*
(ii) *Let $(T, \mathcal{B}_T)$ be any measurable space with reversible Markov kernels $Q_n : S_n \to T$, each with a deterministic right inverse $g_n : T \to S_n$ such that $Q_m = P_{mn} \; ; Q_n$ for all $m \leq n$. There exists a reversible Markov kernel*

$$Q_\omega : S_\omega \to \widehat{T}$$

*such that $Q_n = P_{n\omega} \; ; Q_\omega$, where $\widehat{T}$ is the completion of $T$ with respect to the pseudometric*

$$d_T(t, t') = \begin{cases} 2^{-n}, & \text{if } n \text{ is the least number} \\ & \text{such that } g_n(t) \neq g_n(t'), \\ 0, & \text{if } g_n(t) = g_n(t') \text{ for all } n. \end{cases} \tag{3.4}$$

*Proof.* (i) We have by (3.3) and Theorem 1 that for $k \geq n$,

$$P_{m\omega} \; ; \pi_k = P_{mk} = P_{mn} \; ; P_{nk} = P_{mn} \; ; P_{n\omega} \; ; \pi_k.$$

Because of the composition with $\pi_k$ on the right, $P_{m\omega}$ and $P_{mn} \; ; P_{n\omega}$ agree on the generators of $\mathcal{B}_\omega$, therefore also on all of $\mathcal{B}_\omega$.

(ii) Under the premises of the theorem, $g_m = g_n \; ; f_{nm}$ and $Q_m \; ; g_n = P_{mn}$ for all $m \leq n$. Since $(S_\omega, \mathcal{B}_\omega)$ is the limit of the $(S_n, \mathcal{B}_n)$ in the category of measurable spaces and measurable functions, there is a measurable function $g : T \to S_\omega$ such that $g_n = g \; ; \pi_n$ for all $n$.

For all $n \geq m$, we have

$$Q_m \; ; g \; ; \pi_n = P_{mn} \; ; Q_n \; ; g_n = P_{mn} = P_{m\omega} \; ; \pi_n.$$

Because of the composition with $\pi_n$ on the right, $(Q_m \mathbin{;} g)(s_m, -)$ and $P_{m\omega}(s_m, -)$ agree on the generators $\bigcup_n \mathcal{F}_n$ of $\mathcal{F}_\omega$, therefore also on $\mathcal{F}_\omega$. Thus

$$Q_m \mathbin{;} g = P_{m\omega}. \tag{3.5}$$

We now construct a kernel $Q_\omega$ with right inverse $g$. Unfortunately, the limit construction does not guarantee that $g$ is surjective, which by Lemma 1(iii) it must be in order to be the right inverse of a kernel. However, its image $g(T) = \{g(t) \mid t \in T\}$ is dense in $S_\omega$ with respect to a certain metric, and we can form the completion $\widehat{T}$ of $T$ without affecting the values of the $Q_n$. This will allow $g$ to be extended to a surjective function $\widehat{g} : \widehat{T} \to S_\omega$, which will allow the construction of a kernel $S_\omega \to \widehat{T}$. Moreover, the ideal $\{A \in \mathcal{B}_{S_\omega} \mid A \cap g(T) = \emptyset\}$ contains only $P_{m\omega}$-nullsets, since if $A \cap g(T) = \emptyset$, then by (3.5),

$$P_{m\omega}(s_m, A) = (Q_m \mathbin{;} g)(s_m, A) = Q_m(s_m, g^{-1}(A)) = 0,$$

thus points not in $g(T)$ can be deleted from $S_\omega$ to give a kernel $g(T) \to T$ if desired.

The completion $\widehat{T}$ of $T$ is taken with respect to the pseudometric $d_T$, where

$$d_{S_\omega}(s, s') = \begin{cases} 2^{-n}, & \text{if } n \text{ is the least number such that } \pi_n(s) \neq \pi_n(s'), \\ 0, & \text{if } s = s' \end{cases}$$

$$d_T(t, t') = d_{S_\omega}(g(t), g(t')).$$

Concretely, let $\widehat{T}$ be the disjoint union of $T$ and $S_\omega - g(T)$. Define

$$\widehat{g}(t) = \begin{cases} g(t), & t \in T, \\ t, & t \in S_\omega - g(T), \end{cases}$$

$$\widehat{g}_n(t) = \begin{cases} g_n(t), & t \in T, \\ \pi_n(t), & t \in S_\omega - g(T). \end{cases}$$

Extend $\mathcal{B}_T$ to $\mathcal{B}_{\widehat{T}}$ by including all subsets of $S_\omega - g(T)$. Extend $Q_n : S \to T$ to $\widehat{Q}_n : S \to \widehat{T}$ by taking subsets of $S_\omega - g(T)$ as nullsets; that is, $\widehat{Q}_n(s_n, B) = Q_n(s, B \cap T)$. Note that $g(T)$ is dense in $S_\omega$ under the metric $d_{S_\omega}$. One can show that $\widehat{Q}_n$ is reversible with right inverse $\widehat{g}_n$ and that $P_{mn} \mathbin{;} \widehat{Q}_n = \widehat{Q}_m$. Moreover, $\widehat{g} : \widehat{T} \to S_\omega$ is surjective. Let us therefore assume henceforth that the original $g$ is surjective and that $\widehat{T} = T$.

