Abstract

The aim of this paper is to present an elementary computable theory of probability, random variables and stochastic processes. The probability theory is based on existing approaches using valuations and lower integrals. Various approaches to random variables are discussed, including the approach based on completions in a Polish space. We apply the theory to the study of stochastic dynamical systems in discrete-time, and give a brief exposition of the Wiener process as a foundation for stochastic differential equations. The theory is based within the framework of type-two effectivity, so has an explicit direct link with Turing computation, and is expressed in a system of computable types and operations, so has a clean mathematical description.

1 Introduction

In this paper, we present a computable theory of probability, random variables and stochastic processes, with the aim of providing a theoretical foundation for the rigorous numerical analysis of discrete-time continuous-state Markov chains and stochastic differential equations. The first part of the paper provide an exposition of the approach to probability distributions using valuations and the development of integrals of positive lower-semicontinuous and of bounded continuous functions, and on the approach to random variables as limits of almost-everywhere defined continuous partial functions. In the second part, we show that our approach allows one to very quickly derive computability results for discrete-time stochastic processes. In the third part, we provide a new construction of the Wiener process in which sample paths are effectively computable, and use this to show that the solutions to stochastic differential equations can be effectively computed.

An early approach to constructive measure theory was developed in [BC72]; see also [BB74]. The standard approach to a constructive theory of probability measures, as developed in [JP89, Eda95, SS06, Esc09], is through valuations, which are measures restricted to open sets. The most straightforward approach to integration is the Choquet or horizontal integral, a lower integral introduced within the framework of domain theory in [Tix95]; see also [Kön97, Law04]. The lower integral on valuations in the form used here was given in [Vic08]. Relationships between the constructive and classical approaches were given in [Eda95]. Explicit representations of valuations within the framework of type-two effectivity were given in [Sch07], and representation of probability measures using probabilistic processes were given by [SS06]. In [Esc09], a language EPCL for nondeterministic and probabilistic computation was given, based on the PCL language of [Esc04].
A constructive theory of measurable functions was also developed in [BC72]. However, the approach we use here, in which measurable functions are defined as limits of effectively-converging Cauchy sequences of continuous functions was introduced in [Spi03] and further developed in [Spi06 | CS09]. Random variables over discrete domains were defined in [Mis07], based on work of [Var02]. This was extended to random variables over continuous domains in [GLV11], but the construction allows only for continuous random variables, and is overly-restrictive in practice.

To the best of our knowledge there has been relatively little work on constructive and computable approaches to stochastic processes. An early constructive theory of discrete-time stochastic processes with focusing on stopping times was given in [Cha72]. A fairly comprehensive theory though technically advanced theory based on stochastic relations is developed in [Dob07]; the approach here is considerably simpler. The monadic properties of the lower integral on valuations, as noted by [Vic11], and of the completion construction [OS10].

We use the framework of type-two effectivity (TTE), in which computations are performed by Turing machines working on infinite sequences, as a foundational theory of computability. We believe that this framework is conceptually simpler for non-specialists than the alternative of using a domain-theoretic framework. Since in TTE we work entirely in the class of quotients of countably-based (QCB) spaces, which form a cartesian closed category, many of the basic operations can be carried out using simple type-theoretic constructions such as the λ-calculus.

We assume that the reader has a basic familiarity with classical probability theory (see e.g. [Shi95]) and stochastic processes (see [B72, WS1, GS0]). Much of this article is concerned with giving computational meaning to classical concepts and arguments. The main difficulty lies in the use of σ-algebras in classical probability, which have poor computability properties. Instead, we use only topological constructions, which can usually be effectivised directly. In particular, we define types of measurable functions as a completion of types of continuous functions.

2 Computable Analysis

In the theory of type-two effectivity, computations are performed by Turing machines acting on sequences over some alphabet Σ. A computation performed by a machine M is valid on an input p ∈ Σω if the computation does not halt, and writes infinitely many symbols to the output tape. A type-two Turing machine therefore performs a computation of a partial function η : Σω → Σω; we may also consider multi-tape machines computing η : (Σω)n → (Σω)m. It is straightforward to show that any machine-computable function Σω → Σω is continuous on its domain.

In order to relate Turing computation to functions on mathematical objects, we use representations of the underlying sets, which are partial surjective functions δ : Σω → X. An operation X → Y is (δX; δY )-computable if there is a machine-computable function η : Σω → Σω with dom(η) ⊆ dom(δX) such that δY ◦ η = f ◦ δX on dom(δX). If X is a topological space, we say that a representation δ of X is an admissible quotient representation if (i) whenever f : X → Y is such that f ◦ δ is continuous, then f is continuous, and (ii) whenever φ : Σω → X is continuous, there exists continuous η : Σω → Σω such that φ = δ ◦ η. A computable type is a pair (X, [δ]) where X is a space and [δ] is an equivalence class of admissible quotient representations of X.

The category of computable types with continuous functions is Cartesian closed, and the computable functions yield a Cartesian closed subcategory. For any types X, Y there exist a canonical product type X × Y with computable projections πX : X × Y → X and πY : X × Y → Y, and a canonical exponential type YX such that evaluation ε : YX × X → Y : (f, x) → f(x) is computable. Since objects of the exponential type are continuous function from X to Y, we also
denote \( \mathbb{Y}^X \) by \( X \to \mathbb{Y} \) or \( C(X; \mathbb{Y}) \). There is a canonical equivalence between \( (X \times \mathbb{Y}) \to \mathbb{Z} \) and \( X \to (\mathbb{Y} \to \mathbb{Z}) \) given by \( \tilde{f}(x) : \mathbb{Y} \to \mathbb{Z} : \tilde{f}(x)(y) = f(x, y) \).

There are canonical types representing basic building blocks of mathematics, including the natural number type \( \mathbb{N} \) and the real number type \( \mathbb{R} \). We use a three-valued logical type with elements \( \{ F, T, \bot \} \) representing false, true, and indeterminate or unknowable, and its subtypes the Boolean type \( \mathbb{B} \) with elements \( \{ F, T \} \) and the Sierpinski type \( \mathbb{S} \) with elements \( \{ T, \bot \} \). Given any type \( X \), we can identify the type \( O(X) \) of open subsets \( U \) of \( X \) with \( X \to \mathbb{S} \) via the characteristic function \( \chi_U \). Further, standard operations on these types, such as arithmetic on real numbers, are computable.

We shall also need the type \( \mathbb{H} \equiv \mathbb{R}^+_{<\infty} \) of positive real numbers with infinity under the lower topology. The topology on the lower halfline \( \mathbb{H} \) is the topology of lower convergence, with open sets \( (a, \infty] \) for \( a \in \mathbb{R}^+ \). We note that the operators + and \( \times \) are computable on \( \mathbb{H} \), where we define \( 0 \times \infty = \infty \times 0 = 0 \), as is countable supremum \( \sup : \mathbb{H}^\omega \to \mathbb{H} \), \( \{ x_0, x_1, x_2, \ldots \} \mapsto \sup \{x_0, x_1, x_2, \ldots \} \). Further, abs : \( \mathbb{R} \to \mathbb{H} \) is computable, as is the embedding \( \mathbb{S} \hookrightarrow \mathbb{H} \) taking \( T \mapsto 1 \) and \( \bot \mapsto 0 \). We let \( \mathbb{I}_x \) be the unit interval \( [0, 1] \), again with the topology of lower convergence with open sets \( (a, 1] \) for \( a \in [0, 1) \), and \( \mathbb{I}_x \) the interval with the topology of upper convergence.

A computable metric space is a pair \((X, d)\) where \( X \) is a computable type, and \( d : X \times X \to \mathbb{R}^+ \) is a computable metric, such that the extension of \( d \) to \( X \times A(X) \) defined by \( d(x, A) = \inf \{ d(x, y) \mid y \in A \} \) is computable as a function into \( \mathbb{R}^+_{<\infty} \). This implies that given an open set \( U \) we can compute \( \epsilon > 0 \) such that \( B_\epsilon(x) \subseteq U \), which captures the relationship between the metric and the open sets. The effective metric spaces of [Wei99] are a concrete class of computable metric space.

Throughout this paper we shall use the term “compute” to indicate that a formula or procedure can be effectively carried out in the framework of type-two effectivity. Other definitions and equations may not be possible to verify constructively, but hold from axiomatic considerations.

### 3 Computable Measure Theory

The main difficulty with classical measure theory is that Borel sets and Borel measures have very poor computability properties. Although a computable theory of Borel sets was given in [Bra05], the measure of a Borel set is in general not computable in \( \mathbb{R} \). However, we can consider an approach to measure theory in which we may only compute the measure of open sets. Since open sets are precisely those which can be approximated from inside, we expect to be able to compute lower bounds for the measure of an open set, but not upper bounds. The above considerations suggest an approach which has become standard in computable measure theory, namely that using valuations [JP89, Eda95, SS06, Esc09].

**Definition 1** (Valuation). The type of (continuous) valuations on \( X \) is the subtype of continuous functions \( \nu : O(X) \to \mathbb{H} \) satisfying \( \nu(\emptyset) = 0 \) and the modularity condition \( \nu(U) + \nu(V) = \nu(U \cup V) + \nu(U \cap V) \) for all \( U, V \in O(X) \).

A valuation \( \nu \) on \( X \) is finite if \( \nu(X) \) is finite, effectively finite if \( \nu(X) \) is a computable real number, and locally finite if \( \nu(U) < \infty \) for any pre-compact \( U \). An effectively finite valuation induces an upper-valuation on closed sets \( \nu : A(X) \to \mathbb{R}^+ \) by \( \nu(A) = \nu(X) - \nu(X \setminus A) \). The following proposition gives standard monotonicity and convergence properties for type of valuations.

**Proposition 2.** Let \( \nu : O(X) \to \mathbb{H} \) be continuous. Then \( \nu \) satisfies the monotonicity condition \( \nu(U) \leq \nu(V) \) whenever \( U \subset V \), and the continuity condition \( \nu(\bigcup_{n=0}^\infty U_n) = \lim_{n \to \infty} \nu(U_n) \) whenever \( U_n \) is an increasing sequence of open sets.
The proof is immediate from properties of continuous functions into $\mathbb{H}$. An immediate consequence if that $\nu(U) \leq \bar{\nu}(A)$ whenever $U \subset A$.

An explicit representation of valuations $\mathcal{M}[0,1]$ on the unit interval was given in [Wei99] using the basic open sets $I_{a,b,r} = \{ \nu \in \mathcal{M}[0,1] \mid \nu(a,b) > r \}$ for $a,b,r \in \mathbb{Q}$ with $0 \leq a < b \leq 1$ and $r > 0$. Various representations for arbitrary spaces were given in [Sch07].

In [Eda95, Section 4], a notion of integral $\mathcal{C}_{bd}(X;\mathbb{R}) \to \mathbb{R}$ on continuous bounded functions was introduced based on the approximation by point measures. The following theorem [Eda95, Proposition 5.2] shows that valuations and measures are equivalent on locally-compact Hausdorff spaces.

**Theorem 3.** On a locally compact countably-based Hausdorff space, bounded Borel measures and continuous valuations coincide.

In [AM02], it was shown that any continuous valuation on a locally compact sober space extends to a unique Borel measure. These results provides a link with classical measure theory, but are not needed for a purely constructive approach; valuations themselves are the objects of study, and we only consider the measure of open and closed sets.

The following result shows that the measure of a sequence of small sets approaches zero. We say a type $X$ is effectively regular if for any open set $U$, we can construct a sequence of open sets $V_n$ and closed sets $A_n$ such that $V_n \subset A_n \subset V_{n+1} \subset U$ and $\bigcup_{n=1}^{\infty} V_n = U$.

**Lemma 4.** Let $X$ be an effectively regular space, and $\nu$ a finite valuation on $X$. Then if $U_n$ is any subset of open sets such that $U_{n+1} \subset U_n$ and $\bigcap_{n=0}^{\infty} U_n = \emptyset$, then $\nu(U_n) \to 0$ as $n \to \infty$.

**Proof.** Since $X$ is effectively regular, we can effectively find open sets $V_{n,k}$ and closed sets $A_{n,k}$ such that $V_{n,k} \subset A_{n,k} \subset V_{n,k+1} \cap U_n$ and $\bigcup_{k=0}^{\infty} V_{n,k} = U_n$. Then $\lim_{k \to \infty} \nu(V_{n,k}) = \nu(U_n)$. Suppose $\inf_{n \in \mathbb{N}} \nu(U_n) \epsilon > 0$. Choose a sequence $\delta_n$ such that $\sum_{n=0}^{\infty} \delta_n = \delta < \epsilon$, and sets $V_n \subset A_n \subset U_n$ such that $\nu(V_n) \geq \nu(U_n) - \delta_n$. Then $\bigcap_{n=0}^{\infty} A_n = U_N \setminus \bigcup_{n=0}^{N} (U_n \setminus A_n)$, so $\nu(\bigcap_{n=0}^{\infty} A_n) \geq \nu(U_N) - \sum_{n=0}^{N} \nu(U_n \setminus A_n) \geq \nu(U_N) - \sum_{n=0}^{N} (\nu(U_n) - \nu(A_n)) \geq \nu(U_N) - \sum_{n=0}^{\infty} \delta_n \geq \epsilon - \delta > 0$. Since $\nu$ is upper-continuous on closed sets, $\nu(\bigcap_{n=0}^{\infty} A_n) \geq \epsilon - \delta > 0$. Then $\bigcap_{n=0}^{\infty} A_n \neq \emptyset$, contradicting $\emptyset = \bigcap_{n=0}^{\infty} U_n \supset \bigcap_{n=0}^{\infty} A_n$.

