D

Some Mathematical Background

In this final section we collect together, for ease of reference, many of those mathematical results which we have used in developing our results on Markov chains and their applications: these come from probability and measure theory, topology, stochastic processes, the theory of probabilities on topological spaces, and even number theory.

We have tried to give results at a relevant level of generality for each of the types of use: for example, since we assume that the leap from countable to general spaces or topological spaces is one that this book should encourage, we have reviewed (even if briefly) the simple aspects of this theory; conversely, we assume that only a relatively sophisticated audience will wish to see details of sample path results, and the martingale background provided requires some such sophistication.

Readers who are unfamiliar with any particular concepts and who wish to delve further into them should consult the standard references cited, although in general a deep understanding of many of these results is not vital to follow the development in this book itself.

D.1 Some Measure Theory

We assume throughout this book that the reader has some familiarity with the elements of measure and probability theory. The following sketch of key concepts will serve only as a reminder of terms, and perhaps as an introduction to some non-elementary concepts; anyone who is unfamiliar with this section must take much in the general state space part of the book on trust, or delve into serious texts such as Billingsley [25], Chung [50] or Doob [68] for enlightenment.

D.1.1 Measurable spaces and σ-fields

A general *measurable space* is a pair $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ with

X: an abstract set of points;

$\mathcal{B}(\mathsf{X})$: a σ-field of subsets of X; that is,

(i) $\mathsf{X} \in \mathcal{B}(\mathsf{X})$;

(ii) if $A \in \mathcal{B}(\mathsf{X})$ then $A^c \in \mathcal{B}(\mathsf{X})$;

(iii) if $A_k \in \mathcal{B}(\mathsf{X})$, $k = 1, 2, 3, \dots$ then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{B}(\mathsf{X})$.

A σ-field $\mathcal{B}$ is *generated* by a collection of sets $\mathcal{A}$ in $\mathcal{B}$ if $\mathcal{B}$ is the smallest σ-field containing the sets $\mathcal{A}$, and then we write $\mathcal{B} = \sigma(\mathcal{A})$; a σ-field $\mathcal{B}$ is *countably generated* if it is generated by a countable collection $\mathcal{A}$ of sets in $\mathcal{B}$. The σ-fields $\mathcal{B}(\mathsf{X})$ we use are always assumed to be countably generated.

On the real line $\mathbb{R} := (-\infty, \infty)$ the Borel σ-field $\mathcal{B}(\mathbb{R})$ is generated by the countable collection of sets $\mathcal{A} = (a, b]$ where a, b range over the rationals $\mathbb{Q}$.

When our state space is $\mathbb{R}$ then we always assume it is equipped with the Borel σ-field.

If $(\mathsf{X}_1, \mathcal{B}(\mathsf{X}_1))$ is a measurable space and $(\mathsf{X}_2, \mathcal{B}(\mathsf{X}_2))$ is another measurable space, then a mapping $h\colon \mathsf{X}_1 \to \mathsf{X}_2$ is called a *measurable function* if

$$h^{-1}\{B\} := \{x : h(x) \in B\} \in \mathcal{B}(\mathsf{X}_1)$$

for all sets $B \in \mathcal{B}(\mathsf{X}_2)$.

As a convention, functions on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ which we use are always assumed to be measurable, and in general this is omitted from theorem statements and the like.

D.1.2 Measures

A (*signed*) *measure* μ on the space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ is a function from $\mathcal{B}(\mathsf{X})$ to $(-\infty, \infty]$ which is countably additive: if $A_k \in \mathcal{B}(\mathsf{X})$, $k = 1, 2, 3, \ldots$, and $A_i \cap A_j = \emptyset$, $i \neq j$ then

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

We say that μ is *positive* if $\mu(A) \geq 0$ for any A. The measure μ is called a *probability* (or *subprobability*) measure if it is positive and $\mu(\mathsf{X}) = 1$ (or $\mu(\mathsf{X}) < 1$).

A positive measure μ is σ-finite if there is a countable collection of sets $\{A_k\}$ such that $\mathsf{X} = \cup A_k$ and $\mu(A_k) < \infty$ for each k.

On the real line $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *Lebesgue measure* μ^{Leb} is a positive measure defined for intervals $(a, b]$ by $\mu^{\text{Leb}}(a, b] = b - a$, and for the other sets in $\mathcal{B}(\mathbb{R})$ by an obvious extension technique. Lebesgue measure on higher dimensional Euclidean space $\mathbb{R}^p$ is constructed similarly using the area of rectangles as a basic definition.