We now show that the kernels $Q_n$ give rise to two collections of martingales. For the first collection, fix $B \in \mathcal{B}_T$ and $s_0 \in S_0$. We show that the measurable functions $(\pi_n \mathbin{;} Q_n)(-, B)$ form a martingale with respect to the filtration $\{\mathcal{F}_n \mid n \geq 0\}$ and the ambient measure $P_{0\omega}(s_0, -)$ on $S_\omega$. Let us check the four properties of martingales listed in §2.

(i) $\mathcal{F}_m \subseteq \mathcal{F}_n$ for $m \leq n$,
(ii) $\mathcal{F}_\omega$ is the $\sigma$-algebra generated by $\bigcup_n \mathcal{F}_n$,
(iii) $(\pi_n \mathbin{;} Q_n)(-, B)$ is $\mathcal{F}_n$-measurable,
(iv) $(\pi_m \mathbin{;} Q_m)(-, B) = \mathcal{E}((\pi_n \mathbin{;} Q_n)(-, B) \mid \mathcal{F}_m)$ for $m \leq n$.

Properties (i) and (ii) are immediate from Lemma 3. For (iii), $(\pi_n \mathbin{;} P_{n\omega})(-, B)$ is $\mathcal{F}_n$-measurable because $\pi_n$ is. Finally, for property (iv), by (2.2) it suffices to show that for any $A_m \in \mathcal{B}_m$,

$$P_{0\omega} \mathbin{;} \pi_m^{-1}(A_m) \mathbin{;} \pi_m \mathbin{;} Q_m = P_{0\omega} \mathbin{;} \pi_m^{-1}(A_m) \mathbin{;} \pi_n \mathbin{;} Q_n. \tag{3.6}$$

Let $A_n = f_{nm}^{-1}(A_m)$. By Lemma 3(i), $\pi_m^{-1}(A_m) = \pi_n^{-1}(A_n)$. Using this, Lemma 1(ii), Theorem 1, and the fact $Q_m = P_{mn} \mathbin{;} Q_n$, (3.6) reduces to

$$A_m \mathbin{;} P_{mn} = P_{mn} \mathbin{;} A_n.$$

But this is just Lemma 1(i).

By the martingale convergence theorem, the $(\pi_n \mathbin{;} Q_n)(-, B)$ converge pointwise to an $\mathcal{F}_\omega$-measurable function $Q_\omega(-, B)$ outside a $P_{0\omega}(s_0, -)$-nullset, thus the $\pi_n \mathbin{;} Q_n$ converge pointwise a.e. to $Q_\omega$.

The map $Q_\omega$ will be our desired kernel. However, note that we have not yet shown that $Q_\omega$ is a measure in its second variable nor that it is reversible with right inverse $g$. We will do this below, but we must be careful not to inadvertently use these properties until they are established.

The $Q_k$ factor through $Q_\omega$ as desired: for $k \leq m \leq n$,

$$P_{k\omega} \mathbin{;} \pi_m \mathbin{;} Q_m = P_{k\omega} \mathbin{;} \pi_n \mathbin{;} Q_n = Q_k,$$

therefore

$$P_{k\omega} \mathbin{;} Q_\omega = \lim_n P_{k\omega} \mathbin{;} \pi_n \mathbin{;} Q_n = \lim_n Q_k = Q_k.$$

The second collection of martingales is defined on $T$. Define the filtration

$$\mathcal{G}_n = \{g_n^{-1}(A) \mid A \in \mathcal{B}_n\} = \{g^{-1}(A) \mid A \in \mathcal{F}_n\} \in \mathcal{B}_T$$

and let $\mathcal{G}_\omega \subseteq \mathcal{B}_T$ be the $\sigma$-algebra generated by $\bigcup_n \mathcal{G}_n$. As above, fix $B \in \mathcal{B}_T$ and $s_0 \in S_0$. We claim that the functions $(g_n \mathbin{;} Q_n)(-, B)$ form a martingale with respect to the filtration $\{\mathcal{G}_n \mid n \geq 0\}$ and the ambient measure $Q_0(s_0, -)$ on $T$. The four properties of martingales we must check are

(i) $\mathcal{G}_m \subseteq \mathcal{G}_n$ for $m \leq n$,
(ii) $\mathcal{G}_\omega$ is the $\sigma$-algebra generated by $\bigcup_n \mathcal{G}_n$,
(iii) $(g_n \mathbin{;} Q_n)(-, B)$ is $\mathcal{G}_n$-measurable,
(iv) $(g_m \mathbin{;} Q_m)(-, B) = \mathcal{E}((g_n \mathbin{;} Q_n)(-, B) \mid \mathcal{G}_m)$ for $m \leq n$.