Just as for classical probability, we say (open) sets $U_1,U_2$ are independent if $\nu(U_1 \cap U_2) = \nu(U_1)\nu(U_2)$. A conditional valuation is a function $\nu(\cdot|\cdot)$ such that $\nu(U|V)$ satisfies $\nu(U|V)\nu(V) = \nu(U \cap V)$ for open $U$.

We can define a notion of integration for positive lower-semicontinuous functions by the Choquet or horizontal integral; see [Tix95] [Law04] [Vic08].

**Definition 5** (Lower horizontal integral). Given a valuation $\nu : (X \to \mathbb{S}) \to \mathbb{H}$, define the lower integral $(X \to \mathbb{H}) \to \mathbb{H}$ by $\int_X \psi \, d\nu = \sup \{ \sum_{m=1}^{\infty} (p_m - p_{m-1}) \nu(\psi^{-1}(p_m, \infty)) \mid (p_0, \ldots, p_n) \in \mathbb{Q}^+ \text{ and } 0 = p_0 < p_1 < \cdots < p_n \}$. (1)

Note that we could use any dense set of computable positive real numbers, such as the dyadic rationals $\mathbb{Q}_2$, instead of the rationals in (1). Since each sum is computable, and the supremum of countably many elements of $\mathbb{H}$ is computable, we immediately obtain:

**Proposition 6.** Given names of a valuation $\nu$ in $(X \to \mathbb{S}) \to \mathbb{H}$ and of a function $\psi$ in $X \to \mathbb{H}$, the lower integral $\int_X \psi \, d\nu$ is computable in $\mathbb{H}$. 

4
Note that although an alternative form for the sum is given through the equality
\[\sum_{m=1}^{n}(p_m - p_{m-1})\nu(\phi^{-1}(p_m, \infty)) = \sum_{m=1}^{n}p_m\nu(\phi^{-1}(p_m, p_{m+1}))\]
where \(p_{n+1} = \infty\), the lower integral cannot be computed in this form since \(\nu(\phi^{-1}(p_m, p_{m+1})) = \nu(\phi^{-1}(p_m, \infty)) - \nu(\phi^{-1}(p_{m+1}, \infty))\) is uncomputable in \(\mathbb{R}\).

It is fairly straightforward to show that the integral is linear,
\[\int_X (a_1\psi_1 + a_2\psi_2) \, d\nu = a_1\int_X \psi_1 \, d\nu + a_2\int_X \psi_2 \, d\nu\] for all \(a_1, a_2 \in \mathbb{R}\) and \(\psi_1, \psi_2 : \mathbb{R} \to \mathbb{R}\).

If \(X_U\) is the characteristic function of a set \(U\), then \(\int_X X_U \, d\nu = \nu(U)\), and it follows that if \(\phi = \sum_{i=1}^{n} a_i X_{U_i}\) is a step function, then \(\int_X \phi \, d\nu = \sum_{i=1}^{n} a_i \nu(U_i)\).

Given a linear functional \(\mu : (X \to \mathbb{R}) \to \mathbb{R}\), we can define a function \(\mathcal{O}(X) \to \mathbb{R}\) by \(U \mapsto \mu(X_U)\) for \(U \in \mathcal{O}(X)\). By linearity,
\[\mu(X_U) + \mu(X_V) = \mu(X_{U \cup V}) + \mu(X_{U \cap V}).\]

Hence \(\mu\) induces a valuation on \(X\). We therefore obtain a computable equivalence between the type of valuations and the type of positive linear lower-semicontinuous functionals:

**Theorem 7.** The type of valuations \((X \to S) \to \mathbb{R}\) is computably equivalent to the type of linear functionals \((X \to \mathbb{R}) \to \mathbb{R}\).

Types of the form \((X \to \mathbb{T}) \to \mathbb{T}\) for a fixed type \(\mathbb{T}\) form a monad over \(X\), and are particularly easy to work with.

Our lower integral on positive lower-semicontinuous functions can be extended to bounded functions as follows:

**Definition 8** (Bounded integration). A measure \(\mu\) on \(X\) is effectively finite if there is a (known) computable real \(c \in \mathbb{R}\) such that \(\mu(X) \leq c\).

An upper-semicontinuous function \(f : X \to \mathbb{R}_+\) is effectively bounded if there is a (known) computable real \(b \in \mathbb{R}\) such that \(f(x) < b\) for all \(x \in X\). Then the function \(b - f : X \to \mathbb{R}_+\) is computable (given names of \(b\) and \(f\)), and we define the integral \(\mathcal{C}_{bd}(X; \mathbb{R}_+) \to \mathbb{R}_+\) by
\[\int_X f(x) \, d\mu(x) = bc - \int_X (b - f(x)) \, d\mu(x).\]

Similarly, if \(f : X \to \mathbb{R}_-\) has a computable lower bound \(a\), we define the integral \(\mathcal{C}_{bd}(X; \mathbb{R}_-) \to \mathbb{R}_-\) by
\[\int_X f(x) \, d\mu(x) = \int_X (a + f(x)) \, d\mu(x) - ac.\]

A continuous function \(f : X \to \mathbb{R}\) is effectively bounded if there are a (known) computable reals \(a, b \in \mathbb{R}\) such that \(a < f(x) < b\) for all \(x \in X\). Then we define the integral \(\mathcal{C}_{bd}(X; \mathbb{R}) \to \mathbb{R}\) by
\[\int_X f(x) \, d\mu(x) = \int_X (a + f(x)) \, d\mu(x) - ac = bc - \int_X (b - f(x)) \, d\mu(x).\]

It is clear that the integrals defined above are computable in \(\mathbb{R}_\ge\) and that the lower and upper integrals agree if \(f\) is continuous. where \(a < f(X) < b\) are lower and upper bounds for \(f\), and \(c = \mu(X)\). If \(X\) is compact, then any (semi)continuous function is effectively bounded, so the integrals always exist.

In order to define a valuation given a positive linear functional \(\mathcal{C}_{cpt}(X; \mathbb{R}) \to \mathbb{R}\), we need some way of approximating the characteristic function of an open set by continuous functions. If \(X\) is effectively regular, then given any open set \(U\), we can construct an increasing sequence of closed
sets \( A_n \) such that \( \bigcup_{n \to \infty} A_n = U \). Further, if a type \( X \) is effectively quasi-normal if given closed sets \( A_0 \) and \( A_1 \), we can construct a continuous function \( \phi : X \to [0, 1] \) such that \( \phi(A_0) = \{0\} \) and \( \phi(A_1) = \{1\} \) using an effective Uryshon lemma; see [Sch09] for details. Then a bounded positive linear functional \( L : C_{cpt}(X; \mathbb{R}) \to \mathbb{R} \) effectively induces a valuation \( \nu : \mathcal{O}(X) \to \mathbb{H} \) on \( X \) by
\[
\nu(U) = \sup\{L(\psi) \mid \phi \in \mathcal{C}(X; [0, 1]) \land \phi(X \setminus U) = \{0\}\}.
\]
This construction extends to locally-finite measures and compactly-supported functions \( C_{cpt}(X \to \mathbb{R}) \). We therefore have an effective version of the Riesz representation theorem:

**Theorem 9.** Suppose \( X \) is an effectively locally-compact type. Then type of locally-finite valuations \( (X \to \mathbb{S}) \to \mathbb{H} \) is effectively equivalent to the type of positive linear functionals \( C_{cpt}(X \to \mathbb{R}) \to \mathbb{R} \) on continuous functions of compact support.

We consider lower-semicontinuous functionals \( (X \to \mathbb{H}) \to \mathbb{H} \) to be more appropriate as a foundation for computable measure theory than the continuous functionals \( (X \to \mathbb{R}) \to \mathbb{R} \), since the equivalence given by Theorem 9 is entirely independent of any assumptions on the type \( X \) whereas the equivalence of Theorem 9 requires extra properties of \( X \) and places restrictions on the function space.

A similar approach to probability measures [Esc09] based on type theory identified the type of probability measures on the Cantor space \( \Omega = \{0, 1\}^\omega \) with the type of integrals \((\Omega \to I) \to I\) where \( I = [0, 1] \) is the unit interval.

### 4 Computable Random Variables

A computable theory of random variables should, at a minimum, enable us to perform certain basic operations, including:

(i) Given a random variable \( X \) and open set \( U \), compute lower-approximation to \( P(X \in U) \).

(ii) Given random variables \( X_1, X_2 \), compute the random variable \( X_1 \times X_2 \) giving the joint distribution.

(iii) Given a random variable \( X \) and a continuous function \( f \), compute the image \( f(X) \).

(iv) Given a probability distribution \( \nu \) on a sufficiently nice space \( X \), compute a random variable \( X \) with distribution \( \nu \).

(v) Given random variables \( X_1, X_2 \), compute a random variable \( X_1 \otimes X_2 \) such that \( P(X_1 \otimes X_2) \in (U_1 \times U_2) = P(X_1 \in U_2)P(X_2 \in U_2) \).

Property (iii) states that we can compute the distribution of a random variable. In particular, property (iii) implies that a random variable is more than its distribution; it also allows us to compute its joint distribution with another random variable. Property (iii) also imply that for random variables \( X_1, X_2 \) on a computable metric space \( (X, d) \), the random variable \( d(X_1, X_2) \) is computable in \( \mathbb{R}^+ \), so the probability \( P(d(X_1, X_2) < \epsilon) \) is computable in \( \mathbb{I}_\omega \), and \( P(d(X_1, X_2) \leq \epsilon) \) is computable in \( \mathbb{I}_+ \).

The standard approach to probability theory used in classical analysis is to define random variables as measurable functions over a base probability space. Given types \( X \) and \( Y \), a representation of the Borel measurable functions \( f : X \to Y \) was given in [Bra05], but this does not allow one to compute lower bounds for the measure of \( f^{-1}(V) \) for \( V \in \mathcal{O}(Y) \). Ideally, one would like a representation of bounded measurable functions \( f : X \to \mathbb{R} \) such that for every
finite measure \( \mu \) on \( X \), the integral \( \int_X f(x) \, d\mu(x) \) is computable. But then \( f(y) = \int_X f(x) \, d\delta_y(x) \) would be computable, so \( f \) would be continuous. Any effective approach to measurable functions and integration must therefore take some information about the measure into account.

In the approach of [BC72], a notion of full-measure set was given independently of a specific measure, but this introduces additional technical details. In the approach of [GLV11] a notion of continuous random variable was introduced as a continuous function on \( \text{supp}(\nu) \), where \( \nu \) is a valuation on the Cantor space \( \{0,1\}^\omega \). However, in order to define a joint distribution, we need to fix the measure \( \nu \), but for fixed \( \nu \), the set of continuous functions is not expressive enough. For example, using the standard probability measure \( P \) on \( \{0,1\}^\omega \), there is no continuous total function \( X : \{0,1\}^\omega \to \{0,1\} \) such that \( P(X(\omega) = 1) = 1/3 \). In [Spi03], a type of integrable real-valued functions is defined as the completion of the continuous functions under the metric defined by \( d(f,g) = \int_X |f(x) - g(x)| \, d\mu(x) \), and extended to a type of measurable functions. This approach is natural, constructive, and allows for integrals of measurable functions to be computed; it is this approach we shall use here.

We will consider random variables on a fixed probability space \((\Omega, P)\). Since any probability distribution on a Polish space is equivalent to a distribution on the standard Lesbesgue-Rokhlin probability space, it is reasonable to take the base space \( \Omega \) to be the Cantor space \( \{0,1\}^\omega \).

**Definition 10** (Continuous random variable). An **continuous random variable** on \((\Omega, P)\) with values in \( X \) is a continuous function \( X : \Omega \to X \).

We will sometimes write \( P(X \in U) \) as a shorthand for \( P(\{\omega \in \Omega \mid X(\omega) \in U\}) \). Continuous random variables \( X \) and \( Y \) are considered equal if \( P(\{\omega \in \Omega \mid X(\omega) \neq Y(\omega)\}) = 0 \). In other words, \( X \) and \( Y \) are **almost-surely equal**.

Suppose \( X \) is a Polish space, i.e. a space which is separable and complete under the metric \( d \). Define the **Fan metric** on continuous random variables by

\[
d(X,Y) = \sup\{\varepsilon \in \mathbb{Q}^+ \mid P(\{\omega \in \Omega \mid d(X(\omega), Y(\omega)) > \varepsilon\}) > \varepsilon\} \\
= \inf\{\varepsilon \in \mathbb{Q}^+ \mid P(\{\omega \in \Omega \mid d(X(\omega), Y(\omega)) \geq \varepsilon\}) < \varepsilon\}.
\]

Given the probability distribution (valuation) \( P \) and a computable metric \( d : X \times X \to \mathbb{R}^+ \), the Fan metric on continuous random variables is easily seen to be computable. The convergence relation defined by the Fan metric corresponds to convergence in probability. As an alternative to using the Fan metric, we can consider a uniform structure on \( X \), or, if the metric \( d \) on \( X \) is bounded, the distance \( d(X,Y) : = \int_\Omega d(X(\omega), Y(\omega)) \, dP(\omega) \).