The *total variation norm* of a signed measure is defined as $\|\mu\| := \sup \int f\, d\mu$, where the supremum is taken over all measurable functions f from $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, such that $|f(x)| \leq 1$ for all $x \in \mathsf{X}$.

For a signed measure μ, the state space X may be written as the union of disjoint sets X_+ and X_- where

$$\mu(\mathsf{X}_+) - \mu(\mathsf{X}_-) = \|\mu\|.$$

This is known as the *Hahn decomposition.*

D.1.3 Integrals

Suppose that h is a non-negative measurable function from $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The *Lebesgue integral* of h with respect to a positive finite measure μ is defined in three steps.

Firstly, for $A \in \mathcal{B}(\mathsf{X})$ define $\mathbb{1}_A(x) = 1$ if $x \in A$, and 0 otherwise: $\mathbb{1}_A$ is called the *indicator function* of the set A. In this case we define

$$\int_{\mathsf{X}} \mathbb{1}_A(x)\mu(dx) := \mu(A).$$

Next consider *simple functions* h such that there exist sets $\{A_1, \dots A_N\} \subset \mathcal{B}(\mathsf{X})$ and positive numbers $\{b_1, \dots b_N\} \subset \mathbb{R}_+$ with $h = \sum_{k=1}^N b_k \mathbb{1}_{A_k}$.

If h is a simple function we can unambiguously define

$$\int_{\mathsf{X}} h(x)\mu(dx) := \sum_{k=1}^N b_k \mu\{A_k\}.$$

Finally, since it is possible to show that given any non-negative measurable h, there exists a sequence of simple functions $\{h_k\}_{k=1}^\infty$, such that for each $x \in \mathsf{X}$,

$$h_k(x) \uparrow h(x)$$

we can take

$$\int_{\mathsf{X}} h(x)\mu(dx) := \lim_k \int_{\mathsf{X}} h_k(x)\mu(dx)$$

which always exists, though it may be infinite.

This approach works if h is non-negative. If not, write

$$h = h^+ - h^-$$

where h^+ and h^- are both non-negative measurable functions, and define

$$\int_{\mathsf{X}} h(x)\mu(dx) := \int_{\mathsf{X}} h^+(x)\mu(dx) - \int_{\mathsf{X}} h^-(x)\mu(dx),$$

if both terms on the right are finite. Such functions are called *μ-integrable*, or just integrable if there is no possibility of confusion; and we frequently denote the integral by

$$\int h\, d\mu := \int_{\mathsf{X}} h(x)\mu(dx).$$

The extension to σ-finite measures is then straightforward.

Convergence of sequences of integrals is central to much of this book. There are three results which we use regularly:

Theorem D.1.1 (Monotone Convergence Theorem) *If μ is a σ-finite positive measure on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and $\{f_i : i \in \mathbb{Z}_+\}$ are measurable functions from $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which satisfy $0 \le f_i(x) \uparrow f(x)$ for μ-almost every $x \in \mathsf{X}$, then*

$$\int_{\mathsf{X}} f(x)\mu(dx) = \lim_i \int_{\mathsf{X}} f_i(x)\mu(dx). \tag{D.1}$$

Note that in this result the monotone limit f may not be finite even μ-almost everywhere, but the result continues to hold in the sense that both sides of (D.1) will be finite or infinite together.

Theorem D.1.2 (Fatou's Lemma) *If μ is a σ-finite positive measure on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and $\{f_i : i \in \mathbb{Z}_+\}$ are non-negative measurable functions from $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ then*

$$\int_{\mathsf{X}} \liminf_i f_i(x)\mu(dx) \le \liminf_i \int_{\mathsf{X}} f_i(x)\mu(dx). \tag{D.2}$$

Theorem D.1.3 (Dominated Convergence Theorem) *Suppose that* μ *is a* σ-*finite positive measure on* $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ *and* $g \geq 0$ *is a* μ-*integrable function from* $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ *to* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

If f *and* $\{f_i : i \in \mathbb{Z}_+\}$ *are measurable functions from* $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ *to* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *satisfying* $|f_i(x)| \leq g(x)$ *for* μ-*almost every* $x \in \mathsf{X}$, *and if* $f_i(x) \to f(x)$ *as* $i \to \infty$ *for* μ-*a.e.* $x \in \mathsf{X}$, *then each* f_i *is* μ-*integrable, and*

$$\int_{\mathsf{X}} f(x)\mu(dx) = \lim_i \int_{\mathsf{X}} f_i(x)\mu(dx)$$

D.2 Some Probability Theory

A general *probability space* is an ordered triple $(\Omega, \mathcal{F}, \mathsf{P})$ with Ω an abstract set of points, $\mathcal{F}$ a σ-field of subsets of Ω, and P a probability measure on $\mathcal{F}$.

