0

Probabilistic background

This chapter summarises some aspects of measure theory and discusses the construction of canonical stochastic processes. We then turn to Brownian motion and Poisson processes to motivate some of the results of Chapters 3 and 4. The development of those chapters is independent of these examples, but since they inspired much of the general theory some knowledge of their properties will greatly aid understanding of that theory.

0.1. Measure and probability

The following concepts should be familiar, but are collected here for ease of reference (further details can be found, for example, in [8], [46]).

0.1.1. ***Definition:*** A *measure space* is a triple $(\Omega, \mathscr{F}, P)$, where Ω is a set, $\mathscr{F}$ a *σ-field* of subsets of Ω (that is, $\Omega \in \mathscr{F}$ and $\mathscr{F}$ is closed under the formation of complements and countable unions) and P is a set function $\mathscr{F} \to [0, \infty]$ which is *countably additive*: if $(A_i)_{i \geq 1}$ is a sequence in $\mathscr{F}$ with $A_i \cap A_j = \varnothing$ when $i \neq j$, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. We shall deal almost exclusively with *probability spaces*, where P has range in $[0,1]$, and also $P(\Omega) = 1$.. Unless otherwise indicated, we shall also take $(\Omega, \mathscr{F}, P)$ to be *complete*: this means that if $F \in \mathscr{F}$ has $P(F) = 0$ and $G \subseteq F$, then G must necessarily belong to $\mathscr{F}$ (and, of course, $P(G) = 0$). If $(\Omega, \mathscr{F}, P)$ is complete, a sub-σ-field $\mathscr{G}$ of $\mathscr{F}$ (in our framework this implies that $\Omega \in \mathscr{G}$) is said to be *complete* if it contains all $F \in \mathscr{F}$ with $P(F) = 0$. These sets are referred to as *P-null sets* of $\mathscr{G}$.

It is worth recalling that any probability space $(\Omega, \mathscr{F}, P)$ can be 'completed' as follows: the *completion* $(\Omega, \bar{\mathscr{F}}, P)$ of $(\Omega, \mathscr{F}, P)$ is defined by putting $F \in \bar{\mathscr{F}}$ if there exist F_1, F_2 in $\mathscr{F}$ with $F_1 \subseteq F \subseteq F_2$ and $P(F_1) = P(F_2)$, and defining $P(F) = P(F_1) = P(F_2)$. It is clear that $(\Omega, \bar{\mathscr{F}}, P)$ is then a complete probability space.

The completion is closely related to the (Caratheodory) *inner measure* P_* and *outer measure* P^* induced by P on arbitrary subsets of Ω: if $A \subseteq \Omega$, let $P_*(A) := \sup\{P(F)\colon F \subseteq A,\ F \in \mathscr{F}\}$ and $P^*(A) := \inf\{P(F)\colon F \supseteq A,\ F \in \mathscr{F}\}$. Then $\bar{\mathscr{F}}$ can be characterised as $\bar{\mathscr{F}} = \{A \subseteq \Omega\colon P^*(A) = P_*(A)\}$ and the common value of P^* and P_* at $A \in \bar{\mathscr{F}}$ defines the extension of P to A.

0.1.2. Although we shall discuss martingale theory in the context of complete probability spaces, the reader should be aware that this restriction precludes discussion of some of the subtler concepts and extensions of the theory developed in recent years (see [19]), and we thus do not discuss some of the most interesting facets of Brownian motion, in particular, which have given rise to these extensions (see [83] for further discussion of these matters).

Now fix a complete probability space $(\Omega, \mathscr{F}, P)$.

0.1.3. ***Definition:*** A measurable function $f\colon \Omega \to \mathbf{R}$ or *random variable* satisfies $f^{-1}(B) \in \mathscr{F}$ for all Borel sets $B \subseteq \mathbf{R}$. (The Borel σ-field $\mathscr{B}(\mathbf{R})$ is that generated by the open intervals in $\mathbf{R}$.) Two random variables f and g will normally be identified if the set $\{\omega \in \Omega\colon f(\omega) \neq g(\omega)\}$ is P-null. We say that $f = g$ a.s. (*almost surely*). By abuse of notation we shall identify the random variable f with the equivalence class $\{g : f = g \text{ a.s.}\}$; this is unlikely to cause any confusion. Thus equations, inequalities, etc., between random variables are assumed to hold a.s. without explicit mention. A sequence (f_n) of random variables *converges a.s.* to f (which is then trivially also a random variable) iff $f_n(\omega) \to f(\omega)$ for *almost all* ω (i.e. except possibly on a P-null set). The vector space of all random variables on $(\Omega, \mathscr{F}, P)$ is denoted by $\mathscr{L}^0 := \mathscr{L}^0(\Omega, \mathscr{F}, P)$.