**Definition 11** (Measurable random variable). The type of **measurable random variables** is the effective completion of the type of continuous random variables under the Fan metric (3). We write \( X : \Omega \rightharpoonup X \) if \( X \) is a measurable random variable taking values in \( X \), and let \( \mathcal{R}(X) \) be the type of measurable random variables with values in \( X \).

In other words, a random variable is represented by a sequence \((X_0, X_1, X_2, \ldots)\) of continuous random variables satisfying \( d(X_m, X_n) < 2^{-\min(m,n)} \), and two such sequences are equivalent (represent the same random variable) if \( d(X_{1,n}, X_{2,n}) \to 0 \) as \( n \to \infty \).

By standard results on the completion, the Fan metric on continuous random variables extends computably to measurable random variables. For if \( m > n \), then \( |d(X_m, Y_m) - d(X_n, Y_n)| \leq d(X_m, X_n) + d(Y_m, Y_n) \leq 2 \cdot 2^{-n} \), so is an effective Cauchy sequence converging to a value we define as \( d(\lim_{n \to \infty} X_n, \lim_{n \to \infty} Y_n) \). Further, if \((X_0, X_1, X_2, \ldots)\) is an effective Cauchy sequence converging to \( X_\infty \), then \( d(X_n, X_\infty) \leq 2^{-n} \).

**Remark 12.** Although a measurable random variable \( X \) is defined relative to the underlying space \( \Omega \), we cannot in general actually compute \( X(\omega) \) in any meaningful sense for fixed \( \omega \in \Omega! \) The expression \( X(\omega) \) only makes sense for random variables given as continuous functions \( \Omega \to X \).
We can relax the condition that a random variable is defined on the entire space $\Omega$.

**Definition 13** (Piecewise-continuous random variable). An *piecewise-continuous random variable* on $(\Omega, P)$ with values in $X$ is a continuous partial function $X : \Omega \to X$ such that $\text{dom}(X) \in \mathcal{O}(\Omega)$ and $P(\text{dom}(X)) = 1$.

We use the terminology “piecewise-continuous” since $X : \omega \to X$ may arise as the restriction of a piecewise-continuous function to its continuity set.

**Proposition 14.** Any piecewise-continuous random variable is a measurable random variable.

*Proof.* Let $X$ be a piecewise-continuous random variable. Let $x_\omega$ be an arbitrary point of $X$, which can be effectively constructed as $x_\omega = X(\omega)$ for some $\omega \in \text{dom}(X)$. Compute an increasing sequence of clopen sets $W_n \subset \Omega$ such that $\mu(W_n) > 1 - 2^{-n}$. Define $X_n(\omega) = X(\omega)$ for $\omega \in W_n$, and $X_n(\omega) = x_\omega$ otherwise. Then clearly $P(d(X_m, X_n) > 0) \leq \mu(\Omega \setminus W_n) \leq 2^{-n}$ for $m \leq n$, so $(X_n)_{n \in \mathbb{N}}$ is an effective Cauchy sequence of random variables, and converges pointwise to $X$ on $\text{dom}(X)$. \[\square\]

We now consider the probability distribution of a measurable random variable. Let $X$ be a computable metric space. For a closed set $A$, define $\overline{\mathcal{N}}_\epsilon(A) := \{x \in X \mid d(x, A) \leq \epsilon\}$, and for an open set $U$ define $I_\epsilon(U) := X \setminus (\overline{\mathcal{N}}_\epsilon(X \setminus U)) = \{x \in U \mid \exists \delta > 0, B(x, \epsilon + \delta) \subset U\}$. Since $d(x, A)$ is computable in $R^{+\infty}$ by definition of a computable metric space, $\overline{\mathcal{N}}_\epsilon(A)$ is computable as a closed set, so $I_\epsilon(U)$ is computable as an open set. Note that $I_{\epsilon_1 + \epsilon_2}(U) \subset I_{\epsilon_1}(I_{\epsilon_2}(U))$.

**Definition 15** (Distribution of a measurable random variable). For a measurable random variable $X$, define its *distribution* by

$$P(X \in U) = \sup\{P(Y \in V) - \epsilon \mid \epsilon \in \mathbb{Q}^+, V \subset I_\epsilon(U), d(Y, X) < \epsilon\},$$

where $Y$ ranges over continuous random variables and $V$ over open sets.

**Theorem 16** (Computability of distribution). Suppose $(X, d)$ is a computable metric space. The distribution of a measurable random variable $X$ taking values in $X$ is a valuation, and is computable from a name of $X$. If $X$ is a continuous random variable, then $P(X \in U) = P(\{\omega \in \Omega \mid X(\omega) \in U\})$.

*Proof.* Suppose $X, Y$ are continuous random variables, $d(X, Y) < \epsilon$ and $V \subset I_\epsilon(U)$. Then $P(X \in U) \geq P(Y \in V \land d(X, Y) < \epsilon) \geq P(Y \in V) - P(d(X, Y), Y(\omega)) \geq P(Y(\omega) \in V) - \epsilon$.

Now take $(X_n)_{n \in \mathbb{N}}$ to be any sequence of continuous random variables converging effectively to a measurable random variable $X$. By definition of $P(X \in U)$, we have $P(X \in U) \geq P(X_m \in I_{2^{-m}}) - 2^{-m}$ for all $m$.

Fix $\delta > 0$ and take a continuous random variable $Y$ such that $d(X, Y) < \epsilon$ and $P(Y \in I_\epsilon(U)) - \epsilon > P(X \in U) - \delta$. By taking $m$ sufficiently large, we can ensure that $P(Y \in I_{2^{m+2}}(U)) - \epsilon > P(X \in U) - \delta$. Then since $d(X_m, Y) \leq \epsilon + 2^{-m}$, we have $P(X_m \in I_{2^{-m}}) - 2^{-m} \geq P(Y \in I_{2^{m+2}}(U)) - \epsilon + 2^{-m}) > P(X \in U) - \delta + 2^{-m})$. Since $\delta$ is arbitrary and $m$ may be taken arbitrarily large, we have $P(X \in U) = \sup_{n \in \mathbb{N}} P(X_n \in I_{2^{-n}}(U)) - 2^{-n}$, so is computable in $I_{\epsilon}$.

If $X$ is a continuous random variable, $P(X \in I_\epsilon(U)) \lor P(X \in U)$ as $\epsilon \to 0$ by continuity, so taking $X_n = X_{\infty} = X$ above, we have $P(X \in U) = \lim_{n \to \infty} P(X \in I_\epsilon(U)) - \epsilon = P(X \in U)$. \[\square\]

**Remark 17.** Although the notation $P(X \in U)$ suggests that we can define a measurable random variable $X$ as a *function* from $\Omega$ to $X$, such a function could not be constructed in general, and is not required to compute probabilities.
Corollary 18. If $X$ is a random variable and $A$ is a closed set, then $\mathbb{P}(X \in A)$ is computable in $[0, 1]_\to$.

If $X_1$, $X_2$ are continuous random variables, define the product $X_1 \times X_2$ as the functional product

$$(X_1 \times X_2)(\omega) = (X_1(\omega), X_2(\omega)).$$

Our second main result is that we can also compute products of measurable random variables.

Theorem 19 (Computability of products). If $X_n$ and $Y_n$ are effective Cauchy sequences of continuous random variables, then $X_n \times Y_n$ is an effective Cauchy sequence.

Proof. If $m > n$, then $\mathbb{P}(d(X_{m+1} \times Y_{m+1}, X_{n+1} \times Y_{n+1}) \geq 2^{-n}) = \mathbb{P}(d(X_{m+1}, X_{n+1}) \geq 2^{-n} \lor d(Y_{m+1}, Y_{n+1}) \geq 2^{-n}) \leq \mathbb{P}(d(X_{m+1}, X_{n+1}) \geq 2^{-(n+1)} + \mathbb{P}(d(Y_{m+1}, Y_{n+1}) \geq 2^{-(n+1)})) \leq 2 \cdot 2^{-(n+1)} = 2^{-n}$. Hence $d(X_m \times Y_m, X_n \times Y_n) \leq 2^{-n}$ as required.

If $X : \Omega \to \mathbb{X}$ is a continuous random variable, and $f : \mathbb{X} \to \mathbb{Y}$ is continuous, then $f \circ X$ is a continuous random variable. By taking limits of effective Cauchy sequences, it is clear that we can define the image of a measurable random variable under a uniformly continuous function. However, it is possible to compute the image under an arbitrary continuous function.

Theorem 20 (Computability of continuous images). If $X_n$ is an effective Cauchy sequence of continuous random variables and $f$ is continuous, then $f \circ X_n$ is an effective Cauchy sequence. Further, for any open $V \subset \mathbb{Y}$, $\mathbb{P}(\lim_{n \to \infty} f \circ X_n \in V) = \mathbb{P}(\lim_{n \to \infty} X_n \in f^{-1}(V))$.

The proof is based on a non-effective version of this result from [MW43].

Proof. Consider the open subsets of $E_1$ defined as

$$B_{\delta,\epsilon}(f) = \{x \in E_1 \mid \exists y \in E_1, d(x, y) < \delta \land d(f(x), f(y)) > \epsilon\}.$$

Since $f$ is continuous, for all $\epsilon > 0$, $\bigcap_{\delta>0} B_{\delta,\epsilon}(f) = \emptyset$.

For all $x, y, z \in E_1$,

$$d(f(x), f(y)) > \epsilon \implies d(f(x), f(z)) > \epsilon/2 \lor d(f(y), f(z)) > \epsilon/2 \implies d(x, z) \geq \delta \lor d(y, z) \geq \delta \lor z \in B_{\delta,\epsilon/2}(f).$$

Hence for random variables $X, Y, Z$,

$$\mathbb{P}(d(f(X), f(Y)) > \epsilon) \leq \mathbb{P}(d(X, Z) > \delta) + \mathbb{P}(d(Y, Z) > \delta) + \mathbb{P}(Z \in B_{\delta,\epsilon/2}(f)).$$

Let $X_n$ be an effective Cauchy sequence of continuous random variables with limit $X$. Fix $\epsilon > 0$. By continuity of the distribution of $X$, we have $\mathbb{P}(X \in B_{\delta,\epsilon/2}) < \epsilon/2$ for sufficiently small $\delta$, and by computability of the distribution, we can effectively find such a $\delta$. Take $N(\epsilon)$ so that $2^{-N(\epsilon)} < \min\{\delta, \epsilon/4\}$. Then for $n \geq N(\epsilon)$, we have $\mathbb{P}(d(X, X_n) > \delta) < \mathbb{P}(d(X, X_n) > 2^{-n}) < 2^{-n} < \epsilon/4$. Therefore if $m, n \geq N(\epsilon)$, we have $\mathbb{P}(d(f(X_m), f(X_n)) > \epsilon) \leq \mathbb{P}(d(X_m, X_n) > \delta) + \mathbb{P}(d(X_n, X) > \delta) + \mathbb{P}(X \in B_{\delta,\epsilon/2}(f)) < \epsilon$. Thus $f(X_n)$ is an effective Cauchy sequence, satisfying $d(f(X_n), f(X_m)) < \epsilon$ whenever $m, n \geq N(\epsilon)$.

Fix $\epsilon > 0$, and choose $\delta$ such that $\mathbb{P}(f(X) \in V) < \mathbb{P}(f(X) \in I_\delta(V)) + \epsilon/2$. For $n$ sufficiently large, we have $d(f(X_n), f(X)) < \min\{\delta, \epsilon/2\}$. Then $\mathbb{P}(f(X) \in V) \geq \mathbb{P}(f(X_n) \in I_\delta(V)) - \epsilon = \mathbb{P}(X_n \in f^{-1}(I_\delta(V))) - \epsilon$. Taking $n \to \infty$ gives $\mathbb{P}(f(X) \in V) \geq \mathbb{P}(X \in f^{-1}(I_\delta(V))) - \epsilon$, and taking $\epsilon \to 0$ given $f^{-1}(I_\delta(V)) \to f^{-1}(V)$, so $\mathbb{P}(f(X) \in V) \geq \mathbb{P}(X \in f^{-1}(V))$. For the reverse inequality, for $n$ sufficiently large, we have $d(f(X_n), f(X)) < \min\{\delta, \epsilon/2\}$, so $\mathbb{P}(f(X_n) \in V) \geq \mathbb{P}(f(X) \in I_\delta(V)) - \epsilon/2$. Then $\mathbb{P}(f(X) \in V) < \mathbb{P}(f(X_n) \in V) + \epsilon = \mathbb{P}(X_n \in f^{-1}(V)) + \epsilon$. Taking $n \to \infty$ gives $\mathbb{P}(f(X) \in V) \leq \mathbb{P}(X \in f^{-1}(V)) + \epsilon$, and since $\epsilon$ is arbitrary, $\mathbb{P}(f(X) \in V) \leq \mathbb{P}(X \in f^{-1}(V))$. 

\[\Box\]
Definition 21 (Expectation). If $X : \Omega \to \mathbb{R}$ is a continuous real-valued random variable, define the expectation of $X$ is defined by the integral
\[
E(X) = \int_{\Omega} f(X(\omega)) \, dP(\omega).
\]
If $X : \Omega \to \mathbb{R}$ is a measurable real-valued random variable, then we define
\[
E(X) = \lim_{n \to \infty} E(X_n),
\]
as long as this limit exists for all sequences $(X_n)_{n \in \mathbb{N}}$ of continuous random variables converging to $X$.