If $(\Omega, \mathcal{F}, \mathsf{P})$ is a probability space and $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ is a measurable space, then a mapping $X: \Omega \to \mathsf{X}$ is called a *random variable* if

$$X^{-1}\{B\} := \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

for all sets $B \in \mathcal{B}(\mathsf{X})$: that is, if X is a measurable mapping from Ω to X.

Given a random variable X on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$, we define the σ-field *generated by* X, denoted $\sigma\{X\} \subseteq \mathcal{F}$, to be the smallest σ-field on which X is measurable.

If X is a random variable from a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ to a general measurable space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$, and h is a real valued measurable mapping from $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ to the real line $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ then the composite function $h(X)$ is a real-valued random variable on $(\Omega, \mathcal{F}, \mathsf{P})$: note that some authors reserve the term “random variable” for such real-valued mappings. For such functions, we define the *expectation* as

$$\mathsf{E}[h(X)] = \int_{\Omega} h(X(\omega))\mathsf{P}(dw)$$

The set of real-valued random variables Y for which the expectation is well-defined and finite is denoted $L^1(\Omega, \mathcal{F}, \mathsf{P})$. Similarly, we use $L^\infty(\Omega, \mathcal{F}, \mathsf{P})$ to denote the collection of essentially bounded real-valued random variables Y; that is, those for which there is a bound M and a set $A_M \subset \mathcal{F}$ with $\mathsf{P}(A_M) = 0$ such that $\{\omega : |Y(\omega)| > M\} \subseteq A_M$.

Suppose that $Y \in L^1(\Omega, \mathcal{F}, \mathsf{P})$ and $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-field of $\mathcal{F}$. If $\hat{Y} \in L^1(\Omega, \mathcal{G}, \mathsf{P})$ and satisfies

$$\mathsf{E}[YZ] = \mathsf{E}[\hat{Y}Z] \quad \text{for all } Z \in L^\infty(\Omega, \mathcal{G}, \mathsf{P})$$

then $\hat{Y}$ is called the *conditional expectation* of Y given $\mathcal{G}$, and denoted $\mathsf{E}[Y \mid \mathcal{G}]$. The conditional expectation defined in this way exists and is unique (modulo P-null sets) for any $Y \in L^1(\Omega, \mathcal{F}, \mathsf{P})$ and any sub σ-field $\mathcal{G}$.

Suppose now that we have another σ-field $\mathcal{D} \subset \mathcal{G} \subset \mathcal{F}$. Then

$$\mathsf{E}[Y \mid \mathcal{D}] = \mathsf{E}[\mathsf{E}[Y \mid \mathcal{G}] \mid \mathcal{D}]. \tag{D.3}$$

The identity (D.3) is often called “the smoothing property of conditional expectations”.

D.3 Some Topology

We summarize in this section several concepts needed for chains on topological spaces, and for the analysis of some of the applications on such spaces. The classical texts of Kelley [126] or Halmos [94] are excellent references for details at the level we require, as is the more introductory but very readable exposition of Simmons [240].

D.3.1 Topological spaces

On any abstract space X a topology $\mathcal{T} := \{\text{open subsets of } \mathsf{X}\}$ is a collection of sets containing

(i) arbitrary unions of members of $\mathcal{T}$,

(ii) finite intersections of members of $\mathcal{T}$,

(iii) the whole space X and the empty set $\emptyset$.

Those members of $\mathcal{T}$ containing a point x are called the *neighborhoods* of x, and the complements of open sets are called *closed*.

A set C is called *compact* if any cover of C with open sets admits a finite subcover, and a set D is *dense* if the smallest closed set containing D (the *closure* of D) is the whole space. A set is called *precompact* if it has a compact closure.

When there is a topology assumed on the state spaces for the Markov chains considered in this book, it is always assumed that these render the space locally compact and separable metric: a *locally compact* space is one for which each open neighborhood of a point contains a compact neighborhood, and a *separable* space is one for which a countable dense subset of X exists. A metric space is such that there is a metric d on X which generates its topology.