As above, properties (i)–(iii) are straightforward: (i) is immediate from Lemma 3, (ii) is by definition, and (iii) is from the fact that $g_n$ is $\mathcal{G}_n$-measurable. Finally, for (iv), by (2.2) we must show that for any $A_m \in \mathcal{B}_m$,

$$Q_0 \mathbin{;} g_m^{-1}(A_m) \mathbin{;} g_m \mathbin{;} Q_m = Q_0 \mathbin{;} g_m^{-1}(A_m) \mathbin{;} g_n \mathbin{;} Q_n. \tag{3.7}$$

Let $A_n = f_{nm}^{-1}(A_m)$. By the fact that $g_m = g_n \mathbin{;} f_{nm}$, we have $g_m^{-1}(A_m) = g_n^{-1}(A_n)$. Using this, Lemma 1(ii), and Theorem 1, (3.7) reduces to

$$Q_k \mathbin{;} g_m \mathbin{;} A_m \mathbin{;} Q_m = Q_k \mathbin{;} g_n \mathbin{;} A_n \mathbin{;} Q_n,$$

which follows by equational reasoning from Lemma 1(i) and the properties

$$Q_m = P_{mn} \mathbin{;} Q_n \qquad Q_m \mathbin{;} g_m = 1 \qquad P_{km} \mathbin{;} P_{mn} = P_{kn}.$$

Again using the martingale convergence theorem, the $(g_n \mathbin{;} Q_n)(-, B)$ converge pointwise to a $\mathcal{G}_\omega$-measurable function outside a $Q_0(s_0, -)$-nullset. In this case, the limit is $(g \mathbin{;} Q_\omega)(-, B)$:

$$\lim_n g_n \mathbin{;} Q_n = \lim_n g \mathbin{;} \pi_n \mathbin{;} Q_n = g \mathbin{;} Q_\omega.$$

As above, the $g_n \mathbin{;} Q_n$ converge pointwise a.e. to $g \mathbin{;} Q_\omega$.

Note that none of these calculations required integration with respect to the second argument of $Q_\omega$. As the reader will recall, we have yet to establish that $Q_\omega$ is countably additive in its second argument. We do that now.

**Lemma 4.** *$g_n \mathbin{;} Q_n = \mathcal{E}_{\mathcal{G}_n}$, the conditional expectation with respect to the measure $Q_0(s_0, -)$ on $T$.*

*Proof.* Using Lemma 1,

$$\begin{aligned} Q_0 \mathbin{;} g_n^{-1}(A_n) \mathbin{;} g_n \mathbin{;} Q_n &= Q_0 \mathbin{;} g_n \mathbin{;} A_n \mathbin{;} Q_n \\ &= P_{0n} \mathbin{;} Q_n \mathbin{;} g_n^{-1}(A_n) \\ &= Q_0 \mathbin{;} g_n^{-1}(A_n). \end{aligned}$$

□

By the Lévy zero-one theorem (2.3), $\mathcal{E}_{\mathcal{G}_n}$ converges pointwise a.e. to $\mathcal{E}_{\mathcal{G}_\omega}$. We have already argued that $g \mathbin{;} Q_\omega$ is the a.e. pointwise limit of the $g_n \mathbin{;} Q_n$. Thus by Lemma 4, $g \mathbin{;} Q_\omega = \mathcal{E}_{\mathcal{G}_\omega}$ a.e. This says that for all $t \in T$,

$$Q_\omega(g(t), -) = (g \mathbin{;} Q_\omega)(t, -) = \mathcal{E}_{\mathcal{G}_\omega}(t, -).$$

The kernel $\mathcal{E}_{\mathcal{G}_\omega}$ is a conditional probability, therefore a measure in its second argument. As $g$ is surjective, $Q_\omega(s, -)$ is also measure for all $s \in S_\omega$, therefore it is a Markov kernel.

It remains to show that $Q_\omega$ is reversible with right inverse $g$. Observe that

$$P_{n\omega} \mathbin{;} \pi_n = 1_{S_n} = Q_n \mathbin{;} g_n = P_{n\omega} \mathbin{;} Q_\omega \mathbin{;} g_n.$$

Thus $Q_\omega \mathbin{;} g_n$ and $\pi_n$ agree outside a $P_{n\omega}$-nullset, and this is true for arbitrary $n$. By the universality of $g$ to the projective limit $\mathcal{B}_\omega$ of the spaces $\mathcal{B}_n$, we have $Q_\omega \mathbin{;} g = 1_{S_\omega}$. □

### 3.1 Discussion

The kernel $Q_\omega$ constructed in the proof of Theorem 2 is not unique, as the martingale convergence theorem determines $Q_\omega$ only up to a nullset for each $B \in \mathcal{B}_T$. Moreover, there is some flexibility in the formation of the completion $\widehat{T}$. Thus the construction is at best a weak colimit.