Note that the integral always exists for continuous random variables since $X$ has compact values. By the dominated convergence theorem, the limit of $E(X_n)$ always exists if there is a continuous function $Y : \Omega \to \mathbb{R}^{+\infty}$ such that $|X_n(\omega)| \leq Y(\omega)$ for all $\omega \in \Omega$, and $\int_{\Omega} Y(\omega) \, dP(\omega) < \infty$. Note that finiteness of the integral cannot be effectively checked given only a name of $Y : \Omega \to \mathbb{R}^{+\infty}$, but checking $|X_n| < Y$ is possible.

Theorem 22 (Expectation). Let $X$ be a positive real-valued random variable such that $E(X) < \infty$. Then
\[
E(X) = \int_{0}^{\infty} \mathbb{P}(X > x) \, dx = \int_{0}^{\infty} \mathbb{P}(X \geq x) \, dx.
\]

The proof follows from the definition of the lower integral:

Proof. First assume $X$ is a continuous random variable with positive values. Then by definition, $E(X) = \int_{\Omega} X(\omega) \, dP(\omega)$. Given values $0 = x_0 < x_1 < \cdots < x_n$ such that $x_{n+1} - x_n$, the definition of the lower horizontal integral gives $\int_{\Omega} X(\omega) \, dP(\omega) \geq \sum_{i=0}^{n-1} (x_i - x_{i-1}) P(\{ \omega \mid X(\omega) > x_i \})$.

Suppose $x_i - x_{i-1} < \epsilon$ for all $i$. Then $E(X) + \epsilon = \int_{\Omega} X(\omega) + \epsilon \, dP(\omega) \geq \sum_{i=0}^{n-1} (x_i - x_{i-1}) P(\{ \omega \mid X(\omega) + \epsilon > x_i \}) = \sum_{i=0}^{n-1} \int_{x_{i-1}}^{x_i} \mathbb{P}(X(\omega) > x_i - \epsilon) \, dx_i \geq \sum_{i=0}^{n-1} \int_{x_{i-1}}^{x_i} \mathbb{P}(X(\omega) > x_i) \, dx \geq \sum_{i=0}^{n-1} \int_{x_{i-1}}^{x_i} \mathbb{P}(X(\omega) > x) \, dx = \int_{0}^{\infty} \mathbb{P}(X(\omega) > x) \, dx$. Taking $n \to \infty$ gives $E(X) \geq \int_{0}^{\infty} \mathbb{P}(X > x) \, dx - \epsilon$, and since $\epsilon$ is arbitrary, $E(X) \geq \int_{0}^{\infty} \mathbb{P}(X > x) \, dx$.

We similarly prove $\int_{0}^{\infty} \mathbb{P}(X > x) \, dx \geq E(X)$ and $\int_{0}^{\infty} \mathbb{P}(X > x) \, dx \geq \int_{0}^{\infty} \mathbb{P}(X \geq x) \, dx$; for the former inequality we need boundedness of the values of $X$. The case of measurable random variables follows by taking limits.

By changing variables in the integral, we obtain:

Corollary 23. If $X$ is a real-valued random variable, then for any $\alpha \geq 1$,
\[
E|X|^\alpha = \int_{0}^{\infty} \alpha \, x^{\alpha-1} \, \mathbb{P}(X > x) \, dx = \int_{0}^{\infty} \alpha \, x^{\alpha-1} \, \mathbb{P}(X \geq x) \, dx.
\]

Remark 24 (Expectation of a distribution). Theorem 22 shows that the expectation of a random variable depends only on its distribution. Indeed, we can compute the expectation of a probability valuation $\pi$ on a bounded subset of $[0, \infty)$ by
\[
E(\pi) = \int_{0}^{\infty} \pi([x, \infty]) \, dx = \int_{0}^{\infty} \pi([x, \infty]) \, dx \in \mathbb{R}_+^\ast.
\]

If $X$ is a measurable $\mathbb{R}$-valued random variable and $f : \mathbb{R} \to H$ where $H$ is not a metric space, then $f(X)$ is not defined as a random variable, since we have only a notion of random variables on metric spaces. However, the distribution of $f(X)$ is well-defined and computable, with $\mathbb{P}(f(X) \in V) := \mathbb{P}(X \in f^{-1}(V))$ for $V \in \mathcal{O}(V)$. If $f : \mathbb{R} \to \mathbb{R}_+^\ast$, then we can compute the lower expectation of $f(X)$ by
\[
E_<(f(X)) := \int_{0}^{\infty} \mathbb{P}(X \in f^{-1}([\lambda, \infty])) \, d\lambda.
\]
Similarly \( f : X \to \mathbb{R}^+_1 \), then the upper expectation of \( f(X) \) is computed by
\[
E_>(f(X)) := \int_0^\infty \mathbb{P}(X \in f^{-1}([\lambda, \infty])) d\lambda.
\]

We can effectivise Lesbegue spaces \( L^\vee(X) \) of integrable random variables through the use of effective Cauchy sequences in the natural way:

**Definition 25** (Integrable random variable). Let \((X, |\cdot|)\) be a normed space. Then the type of \(p\)-integrable random variables with values in \(X\) is the completion of the type of \(p\)-integrable piecewise-continuous random variables under the metric \(d_p(X,Y) = ||X - Y||_p\) induced by the norm
\[
||X||_p = \left(\int_\Omega |X(\omega)|^p d\mathbb{P}(\omega)\right)^{1/p} = \left(\mathbb{E}(|X|^p)\right)^{1/p}.
\]

If \(X_n\) is a sequence of continuous random variables with \(||X_m - X_n||_2 \leq 2^{-\min(m,n)}\) and \(\lim_{n \to \infty} X_n = X\), then it is easy to show that \(||X||_2 - ||X_n||_2 \leq 2^{-n}\). We can then easily prove the Cauchy-Schwarz and triangle inequalities for measurable random variables \(\mathbb{E}(|XY|) \leq ||X||_2 \cdot ||Y||_2\) and \(||X + Y||_2 \leq ||X||_2 + ||Y||_2\).

In classical measure theory, it is often useful to consider the indicator (characteristic) function of a set of values of a random variable, defined as
\[
I[X \in S] : \omega \to \{0, 1\} : \omega \mapsto \begin{cases} 1 & \text{if } X(\omega) \in S; \\ 0 & \text{if } X(\omega) \notin S. \end{cases}
\]  

(4)

However, indicator functions taking values in \(\{0, 1\} \subset \mathbb{R}\) are only computable for clopen sets as the following example shows:

**Example 26** (Uncomputability of indicator functions). Let \(X\) be a random variable and \(A\) a closed set. Suppose \(I[X \in A]\) were to be computable as a measurable random variable from \(X\) and \(A\). Then \(\mathbb{P}(X \in A) = \mathbb{P}(I[X \in A] > 0)\) would be computable in \([0, 1]_<\), so \(\mathbb{P}(X \in A)\) would be computable in \([0, 1]\). Taking \(\delta_x\) gives \(\mathbb{P}(X \in A) = 1\) if \(x \in A\) and \(0\) if \(x \notin A\), so \(A\) would be effectively open.

If \(X\) is a continuous random variable, If \(U\) is open, then \(I[X \in U]\) is computable as a function \(\Omega \to [0, 1]_<\), and if \(A\) is closed, then \(I[X \in A]\) is computable in \(\Omega \to [0, 1]_>\). These indicator functions cannot be computed for measurable random variables as the range spaces are not Hausdorff. Instead, for an open set \(U\), the characteristic function \(\chi_U\) is computable as a function \(X \to [0,1]_<\), and since \(I[X \in U] = \chi_U(X)\), the distribution of \(I[X \in U]\) is computable. This makes indicator functions useful when we are only interested in information about probabilities and expectations, such as the submartingale inequality \([5]\).

We now consider independence for random variables and conditional random variables. Following classical probability theory, we say random variables \(X_1, X_2\) taking values in \(E_1, E_2\) are independent if for all open \(U_1 \subset E_1\) and \(U_2 \subset E_2\)
\[
\mathbb{P}(X_1 \in U_1 \land X_2 \in U_2) = \mathbb{P}(X_1 \in U_1) \cdot \mathbb{P}(X_2 \in U_2).
\]

Clearly if \(X_1\) and \(X_2\) are independent real-valued random variables, then \(\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)\).

Given a topology \(\mathcal{F}\) on \(\Omega\) (which is coarser than the standard topology), we say a random variable \(X : \Omega \to X\) is independent of \(\mathcal{F}\) if
\[
\mathbb{P}(X \times J \in U \times V) = \mathbb{P}(X \in U)P(V)
\]
(5)

whenever \(U \in \mathcal{O}(X)\) and \(V \in \mathcal{F}\), where \(J : \Omega \to \Omega\) is the identity random variable. If \(X : \Omega \to X\) is continuous, this reduces to
\[
P(\{\omega \in \Omega \mid X(\omega) \in U \land \omega \in V\} = \mathbb{P}(X \in U)P(V)
\]

11
If $\mathcal{T}$ is the topology generated by a continuous random variable $Y$, then $X$ is independent of $Y$, since
\[
\mathbb{P}(X \in U \land Y \in V) = \mathbb{P}(X \in U \land I \in Y^{-1}(V)) = \mathbb{P}(X \in U)\mathbb{P}(Y^{-1}(V)) = \mathbb{P}(X \in U)\mathbb{P}(Y \in V).
\]

Since $\Omega$ is canonically isomorphic to $\Omega \times \Omega$, given random variables $X_1$, $X_2$, we can always construct independent random variables $Y_1, Y_2$ with the same distribution as $X_1, X_2$ by taking $Y_i(\omega_1, \omega_2) = X_i(\omega_i)$ for $i = 1, 2$. We call the random variable $Y_1 \times Y_1$ an independent product of $X_1$ and $X_2$. Countable independent products can be constructed since $\Omega$ is isomorphic to $\Omega^\infty$. We say random variables $X_1$ and $X_2$ are strongly independent if they are formed as an independent product.

The concept of conditional random variable is subtle even in classical probability theory. Here we use the following definition.

**Definition 27** (Conditional random variable). Given random variables $X$, $Y$, taking values in $\mathbb{R}$, a conditional random variable is a function $X : \Omega \rightarrow (\Omega \hookrightarrow \mathbb{R})$, denoted $Y|X$ with values $(Y|X)(x) : \Omega \rightarrow \mathbb{R}$ denoted $Y|X = x$, such that $Y|X = x$ is independent of $X$ for all $x$, and $E(Y|X) = \mathbb{E}(Y|X, X)$.

If $Y = \mathbb{R}$, then the conditional expectation $E(Y|X)$ is the function $X \rightarrow \mathbb{R}$ defined by
\[
E(Y|X) : x \mapsto E(Y|X = x).
\]

If $X$ is a real-valued random variable and $\mathcal{F}$ is a topology on $\Omega$, then a conditional expectation $E(X|\mathcal{F})$ is a real-valued random variable such that for any open $U \in \mathcal{F}$, $\int_U E(X|\mathcal{F})(\omega)dP(\omega) = \int_U X(\omega)dP(\omega)$.

If $V \in \mathcal{O}(\mathbb{R})$, then the conditional probability $\mathbb{P}(Y|X)$ is the function $X \times \mathcal{O}(\mathbb{R}) \rightarrow [0, 1]$, defined by $(x, V) \mapsto \mathbb{P}(Y|X = x \in V)$. We define $\mathbb{P}(Y \in V|X) : X \rightarrow [0, 1]$, $\mathbb{P}(Y|X = x) : \mathcal{O}(\mathbb{R}) \rightarrow [0, 1]$ in the natural way.

Clearly, the expectation of $Y$ can be computed given $X$ and the conditional expectation of $Y$ given $X$.

**Proposition 28** (Conditional probability and expectation). The joint probabilities $\mathbb{P}(X \in U \land Y \in V)$ are computable from $X$ and $\mathbb{P}(Y|X)$. The expectation $E(Y)$ is computable given $X$ and $E(Y|X)$.

**Proof.** Given $U, V$, define a function $f : x \rightarrow [0, 1]$ by $f(x) = \chi_U(x)\mathbb{P}(Y \in V|X = x)$. Then
\[
\mathbb{P}(X \in U \land Y \in V) = \mathbb{E}(f(X)).
\]

$E(Y|X)$ is a continuous function $X \rightarrow \mathbb{R}$, and $E(Y) = E(Y|X)(X)$, so the result follows from Theorem 20.

If $X : \Omega \rightarrow \mathbb{R}$ and $Y|X : \mathbb{R} \rightarrow (\Omega \hookrightarrow \mathbb{R})$ are both continuous random variables, then clearly we can compute $Y(\omega) = Y|X(\omega, X(\omega))$ using the isomorphism $X \rightarrow (\Omega \hookrightarrow \mathbb{R}) \equiv \mathbb{X} \times \Phi \rightarrow \mathbb{Y} \equiv \Omega \rightarrow (\Omega \hookrightarrow \mathbb{Y})$. In general, we do not know whether this extends to measurable random variables. For our purposes, it suffices to consider strongly independent random variables i.e. $\Omega = \Omega_1 \times \Omega_2$ where $X : \Omega_1 \hookrightarrow \mathbb{X}$ and $Y|X : \mathbb{X} \rightarrow (\Omega_2 \hookrightarrow \mathbb{Y})$.