For the topological spaces we consider, Lindelöf's Theorem holds:

Theorem D.3.1 (Lindelöf's Theorem) *If X is a separable metric space, then every cover of an open set by open sets admits a countable subcover.*

If X is a topological space with topology $\mathcal{T}$, then there is a natural σ-field on X containing $\mathcal{T}$. This σ-field $\mathcal{B}(\mathsf{X})$ is defined as

$$\mathcal{B}(\mathsf{X}) := \bigcap \{G : \mathcal{T} \subset G,\ G \text{ a } \sigma\text{-field on } \mathsf{X}\}$$

so that $\mathcal{B}(\mathsf{X})$ is generated by the open subsets of X.

Extending the terminology from $\mathbb{R}$, this is often called the *Borel σ-field* of X: throughout this book, we have assumed that on a topological space the Borel σ-field is being addressed, and so our general notation $\mathcal{B}(\mathsf{X})$ is consistent in the topological context with the conventional notation.

A measure μ is called *regular* if for any set $E \in \mathcal{B}(\mathsf{X})$,

$$\mu(E) = \inf\{\mu(O) : E \subseteq O,\ O \text{ open}\} = \sup\{\mu(C) : C \subseteq E,\ C \text{ compact}\}$$

For the topological spaces we consider, measures on $\mathcal{B}(\mathsf{X})$ are regular: we have ([233] p. 49)

Theorem D.3.2 *If X is locally compact and separable, then every σ-finite measure on $\mathcal{B}(\mathsf{X})$ is regular.*

D.4 Some Real Analysis

A function $f : \mathsf{X} \to \mathbb{R}$ on a space X with a metric d is called *continuous* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that for any two $x, y \in \mathsf{X}$, if $d(x,y) < \delta$ then $|f(x) - f(y)| < \varepsilon$. The set of all bounded continuous functions on the locally compact and separable metric space X forms a metric space denoted $\mathcal{C}(\mathsf{X})$, whose metric is generated by the supremum norm

$$|f|_c := \sup_{x \in \mathsf{X}} |f(x)|.$$

A function $f : \mathsf{X} \to \mathbb{R}$ is called *lower semicontinuous* if the sublevel set $\{x : f(x) \le c\}$ is closed for any constant c, and *upper semicontinuous* if $\{x : f(x) < c\}$ is open for any constant c.

Theorem D.4.1 *A real-valued function f on X is continuous if and only if it is simultaneously upper semicontinuous and lower semicontinuous.*

If the function f is positive, then it is lower semicontinuous if and only if there exists a sequence of continuous bounded positive functions $\{f_n : n \in \mathbb{Z}_+\} \subset \mathcal{C}(\mathsf{X})$, each with compact support, such that for all $x \in \mathsf{X}$,

$$f_n(x) \uparrow f(x) \qquad \text{as } n \to \infty.$$

A sequence of functions $\{f_i : i \in \mathbb{Z}_+\} \subset \mathcal{C}(\mathsf{X})$ is called *equicontinuous* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that for any two $x, y \in \mathsf{X}$, if $d(x,y) < \delta$ then $|f_i(x) - f_i(y)| < \varepsilon$ for all i.

Theorem D.4.2 (Ascoli's Theorem) *Suppose that the topological space X is compact. A collection of functions $\{f_i : i \in \mathbb{Z}_+\} \subset \mathcal{C}(\mathsf{X})$ is precompact as a subset of $\mathcal{C}(\mathsf{X})$ if and only if the following two conditions are satisfied:*

(i) *The sequence is uniformly bounded: i.e. for some $M < \infty$, and all $i \in \mathbb{Z}_+$,*

$$|f_i|_c = \sup_{x \in \mathsf{X}} |f_i(x)| \le M.$$

(ii) *The sequence is equicontinuous.*

Finally, in our context one of the most frequently used of all results on continuous functions is that which assures us that the convolution operation applied to any pair of $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu^{\text{Leb}})$ and $L^\infty(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu^{\text{Leb}})$ functions is continuous.

For two functions $f, g : \mathbb{R} \to \mathbb{R}$, the *convolution* $f * g$ is the function on $\mathbb{R}$ defined for $t \in \mathbb{R}$ by

$$f * g\,(t) = \int_{-\infty}^{\infty} f(s) g(t-s)\, ds.$$

This is well defined if, for example, both f and g are positive. We have (see [233], p. 196)

Theorem D.4.3 *Suppose that f and g are measurable functions on $\mathbb{R}$, that f is bounded, and that $\int |g|\, dx < \infty$. Then the convolution $f * g$ is a bounded continuous function.*

D.5 Some Convergence Concepts for Measures

In this section we summarize various forms of convergence of probability measures which are used throughout the book. For further information the reader is referred to Parthasarathy [213] and Billingsley [24].