If $g$ is not surjective, the kernel $Q_\omega$ does not give a universal arrow in the strict sense of the word, as it is not necessarily of type $S_\omega \to T$. An extension of $T$ to $\widehat{T}$ may be required to accommodate the orphans $s \in S_\omega$. This can always be done in a straightforward way as we have done in the proof of Theorem 2, but the type of the arrow is then $S_\omega \to \widehat{T}$, not $S_\omega \to T$. As we have noted, the orphans can be omitted, giving a kernel of type $S'_\omega \to T$ for a dense subset $S'_\omega \subseteq S_\omega$, but this is not of the correct type either. However, under the assumption that $T$ is complete with respect to the pseudometric (3.4), the construction becomes a genuine weak colimit.

We made use of inner regularity in the construction of the $P_{k\omega}$. Moy [23, p. 907] seems to suggest that this assumption is not necessary. But Moy is working in the space of real sequences, which is implicitly inner regular. The claim does not hold more generally, as the following counterexample shows. Let $\mathcal{F}_n$ be the $\sigma$-algebra on $\mathbb{N}$ generated by the sets $\{0\}, \{1\}, \ldots, \{n-1\}$ and $\{n, n+1, n+2, \ldots\}$. Then $\mathcal{F}_n$ is finite and $\bigcup_n \mathcal{F}_n$ consists of all finite and cofinite sets. The $\sigma$-algebra $\mathcal{F}_\omega$ generated by $\bigcup_n \mathcal{F}_n$ is the full powerset of $\mathbb{N}$, as every set is a countable union of singletons.

Now let $U$ be a nonprinciple ultrafilter and let

$$\mu(A) = \begin{cases} \dfrac{1}{2} + \displaystyle\sum_{n \in A} 2^{-(n+2)}, & A \in U, \\ \displaystyle\sum_{n \in A} 2^{-(n+2)}, & A \notin U. \end{cases}$$

Then $\mu$ is nonnegative, finite-valued, and countably additive on every $\mathcal{F}_n$, but not countably additive on $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n$, since

$$\sum_{n \in \mathbb{N}} \mu(\{n\}) = \frac{1}{2} \qquad \mu(\mathbb{N}) = 1.$$

The space is not inner regular, as any set in $U$ is at least $1/2$ heavier than any compact subset.

### 3.2 Encoding Kolmogorov Extension

The standard Kolmogorov extension theorem is a special case involving measures on product spaces

$$S_n = \prod_{n=0}^{n} S'_n \qquad S_\omega = \prod_{n=0}^{\infty} S'_n.$$

The functions $f_{nm} : S_n \to S_m$ for $m \le n$ and $\pi_n : S_\omega \to S_n$ are simply the projections onto lower-dimensional products:

$$f_{nm}(s_0, \ldots, s_n) = (s_0, \ldots, s_m),\ m \le n$$
$$\pi_n(s_0, s_1, \ldots) = (s_0, \ldots, s_n).$$

The generalization to projective limits of spaces connected by measurable functions $f_{nm}$ has been observed by several authors [2, 5, 13, 21, 26, 27].

In the classical treatment, we are given component probability measures $\mu_n$ on the $S_n$ satisfying the consistency condition $\mu_m = \mu_n \circ f_{nm}^{-1}$ for all $m \le n$. The Kolmogorov extension theorem guarantees the existence of a unique probability measure $\mu$ on $S_\omega$ such that $\mu_n = \mu \circ \pi_n^{-1}$ for every $n$.

In our framework, the kernels $P_{mn} : S_m \to S_n$ are the conditional expectations $P_{mn}(s, A) = \mathcal{E}_m(s, A)$ for $s \in S_m$ and $A$ a measurable subset of $S_n$. The kernels compose properly by virtue of (2.1). The necessary consistency condition among the component measures is given by Lemma 2: for $k \le m \le n$, $s \in S_k$, and $A$ a measurable subset of $S_m$, $\mathcal{E}_k(s, A) = \mathcal{E}_k(s, f_{nm}^{-1}(A))$.

### 3.3 Encoding Martingales

The martingale convergence theorem is also a special case. Given a $[0,1]$-valued martingale $(X_n, \mathcal{F}_n)$ on a space $(S, \mathcal{F}_\omega, \mu)$, we can encode it as a cocone on a chain of measurable spaces and reversible kernels. This can be done in two distinct but equivalent ways, the first closer in spirit to classical martingale theory on a single space with a filtration of $\sigma$-algebras, the second closer to probabilistic semantics involving a chain of state transition systems.