**Theorem 29** (Conditional random variables). Suppose $X : \Omega \hookrightarrow \mathbb{X}$, and $Y|X : \mathbb{X} \rightarrow (\Omega \hookrightarrow \mathbb{Y})$ are such that $X$ and each $Y|X = x$ are strongly independent. Then there is a random variable $Y$ which is computable from $X$ and $Y|X$, such that $\mathbb{P}(Y \in V|X) =$
Proof. Taking $F = \{Y|X|X\}$, so $F : \Omega_1 \rightarrow (\Omega_2 \rightarrow \mathcal{X})$. Since random variables are defined as a subtype of sequences $\mathcal{R} = \{\Omega_1 \times \mathbb{N} \rightarrow \mathbb{X}\}$, we have $F : (\Omega_1 \times \mathbb{N}_1) \rightarrow (\Omega_2 \times \mathbb{N}_2 \rightarrow \mathbb{X})$, or $X : \Omega_1 \times \Omega_2 \times \mathbb{N}_1 \times \mathbb{N}_2 \rightarrow \mathbb{X}$. We claim that the function $X : (\Omega_1 \times \Omega_2) \times \mathbb{N} \rightarrow \mathbb{X}$ given by $X(\omega_1, \omega_2, n) = F(\omega_1; 1, \omega_2; n)$ is a random variable.

For fixed $\mu$, $F_\mu$ is a continuous function $\Omega \rightarrow \mathcal{R}(\mathbb{X})$. Further, there is a set $W_1 \subseteq \Omega_1$ such that for $\mu_1 \in W_1$, $d(F_{\mu_1}(\omega_1), F_{\mu_2}(\omega_1)) \leq 2^{-\mu_1}$ whenever $\mu_1 \geq n$ and $\mu_2 \in W_1$. Further, by shiriking $W_1$ slightly, we can ensure $P(W_1) \geq 1 - 2^{-\mu}$ and $W_1$ is a clopen set.

For fixed $\mu$, $d(F_{\mu_1}(\omega_1), F_{\mu_2}(\omega_1)) \leq 2^{-\mu_1}$. Since $W_1$ is compact, $F_{\mu_1}$ is uniformly continuous on $W_1$, so we can partition $W_1$ into sets $B_{\mu, k}$ such that $d(F_{\mu_1}(\nu_1), F_{\mu_2}(\nu_2)) \leq 2^{-\mu_1}$ whenever $\nu_1, \nu_2 \in B_{\mu, k}$ for some $k$. Then $d(F_{\mu_1}(\omega_1), F_{\mu_2}(\nu_1)) \leq 2^{-\mu_1}$ whenever $\nu_1, \omega_1$ lie in the same $B_{\mu, k}$ and $\nu_1 \geq \mu_1 \geq 0$.

Let $n_2 = n_1$. Then for each $B_{\mu, k}$, there is a set $W_{\mu, 2k} \subseteq \Omega_2$ with $P(W_{\mu, 2k}) \geq 1 - 2^{-\mu_2}$ such that for $\mu_1, \nu_1 \in B_{\mu, k}$ and $\nu_2 \in B_{\mu, 2k}$, we have $d(F_{\mu_1}(\omega_1, \omega_2), F_{\mu_1}(\nu_1, \nu_2)) \leq 2^{-\mu_2}$ whenever $\nu_2 \geq n_2$. Hence if $\mu_1 \in B_{\mu, k}$ and $\omega_2 \in W_{\mu, 2k}$, then $d(F_{\mu_1, n_2}(\omega_1, \omega_2), F_{\mu_1, n_2}(\nu_1, \omega_2)) \leq 2^{-\mu_1 + \mu_2}$. Each $B_{\mu, k} \times W_{\mu, 2k}$ has probability $P(B_{\mu, k}) \times P(W_{\mu, 2k}) \geq 1 - 2^{-\mu_1 - \mu_2}$. Taking $n_1 = n_2 = n$ and $m_1 = m_2 = m$ gives $d(F_{\mu,n}(\omega_1, \omega_2), F_{\mu,n}(\omega_1, \omega_2)) \leq 2^{-m - 1}$ on a subset of $\Omega_1 \times \Omega_2$ of probability at least $1 - 2^{-m-2}$. Taking $X_n(\omega_1, \omega_2) = F_{n+n_2}(\omega_1, \omega_2)$ defined $d(X_n, \mathbb{M}) \leq 2^{-m-2}$ for $m > n$, so the sequence $X_n$ is an effective Cauchy sequence and converges to a random variable.

We finally show that we can construct a random variable with a given distribution.

**Theorem 30.** Let $\mathbb{X}$ be a computable metric space, and $\nu$ be a valuation on $\mathbb{X}$. Then we can compute a measurable random variable $X$ such that for any open $U$, $\mathbb{P}(X \in U) = \nu(U)$.

**Proof.** We construct a sequence of random variables $X_n$ as follows. For each $n$, construct a be a finite collection $\mathcal{B}_n$ of mutually-disjoint closed balls such that each $B \in \mathcal{B}_n$ has radius at most $2^{-n}$, and such that $\nu(\{B \in \mathcal{B}_n\}) = \sum_{B \mathcal{B}_n} \nu(B) > 1 - 2^{-n}$. We further require that for every $k < n$, a subcollection $\mathcal{B}_{n,k}$ of $\mathcal{B}_n$ with measure at least $1 - 2^{-k}$ is a refinement of $\mathcal{B}_k$. For each ball $B_{n,m} \in \mathcal{B}_n$ with centre $x_{n,m}$ and radius $r_{n,m}$, compute a dyadic number $p_{n,m}$ such that $p_{n,m} > \nu(B_{n,m})$ and $\sum_{n,m} p_{n,m} = 1$. Take $X_n$ to be the sum of point-measures supported at the centre $x$ of each $B \in \mathcal{B}_n$, with mass $p_{n,m}$, and such that if $B_{n,k} \supseteq B_{n,m}$ for $m > k$, the $X_m(\omega) \in B_{n,m}$ only if $X_n(\omega) \in B_{n,k}$.

Then directly from the construction we see that $X_n$ is an effective Cauchy sequence since for any $n > m$, $X_m(\omega)$ and $X_n(\omega)$ agree to within $2^{-m}$ on a set of measure at least $1 - 2^{-n}$, and that $\nu(U) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n \in U)$.

5 Discrete-Time Stochastic Processes

A **discrete-time stochastic process** with state space $\mathbb{X}$ is a random variable $\bar{X} = (X_0, X_1, X_2, \ldots)$ taking values in $\mathbb{X}^\infty$. A **Markov process** is a stochastic process such that $X_{n+1}$ depends only on the previous state $X_n$, so is determined by the conditional value $X_{n+1}|X_n$, such that $X_{n+1}|X_n = x_n$ is independent of $(X_0, X_1, \ldots, X_{n-1})$. A Markov process is **stationary** if the distributions of $X_{n+1}|X_n$, i.e. $\mathbb{P}(X_{n+1}|X_n = x_n) \in U)$, are equal. In this case we can write $\mathbb{P}(X_{n+1}|X_n) = F_n : \Omega \rightarrow (\Omega \rightarrow \mathbb{X})$, where $F_n(x, \omega) = F_n(x_0, \omega_1, \ldots) = F(\omega_n)$. Hence the process is defined by $F : \Omega \rightarrow \mathcal{R}(\mathbb{X})$.

Typically, we are only interested in the **distribution** of the states $X_n$, and so rather than treating $X_n$ as a random variable $X_n : \Omega \rightarrow \mathbb{X}$, we consider $X_n \in \mathcal{P}(\mathbb{X})$. Then the Markov process is defined by $F : \mathbb{X} \rightarrow \mathcal{P}(\mathbb{X})$. 

13
When working in Cartesian-closed categories, objects of the form \((X \to T) \to T\) for some fixed type \(T\) are an example of a **monad** \([Str72]\). They support standard manipulations which make them ideal for the representation of dynamic systems. When \(T = \mathbb{S}\), the Sierpinski type, we obtain categories of overt and compact sets \([Esc04]\), which form a basis for discrete-time nondeterministic systems \([Col09]\). Since \(\mathcal{P}(X)\) is a subtype of \((X \to \mathbb{H}) \to \mathbb{H}\) we can take \(T = \mathbb{H}\) and obtain the same operators for discrete-time stochastic systems.

**Proposition 31.** Let \(T\), \(X\) and \(Y\) be elements of the category of computable types. Then the following operators are computable:

1. The embedding of \(X\) in \((X \to T) \to T\) given by \(\delta_x(\phi) = \phi(x)\) for \(x \in X\) and \(\phi : X \to T\).
2. The canonical equivalence between \(X \to ((Y \to T) \to T)\) and \((Y \to T) \to (X \to T)\) given by \(F^* \psi(x) = F(x)(\psi)\) for \(F : X \to ((Y \to T) \to T)\) and \(\psi : Y \to T\).
3. An element \(f\) of \(X \to Y\) lifts to an operator \(F^*\) from \((X \to T) \to T\) to \(((Y \to T) \to T)\) defined by \(f_\ast \mu(\psi) = \mu(\psi \circ f)\) for \(\mu : (X \to T) \to T\) and \(\psi : Y \to T\).
4. An element \(F\) of \((X \to T) \to T\) lifts to an operator \(F^*\) from \((X \to T) \to T\) to \(((Y \to T) \to T)\) defined by \(F_\ast \mu(\psi) = \mu(\lambda x.F(x)(\psi))\).
5. Given \(F : (X \to T) \to T\) and \(G : (Y \to T) \to T\), the skew-products \(F \times G : (X \times Y \to T) \to T\) defined by \((F \times G)(\psi) = F(\lambda x.G(\lambda y.\psi(x, y)))\) and \((F \times G)(\psi) = G(\lambda y.F(\lambda x.\psi(x, y)))\).

We write \(F \times G\) for the product if \(F \times G = F \times G\) for all \(F, G\) in some restricted class of interest.

For the case of set types, the embedding \(X \mapsto ((X \to T) \to T)\) a singleton set; for measures, the point-measure \(\delta_x\). Note that if \(F : (X \to T) \to T\) and \(G : (Y \to T) \to T\), then in general \(F \times G\) and \(F \times G\) are not equal. Equality (i.e. commutativity of the product) does hold in many important cases, including products of measures. The generalisation to monads \(\mathcal{M}(X)\) requires canonical operators \(X \to \mathcal{M}(X)\) and \((X \to \mathcal{M}(Y)) \to (\mathcal{M}(X) \to \mathcal{M}(Y))\).

We now apply the standard push-forward operators of Proposition 31 to the case of probability measures. Computability of the operators on \((X \to \mathbb{H}) \to \mathbb{H}\) is clear, it remains to check the linearity properties and the unit total measure.

**Lemma 32.** There is a computable point-measure operator taking \(x \in X\) to \(\delta_x \in \mathcal{P}(X)\).

**Proof.** For \(\psi : X \to \mathbb{H}\), define \(\delta_x(\psi) = \psi(x)\). Then \(\delta_x(\alpha_1 \psi_1 + \alpha_2 \psi_2) = (\alpha_1 \psi_1 + \alpha_2 \psi_2)(x) = \alpha_1 \psi_1(x) + \alpha_2 \psi_2(x) = \alpha_1 \delta_x(\psi_1) + \alpha_2 \delta_x(\psi_2)\), and if \(\psi \equiv 1\), then \(\delta_x(\psi) = 1\), so \(\delta_x\) is a probability measure.

**Proposition 33.** There is a computable push-forward operator taking a function \(F : X \to \mathcal{P}(Y)\) and \(\mu \in \mathcal{P}(X)\) to the push-forward distribution \(F_\ast \mu \in \mathcal{P}(Y)\) is computable.

**Proof.** For \(\psi : Y \to T\), we have \(F_\ast \mu(\psi) = \mu(\lambda x.F(x)(\psi))\) is computable. We need to check that \(F_\ast \mu\) is a probability measure. It is easy to verify that \(F^* (\alpha_1 \psi_1 + \alpha_2 \psi_2) = \alpha_1 F^*(\psi_1) + \alpha_2 F^*(\psi_2)\), and hence \(F^*_\ast \mu(\psi_1 + \alpha_2 \psi_2) = \alpha_1 F_\ast \mu(\psi_1) + \alpha_2 F_\ast \mu(\psi_2)\). If \(\psi \equiv 1\), then \(\phi = F^* \psi = \lambda x. F(x)(\psi) \equiv 1\) since for all \(x \in X\), \(\phi(x) = F(x)(\psi)\) and \(F(x)\) is a probability measure. Then \(\mu(\phi) = 1\) as \(\mu\) is a probability measure.

**Corollary 34.** If \(f : X \to Y\), then \(f\) induces a computable operator \(f_\ast : \mathcal{P}(X) \to \mathcal{P}(Y)\) by \(f_\ast \mu = F_\ast \mu\) where \(F(x) = \delta_{f(x)}\). Explicitly, \(f_\ast \mu(\psi) = \mu(\psi \circ f)\) for \(\psi : Y \to \mathbb{H}\).
Proposition 35. The push-forward operator taking a function $F : \mathbb{X} \to \mathcal{P}(\mathbb{Y})$ and probability measure $\mu \in \mathcal{P}(\mathbb{X})$ to the joint distribution $(\mu, F_* \mu) = (\id \times F)_* \mu$ on $\mathbb{X} \times \mathbb{Y}$ is computable.