The space of equivalence classes of functions in $\mathscr{L}^0$, under the equivalence relation '$f \sim g$ iff $f = g$ a.s.' is denoted by $L^0 := L^0(\Omega, \mathscr{F}, P)$. We shall treat $f \in L^0$ as if it were a random variable (which need only be defined a.s. or can take the values $+\infty$ or $-\infty$ on some P-null set). L^0 can be equipped with the metric d of *convergence in probability*: if $f, g \in L^0$, define $d(f,g) = \int_\Omega \min(1, |f(\omega) - g(\omega)|)\,dP(\omega)$. Then (L^0, d) is a complete metric space and d-convergence of a sequence (f_n) in L^0 to f is equivalent to the statement: for any given $\varepsilon > 0$ we can find N such that $P\{|f_n - f| \geqslant \varepsilon\} < \varepsilon$ for all $n \geqslant N$. (Here $\{|f_n - f| \geqslant \varepsilon\}$ is the set $|\omega \in \Omega : |f_n(\omega) - f(\omega)| \geqslant \varepsilon\}$. Abbreviations of this type will be used freely in the sequel.)

The Banach spaces $L^p := L^p(\Omega, \mathscr{F}, P)$, for $1 \leqslant p \leqslant \infty$, are defined via the norms $\|f\|_p = (\int_\Omega |f|^p\,dP)^{1/p}$ for $1 \leqslant p < \infty$ and $\|f\|_\infty = \text{ess sup}_{\omega \in \Omega} |f(\omega)|$. Note that $L^p = \{f \in L^0 : \|f\|_p < \infty\}$. It is not hard to show that norm-convergence of a sequence $(f_n) \subseteq L^p$ to f (meaning that $\|f_n - f\|_p \to 0$) implies convergence of (f_n) to f in probability.

The integral $\int_\Omega f dP$ of a random variable $f \in L^1$ is called the *expectation* of f and denoted by $\mathbf{E}(f)$. If we allow $\mathbf{E}(f)$ to take the value $+\infty$, generalised expectations can be defined on L^0_+. (For $p=0$ or $1 \leqslant p < \infty$, $L^p_+ = \{f \in L^p: f \geqslant 0\}$.) Since $P(\Omega)=1$, $\mathbf{E}(f)$ represents the 'average value' or *mean* of f over Ω. For $1 \leqslant p < \infty$, $\|f\|_p^p$ represents the pth moment of f. Of particular importance is the second moment $\|f\|_2^2 = (\int_\Omega |f|^2 \, dP)$.

The *variance* $\sigma^2 = \mathbf{E}((f - \mathbf{E}(f))^2)$ measures the dispersion of f about the mean $\mathbf{E}(f)$, distances being taken in the Hilbert space L^2.

Finally, we recall three well-known convergence theorems for sequences in L^1; these will be in constant use throughout this book. For proofs, see [77].

Monotone convergence theorem: If (f_n) is a monotone increasing sequence in L^1 with a.s. limit f and such that $(\mathbf{E}(f_n))$ is bounded above, then f is in L^1 and $\|f_n - f\|_1 \to 0$. Hence also $\mathbf{E}(f_n) \uparrow \mathbf{E}(f)$.

This result extends to L^0_+ if we allow $\mathbf{E}(f)$ to take the value $+\infty$. In that case the boundedness condition is superfluous.

Fatou's lemma: If (f_n) is in L^0_+ then $\mathbf{E}(\liminf_{n\to\infty} f_n) \leqslant \liminf_{n\to\infty} \mathbf{E}(f_n)$.

Dominated convergence theorem: If $f_n \to f$ a.s. and there exists g in L^1 such that $|f_n| \leqslant g$ for all n, then $f \in L^1$ and $\mathbf{E}(f_n) \to \mathbf{E}(f)$.

Exercises:

(1) Let (f_n), $n \geqslant 1$, and f be functions in L^0.
   (i) Show that if $f_n \to f$ in L^p-norm, then $f_n \to f$ in probability.
   (ii) Show that if $f_n \to f$ in probability, then there is a subsequence (f_{n_k}) converging to f a.s.

