Appendix

We collect here some definitions and results (largely without proofs) that are usually included in first year graduate analysis and probability courses. Proofs that are omitted can be found in standard texts for such courses, such as [**9**], [**11**], [**18**], [**23**], [**26**], and [**41**]. Precise locations of these proofs are given following the statements of the results.

A.1. Commonly used notation

Here are some notational conventions that we will adopt:

(a) The complement and closure of a set A are denoted by A^c and $\overline{A}$ respectively.

(b) If A is an event, 1_A will denote the indicator random variable that takes the value 1 on A and 0 on A^c.

(c) If $\mathcal{F}$ is a σ-algebra of subsets of Ω, $\xi \in \mathcal{F}$ means that ξ is measurable on $(\Omega, \mathcal{F})$.

(d) If ξ_k are random variables, then $\sigma(\xi_1, \xi_2, \dots)$ is the smallest σ-algebra with respect to which they are all measurable.

(e) The expected value of a random variable X is denoted by EX.

(f) If ξ is a random variable and A is an event, then $E(\xi, A)$ means $E(\xi 1_A)$.

(g) The maximum and minimum of two numbers are denoted by $\vee$ and $\wedge$, respectively.

(h) The rationals, the d-dimensional integer lattice, and n-dimensional Euclidean space are denoted by Q, Z^d, and R^n, respectively. The nonnegative rationals are denoted by Q^+.

(i) The central limit theorem is often abbreviated CLT.

(j) The restriction of the function f to the set A is written $f|_A$.

(k) Weak convergence of probability measures is denoted by $\Rightarrow$.

(l) The product measure on $\{0,1\}^S$ with density ρ, where S is countable, is denoted by ν_ρ.

(m) An equality preceded by a colon (:=) indicates that the right side defines the left side.

A.2. Some measure theory

The main result that is used to construct probability spaces that support random variables with prescribed finite-dimensional distributions is due to Kolmogorov. Suppose that we would like the random variables $X(t)$ indexed by $t \in [0, \infty)$ to have certain finite-dimensional distributions. These of course must be consistent, in the sense that if $\mu_{t_1,\dots,t_n}$ is the probability measure on R^n that is the desired distribution of $(X(t_1), \dots, X(t_n))$, then

$$\mu_{t_1,\dots,t_n}(A \times R^1) = \mu_{t_1,\dots,t_{n-1}}(A)$$

for all Borel sets $A \subset R^{n-1}$. Let $\Omega = R^{[0,\infty)}$, endowed with the smallest σ-algebra such that the projection $\omega \to \omega(t)$ from Ω to R^1 is measurable for each t.

Theorem A.1 (Kolmogorov). *Given a consistent family of probability measures $\mu_{t_1,\dots,t_n}$, there exists a unique probability measure μ on Ω so that the induced measure generated by the projection $\omega \to (\omega(t_1), \dots, \omega(t_n))$ from Ω to R^n is $\mu_{t_1,\dots,t_n}$ for each choice of $n \geq 1$ and $t_1, \dots, t_n$.* [**18**, page 471]

Once we have such a μ, we can define random variables $X(t)$ on Ω by $X(t, \omega) = \omega(t)$. They have the desired joint distributions. Note that every event in Ω is of the form

$$\{\omega : (\omega(t_1), \omega(t_2), \dots) \in A\}$$

for some $A \subset R^\infty$ and some countable collection of t_i's, since such sets form a σ-algebra with respect to which the projections $\omega \to \omega(t)$ are measurable. This means that Kolmogorov's theorem for an uncountable collection of random variables is no deeper than it is for a countable collection of random variables.

Quite often, one needs to verify that probabilities or expectations satisfy certain identities that are easy to check for special events or random variables. In order to extend these identities to more general events or random variables, one often uses the following results. The relevant definitions are:

Definition A.2. Suppose $(\Omega, \mathcal{F})$ is a measurable space.

(a) A collection of events $\mathcal{P} \subset \mathcal{F}$ is a π-system if

$$A, B \in \mathcal{P} \quad \text{implies} \quad A \cap B \in \mathcal{P}.$$

(b) A collection of events $\mathcal{L} \subset \mathcal{F}$ is a λ-system if it satisfies the following three properties:

(i) $\Omega \in \mathcal{L}$,

(ii) $A, B \in \mathcal{L}$ and $A \subset B$ implies that $B \backslash A \in \mathcal{L}$,

and

(iii) $A_n \in \mathcal{L}$ and $A_n \uparrow A$ implies that $A \in \mathcal{L}$.