Proof. We have $F : \mathbb{X} \to ((\mathbb{Y} \to \mathbb{H}) \to \mathbb{H})$ and $\mu : ((\mathbb{X} \to \mathbb{H}) \to \mathbb{H})$. Define $(\id \times F)$ to be the function $X \to \mathcal{P}(X \times \mathbb{Y})$ given by $(\id \times F)(x)(\psi) = F(x)(\lambda y. \psi(x, y))$. Note that if $\psi : X \times Y \to H$ is the constant function 1, then $\lambda y. \psi(x, y) \equiv 1$, so $(\id \times F)(x)(\psi) = F(x)(\phi) = 1$ since $F(x)$ is a probability distribution. Then by Proposition 33, $(\id \times F)_* \mu$ is computable in $\mathcal{P}(\mathbb{X} \times \mathbb{Y})$. \hfill \Box

We first consider the simplest approach to stochastic processes, where we only compute the distribution of the states. A Markov process is then defined by a stochastic update rule.

Theorem 37. Let $F : \mathbb{X} \to \mathcal{P}(\mathbb{X})$ be a Markov process. Then given a probability distribution $\mu_0$ of the initial state $x_0$, the probability distributions $\mu_n$ of the state $x_n$ at time $n$, and the joint probability distribution $\gamma_n$ of the states $(x_0, \ldots, x_n)$ up to time $n$, are computable.

The proof is trivial given the categorical constructions of Proposition 31.

Proof. Compute $\mu_n \in \mathcal{P}(X)$ recursively by $\mu_n = F_* \mu_{n-1}$, which are computable by Proposition 33. Compute the joint distributions $\gamma_n \in \mathcal{P}(X^{n+1})$ recursively by $\gamma_0 = \mu_0$ and $\gamma_n = (\id \times F)_* \gamma_{n-1}$, which are computable by Proposition 33. \hfill \Box

Note that $\mu_n = (\pi_n)_* \gamma_n$, where $\pi_n : X^{n+1} \to X$ is given by $\pi_n(x_0, \ldots, x_n) = x_n$; in other words, the distribution at time $n$ can be extracted from the joint distribution up to time $n$.

We can also consider the state as a random variable on the base probability space $\Omega$. This approach yields a random variable $X_n$ for the state at time $n$.

Definition 38. A parameterised Markov process on a type $\mathbb{X}$ is defined by a conditional random variable $F : \mathbb{X} \to \mathcal{R}(\mathbb{X})$ and a random variable $X_0 : \mathcal{R}(\mathbb{X})$.

Given a parameterised Markov process, we can trivially extract the distribution of $X_0$ and the conditional distribution function $\Phi : \mathbb{X} \to \mathcal{P}(\mathbb{X})$ by $\Phi(x)(\psi) = \mathbb{P}(\lambda \omega. \psi(F(x, \omega)))$.

The following result shows that a parameterised Markov process gives rise to random variables $X_0, X_1, X_2, \ldots$ over the probability space $(\Omega, \mathcal{F}, \mu)$.

Theorem 39. If $F : \mathbb{X} \to (\Omega \rightsquigarrow \mathbb{X})$ is a parameterised Markov process, and $X_{\text{init}} : \Omega \rightsquigarrow \mathbb{X}$ is a random variable giving the initial probability distribution, then we can compute the stochastic process $(X_0, X_1, X_2, \ldots)$ as a random variable $\Omega \rightsquigarrow \mathbb{X}^\omega$.

Proof. Let $\Omega_i$ be a copy of $\Omega$ for each $i \in \mathbb{N}$, and define $X_n : \Omega_0 \times \Omega_1 \times \cdots \times \Omega_n \times \cdots \to \mathbb{X}$ recursively by $X_0(\omega_0, \omega_1, \omega_2, \ldots) = X_{\text{init}}(\omega_0)$ and $X_n(\omega_0, \omega_1, \ldots, \omega_n, \ldots) = F(X_{n-1}(\omega_0, \omega_1, \ldots, \omega_{n-1}, \ldots))(\omega_n)$. Then each $X_n$ is computable by computability of random variables from conditional random variables given by Theorem 29. Further, $X_n$ is dependent on $(\omega_0, \omega_1, \ldots, \omega_n)$ only. \hfill \Box
6 The Wiener process

The Wiener process $W(t)$ or $W_t$ is a random process such that $W(0) = 0$, the distribution function $t \mapsto W(t)$ is almost surely continuous in the weak topology, and $W(t)$ has independent increments with $W(t) - W(s) \sim N(0, t - s)$ for $0 \leq s < t$, where $N(\mu, \sigma^2)$ is the normal distribution with mean $\mu$ and variance $\sigma^2$. The Wiener process is used in the definition of a stochastic differential equation

$$dX(t) = f(X(t), t) \, dt + g(X(t), t) \, dW(t).$$

There are many comprehensive books available for continuous-time stochastic processes, notably [Eva13].

**Theorem 40.** A sample path of the Wiener process is almost-surely $\alpha$-Hölder continuous for all $\alpha < 1/2$.

The following result on the maximum of the Wiener process up to a given time is based on the André reflection principle.

**Theorem 41.** Denote by $M(t)$ the maximum of the Wiener process up to time $t$. Then

$$\mathbb{P}(M(t) \geq X) = 2\mathbb{P}(W(t) > X).$$

There are two main constructions of a Wiener process. The Paley-Wiener construction yields a Wiener process on $[0, 2\pi]$ as

$$W(t) = A_0 + \sum_{n=1}^{\infty} A_n \cos(nt) + B_n \sin(nt)$$

where the $A_n$ and $B_n$ are independent $N(0, 1)$ random variables. The simpler Lévy-Ciesielski construction uses wavelets. Let $h_{n,k}$ be the $(n,k)$-th Haar function, defined for $0 \leq k < 2^n$ by

$$h_{n,k}(x) = \begin{cases} 
+2^{n/2} & \text{for } \frac{k}{2^n} \leq t \leq \frac{k+1/2}{2^n}, \\
-2^{n/2} & \text{for } \frac{k+1/2}{2^n} \leq t \leq \frac{k+1}{2^n}, \\
0 & \text{otherwise}. 
\end{cases}$$

Let be $s_{n,k}$ be the $(n,k)$-th Schauder function defined by

$$s_{n,k}(t) = \int_0^t h_{n,k}(\tau) \, d\tau.$$ 

Note that $\sup_{t \in [0,1]} s_{n,k}(t) = 2^{-n/2}$. Let $A_{n,k}$ be a sequence of independent $N(0, 1)$ random variables on a probability space $(\Omega, \mathcal{P})$. Then

$$W(t) = \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} A_{n,k} s_{n,k}(t)$$

is a Wiener process on $[0,1]$.

It should be noted that the the sum $\sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} A_{n,k}(\omega) s_{n,k}(t)$ does not converge for all values of the random variables $A_{n,k}$. However, if the $A_{n,k}$ have growth bounded by $\alpha^{n/2}$ where $\alpha < 2$, then $\sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} A_{n,k} s_{n,k}(t)$ converges uniformly. By the Borel-Cantelli lemma, $\mathbb{P}(A_{n,k} \geq 2^{n/2} \ i.o.) = 0$. 

16
However, given only finitely many values of $A_{n,k}(\omega)$, we cannot compute a uniform approximation to the sample path $W(\omega)$, or even an approximation in $L^2([0,1])$. In other words, the function $\omega \to \sum A_{n,k}(\omega) s_{n,k}$ is not a computable function from $\Omega$ to $C([0,1])$ or $L^2([0,1])$. However, it is the case that for any open subset $U$ of $C([0,1])$, the probability $P(\{|\omega| \sum A_{n,k}(\omega) s_{n,k} \in U\})$ is computable in $\mathbb{H}$. Further, there is a sequence of closed compact subsets $K_n$ of $\Omega$ such that $\mathbb{P}(K_n) \to 1$ as $n \to \infty$ and $W$ is computable on each $K_n$.

We now give a modification of the Lévy-Ciesielski construction with base space $\Omega = \{0,1\}^\omega$ for which the Wiener process is a continuous function $W : \Omega \to C([0,1])$. In fact, we obtain sample paths which are Hölder-continuous in $C^\alpha$ for any $\alpha < 1/2$, though we shall only prove the continuous case.

**Theorem 42** (Computable Wiener process). Let $\Omega = \{0,1\}^\omega$ and $P$ be the standard probability measure on $\Omega$. Then there exists a computable Wiener process $W : \Omega \to C([0,1])$ with open full measure domain.

**Sketch of proof.** The basic idea is to modify the Lévy-Ciesielski construction so that after a finite number of bits of information we can bound the size of $A_{n,k}$ for all sufficiently large $n$.

For the event described by $|A_{n,k}| < n$ whenever $n \geq m$, we have

$$
\prod_{n=m}^{\infty} \prod_{k=1}^{2^n} \mathbb{P}(|A_{n,k}| < n) \geq 1 - \sum_{n=m}^{\infty} 2^n \mathbb{P}(|A_{n,k}| \geq n) \geq 1 - \sum_{n=m}^{\infty} 2^n \cdot 2 \cdot \frac{1}{\sqrt{2\pi}} \int_n^{\infty} e^{-t^2/2} dt
$$

whenever $m \geq 6$, since for $n \geq 6$ we have $e^{-n^2/4} < 4^{-n}$ and $\frac{1}{\sqrt{2\pi}} \int_n^{\infty} e^{-t^2/4} < 1/4$.

We can therefore construct numbers $\beta_{m,n}$ such that $\beta_{m,n} = n$ whenever $n > m$, $\beta_{m+1,n} \geq \beta_{m,n}$ for all $m,n$, and

$$
\mathbb{P}(\forall n = 0, \ldots, \infty, \forall k = 0, \ldots, 2^n - 1, |A_{n,k}| < \beta_{m,n}) = 1/2^m.
$$

We now partition a full-measure open subset of $\Omega$ into sets $\Omega_m$ of measure $1/2^{m+1}$ such that $|A_{n,k}(\omega)| < \beta_{m,n}$ but not every $|A_{n,k}(\omega)| < \beta_{m-1,n}$ whenever $\omega \in \Omega_m$. On each $\Omega_m$ we can computably construct the corresponding values of $A_{n,k}(\omega)$. In particular, on $\Omega_m$, every $A_{n,k}$ is bounded and $|A_{n,k}| < n$ whenever $n > m$, so

$$
|W(\omega,t) - \sum_{n=0}^{m} \sum_{k=0}^{2^n-1} A_{n,k}(\omega)s_{n,k}(t)| = |\sum_{n=m+1}^{\infty} \sum_{k=0}^{2^n-1} A_{n,k}(\omega)s_{n,k}(t)| \\
\leq \sum_{n=m+1}^{\infty} n \cdot 2^{-n/2} \to 0 \text{ as } m \to \infty.
$$

\[\square\]

7 **Stochastic integration**

A continuous-time real-valued stochastic process defined over the interval $[0,T]$ is a random variable taking values in $C([0,T];\mathbb{R})$. Since the indefinite integral $C([0,T];\mathbb{R}) \to C([0,T];\mathbb{R})$ taking $\xi$ to the function $t \mapsto \int_0^t \xi(s) \, ds$ is computable, so is the integral $t \mapsto \int_0^t X(s) \, ds$. In stochastic integration, we aim to give a meaning to the integral

$$
\int_0^t X(s) dW(s)
$$

for a process $X$ with respect to the Wiener process.
We say that a process $X(t)$ is nonanticipative with respect to the Wiener process if $X(t)$ depends only on $W^i|_{[0,t]}$, the restriction of $W$ to $[0,t]$. Formally, letting $\mathcal{F}_t$ be the topology on $\Omega$ generated by $W^i|_{[0,t]}$, then $X|_{[0,t]}$ is a limit of $\mathcal{F}_t$-continuous functions $X_n : \Omega \to C([0,t];\mathbb{R})$.

It turns out that this integral cannot be computed pathwise by the Stieltjes integral. Instead, one uses the Itô integral, which is first defined for step processes, and then extended to continuous processes. In this section, we prove that the standard construction of the Itô integral effectivises.

A stochastic process $X(\cdot)$ is a step process if there are random variables $X_i$, $i = 0, \ldots, n$ and times $0 = t_0 < t_1 < \cdots < t_n = T$ such that $X(t) = X_i$ for $t \in [t_i, t_{i+1})$. We formally write $X(t) = X_i I[t \in [t_i, t_{i+1})]$, where $I[t \in [t_i, t_{i+1})]$ is the indicator function with value 1 if $t \in [t_i, t_{i+1})$ and 0 otherwise. It is straightforward to show that if $\mathbb{E}(X_i^2) < \infty$ for all $i$, then the step process is well-defined as an element of $M^2(L^2([0,T];\mathbb{R}))$, where $L^2(L^2([0,T];\mathbb{R}))$ is the space of $\mathbb{L}$-integrable random variables.

We first show that given $\xi \in C([0,T];\mathbb{R})$, we can compute step functions $\eta$ taking values in the $\mathbb{L}$-space $L^2([0,T];\mathbb{R})$.

**Theorem 43.** Given $\xi : C([0,T];\mathbb{R})$, we can compute a sequence of step function $\eta_n : [0,T] \to \mathbb{R}$ such that $\eta_n \to \xi$ effectively in $L^2([0,T];\mathbb{R})$.

**Proof.** Choose a sequence $\delta_n > 0$ effectively converging to 0, an choose partitions $\mathcal{T}_n = \{0 = t_{n,0} < t_{n,1} < \cdots < t_{n,m_n} = T\}$ where each $t_{n,i}$ is computable and $t_{n,i+1} - t_{n,i} < \delta_n$ for all $n, i$. Compute $\eta_{n,i} = \xi(t_{n,i})$ and define $\eta_n(t) = \eta_{n,i}$ for $t_{n,i} \leq t < t_{n,i+1}$. Clearly $\eta_n \in L^2([0,T];\mathbb{R})$.