In the first approach, we define $S_n = S$, $\mathcal{B}_n = \mathcal{F}_n$, and for $m \le n$, $s \in S$, and $A \in \mathcal{F}_n$,

$$P_{mn}(s, A) = \mathcal{E}_m(s, A) \qquad f_{nm}(s) = s.$$

As observed in §2, the conditional expectation $\mathcal{E}_m(s, A)$ is a measurable function in $s$ and a measure in $A$, thus a Markov kernel. Note that $P_{mn}(s, A)$ does not depend on $n$. The standard property (2.1) of conditional expectations implies that composition works correctly: for $k \le m \le n$ and $A \in \mathcal{F}_n$,

$$\begin{aligned}(P_{km} \mathbin{;} P_{mn})(s, A) &= \int_{t \in S} \mathcal{E}_k(s, dt) \cdot \mathcal{E}_m(t, A) \\ &= \mathcal{E}_k(s, A) = P_{kn}(s, A).\end{aligned}$$

The function $f_{nm}$ is a measurable function with respect to the measurable sets $\mathcal{F}_n$ on its domain and $\mathcal{F}_m$ on its range, since $\mathcal{F}_m \subseteq \mathcal{F}_n$. Moreover, $f_{nm}$ is the right inverse of $P_{mn}$: for $A_m \in \mathcal{F}_m$,

$$\begin{aligned}(P_{mn} \mathbin{;} f_{nm})(s, A_m) &= \int_{t \in S} \mathcal{E}_m(s, dt) \cdot 1(t, A_m) \\ &= \mathcal{E}_m(s, A_m) = 1(s, A_m).\end{aligned}$$

The projective limit $S_\omega$ of the $S_n$ is just $S$ itself, and $\pi_n(s) = s$. This gives

$$\begin{aligned}P_{m\omega}(s, A) &= P_{m\omega}(s, \pi_n^{-1}(A)) \\ &= P_{mn}(s, A) = \mathcal{E}_m(s, A),\ A \in \mathcal{F}_n.\end{aligned}$$

Since $P_{m\omega}(s, -)$ and $\mathcal{E}_m(s, -)$ agree on $\bigcup_n \mathcal{F}_n$, they agree on all of $\mathcal{F}_\omega$.

Now to encode the martingale $X_n$, let $X_\omega$ be the pointwise limit of the $X_n$ as guaranteed by the martingale convergence theorem. Let

$$\begin{aligned}T &= S \times \{0,1\} \qquad g(s,0) = g(s,1) = s \\ \mathcal{G}_\alpha &= \{g^{-1}(A) \mid A \in \mathcal{F}_\alpha\} \\ &= \{A \times \{0,1\} \mid A \in \mathcal{F}_\alpha\},\ \alpha \in \omega \cup \{\omega\} \\ \mathcal{B}_T &= \{(A \times \{1\}) \cup (B \times \{0\}) \mid A, B \in \mathcal{G}_\omega\}.\end{aligned}$$

The set $\mathcal{B}_T$ is the $\sigma$-algebra generated by $\mathcal{G}_\omega \cup \{S \times \{1\}\}$. Define the kernel $Q : S \to T$ by

$$\begin{aligned} &Q(s, (A \times \{1\}) \cup (B \times \{0\})) \\ &= X_\omega(s) \cdot 1(s, A) + (1 - X_\omega(s)) \cdot 1(s, B) \end{aligned}$$

for $A, B \in \mathcal{G}_\omega$. In other words, $Q(s, -)$ is a weighted sum of Dirac measures on $(s, 1)$ and $(s, 0)$ with weights $X_\omega(s)$ and $1 - X_\omega(s)$, respectively:

$$\begin{aligned} Q(s, S \times \{1\}) &= Q(s, \{(s, 1)\}) = X_\omega(s) \\ Q(s, S \times \{0\}) &= Q(s, \{(s, 0)\}) = 1 - X_\omega(s). \end{aligned}$$

Intuitively, from state $s$, flip an $X_\omega(s)$-biased coin and enter state $(s, 1)$ on heads and $(s, 0)$ on tails. This $Q$ will turn out to be $Q_\omega : S_\omega \to T$ for the sequence $Q_n : S_n \to T$ we are about to define.

For $A \in \mathcal{F}_\omega$, we have

$$\begin{aligned} (Q \mathbin{;} g)(s, A) &= \int_{t \in T} Q(s, dt) \cdot g(t, A) \\ &= \sum_{i \in \{0,1\}} Q(s, \{(s, i)\}) \cdot 1(g(s, i), A) \\ &= X_\omega(s) \cdot 1(s, A) + (1 - X_\omega(s)) \cdot 1(s, A) \\ &= 1(s, A), \end{aligned}$$

therefore $g$ is the right inverse of $Q$.