(2) The following basic facts from elementary probability theory will be useful on occasion. Prove them.
   (i) Chebychev's inequality: Let $f \in L^2$ and $t \in \mathbf{R}$ be given. Then
   $$P(|f| > t) \leqslant \frac{\mathbf{E}(f^2)}{t^2}.$$
   (ii) Borel–Cantelli lemmas: if $(A_n) \subset \mathscr{F}$ and $\sum_n P(A_n) < \infty$, then $P(\bigcap_k \bigcup_{n \geqslant k} A_n) = 0$. If (A_n) are independent events (see 0.1.4) and $\sum_n P(A_n) = +\infty$, then $P(\bigcap_k \bigcup_{n \geqslant k} A_n) = 1$.

0.1.4. *Conditioning:* In attempting to model 'reality' by means of the probability space $(\Omega, \mathscr{F}, P)$ we can think of the sets in $\mathscr{F}$ as possible 'events', and $P(A)$ is then our assignment of the probability that A occurs.

Our further assignment of probabilities may be influenced by the knowledge that A has occurred (think of the effect of election results upon the stock market!). We define the *conditional probability* of $B \in \mathscr{F}$, given that A has occurred, and $P(A)>0$, as

$$P(B|A)=\frac{P(A\cap B)}{P(A)}.$$

For example, given that a family with exactly two children has at least one boy, what are the chances both children are boys? Here event $A=\{$the family has at least one boy$\}$ has probability $\frac{3}{4}$, assuming that the possible combinations of sexes are all equally likely. On the other hand, if $B=\{$both are boys$\}$, then $P(B\cap A)=P(B)=\frac{1}{4}$, so $P(B|A)=\frac{1}{4}/\frac{3}{4}=\frac{1}{3}$. (If this result seems surprising, consider the respective lengths of file indexes of families with at least one boy, and that of families with two boys. See [31] for a further discussion of such examples.)

Taking the conditional probability with respect to A amounts to choosing A as the new sample space (instead of Ω) and normalising to make the probability of A equal to 1. This indicates that all general theorems for probabilities will have counterparts for conditional probabilities. The distinctive nature of probability theory lies in the study of *independent events*, that is, events A and B for which $P(A|B)=P(B)$, or in other words, where $P(A)\cdot P(B)=P(A\cap B)$. Here the restriction of our 'universe' to A does not alter the likelihood that B occurs. (See [31; Ch. V] for detailed discussions.)

Now if $f \in L^1$ we can define the *conditional expectation* of f, given A in $\mathscr{F}$, as the 'average value'

$$\mathbf{E}(f|A)=\frac{1}{P(A)}\int_A f\,\mathrm{d}P$$

of f on A, by analogy with the definitions of $\mathbf{E}(f)$ and $P(B|A)$. Note that

$$\mathbf{E}(1_B|A)=\frac{1}{P(A)}\int_A 1_B\,\mathrm{d}P=\frac{P(A\cap B)}{P(A)}=P(B|A).$$

We can interpret $\mathbf{E}(f|A)$ as our 'best estimate' of the values of f, given only the 'information' contained in A (and hence in its complement, A^c). In a finite sample space Ω, this information amounts to knowing whether a given $\omega \in \Omega$ belongs to A or not. Now the event A generates the σ-field $\{\varnothing, A, A^c, \Omega\}$. More generally, we can regard any sub-σ-field $\mathscr{G}$ of $\mathscr{F}$ as containing some information – whether relevant to f or not. This also enables us to measure the 'amount of information' given: the larger the σ-

field the more information it contains (think of the σ-fields generated by ever finer partitions of Ω). The conditional expectation $\mathbf{E}(f|\mathscr{G})$ will then represent our 'best guess' at the values of f, given only the information in $\mathscr{G}$. The usual construction of $\mathbf{E}(f|\mathscr{G})$ for $f \in L^1$ (or even $f \in L^0_+$) as the unique $\mathscr{G}$-measurable integrable function (write $L^1(\mathscr{G})$ for $L^1 \cap L^0(\mathscr{G})$) such that $\int_G f \mathrm{d}P = \int_G \mathbf{E}(f|\mathscr{G}) \mathrm{d}P$ for all $G \in \mathscr{G}$, is via the *Radon–Nikodym theorem*. This states that if μ is a bounded measure on $(\Omega, \mathscr{F})$ which is absolutely continuous with respect to P (i.e. $\mu(A)=0$ whenever $A \in \mathscr{F}$ satisfies $P(A)=0$), then there exists a unique $X \in L^1_+$ with $\int_F X \,\mathrm{d}P = \mu(F)$ for all $F \in \mathscr{F}$. It is easy to extend this result to bounded signed measures (countably additive real-valued set functions), where $X \in L^1$ need no longer be positive. Apply this with P restricted to the sub-σ-field $\mathscr{G}$ of $\mathscr{F}$ and μ on $\mathscr{G}$ defined by $\mu(G) = \int_G f \,\mathrm{d}P$, to obtain $X = \mathbf{E}(f|\mathscr{G}) \in L^1(\Omega, \mathscr{G}, P)$ such that $\mu(G) = \int_G X \,\mathrm{d}P$ for all $G \in \mathscr{G}$.