The integral $\int_0^T (\xi(t) - \eta_n(t))^2 \, dt = \sum_{i=0}^{m_n-1} \int_{t_i}^{t_{i+1}} (\xi(t) - \xi(t_{n,i}))^2 \, dt$ is computable, and converges to 0 as $n \to \infty$ by continuity of $\xi$, so convergence is effective.

Note that continuity of $\xi$ is required to compute $\xi(t_{n})$, but we do not need to know the modulus of continuity to compute the rate of convergence of $\eta_n$. By Theorem 20, this pathwise computation extends to random variables, and it is clear that if $X$ is nonanticipative with respect to $W$, then so are the step processes $X_n$.

**Definition 44** (Itô integral for step processes). Given a step process $X = \sum_{i=0}^{n-1} X_i I[t \in [t_i, t_{i+1})]$, we define the Itô integral as

$$\int_0^T X(t) \, dW(t) = \sum_{i=0}^{n-1} X_i (W(t_{i+1}) - W(t_i)).$$

This definition can be extended to an indefinite integral: Take $m(s) = \max\{i \mid t_i < s\}$ and define

$$\int_0^t X(t) \, dW(t) = \sum_{i=0}^{m(s)-1} X_i (W(t_{i+1}) - W(t_i)) + X_{m(s)} (W(s) - W(t_{m(s)})).$$

**Lemma 45.** The Itô integral of a step process is computable as a continuous process. Further, if $X(\cdot)$ is nonanticipative with respect to the Wiener process, then so is its Itô integral.

**Proof.** By the defining equation (6), $W(t)$ is continuous when restricted to each interval $[t_i, t_{i+1}]$, and clearly the integral is continuous over the step boundaries. It is also clear that the Itô integral is nonanticipative, at time $t$ since it depends only on $W(s)$ for $s \leq t$.

The following Itô equality is crucial, since it relates the stochastic integral with an ordinary integral.

18
Lemma 46. If \( X = \sum X_k X_{|t_k,t_{k+1}} \) is a step process, and \( X_k \) is independent of \( W(t) - W(s) \) for all \( t > s > t_k \), then
\[
E \left( \int_0^T X(t) \, dW(t) \right)^2 = E \int_0^T X(t)^2 \, dt. \tag{7}
\]

Proof.
\[
E \left( \int_0^T X(t) \, dW(t) \right)^2 = E \left( \int_0^T X(s) \, dW(s) \int_0^T X(t) \, dW(t) \right)
\]
\[
= E \left( \sum_{i=0}^{m-1} X_i(W(t_{i+1}) - W(t_i)) \sum_{j=0}^{m-1} X_j(W(t_{j+1}) - W(t_j)) \right)
\]
\[
= \sum_{i,j=0}^{m-1} E(X_i X_j(W(t_{i+1}) - W(t_i))(W(t_{j+1}) - W(t_j)))
\]
If \( i < j \), then since \( X_i, X_j \) are independent of \( (W(t_{j+1}) - W(t_j)) \),
\[
E(X_i X_j(W(t_{i+1}) - W(t_i))(W(t_{j+1}) - W(t_j)))
\]
\[
= E(X_i X_j(W(t_{i+1}) - W(t_i))) E(W(t_{j+1}) - W(t_j)) = 0
\]
and a similar estimate holds for \( i > j \). Hence
\[
E \left( \int_0^T X(t) \, dW(t) \right)^2 = \sum_{i=0}^{m-1} E(X_i^2(W(t_{i+1}) - W(t_i))^2) = \sum_{i=0}^{m-1} E(X_i^2)E(W(t_{i+1}) - W(t_i))^2
\]
\[
= \sum_{i=0}^{m-1} E(X_i^2)(t_{i+1} - t_i) = E \left( \sum_{i=0}^{m-1} X_i^2(t_{i+1} - t_i) \right)
\]
\[
= E \left( \int_0^T X(t)^2 \, dt \right). \tag*{\Box}
\]

In order to bound the expected maximum value along a path, we will use martingale properties of the integrated process.

Definition 47 ((Sub)martingale). A discrete stochastic process \( X \) is a martingale if for all \( k, E(|X_k|) < \infty \) and \( E(X_{k+1}|F_k) = E(X_k) \), and a submartingale if for all \( k, E(|X_k|) < \infty \) and \( E(X_{k+1}|F_k) \geq E(X_k) \).

A stochastic process \( X \) is a martingale if for all \( t, E(|X_t|) < \infty \) and for all \( t > s, E(X_t|X_s) = E(X_s) \), and a submartingale if \( E(X_t|X_s) \geq E(X_s) \).

We can give sufficient conditions for a process to be a (sub)martingale avoiding the use of conditional expectation. We say \( X \) has independent increments if \( X_t - X_s \) is independent of \( X_s \) whenever \( t > s \). If \( X \) has independent increments, then \( X \) is a martingale if \( E(X_t - X_s) = 0 \) and a submartingale if \( E(X_t - X_s) \geq 0 \).

Lemma 48. The Itô integral of a step process has independent increments with zero expectation, so is a martingale.

Proof. Let \( Y \) be the integrated process \( Y(t) = \int_0^t X(s) \, dW(s) \). Then \( Y(t) = Y(s) + \int_s^t X(r) \, dW(r) = Y(s) + X_m(s)(W(t_{m(s)+1}) - W(s)) + \sum_{i=m(s)}^{m(t)-1} X_i(W(t_{i+1}) - W(t_i)) + X_m(t)(W(t) - W(t_{m(t)})) \).

Since \( W(t_3) - W(t_2) \) is independent of \( X(t_1) \) whenever \( t_1 < t_2 < t_3 \), we have \( E(Y(t)) = E(Y(s)) \) or \( E(Y(t) - Y(s)) = 0 \). \( \Box \)
Lemma 49. If \((X_k)_{k=1,\ldots,m}\) is a martingale and \(\phi\) is convex, then \((\phi(X_k))_{k=1,\ldots,m}\) is a submartingale. Similarly, if \(X()\) is a martingale, then \(\phi(X())\) is a submartingale.

Proof. By Jensen’s inequality, \(\mathbb{E}(\phi(X_{k+1})|\mathcal{F}_k) \geq \phi(\mathbb{E}(X_{k+1}|\mathcal{F}_k)) = \phi(\mathbb{E}(X_k))\).

We now give some estimates known as martingale inequalities which will allow us to compute limits of processes in \(C([0,T];\mathbb{R})\).

Lemma 50 (Discrete submartingale inequality). Let \((X_k)\) be a discrete positive submartingale, and \(Y_n = \max_{k=1,\ldots,n} X_k\). Then for any \(\lambda > 0\),

\[
\lambda P(Y_n \geq \lambda) \leq \mathbb{E}(X_n I[Y_n \geq \lambda]) \leq \mathbb{E}X_n.
\]

(8)

Note that the expression \(\mathbb{E}(X_n I[Y_n \geq \lambda])\) makes sense since the function \(\max : \mathbb{R}^n \to \mathbb{R}\) is computable and the characteristic function \(\xi_{[\lambda,\infty]}\) is computable \(\mathbb{R} \to [0,1]\), for a given \(\lambda\). Hence \(\mathbb{E}(X_n I[Y_n \geq \lambda])\) is computable as the upper integral \(\mathbb{E}_{\geq}(f(X_1,\ldots,X_n))\) where \(f = \chi_{[\lambda,\infty]} \circ \max\).

The basic idea of the proof is to consider events \(X_k \geq \lambda\) and for all \(i < k\), \(X_i < \lambda\). However, these sets of events are neither closed nor open, so we cannot directly probabilities or expectations.

Proof. Let \(A\) be the event \(\max_{k=1,\ldots,n} X_k \geq \lambda\). For fixed \(\delta > 0\), let \(A_{\delta,k}\) be the event \(\bigwedge_{i=1}^k (X_i \leq \lambda - \delta) \wedge (X_k \geq \lambda)\), and \(A_{\delta}\) the event \(\bigcup_{n=1}^\infty A_{\delta,n}\). Note that \(A = \bigcup_{\delta > 0} A_{\delta}\). Since \(A_{\delta,k}\) holds on a closed set, the characteristic function is upper semicontinuous and we can consider the upper horizontal integral of \(\chi_{A_{\delta,k}}\). Since \(X_n - X_k\) is independent of \(X_1,\ldots,X_k\) for \(k < n\), and hence independent of \(A_{\delta,k}\), we have

\[
\mathbb{E}(X_n \chi_{A_{\delta,k}}) = \mathbb{E}((X_n - X_k)\chi_{A_{\delta,k}}) + \mathbb{E}(X_k \chi_{A_{\delta,k}})
\]

\[
= \mathbb{E}(X_n - X_k)\mathbb{E}(\chi_{A_{\delta,k}}) + \mathbb{E}(X_k \chi_{A_{\delta,k}}) \geq \mathbb{E}(X_k \chi_{A_{\delta,k}}).
\]

Then summing probabilities over each \(A_{\delta,k}\) gives

\[
\lambda P(X \in A_{\delta}) \leq \sum_{k=1}^n \lambda P(A_{\delta,k}) = \sum_{k=1}^n \mathbb{E}(\lambda \chi_{A_{\delta,k}})
\]

\[
\leq \sum_{k=1}^n \mathbb{E}(X_k \chi_{A_{\delta,k}}) \leq \sum_{k=1}^n \mathbb{E}(X_n \chi_{A_{\delta,k}}) = \mathbb{E}(\sum_{k=1}^n X_n \chi_{A_{\delta,k}}) = \mathbb{E}(X_n \chi_{A_{\delta}})
\]

By Lemma \(\blacksquare\) \(P(X \in A_{\delta}) \to P(X \in A)\) as \(\delta \to 0\), and the result follows. \(\blacksquare\)

We can also show the standard result that if \(X_t\) is a martingale, and \(\mathbb{E}|X_T|^\alpha < \infty\) for some \(\alpha \geq 1\), then \(\mathbb{P}(\max_{t \leq T}|X_t| \geq \epsilon) \leq \frac{\epsilon}{\alpha} \mathbb{E}|X_T|^\alpha\), but will not need this in the sequel. Instead, we use the following extension to square-integrable martingales:

Lemma 51 (Discrete integrable martingale inequality). Let \((X_k)_{k=1,\ldots,n}\) be a discrete martingale. Then

\[
\mathbb{E}(\max_{k=1,\ldots,n}|X_k|^2) \leq 4\mathbb{E}(X_n^2).
\]

(9)

Proof. Define \(Y = \max_{k=1,\ldots,n}|X_k|\) and \(Z = X_n\). By the stronger form of the submartingale inequality of Lemma \(\blacksquare\), we have \(\lambda P(Y \geq \lambda) \leq \mathbb{E}(I[Y \geq \lambda] Z)\). Since \(Y = \int_0^\infty I[Y \geq \lambda] d\lambda\) and \(Y^2 = 2\int_0^\infty \lambda I[Y \geq \lambda] d\lambda\), we have

\[
\mathbb{E}Y^2 = 2\mathbb{E}\int_0^\infty \lambda I[Y \geq \lambda] d\lambda \leq 2\mathbb{E}\int_0^\infty I[Y \geq \lambda] Z d\lambda = 2\mathbb{E}\left(\int_0^\infty I[Y \geq \lambda] d\lambda\right) = 2\mathbb{E}(ZY).
\]

Hölder’s inequality gives \(\mathbb{E}(YZ) \leq (\mathbb{E}(Y^2))^{1/2}(\mathbb{E}(Z^2))^{1/2}\). Therefore we have \(\mathbb{E}(Y^2) \leq 2\mathbb{E}(YZ) \leq 2(\mathbb{E}(Y^2))^{1/2}(\mathbb{E}(Z^2))^{1/2}\). Hence \((\mathbb{E}Y^2)^{1/2} \leq 2(\mathbb{E}Z^2)^{1/2}\) yielding \(\mathbb{E}Y^2 \leq 4\mathbb{E}Z^2\). \(\blacksquare\)
The results above on discrete (sub)martingales extend to continuous processes. We will require:

**Theorem 52** (Integrable martingale inequality). Let \( X(\cdot) \) be a martingale. Then

\[
\mathbb{E}(\max_{t \in [0,T]} |X(t)|^2) \leq 4\mathbb{E}(X(T)^2). \tag{10}
\]

*Proof.* Let \( \{t_1, t_2, \ldots \} \) be a dense subset of \([0,T]\). Then by Lemma 51, \( \mathbb{E}(\max_{t \in [0,T]} |X(t)|^2) = \lim_{n \to \infty} \mathbb{E}(\max_{k=1,\ldots,n} |X(t_k)|^2) \leq \lim_{n \to \infty} 4\mathbb{E}(X(T)^2) = 4\mathbb{E}(X(T)^2). \)

Combining these results, we see that for nonanticipative step processes \( X(\cdot) \), the Itô integral \( \int_{s=0}^t X(s)dW(s) \) is computable and is a continuous martingale. The Itô equality

\[
\mathbb{E}(\int_0^T X(t)dW(t))^2 = \mathbb{E}(\int_0^T X(t)dt)^2.
\]

shows that the integrated process is a square-integrable random variable, and the martingale inequality

\[
\mathbb{E}(\max_{t \in [0,T]} |\int_0^t X(s)dW(s)|^2) \leq 4\mathbb{E}(\int_0^T X(t)dW(t))^2
\]
gives uniform bounds on the integrated process. Combining these inequalities gives

\[
\mathbb{E}(\max_{t \in [0,T]} |\int_0^t X(s)dW(s)|^2) \leq 4\mathbb{E}(\int_0^T X(t)dt)^2.
\]