Now define

$$Q_n = P_{n\omega} \mathbin{;} Q \qquad\qquad g_n(s, 0) = g_n(s, 1) = s.$$

Then

$$Q_m = P_{m\omega} \mathbin{;} Q = P_{mn} \mathbin{;} P_{n\omega} \mathbin{;} Q = P_{mn} \mathbin{;} Q_n,$$

and for $A \in \mathcal{F}_n$,

$$(Q_n \mathbin{;} g_n)(s, A) = (P_{n\omega} \mathbin{;} Q \mathbin{;} g)(s, A) = P_{n\omega}(s, A),$$

and since $P_{n\omega}$ agrees with $P_{nn}$ on $A \in \mathcal{F}_n$, this is $1(s, A)$, thus $g_n$ is the right inverse of $Q_n$.

Finally, to show that $Q_n$ encodes the martingale,

$$\begin{aligned} Q_n(s, S \times \{1\}) &= (P_{n\omega} \mathbin{;} Q)(s, S \times \{1\}) \\ &= \int_{t \in S} P_{n\omega}(s, dt) \cdot Q(t, S \times \{1\}) \\ &= \int_{t \in S} \mathcal{E}_n(s, dt) \cdot X_\omega(t) \\ &= \mathcal{E}(X_\omega \mid \mathcal{F}_n)(s) = X_n(s). \end{aligned}$$

### 3.4 An Alternative Construction

There is another construction equivalent to the one of §3.3 but closer in spirit to state transition systems as they arise in programming language semantics. As above, suppose we are given a martingale $(X_n, \mathcal{F}_n)$ on a probability space $(S, \mathcal{F}_\omega, \mu)$. For $s, t \in S$, define

$$\begin{aligned} s \equiv_n t &\Leftrightarrow \forall A \in \mathcal{F}_n \ (s \in A \Leftrightarrow t \in A) \\ [s]_n &= \{t \in S \mid s \equiv_n t\} \\ A/{\equiv_n} &= \{[s]_n \mid s \in A\}, \ A \in \mathcal{F}_n \\ S_n &= S/{\equiv_n} \\ \mathcal{B}_n &= \{A/{\equiv_n} \mid A \in \mathcal{F}_n\}. \end{aligned}$$

The Boolean operations on $\mathcal{F}_n$ respect the equivalence relation $\equiv_n$, therefore $(S_n, \mathcal{B}_n)$ is a measurable space.

For $A/{\equiv_n} \in \mathcal{B}_n$ and $m \le n$, we define

$$P_{mn}([s]_m, A/{\equiv_n}) = \mathcal{E}_m(s, A) \qquad f_{nm}([s]_n) = [s]_m.$$

The standard definition of conditional expectation as a Radon–Nikodým derivative ensures that $\mathcal{E}(A_n \mid \mathcal{F}_m)$ is $\mathcal{F}_m$-measurable, so if $s \equiv_n t$, then $\mathcal{E}(A_n \mid \mathcal{F}_m)$ takes the same value on $s$ and $t$, thus $P_{mn}$ is well defined up to a $\mu$-nullset. Also, $f_{nm}$ is well defined, since $\mathcal{F}_m \subseteq \mathcal{F}_n$, therefore $\equiv_n$ refines $\equiv_m$.

One can define $T$, $g$, $\mathcal{B}_T$, $\mathcal{G}_n$, $Q_n$, $g_n$, and $Q$ as in §3.3. The collection of maps $[\,\cdot\,]_\alpha : (S, \mathcal{F}_\alpha) \to (S_\alpha, \mathcal{B}_\alpha)$ for $\alpha \in \omega \cup \{\omega\}$ constitute a natural isomorphism between the two cocones $(S, \mathcal{F}_\alpha)$ and $(S_\alpha, \mathcal{B}_\alpha)$. As such, they preserve all relevant measure-theoretic structure.

## 4. An Application

In [12], operational and denotational semantics are given for probabilistic NetKAT, a probabilistic language for reasoning about packet switching networks. In this language, programs are interpreted as packet filters that consume an input set of packet histories (nonnull sequences of packets) and produce an output set of packet histories according to some probability distribution. There are atomic actions for querying fields of the packet header, modifying these fields, and duplicating the head packet.

Denotationally, programs $p$ are interpreted as Markov kernels $[\![p]\!] : 2^H \to 2^H$, where $H$ is the set of packet histories. The relevant measurable space is $(2^H, \mathcal{B})$ whose points $a \in 2^H$ are sets of packet histories and whose events $A \in \mathcal{B}$ are the Borel sets of the Cantor space $2^H$.

Programs can be composed using parallel composition (&), sequential composition (;, typically elided in expressions), probabilistic choice ($\oplus_r$), and iteration ($^*$). The parallel composition operator & supplants the choice operator + of Kleene algebra. Operationally, to determine $[\![p \mathbin{\&} q]\!](a, A)$, we sample $[\![p]\!](a, -)$ and $[\![q]\!](a, -)$ independently, then take the union of the two outcomes and ask whether it is in $A$. Denotationally,

$$[\![p \mathbin{\&} q]\!] = ([\![p]\!] \times [\![q]\!]) \mathbin{;} \bigcup,$$

Where $[\![p]\!] \times [\![q]\!] : 2^H \to (2^H)^2$ is the product kernel defined in §2 and $\bigcup : (2^H)^2 \to 2^H$ is the deterministic union kernel $\bigcup(a, b) = a \cup b$.