We shall deduce the Radon–Nikodym theorem as a consequence of the martingale convergence theorem in Chapter 2. For this reason we include in Chapter 2 a definition of $\mathbf{E}(f|\mathscr{G})$ which does not require the Radon–Nikodym theorem, but is based instead upon the characterisation of the operator $\mathbf{E}(\cdot|\mathscr{G})$ in L^2 as the orthogonal projection onto the subspace $L^2(\mathscr{G})$. This will exhibit $\mathbf{E}(f|\mathscr{G})$ as the $\mathscr{G}$-measurable function 'nearest' to f in the least-squares sense. Thus $\mathbf{E}(f|\mathscr{G})$ represents our 'best estimate' of f given only the information contained in $\mathscr{G}$.

0.1.5. ***The Monotone Class Theorem:*** Suppose that we wish to prove that all sets or functions in some class $\mathscr{C}$ have a property (*). One way of doing this is to find a collection $\mathscr{C}_0$ of sets or functions which 'generates' $\mathscr{C}$, so that each element of $\mathscr{C}$ can be constructed from $\mathscr{C}_0$ using certain operations. If each element of $\mathscr{C}_0$ has (*) and the class of all sets or functions which have (*) is closed under these operations, then each element of $\mathscr{C}$ has (*). We shall repeatedly use this procedure for σ-fields of sets and vector spaces of measurable functions using the following two versions of the *Monotone Class Theorem* (there are many versions with this name: see [19; Ch. I]):

Let Ω be a set, $\mathscr{S}$ a collection of subsets of Ω, closed under finite intersections.

(*a*) Let $\mathscr{M}(\mathscr{S})$ be the smallest collection of subsets of Ω which contains $\mathscr{S}$ and satisfies

(i) $\Omega \in \mathscr{M}(\mathscr{S})$,

(ii) if $A, B \in \mathscr{M}(\mathscr{S})$ and $A \subseteq B$, then $B \setminus A \in \mathscr{M}(\mathscr{S})$,

(iii) if (A_n) is an increasing sequence in $\mathscr{M}(\mathscr{S})$, then $\bigcup_n A_n \in \mathscr{M}(\mathscr{S})$.

Under these conditions $\mathscr{M}(\mathscr{S})$ is the smallest σ-field containing $\mathscr{S}$.

(*b*) Let $\mathscr{H}$ be a vector space of functions from Ω to $\mathbf{R}$ satisfying

(i) $1 \in \mathscr{H}$ and $1_A \in \mathscr{H}$ for $A \in \mathscr{S}$,

(ii) if (f_n) is an increasing sequence of non-negative functions in $\mathscr{H}$ with bounded supremum, then $\sup_n f_n \in \mathscr{H}$.

Then $\mathscr{H}$ contains all bounded $\sigma(\mathscr{S})$-measurable real functions on Ω.

Proof: (*a*) If $\sigma(\mathscr{S})$ is the σ-field generated by $\mathscr{S}$, it satisfies (i)–(iii) trivially and contains $\mathscr{S}$, hence $\sigma(\mathscr{S}) \supseteq \mathscr{M}(\mathscr{S})$. To prove the converse inclusion, it will be enough to show that $\mathscr{M}(\mathscr{S})$ is closed under *finite* intersections. For then we can express any countable union of sets (M_i) in $\mathscr{M}(\mathscr{S})$ as follows: set $N_k = \bigcup_{i=1}^k M_i$, which is in $\mathscr{M}(\mathscr{S})$ since $N_k = \Omega \backslash (\Omega \backslash \bigcup_{i=1}^k M_i) = \Omega \backslash \bigcap_{i=1}^k (\Omega \backslash M_i)$. So by (iii), $\bigcup_{i=1}^\infty M_i = \bigcup_{k=1}^\infty N_k \in \mathscr{M}(\mathscr{S})$. Thus $\mathscr{M}(\mathscr{S})$ is a σ-field.