Assume X to be a locally compact and separable metric space. Letting $\mathcal{M}$ denote the set of probability measures on $\mathcal{B}(\mathsf{X})$, we can construct a number of natural topologies on $\mathcal{M}$.

As is obvious in Part III of this book, we are frequently concerned with the very strong topology of convergence in total variation norm. However, for individual sequences of measures, the topologies of weak or vague convergence prove more natural in many respects.

D.5.1 Weak Convergence

In the topology of *weak convergence* a sequence $\{\nu_k : k \in \mathbb{Z}_+\}$ of elements of $\mathcal{M}$ converges to ν if and only if

$$\lim_{k\to\infty} \int f\, d\nu_k = \int f\, d\nu \tag{D.4}$$

for every $f \in \mathcal{C}(\mathsf{X})$.

In this case we say that $\{\nu_k\}$ converges *weakly* to ν as $k \to \infty$, and this will be denoted $\nu_k \xrightarrow{\text{w}} \nu$.

The following key result is given as Theorem 6.6 in [213]:

Proposition D.5.1 *There exists a sequence of uniformly continuous, uniformly bounded functions $\{g_n : n \in \mathbb{Z}_+\} \subset \mathcal{C}(\mathsf{X})$ with the property that*

$$\mu_k \xrightarrow{\text{w}} \mu_\infty \quad \Longleftrightarrow \quad \forall n \in \mathbb{Z}_+,\ \lim_{k\to\infty} \int g_n\, d\mu_k = \int g_n\, d\mu_\infty. \tag{D.5}$$

It follows that $\mathcal{M}$ can be considered as a metric space with metric $|\cdot|_w$ defined for $\nu,\, \mu \in \mathcal{M}$ by

$$|\nu - \mu|_w := \sum_{k=1}^{\infty} 2^{-k} \Big| \int g_k\, d\nu - \int g_k\, d\mu \Big|$$

Other metrics relevant to weak convergence are summarized in, for example, [119].

A set of probability measures $\mathcal{A} \subset \mathcal{M}$ is called *tight* if for every $\varepsilon \geq 0$ there exists a compact set $C \subset \mathsf{X}$ for which

$$\nu\{C\} \geq 1 - \varepsilon \qquad \text{for every } \nu \in \mathcal{A}.$$

The following result, which characterizes tightness with $\mathcal{M}$ viewed as a metric space, follows from Proposition D.5.6 below.

Proposition D.5.2 *The set of probabilities $A \subset \mathcal{M}$ is precompact if and only if it is tight.*

A function $V\colon \mathsf{X} \to \mathbb{R}_+$ is called *norm-like* if there exists a sequence of compact sets, $C_n \subset \mathsf{X}$, $C_n \uparrow \mathsf{X}$ such that

$$\lim_{n\to\infty}\left(\inf_{x\in C_n^c} V(x)\right) = \infty$$

where we adopt the convention that the infimum of a function over the empty set is infinity. If X is a closed and unbounded subset of $\mathbb{R}^k$ it is evident that $V(x) = |x|^p$ is norm-like for any $p > 0$. If X is *compact* then our convention implies that any positive function V is norm-like because we may set $C_n = \mathsf{X}$ for all $n \in \mathbb{Z}_+$.

It is easily verified that a collection of probabilities $\mathcal{A} \subset \mathcal{M}$ is tight if and only if a norm-like function V exists such that

$$\sup_{\nu\in\mathcal{A}} \int V\, d\nu < \infty.$$

The following simple lemma will often be needed.

Lemma D.5.3 **(i)** *A sequence of probabilities $\{\nu_k : k \in \mathbb{Z}_+\}$ is tight if and only if there exists a norm-like function V such that*

$$\limsup_{k\to\infty} \nu_k(V) < \infty.$$

(ii) *If for each $x \in \mathsf{X}$ there exists a norm-like function $V_x(\,\cdot\,)$ on X such that*

$$\limsup_{k\to\infty} \mathsf{E}_x[V_x(\Phi_k)] < \infty,$$

then the chain is bounded in probability. □

The next result can be found in [24] and [213].