In terms of norms on the processes, we have

\[
\| \int XdW \|_{\infty,2} \leq 2\|X\|_{2,2}.
\]

Now if \( X_n \) is a Cauchy sequence of step processes converging effectively to \( X_\infty \) in the \( \| \cdot \|_{2,2} \) norm, we have form \( m \geq n \),

\[
\| \int X_m - X_n \, dW \|_{\infty,2} \leq 2\| X_m - X_n \|_{2,2} \leq 2(\| X_m - X_\infty \|_{2,2} + \| X_n - X_\infty \|_{2,2}).
\]

Hence \( \int X_n \, dW \) converges effectively in the \( \| \cdot \|_{\infty,2} \) norm. Further, for continuous processes

\[
\int_0^T X(t)^2 \, dt \leq T \max_{t \in [0,T]} (X(t))^2 = T(\max_{t \in [0,T]} |X(t)|)^2,
\]

**Theorem 53** (Computability of the Itô integral). If \( X \) is a square-integrable step process, then \( \int XdW \) a continuous process computable from \( X \). If \( (X_n)_{n \in \mathbb{N}} \) a sequence of step processes converging effectively to \( X_\infty \) in the \( \| \cdot \|_{2,2} \) norm, then \( \int X_n \, dW \) is a Cauchy sequence converging effectively in the \( \| \cdot \|_{\infty,2} \) norm to a process which we define as \( \int X_\infty \, dW \). Further, if \( X \) is a continuous process, then

\[
\| \int XdW \|_{\infty,2} \leq 2\sqrt{T} \|X\|_{\infty,2}.
\]

### 8 Stochastic differential equations

We now consider stochastic differential equations

\[
dX(t) = f(X(t)) \, dt + g(X(t)) \, dW(t); \quad X(0) = X_0.
\]

The integral form is

\[
X(t) = X_0 + \int_0^t f(X(s)) \, ds + \int_0^t g(X(s)) \, dW(s) =: J[X](t).
\]

We assume \( f, g : \mathbb{R} \to \mathbb{R} \) are Lipschitzian functions with constants \( K \) and \( L \), respectively. We first assume \( X_0 \in M^2(\mathbb{R}) \) is square-integrable, though the case of a constant \( X_0 = x_0 \) will actually suffice.
**Lemma 54.** Suppose $f : \mathbb{R} \to \mathbb{R}$ is a Lipschitzian function with constant $K$, and suppose $Y,Z$ are stochastic processes with $Y(0) = Z(0)$. Then

$$d_{2,\infty}(f(Y), f(Z)) \leq K d_{2,\infty}(Y,Z).$$

**Proof.**

$$\mathbb{E}(\max_{t \in [0,T]}|f(Y(t)) - f(Z(t))|^2) \leq \mathbb{E}(\max_{t \in [0,T]}K|Y(s) - Z(s)|^2) \leq K^2 \mathbb{E}(\max_{t \in [0,T]}|Y(s) - Z(s)|^2). \quad \Box$$

**Lemma 55** (Computability of non-stochastic integrals). The non-stochastic integral $t \mapsto \int_0^t X(s) \, ds$ of a continuous stochastic process is computable. Further,

$$d_{2,\infty}(\int Y dt, \int Z dt) \leq T d_{2,\infty}(Y,Z).$$

**Proof.** Computability is immediate from computability of non-random integrals. Further, we have

$$\mathbb{E}\left(\max_{t \in [0,T]}\left|\int_0^t X(s) \, ds\right|\right)^2 \leq \mathbb{E}\left(\max_{t \in [0,T]}\left|\int_0^t X(s) \, ds\right|\right)^2 \leq \mathbb{E}\left(\int_0^T |X(t)| \, dt\right)^2$$

$$\leq \mathbb{E}(T \max_{t \in [0,T]}|X(t)|)^2 = T^2 \mathbb{E}(\max_{t \in [0,T]}|X(t)|)^2. \quad \Box$$

For stochastic integrals, we recall from Section 7 that

$$\mathbb{E}\left(\max_{t \in [0,T]}\left|\int_0^t X(s) \, dW(s)\right|\right)^2 \leq 4T \mathbb{E}(\max_{r \in [0,T]}|X(r)|)^2 \, ds$$

so

$$d_{2,\infty}(\int Y \, dW, \int Z \, dW) \leq 2\sqrt{T} d_{2,\infty}(Y,Z).$$

Now

$$J[Y](t) - J[Z](t) = \int_0^t f(Y(s)) - f(Z(s)) \, ds + \int_0^t g(Y(s)) - g(Z(s)) \, dW(s)$$

The expected square of the supremum of this quantity is

$$d_{2,\infty}(J[Y], J[Z]) = ||J[Y] - Z||_{2,\infty} \leq ||f(Y) - f(Z)||_{2,\infty} + ||g(Y) - g(Z)||_{2,\infty} d_{2,\infty}(Y,Z)$$

$$= d_{2,\infty}(\int f(Y) dt, \int f(Z) dt) + d_{2,\infty}(\int g(Y) \, dW, \int g(Z) \, dW)$$

$$\leq T d_{2,\infty}(f(Y), f(Z)) + 2\sqrt{T} d_{2,\infty}(g(Y), g(Z)) \leq (TK + 2\sqrt{T}L) d_{2,\infty}(Y,Z).$$

Hence we have

$$d_{2,\infty}(J[Y] - J[Z]) \leq (KT + 2L\sqrt{T}) d_{2,\infty}(Y,Z).$$

Taking $T < \min(1/2K, 1/16L^2)$ gives $(KT + 2L\sqrt{T}) < 1$, so

$$d_{2,\infty}(J[Y], J[Z]) \leq \kappa d_{2,\infty}(Y,Z)$$

where $\kappa < 1$.

The integral form of the equation i.e. the Picard operator, is therefore a contraction map for small enough $T$. Taking $X_0$ to be the constant process, $X_0(t) \equiv X_0$, and setting $X_{n+1} = J[X_n]$ for all $n \in \mathbb{N}$ gives an effectively convergent subsequence. The initial difference is given by

$$d_{2,\infty}(X_1 - X_0) \leq \kappa (\mathbb{E}X_0^2)^{1/2}.$$

By joining together computations of the solution for small enough $T$, we obtain:
Theorem 56 (Computability of Lipschitz stochastic differential equations). Consider the stochastic differential equation
\[ dX(t) = f(X(t)) \, dt + g(X(t)) \, dW(t); \quad X(0) = X_0 \]
where \( f, g : \mathbb{R} \to \mathbb{R} \) are Lipschitz and \( X_0 \in R(\mathbb{R}) \). Then the solution \( X(t) \) is computable as a random variable taking values in \( C([0, \infty); \mathbb{R}) \).

Proof. Let \( K \) be the Lipschitz constant for \( f, g \), and choose \( T < \min(4, 1/16K^2) \). For a given \( x_0 \), the solution with initial condition \( X(0) = x_0 \) is computable in \( M^2(C([0,T]; \mathbb{R})) \), and hence in \( R(C([0,T]; \mathbb{R})) \). Hence the solution operator given initial condition \( x_0 \in \mathbb{R} \) is computable \( \mathbb{R} \to R(C([0,T]; \mathbb{R})) \). For an initial condition which is a probability distribution over \( \mathbb{R} \), the solution is computable over \([0,T]\) by Theorem 53. Then the random variable \( X(T) \) is computable by projection. The result follows by recursively computing \( X \) over the intervals \([kT, (k+1)T]\).

9 Conclusions

In this paper, we have developed a theory of probability, random variables and stochastic processes which is sufficiently powerful to effectively compute the solution of stochastic differential equations. The theory uses type-two effectivity to provide an underlying machine model of computation, but is largely developed using type theory in the cartesian-closed category of quotients of countably-based spaces, which has an effective interpretation. The approach extends existing work on probability via valuations and random variables in metric spaces via limits of Cauchy sequences. Ultimately, we hope that this work will form a basis for practical software tools for the rigorous computational analysis of stochastic systems.

Acknowledgement: The author would like to thank Bas Spitters for many interesting discussions on measurable functions and type theory, and pointing out the connection with monads.

References

[AM02] Mauricio Alvarez-Manilla. Extension of valuations on locally compact sober spaces. *Topology Appl.*, 124:397–433, 2002.

[BB85] Errett Bishop and Douglas Bridges. *Constructive analysis*, volume 279 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 1985.

[BC72] Errett Bishop and Henry Cheng. *Constructive measure theory*. American Mathematical Society, 1972.

[Bra05] Vasco Brattka. Effective Borel measurability and reducibility of functions. *Math. Logic Quarterly*, 51:19–44, 2005.

[Cha72] Yuen-Kwok Chan. A constructive approach to the theory of stochastic processes. *Transactions of the American Mathematical Society*, 165:37–44, 1972.

[Col09] Pieter Collins. Computable types for dynamic systems. In *Proceedings of the Fifth Conference on Computability in Europe*, 2009.

[CS09] Thierry Coquand and Bas Spitters. Integrals and valuations. *J. Logic Analysis*, 1(3):1–22, 2009.
[Dob07] Ernst-Erich Doberkat. *Stochastic Relations: Foundations for Markov Transition Systems*. Chapman & Hall, 2007.

[Eda95] Abbas Edalat. Domain theory and integration. *Theor. Comput. Sci.*, 151:163–193, November 1995.

[Esc04] Martín Escardó. Synthetic topology. *Electron. Notes Theor. Comput. Sci.*, 87:21–156, November 2004.

[Esc09] Martín Escardó. Semi-decidability of may, must and probabilistic testing in a higher-type setting. *Electron. Notes Theor. Comput. Sci.*, 249:219–242, August 2009.

[Eva13] Craig Evans. *An Introduction to Stochastic Differential Equations*. AMS, 2013.

[Fri75] Avner Friedman. *Stochastic Differential Equations*. Dover, 1975.

[GLV11] Jean Goubault-Larrecq and Daniele Varacca. Continuous random variables. In *Proceedings of the 2011 IEEE 26th Annual Symposium on Logic in Computer Science*, pages 97–106, Washington, DC, USA, 2011.

[GS04] I. Gikhman and A. Skorohod. *The Theory of Stochastic Processes 1*. Springer, 2004.

[JP89] C. Jones and G. Plotkin. A probabilistic powerdomain of evaluations. In *Proceedings of the Fourth Annual Symposium on Logic in computer science*, pages 186–195, Piscataway, NJ, USA, 1989.

[Kön97] H. König. *Measure and Integration*. Springer-Verlag, 1997.

[Law04] Jimmie D. Lawson. Domains, integration and ‘positive analysis’. *Mathematical Structures in Comp. Sci.*, 14:815–832, December 2004.

[Mis07] Michael Mislove. Discrete random variables over domains. *Theor. Comput. Sci.*, 380:181–198, July 2007.

[MW43] H.B. Mann and A. Wald. On stochastic limit and order relationships. *Ann. Math. Statistics*, 14(3):217–226, 1943.

[OS10] Russell O’Connor and Bas Spitters. A computer-verified monadic functional implementation of the integral. *Theor. Comput. Sci.*, 411(37):3386–3402, 2010.

[Sch07] Matthias Schröder. Admissible representations of probability measures. *Electron. Notes Theor. Comput. Sci.*, 167:61–78, January 2007.

[Sch09] Matthias Schröder. An effective Tietze-Urysohn theorem for QCB-spaces. *J. Univers. Comput. Sci.*, 15(6):1317–1336, 2009.

[Shi95] Al’bert Nikolaevich Shiryaev. *Probability*. Springer, 1995.

[Spi03] Bas Spitters. *Constructive and intuitionistic integration theory and functional analysis*. PhD thesis, Katholieke Universiteit Nijmegen, 2003.

[Spi06] Bas Spitters. Constructive algebraic integration theory. *Ann. Pure Appl. Logic*, 137(1-3):380–390, 2006.

[SS06] Matthias Schröder and Alex Simpson. Probabilistic observations and valuations. *Electron. Notes Theor. Comput. Sci.*, 155:605–615, May 2006.
[Str72] Ross Street. The formal theory of monads. *J. Pure Appl. Math.*, 2:149–168, 1972.

[Tix95] R. Tix. *Stetige Bewertungen auf topologischen Räumen*. PhD thesis, Master’s Thesis, Technische Universität Darmstadt, 1995.

[Var02] Daniele Varacca. The powerdomain of indexed valuations. In *Proceedings of the 17th Annual IEEE Symposium on Logic in Computer Science*, pages 299–, Washington, DC, USA, 2002.

[Vic08] Steven Vickers. A localic theory of lower and upper integrals. *Math. Log. Quart.*, 54(1):109–123, 2008.

[Vic11] Steven Vickers. A monad of valuation locales. [http://www.cs.bham.ac.uk/~sjv/Riesz.pdf](http://www.cs.bham.ac.uk/~sjv/Riesz.pdf), 2011.

[Wei99] Klaus Weihrauch. Computability on the probability measures on the Borel sets of the unit interval. *Theor. Comput. Sci.*, 219:421–437, May 1999.

[WI81] S. Watanabe and N. Ikeda. *Stochastic Differential Equations and Diffusion Processes*. North-Holland, 1981.