To prove that $\mathscr{M}(\mathscr{S})$ is closed under finite intersections, first set $\mathscr{D}_1 = \{B \in \mathscr{M}(\mathscr{S}) \colon B \bigcap A \in \mathscr{M}(\mathscr{S})$ for all $A \in \mathscr{S}\}$. Since $\mathscr{S}$ is closed under finite intersections by hypothesis, $\mathscr{D}_1 \supset \mathscr{S}$. We can now check that $\mathscr{D}_1$ satisfies (i)–(iii) to conclude that $\mathscr{D}_1 = \mathscr{M}(\mathscr{S})$. (Exercise!) Finally, let $\mathscr{D}_2 = \{B \in \mathscr{M}(\mathscr{S}) \colon B \bigcap A \in \mathscr{M}(\mathscr{S})$ for all $A \in \mathscr{M}(\mathscr{S})\}$. Again one may check easily that D_2 satisfies (i)–(iii). Moreover, if $A \in \mathscr{S}$, $B \bigcap A \in \mathscr{M}(\mathscr{S})$ for all $B \in \mathscr{D}_1 = \mathscr{M}(\mathscr{S})$, so $\mathscr{S} \subseteq \mathscr{D}_2$. Hence $\mathscr{D}_2 = \mathscr{M}(\mathscr{S})$, and this means that $\mathscr{M}(\mathscr{S})$ is closed under finite intersections.

(*b*) Let $\mathscr{M} = \{A \colon 1_A \in \mathscr{H}\}$. Then $\mathscr{S} \subseteq \mathscr{M}$, $\Omega \in \mathscr{M}$ and $\mathscr{M}$ is closed under relative complements (if $A, B \in \mathscr{M}$, $A \subseteq B$, then $1_{B \backslash A} = 1_B - 1_A \in \mathscr{H}$). Also, if (A_i) is an increasing sequence in $\mathscr{M}$, and $A = \bigcup_{i=1}^\infty A_i$, then $1_A = \sup_{i \geqslant 1} 1_{A_i} \in \mathscr{H}$. By part (*a*), $\mathscr{M} = \sigma(\mathscr{S})$. Now if $f \colon \Omega \to \mathbf{R}$ is $\sigma(\mathscr{S})$-measurable and bounded, let $f = f^+ - f^-$. Each of f^+ and f^- is the supremum of a sequence of $\mathscr{M}$-simple functions, hence belongs to $\mathscr{H}$ by (iii). So $f \in \mathscr{H}$ as required.

0.1.6. ***Stochastic processes and their distributions:*** A random variable X induces a probability measure P_X on $(\mathbf{R}, \mathscr{B}(\mathbf{R}))$, the *distribution* of X, by $P_X(B) = P(X^{-1}(B))$ for $B \in \mathscr{B}(\mathbf{R})$. This Lebesgue–Stieltjes measure is generated by the increasing right-continuous function F_X, the *distribution function* of X, by $F_X(t) = P\{X \leqslant t\}$. If $X \in L^1(\Omega, \mathscr{F}, P)$, $\mathbf{E}(X) = \int_{\mathbf{R}} x \, dP_X(x)$.

Given a finite sequence $X_1, X_2, \ldots, X_n$ of random variables, let $Z(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$ for all $\omega \in \Omega$. This defines a measurable function $Z \colon \Omega \to \mathbf{R}^n$, where $\mathbf{R}^n$ is given the Borel σ-field $\mathscr{B}(\mathbf{R}^n)$. Hence Z induces a probability measure $P_Z = P_{X_1, \ldots, X_n}$ on $(\mathbf{R}^n, \mathscr{B}(\mathbf{R}^n))$, the *n-dimensional joint distribution* of $X_1, \ldots, X_n$, by $P_Z(B) = P(Z^{-1}(B))$.

We can think of a stochastic process $X = (X_t)_{t \in \mathbf{T}}$ as a family of random variables indexed by some $\mathbf{T} \subseteq \mathbf{R}$. (But see also section 3.1.) Usually we take $\mathbf{T} = \mathbf{N}$ or as an interval in $\mathbf{R}^+$. If $\mathbf{T}$ models the passage of time and X models the time-evolution of some observed system, an immediate practical difficulty is that we can only make finitely many observations. Thus we only observe $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ for some $t_1, \ldots, t_n$ in $\mathbf{T}$. The question arises to what extent these observations determine X, i.e. how many different models can be built upon the same sets of observations? Writing $T = (t_1, \ldots, t_n)$ we can define the measurable function $X_T = (X_{t_1}, \ldots, X_{t_n})$ as above and determine the joint distribution P_{X_T}. Doing this for all possible choices of n and T then yields the set of all *finite-dimensional distributions* of X. We can now rephrase our question: can a process X be constructed uniquely to have a given set of finite-dimensional distributions?