Theorem D.5.4 *The following are equivalent for a sequence of probabilities $\{\nu_k : k \in \mathbb{Z}_+\} \subset \mathcal{M}$*

(i) $\nu_k \xrightarrow{\mathrm{w}} \nu$

(ii) *for all open sets $O \subset \mathsf{X}$, $\liminf_{k\to\infty} \nu_k\{O\} \ge \nu\{O\}$*

(iii) *for all closed sets $C \subset \mathsf{X}$, $\limsup_{k\to\infty} \nu_k\{C\} \le \nu\{C\}$*

(iv) *for every uniformly bounded and equicontinuous family of functions $\mathcal{C} \subset \mathcal{C}(\mathsf{X})$,*

$$\lim_{k\to\infty} \sup_{f\in\mathcal{C}} \left|\int f d\nu_k - \int f d\nu\right| = 0.$$

□

D.5.2 Vague Convergence

Vague convergence is less stringent than weak convergence. Let $\mathcal{C}_0(\mathsf{X}) \subset \mathcal{C}(\mathsf{X})$ denote the set of continuous functions on X which converge to zero on the "boundary" of X: that is, $f \in \mathcal{C}_0(\mathsf{X})$ if for some (and hence any) sequence $\{C_k : k \in \mathbb{Z}_+\}$ of compact sets which satisfy

$$C_k \subset C_{k+1}, \qquad \text{and} \quad \bigcup_{k=0}^{\infty} C_k = \mathsf{X},$$

we have

$$\lim_{k\to\infty} \sup_{x\in C_k^c} |f(x)| = 0.$$

The space $\mathcal{C}_0(\mathsf{X})$ is simply the closure of $\mathcal{C}_c(\mathsf{X})$, the space of continuous functions with compact support, in the uniform norm.

A sequence of subprobability measures $\{\nu_k : k \in \mathbb{Z}_+\}$ is said to converge *vaguely* to a subprobability measure ν if for all $f \in \mathcal{C}_0(\mathsf{X})$

$$\lim_{k\to\infty} \int f\, d\nu_k = \int f\, d\nu,$$

and in this case we will write

$$\nu_k \xrightarrow{\mathrm{v}} \nu \qquad \text{as } k \to \infty.$$

In this book we often apply the following result, which follows from the observation that positive lower semicontinuous functions on X are the pointwise supremum of a collection of positive, continuous functions with compact support (see Theorem D.4.1).

Lemma D.5.5 *If* $\nu_k \xrightarrow{\mathrm{v}} \nu$ *then*

$$\liminf_{k\to\infty} \int f\, d\nu_k \geq \int f\, d\nu \tag{D.6}$$

for any positive lower semicontinuous function f *on* X.

It is obvious that weak convergence implies vague convergence. On the other hand, a sequence of probabilities converges weakly if and only if it converges vaguely and is tight.

The use and direct verification of boundedness in probability will often follow from the following results: the first of these is a consequence of our assumption that the state space is locally compact and separable (see Billingsley [24] and Revuz [223]).

Proposition D.5.6 **(i)** *For any sequence of subprobabilities* $\{\nu_k : k \in \mathbb{Z}_+\}$ *there exists a subsequence* $\{n_k\}$ *and a subprobability* ν_∞ *such that*

$$\nu_{n_k} \xrightarrow{\mathrm{v}} \nu_\infty, \qquad k \to \infty.$$

(ii) *If* $\{\nu_k\}$ *is tight and each* ν_k *is a probability measure, then* $\nu_{n_k} \xrightarrow{\mathrm{w}} \nu_\infty$ *and* ν_∞ *is a probability measure.*

□

D.6 Some Martingale Theory

D.6.1 The Martingale Convergence Theorem

A sequence of integrable random variables $\{M_n : n \in \mathbb{Z}_+\}$ is called *adapted* to an increasing family of σ-fields $\{\mathcal{F}_n : n \in \mathbb{Z}_+\}$ if M_n is $\mathcal{F}_n$-measurable for each n. The sequence is called a *martingale* if $\mathsf{E}[M_{n+1} \mid \mathcal{F}_n] = M_n$ for all $n \in \mathbb{Z}_+$, and a *supermartingale* if $\mathsf{E}[M_{n+1} \mid \mathcal{F}_n] \leq M_n$ for $n \in \mathbb{Z}_+$.

A *martingale difference sequence* $\{Z_n : n \in \mathbb{Z}_+\}$ is an adapted sequence of random variables such that the sequence $M_n = \sum_{k=0}^{n} Z_k$ is a martingale.