Of course, the most interesting part of the language is iteration. Intuitively, $p^*$ says "do $p$ zero or more times." The `while` loop is defined in terms of star as in propositional dynamic logic or Kleene algebra with tests:

$$\texttt{while}\ b\ \texttt{do}\ p = (bp)^* \bar{b},$$

where the test $b$ is the deterministic kernel $[\![b]\!](a, -) = \delta_{a \cap b}$. As noted in [12], the usual definition of star as a sum of powers does not work in the probabilistic case. Instead, we define an operational semantics in terms of an infinite stochastic process. To determine $[\![p^*]\!](c_0, A)$, we start with the input set $c_0 \in 2^H$ and create a sequence $c_0, c_1, c_2, \ldots$ inductively. After constructing $c_0, \ldots, c_n$, let $c_{n+1}$ be the outcome obtained by sampling $2^H$ according to the distribution $[\![p]\!](c_n, -)$. We continue this process forever to get an infinite sequence $c_0, c_1, c_2, \ldots \in (2^H)^\omega$. We take the union of the resulting sequence $\bigcup_n c_n$ and ask whether it is in $A$. The probability of this event is defined to be $[\![p^*]\!](c_0, A)$.

Unlike KAT and NetKAT, $[\![p^*]\!]$ is *not* the same as the infinite sum of powers $[\![\&_n\, p^n]\!]$. The latter fails to capture the sequential nature of iteration in the presence of probabilistic choice.

This process is shown to satisfy the crucial fixpoint equation

$$p^* = \texttt{skip} \mathbin{\&} pp^*, \tag{4.8}$$

where `skip` is the identity kernel, although it is by no means the only solution.

The operational definition of the star operator can be justified denotationally, but the formal development as given [12] is quite

involved. Now with Theorem 1, we can give a more streamlined account.

Let $\mathsf{hd} : (2^H)^\omega \to 2^H$ and $\mathsf{tl} : (2^H)^\omega \to (2^H)^\omega$ be the deterministic kernels

$$\mathsf{hd}(c_0, c_1, c_2, \ldots) = c_0 \qquad \mathsf{tl}(c_0, c_1, c_2, \ldots) = c_1, c_2, \ldots$$

The kernel $[\![p]\!] : 2^H \to 2^H$ gives rise to reversible kernels

$$[\![p_{mn}]\!] : (2^H)^{m+1} \to (2^H)^{n+1},\ m \leq n$$

defined inductively as follows:

$$[\![p_{0,n+1}]\!] = 1 \times ([\![p]\!] \,;\, [\![p_{0n}]\!]) \quad [\![p_{m+1,n+1}]\!] = \mathsf{hd} \times (\mathsf{tl} \,;\, [\![p_{mn}]\!])$$

and $[\![p_{00}]\!] = 1$. One can show that the composition property (3.1) is satisfied. By Theorem 1, we have kernels $[\![p_{m\omega}]\!] : (2^H)^{m+1} \to (2^H)^\omega$. Moreover,

$$\begin{aligned}[\![p_{0\omega}]\!] &= [\![p_{01}]\!] \,;\, [\![p_{1\omega}]\!] \\ &= (1 \times [\![p]\!]) \,;\, (\mathsf{hd} \times (\mathsf{tl} \,;\, [\![p_{0\omega}]\!])) \\ &= 1 \times ([\![p]\!] \,;\, [\![p_{0\omega}]\!]).\end{aligned}$$

Composing on the right with the (continuous) union operator $\bigcup : (2^H)^\omega \to 2^H$ gives $[\![p^*]\!]$ by definition:

$$[\![p^*]\!] = [\![p_{1\omega}]\!] \,;\, \bigcup : 2^H \to 2^H.$$

The fixpoint equation (4.8) is satisfied:

$$\begin{aligned}[\![p^*]\!] &= [\![p_{0\omega}]\!] \,;\, \bigcup \\ &= (1 \times ([\![p]\!] \,;\, [\![p_{0\omega}]\!])) \,;\, \bigcup \\ &= (1 \times ([\![p]\!] \,;\, [\![p_{0\omega}]\!] \,;\, \bigcup)) \,;\, \bigcup \\ &= (1 \,\&\, ([\![p]\!] \,;\, [\![p_{0\omega}]\!] \,;\, \bigcup)) \\ &= [\![\mathtt{skip} \,\&\, pp^*]\!].\end{aligned}$$

## 5. Conclusion

We have characterized the Kolmogorov extension theorem as a colimit-like construction in a category of Radon spaces and reversible Markov kernels. The Doob martingale convergence theorem is used to establish universality. These results provide a compositional denotational semantics for standard iteration operators in programming languages as a limit of finite approximants, even in the lossless case in which there is no natural approximation order. This is the case, for example, with the system reported in [12].