Kolmogorov's extension theorem provides an explicit canonical construction of X on the product space $\mathbf{R}^{\mathbf{T}}$ when we have a *projective system* of probability measures: for each pair of finite subsets $S \subseteq T$ of $\mathbf{T}$, $P_S = P_T \circ \Pi_{T,S}^{-1}$, where $\Pi_{T,S}: \mathbf{R}^T \to \mathbf{R}^S$ is the natural projection map. This allows us to construct a unique probability measure μ on $\Lambda = \mathbf{R}^{\mathbf{T}}$ as the *projective limit* of the system $\{P_S : S \subset \mathbf{T}, \text{ finite}\}$, so that for each finite $S \subseteq \mathbf{T}$, $P_S = \mu \circ \Pi_S^{-1}$, where $\Pi_S : \Lambda \to \mathbf{R}^S$ is the natural projection map.

The construction of such a projective limit measure μ proceeds from the *Caratheodory extension theorem* for measures: if $\mathscr{E}$ is a *field* of subsets of Ω (replacing countable unions by finite unions in the definition of a σ-field yields the definition of a field) and μ is a probability measure on $\mathscr{E}$ (so if $\bigcup_{i=1}^{\infty} E_i \in \mathscr{E}$ for disjoint E_i, then $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu E_i$), then μ extends uniquely to the σ-field $\sigma(\mathscr{E})$ generated by $\mathscr{E}$. (See [37] for a proof.)

To use this result, we define the field $\mathscr{C}$ of *cylinder sets* of $\Lambda = \mathbf{R}^{\mathbf{T}}$, given by the finite-dimensional projection maps: given a finite set $S \subseteq \mathbf{T}$, let $\mathscr{C}_S = \Pi_S^{-1}(\mathscr{B}(\mathbf{R}^S))$, i.e. $C \in \mathscr{C}_S$ iff $C = \{\omega \in \Lambda : \Pi_S(\omega) \in B\}$ for some Borel set $B \subseteq \mathbf{R}^S$. Then each $\mathscr{C}_S$ is a σ-field (Exercise!). We set $\mathscr{C} = \{C \in \mathscr{C}_S : S \subseteq \mathbf{T}, \text{ finite}\}$.

To prove that $\mathscr{C}$ is a field one obviously requires consistency conditions. Thus, given a family $\{P_S : S \subseteq \mathbf{T}, \text{ finite}\}$ of finite-dimensional probability distributions, we require that

(i) if $S_1 = \sigma(S)$ is a permutation of the elements of S, then $P_{S_1}(B) = P_S(f_\sigma^{-1}(B))$ for any Borel set $B \subseteq \mathbf{R}^S$, where $f_\sigma(x_1, \ldots, x_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$.

(ii) if $S = \{s_1, \ldots, s_n\}$ and $T = \{s_1, \ldots, s_n, t_{n+1}\}$, then $P_S(B) = P_T(B \times \mathbf{R})$ for all Borel sets $B \subseteq \mathbf{R}^S$.

(This is of course just an explicit statement of the requirement that the probability distributions form a projective system.)

The measure μ on $\mathscr{C}$ is now defined by setting $\mu = P_S \circ \Pi_S$ for each finite $S \subseteq \mathbf{T}$. The consistency conditions ensure that μ is well-defined, since any two representations of a cylinder set can be related by projections and permutation of indices. To show that μ is countably additive on $\mathscr{C}$ we need only prove that $\mu(C_n)\to 0$ when $(C_n)\subseteq\mathscr{C}$ is a decreasing sequence with empty intersection. But this follows because for each Borel set $B\subseteq\mathbf{R}^S$ we can find a compact set $K\subset B$ such that $P_S(B\setminus K)$ is arbitrarily small (this expresses the fact that each P_S is *tight* – see [4], [83, p. 25ff]). For if $\mu(C_n)\to\alpha>0$ we can assume that each $B_n=\Pi_{S_n}(C_n)$ is compact and that the index sets S_n defining C_n increase with n. Taking $\omega_n\in C_n$ we can find convergent subsequences $\{\Pi_\alpha(\omega_n)\}$ for each $\alpha\in\mathbf{T}$, and a diagonal argument provides a point $\omega\in\bigcap_{n=1}^{\infty}C_n$. (The details may be found in [52], a more sophisticated proof in [67].)