The following result is basic:

Theorem D.6.1 (The Martingale Convergence Theorem) *Let M_n be a supermartingale, and suppose that*

$$\sup_n \mathsf{E}[|M_n|] < \infty.$$

Then $\{M_n\}$ converges to a finite limit with probability one.

If $\{M_n\}$ is a positive, real valued supermartingale then by the smoothing property of conditional expectations (D.3),

$$\mathsf{E}[|M_n|] = \mathsf{E}[M_n] \leq \mathsf{E}[M_0] < \infty, \qquad n \in \mathbb{Z}_+$$

Hence we have as a direct corollary to the Martingale Convergence Theorem

Theorem D.6.2 *A positive supermartingale converges to a finite limit with probability one.*

Since a positive supermartingale is convergent, it follows that its sample paths are bounded with probability one. The following result gives an upper bound on the magnitude of variation of the sample paths of both positive supermartingales, and general martingales.

Theorem D.6.3 (Kolmogorov's Inequality) **(i)** *If M_n is a martingale then for each $c > 0$ and $p \geq 1$,*

$$\mathsf{P}\{\max_{0\leq k\leq n} |M_k| \geq c\} \leq \frac{1}{c^p}\mathsf{E}[|M_n|^p]$$

(ii) *If M_n is a positive supermartingale then for each $c > 0$*

$$\mathsf{P}\{\sup_{0\leq k\leq \infty} M_k \geq c\} \leq \frac{1}{c}\mathsf{E}[M_0]$$

These results, and related concepts, can be found in Billingsley [25], Chung [50], Hall and Heyde [93], and of course Doob [68].

D.6.2 The functional CLT for martingales

Consider a general martingale $(M_n, \mathcal{F}_n)$. Our purpose is to analyze the following sequence of continuous functions on $[0,1]$:

$$m_n(t) := M_{\lfloor nt\rfloor} + (nt - \lfloor nt\rfloor)\Big[M_{\lfloor nt\rfloor+1} - M_{\lfloor nt\rfloor}\Big], \qquad 0 \leq t \leq 1. \tag{D.7}$$

The function $m_n(t)$ is piecewise linear, and is equal to M_i when $t = i/n$ for $0 \leq t \leq 1$. In Theorem D.6.4 below we give conditions under which the normalized sequence $\{n^{-1/2}m_n(t) : n \in \mathbb{Z}_+\}$ converges to a continuous process (Brownian motion) on $[0,1]$. This result requires some care in the definition of convergence for a sequence of stochastic processes.

Let $\mathcal{C}[0,1]$ denote the normed space of all continuous functions $\phi\colon [0,1] \to \mathbb{R}$ under the uniform norm, which is defined as

$$|\phi|_c = \sup_{0\leq t\leq 1} |\phi(t)|.$$

The vector space $\mathcal{C}[0,1]$ is a complete, separable metric space, and hence the theory of weak convergence may be applied to analyze measures on $\mathcal{C}[0,1]$.

The stochastic process $m_n(t)$ possesses a distribution μ_n, which is a probability measure on $\mathcal{C}[0,1]$. We say that $m_n(t)$ *converges in distribution* to a stochastic process $m_\infty(t)$ as $n \to \infty$, which is denoted $m_n \xrightarrow{\mathrm{d}} m_\infty$, if the sequence of measures μ_n converge weakly to the distribution μ_∞ of m_∞. That is, for any bounded continuous functional h on $\mathcal{C}[0,1]$,

$$\mathsf{E}[h(m_n)] \to \mathsf{E}[h(m_\infty)] \qquad \text{as } n \to \infty.$$

The limiting process, standard Brownian motion on $[0,1]$, which we denote by B, is defined as follows:

Standard Brownian Motion

Brownian motion $B(t)$ is a real-valued stochastic process on $[0,1]$ with $B(0) = 0$, satisfying

(i) The sample paths of B are continuous with probability one;

(ii) The increment $B(t) - B(s)$ is independent of $\{B(r) : r \le s\}$ for each $0 \le s \le t \le 1$;

(iii) The distribution of $B(t) - B(s)$ is Gaussian $N(0, |t-s|)$.

To prove convergence we use the following key result which is a consequence of Theorem 4.1 of [93].