In Theorem 2(ii), the function $g$ would already be surjective and one would not need to take the completion $\widehat{T}$ if $T$ were already complete with respect to the pseudometric (3.4). It would be interesting to identify the weakest possible completeness assumptions on $T$ that guarantee this. Another intriguing question is whether a point-free approach as in [15] might yield a true colimit.

We have forgone several possible generalizations: continuous time, signed measures, and more general colimits. Such matters present themselves as interesting topics for future investigation.

## Acknowledgments

Sincere thanks to Nate Foster, Jean Goubault-Larrecq, Lionel Levine, Konstantinos Mamouras, Prakash Panangaden, John Pike, Mark Reitblatt, Alexandra Silva, Alex Simpson, Steffen Smolka, Ana Sokolova, and Harald Woracek.

## References

[1] P. Billingsley. *Convergence of probability measures*. Wiley, 2nd edition, 1999.

[2] S. Bochner. *Harmonic Analysis and the Theory of Probability*. Univ. California Press, 1955.

[3] A. Chakarov and S. Sankaranarayanan. Probabilistic program analysis with martingales. In N. Sharygina and H. Veith, editors, *25th Int. Conf. Computer Aided Verification (CAV 2013)*, volume 8044 of *Lecture Notes in Computer Science*, pages 511–526. Springer, 2013.

[4] A. Chakarov and S. Sankaranarayanan. Expectation invariants as fixed points of probabilistic programs. In *Static Analysis Symposium (SAS 2014)*, volume 8723 of *Lecture Notes in Computer Science*, pages 85–100. Springer, 2014.

[5] J. R. Choksi. Inverse limits of measure spaces. *Proc. London Math. Soc.*, 8:321–342, 1958.

[6] K. L. Chung. *A Course in Probability Theory*. Academic Press, 2nd edition, 1974.

[7] E.-E. Doberkat. *Stochastic Relations: Foundations for Markov Transition Systems*. Studies in Informatics. Chapman Hall, 2007.

[8] J. L. Doob. *Stochastic Processes*, volume 2. Wiley, New York; Chapman & Hall, London, 1953.

[9] D. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[10] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, 2nd edition, 1971.

[11] L. M. F. Fioriti and H. Hermanns. Probabilistic termination: Soundness, completeness, and compositionality. In S. K. Rajamani and D. Walker, editors, *Proc. 42nd ACM Symp. Principles of Programming Languages (POPL 2015)*, pages 489–501. ACM, January 2015.

[12] N. Foster, D. Kozen, K. Mamouras, M. Reitblatt, and A. Silva. Probabilistic NetKAT. Technical Report `http://hdl.handle.net/1813/40335`, Computing and Information Science, Cornell University, July 2015. 25th European Symposium on Programming (ESOP 2016), to appear.

[13] Z. Frolík. Projective limits of measure spaces. In L. L. Cam, J. Neyman, and E. L. Scott, editors, *Proc. 6th Berkeley Symp. Math. Statistics and Probability*. Univ. California Press, June–July 1970.

[14] P. R. Halmos. *Measure Theory*. Van Nostrand, 1950.

[15] M. Jackson. *A Sheaf Theoretic Approach to Measure Theory*. PhD thesis, University of Pittsburgh, 2006.

[16] A. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea, 2nd edition, 1956.

[17] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Prentice Hall, 1970.

[18] D. Kozen. Semantics of probabilistic programs. *J. Comput. Syst. Sci.*, 22:328–350, 1981.

[19] D. Kozen. A probabilistic *PDL. J. Comput. Syst. Sci.*, 30(2):162–178, April 1985.

[20] C. W. Lamb. Shorter notes: A short proof of the martingale convergence theorem. *Proceedings of the American Mathematical Society*, 38(1):215–217, 1973.

[21] J. D. Mallory and M. Sion. Limits of inverse systems of measures. *Annales de l'Institut Fourier*, 21(1):25–57, 1971.

[22] A. McIver and C. Morgan. *Abstraction, Refinement And Proof For Probabilistic Systems*. Springer, 2004.

[23] S.-T. C. Moy. Measure extensions and the martingale convergence theorem. *Proc. Amer. Math. Soc.*, 4:902–907, December 1953.

[24] P. Panangaden. The category of Markov kernels. *Electronic Notes in Theoretical Computer Science*, 22:171–187, 1999.

[25] P. Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.

[26] Y. V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.*, 1(2):157–214, 1956.

[27] M. Rao. Projective limits of probability spaces. *Journal of Multivariate Analysis*, 1(1):28–57, 1971.