So μ is a probability measure on $\mathscr{C}$, hence extends to a probability measure on $\sigma(\mathscr{C})$ by Caratheodory's theorem. Thus we have constructed the probability space $(\Lambda, \sigma(\mathscr{C}), \mu)$. Finally we define the stochastic process X on $(\Lambda, \sigma(\mathscr{C}), \mu)$ by setting $X_t(\omega)=\omega(t)$ for $t\in\mathbf{T}$, $\omega\in\Lambda$, where $\omega(t)=\Pi_{\{t\}}(\omega)$. It is then clear that X_t is $\sigma(\mathscr{C})$-measurable and that for S finite, $P_{X_S}=\mu\circ\Pi_S^{-1}=P_S$. We have 'proved' the following result!

0.1.7. ***Theorem (Daniell–Kolmogorov):*** Given a projective system of finite-dimensional probability distributions $\Phi=\{P_S: S\subseteq\mathbf{T}, \text{finite}\}$ there is a stochastic process X having Φ as its system of finite-dimensional distributions. Moreover, the process X can be defined uniquely on the probability space $(\mathbf{R}^\mathbf{T},\sigma(\mathscr{C}),\mu)$, by setting $X_t(\omega)=\omega(t)$ for $\omega\in\mathbf{R}^\mathbf{T}$, $t\in\mathbf{T}$. Thus if $Y=(Y_t)_{t\in\mathbf{T}}$ is any stochastic process on a probability space $(\Omega',\mathscr{F},P)$ with Φ as its system of finite-dimensional distributions, then Y has a *canonical representation* X on $(\mathbf{R}^\mathbf{T},\sigma(\mathscr{C}),\mu)$.

It is clear that Theorem 0.1.7 is fundamental in the construction of stochastic processes. It is now natural to say that two stochastic processes are *equivalent* if they have the same system of finite-dimensional distributions, since this will ensure that they have the same canonical representation on the function space $\mathbf{R}^\mathbf{T}$. Of particular interest is the case when the canonical process 'lives' on a particular subset of $\mathbf{R}^\mathbf{T}$, i.e. its *paths* $t\to\omega(t)$ μ-almost surely possess a certain property, such as continuity. The verification of such properties requires much more sophisticated techniques and relies heavily on the form of the given system of finite-dimensional distributions, as we shall see below.

The discussion of the *paths* $t\to X_t(\omega)$ of a stochastic process X will in general require rather stronger notions of equivalence of process than the above. We define two such notions in Exercise 0.1.8. They will be discussed further in Chapter 3.

To what extent these finer distinctions accord with 'reality' naturally remains debatable.

0.1.8. ***Exercises:***

(1) Let X and Y be stochastic processes on $(\Omega, \mathscr{F}, P)$, with parameter set $\mathbf{T}=[0, \infty[$. Suppose that $X_t = Y_t$ a.s. (P) for all $t \in \mathbf{T}$. (We say that Y is a *modification* of X; see Chapter 3.) Show that X and Y are equivalent.

(2) Show that if X and Y have a.s. *continuous* paths so that $t \to X(t,\omega)$, $t \to Y(t,\omega)$ are continuous except on some P-null set) and X is a modification of Y, then there is a single P-null set N such that the paths $t \to X(t,\omega)$ and $t \to Y(t,\omega)$ are identical for all $\omega \notin N$. (First consider $t \in \mathbf{Q}^+$.)

This result is extended to right-continuous processes in section 3.1. Quite generally, we say that two processes X and Y are *indistinguishable* if, for almost all ω, the paths $t \to X(t, \omega)$ and $t \to Y(t, \omega)$ are identical.

0.1.9. ***Definition:*** Let $X = (X_t)_{t \in \mathbf{T}}$ be a stochastic process, with finite-dimensional distributions $\Phi = \{P_{X_S} : S \subseteq \mathbf{T}, \text{ finite}\}$. The measure $\mu_X := \mu$ defined by Φ on $(\mathbf{R}^{\mathbf{T}}, \sigma(\mathscr{C}))$ is called the *distribution* of X. If we view X as a *random function*, that is as a map $X : \Omega \to \mathbf{R}^{\mathbf{T}}$ given by $\omega \to X(\cdot, \omega)$, then for $E \in \mathscr{C}$ we have $X^{-1}(E) = X^{-1}(\Pi_S^{-1}(B) = (\Pi_S X)^{-1}(B)$ for some $B \in \mathbf{R}^S$ ($S \subseteq \mathbf{T}$, finite), so that $\mu_X(E) = P((\Pi_S X)^{-1}(B)) = P(X^{-1}(E))$. This identity extends to $\sigma(\mathscr{C})$: since the probability measures μ_X and $P \circ X^{-1}$ agree on $\mathscr{C}$, they also agree on $\sigma(\mathscr{C})$.