Theorem D.6.4 *Let $(M_n, \mathcal{F}_n)$ be a square integrable martingale, so that for all $n \in \mathbb{Z}_+$*

$$\mathsf{E}[M_n^2] = \mathsf{E}[M_0^2] + \sum_{k=1}^{n} \mathsf{E}[(M_k - M_{k-1})^2] < \infty,$$

and suppose that the following conditions hold:

(i) *For some constant $0 < \gamma^2 < \infty$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1}] = \gamma^2 \qquad \text{a.s.} \tag{D.8}$$

(ii) *For all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}[(M_k - M_{k-1})^2 \mathbb{1}\{(M_k - M_{k-1})^2 \ge \varepsilon n\} | \mathcal{F}_{k-1}] = 0 \qquad \text{a.s.} \tag{D.9}$$

Then $(\gamma^2 n)^{-1/2} m_n \xrightarrow{\mathrm{d}} B$. □

Function space limits of this kind are often called *invariance principles*, though we have avoided this term because *functional CLT* seems more descriptive.

D.7 Some Results on Sequences and Numbers

We conclude with some useful lemmas on sequences and convolutions. The first gives an interaction between convolutions and limits. Recall that for two series a, b on $\mathbb{Z}_+$, the convolution is defined as

$$a * b\,(n) := \sum_{j=0}^{n} a(j) b(n-j)$$

Lemma D.7.1 *If* $\{a(n)\}, \{b(n)\}$ *are non-negative sequences such that* $b(n) \to b(\infty) < \infty$ *as* $n \to \infty$, *and* $\sum a(j) < \infty$, *then*

$$a * b\,(n) \to b(\infty) \sum_{j=0}^{\infty} a(j) < \infty, \qquad n \to \infty. \tag{D.10}$$

Proof Set $b(n) = 0$ for $n < 0$. Since $b(n)$ converges it is bounded, and so by the Dominated Convergence Theorem

$$\lim_{n\to\infty} a * b\,(n) = \sum_{j=0}^{\infty} a(j) \lim_{n\to\infty} b(n-j) = b(\infty) \sum_{j=0}^{\infty} a(j) \tag{D.11}$$

as required. □

The next lemma contains two valuable summation results for series.

Lemma D.7.2 **(i)** *If* $c(n)$ *is a non-negative sequence then for any* $r > 1$,

$$\sum_{n\ge 0} \Big[\sum_{m \ge n} c(m) \Big] r^n \le \frac{r}{r-1} \sum_{m\ge 0} c(m) r^m$$

and hence the two series

$$\sum_{n\ge 0} c(n) r^n, \qquad \sum_{n\ge 0} \Big[\sum_{m\ge n} c(m) \Big] r^n$$

converge or diverge together.

(ii) *If* a, b *are two non-negative sequences and* $r \ge 0$ *then*

$$\sum a * b\,(n) r^n = \Big[\sum a(n) r^n \Big] \Big[\sum b(n) r^n \Big].$$

PROOF By Fubini's Theorem we have

$$\begin{aligned}\sum_{n\ge 0}[\sum_{m\ge n} c(m)]r^n &= \sum_{m\ge 0} c(m) \sum_{n\le m} r^n \\ &= \sum_{m\ge 0} c(m)[r^{m+1}-1]/[r-1]\end{aligned}$$

which gives the first result. Similarly, we have

$$\begin{aligned}\sum_{n\ge 0} a*b\,(n)r^n &= \sum_{n\ge 0}[\sum_{m\le n} a(m)b(n-m)]r^n \\ &= \sum_{m\ge 0} a(m)r^m \sum_{n\ge m} b(n-m)r^{n-m} \\ &= \sum_{m\ge 0} a(m)r^m \sum_{n\ge 0} b(n)r^n\end{aligned}$$

which gives the second result. □

An elementary result on the greatest common divisor is useful for periodic chains.

Lemma D.7.3 *Let d denote the greatest common divisor (g.c.d) of the numbers m, n. Then there exist integers a, b such that*

$$am + bn = d$$

For a proof, see the corollary to Lemma 1.31 in Herstein [97].

Finally, in analyzing the periodic behavior of Markov chains, the following lemma is invaluable on very many occasions in ensuring positivity of transition probabilities:

Lemma D.7.4 *Suppose that $\mathcal{N} \subset \mathbb{Z}_+$ is a subset of the integers which is closed under addition: for each $j, k \in \mathcal{N}$, $j + k \in \mathcal{N}$. Let d denote the greatest common divisor of the set $\mathcal{N}$. Then there exists $n_0 < \infty$ such that $nd \in \mathcal{N}$ for all $n \ge n_0$.*

For a proof, see p. 569 of Billingsley [25].