Although the canonical representation of X on $(\mathbf{R}^{\mathbf{T}}, \sigma(\mathscr{C}), \mu_X)$ has the advantage that $\sigma(\mathscr{C})$ is defined without reference to the probability space $(\Omega, \mathscr{F}, P)$ on which X was originally defined, it also has serious limitations. First note that each set in $\sigma(\mathscr{C})$ is a *σ-cylinder*, i.e. has the form $\{\omega \in \mathbf{R}^{\mathbf{T}}: (\omega(t_1), \omega(t_2), \ldots) \in B\}$ where $T = (t_1, t_2, \ldots)$ is a sequence in $\mathbf{T}$ and $B \in \mathscr{B}(\mathbf{R})^T$: It is easy to see that $\mathscr{B}(\mathbf{R})^T = \mathscr{B}(\mathbf{R}^T)$ as T is countable, so that $\mathscr{B}(\mathbf{R})^T$ is the σ-field generated by all finite-dimensional rectangles in $\mathbf{R}^T$. It follows (see, e.g., [8; Prop. 12.8]) that the σ-cylinders form a σ-field, which must therefore be $\sigma(\mathscr{C})$.

But this means that the only sets we can guarantee to be measurable in our standard space $(\mathbf{R}^{\mathbf{T}}, \sigma(\mathscr{C}))$ are those which depend on countably many values of the process X, i.e.

$$\{\omega : (\omega(t_1), \ldots, \omega(t_n), \ldots) \in B\} = \{\omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega), \ldots) \in B\}$$

for some Borel set B in $\mathbf{R}^{\mathbf{T}}$. We can complete $\sigma(\mathscr{C})$ relative to μ_X, but this will still exclude sets such as

$$A = \{\omega : X_t(\omega) = 0 \text{ for some } t \in \mathbf{T}\} = \bigcup_{t \in \mathbf{T}} \{\omega : \omega(t) = 0\}$$

or

$$W = \{\omega : t \to X_t(\omega) \text{ is continuous on } \mathscr{T}\}.$$

Thus if we want to study *path properties* of X, the canonical representation is not very useful. We need to identify subsets of $\mathbf{R}^{\mathrm{T}}$ which 'carry' the measure μ_X when X satisfies path regularity properties. In the case we shall study in the next section, such properties derive in turn from the special nature of the finite-dimensional distributions of the process.

0.2. Brownian motion and the Itô integral

There are many excellent treatises on this most important stochastic process, and we shall not attempt to duplicate them. Conceived as a mathematical model for the highly irregular motions of particles in colloidal suspensions, which experience so many collisions that their motion appears to be quite random, Brownian motion has become the paradigm of a 'random process', and finds applications in fields as disparate as filtering theory in electronics and the fluctuations of the stock market. We shall not attempt to discuss how 'realistic' these models are, but discuss, with only sketched proofs, the construction and basic properties of this process. For fuller treatments we refer to [33], [40], [53], [57], [83].

0.2.1: A discrete model for random behaviour (e.g. coin-tossing) is given by symmetric simple *random walk* on the line: let $\Omega = [0, 1]$, $\mathscr{F} = \mathscr{B}([0,1])$ and let P be Lebesgue measure. Each $\omega \in \Omega$ is described by its binary expansion

$$\omega = \sum_{i=1}^{\infty} \frac{\varepsilon_i}{2^i},$$

where $\varepsilon_i = 0$ or 1. Define a sequence (R_i) on Ω by

$$R_i(\omega) = \begin{cases} 1 & \text{if } \varepsilon_i = 1 \\ -1 & \text{if } \varepsilon_i = 0 \end{cases}$$

and let $S_0 = 0$, $S_n(\omega) = \sum_{i=1}^{n} R_i(\omega)$. If ω has two binary expansions, set $R_i(\omega) = 0$ for all i. This only involves a P-null set of ω's. The functions (R_i) give a model for coin-tossing. Alternatively we can regard (S_n) as giving the position of a particle, starting at the origin, which at each time $t = 1,2,3,\ldots$ moves instantaneously to the right or left with probability $\frac{1}{2}$. The (R_i) are obviously independent and identically distributed, and have mean 0 and variance 1. The simplest form of the *Central Limit Theorem* (see e.g. [8; Th. 1.17]) then states that

$$\lim_n P\left(\frac{S_n}{\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \mathrm{d}y.$$