Theorem A.3 ($\pi - \lambda$ theorem)**.** *Suppose $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system satisfying $\mathcal{P} \subset \mathcal{L}$. Then $\mathcal{L}$ contains the σ-algebra $\sigma(\mathcal{P})$ generated by $\mathcal{P}$.* [**18**, page 444]

Theorem A.4 (Monotone class theorem)**.** *Suppose that $\mathcal{P}$ is a π-system that contains Ω, and that $\mathcal{H}$ is a vector space of random variables satisfying the following properties:*

(i) $\quad$ *$A \in \mathcal{P}$ implies $1_A \in \mathcal{H}$.*

(ii) $\quad$ *$X_n \in \mathcal{H}, X$ bounded, and $X_n \uparrow X$ implies $X \in \mathcal{H}$.*

Then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{P})$-measurable random variables. [**18**, page 277]

A.3. Some analysis

A metric space (S, ρ) is said to be separable if it contains a countable dense set. It is complete if every Cauchy sequence has a limit. It is locally compact if for every $x \in S$ there is an $\epsilon > 0$ so that the closure of the open ball

$$\{y \in S : \rho(x, y) < \epsilon\}$$

is compact. Typical examples with all three properties are the Euclidean spaces R^n. A function h on a locally compact metric space is said to vanish at infinity, written

$$\lim_{x \to \infty} h(x) = 0,$$

if for every $\epsilon > 0$ there is a compact set $K \subset S$ so that $|h(x)| < \epsilon$ for all $x \notin K$. Every continuous function that vanishes at infinity is uniformly continuous.

A set $A \subset S$ is said to be nowhere dense if $(\overline{A})^c$ is dense.

Theorem A.5 (Baire category theorem). *A complete metric space is not the union of countably many nowhere dense sets.* [**41**, page 139]

If S is a linear space, the metric is often given in terms of a norm: $\rho(x, y) = ||x - y||$, and in this case, the space is said to be normed. A Banach space is a complete normed linear space. Its dual S^* is the Banach space of all bounded linear functions on S. The topology on S determined by the norm is called the strong topology. The weak topology is the weakest topology that makes all elements of S^* continuous. These topologies are generally not the same. However, the following statement is true, and its corollary is used in Chapter 3. The proof of the theorem is an application of the Hahn-Banach theorem.

Theorem A.6. *A linear subset of a Banach space is weakly closed if and only if it is strongly closed.* [**41**, page 201]

Corollary A.7. *If L is a linear subset of a Banach space and $\overline{L}_s$ and $\overline{L}_w$ are the strong and weak closures of L respectively, then $\overline{L}_s = \overline{L}_w$.*

Proof. Since the weak topology is weaker than the strong topology (i.e., has fewer closed sets), $\overline{L}_s \subset \overline{L}_w$. By Theorem A.6, $\overline{L}_s$ is weakly closed, so $\overline{L}_w \subset \overline{L}_s$. □

In Chaper 3, the calculus for Banach space-valued functions plays an important role. There are analogues of both the Riemann and Lebesgue integral for Banach space-valued functions. The analogue of the Lebesgue integral is known as the Bochner integral. For the purposes of this book, the analogue of the Riemann integral for continuous functions suffices. For a treatment of this integral, see Chapter 5 of [**29**]. Here are the basic facts. The proofs are essentially the same as the ones for real-valued functions.

For a continuous function $h : [a, b] \to S$, where S is a Banach space, define the modulus of continuity by

$$\delta_h(\epsilon) = \sup_{|s-t|\leq\epsilon} ||h(s) - h(t)||.$$

The function h is differentiable at $t \in (a, b)$ if

$$h'(t) := \lim_{s\to t} \frac{h(s) - h(t)}{s - t}$$

exists strongly. It is continuously differentiable if the derivative is continuous.

For a partition $\pi = \{a = t_0 < t_1 < \cdots < t_n = b\}$ of $[a, b]$ and intermediate points $s_i \in [t_i, t_{i+1}]$, define the Riemann sum by

$$\mathcal{R}(h, \pi) = \sum_{i=0}^{n-1} h(s_i)(t_{i+1} - t_i).$$

The mesh of π is defined by $|\pi| = \max_i(t_{i+1} - t_i)$.

Theorem A.8. *Suppose $h : [a, b] \to S$ is continuous. Then the following statements hold:*

(a) h is uniformly continuous, i.e.,

$$\lim_{\epsilon \to 0} \delta_h(\epsilon) = 0.$$

(b) Given two partitions π_1 and π_2 and corresponding sets of intermediate points, if π_2 is a refinement of π_1, then

$$|\mathcal{R}(h, \pi_1) - \mathcal{R}(h, \pi_2)| \leq \delta_h(|\pi_1|)(b - a).$$

(c) The following limit exists:

$$\int_a^b h(t)\, dt := \lim_{|\pi| \to 0} \mathcal{R}(h, \pi).$$

(d) $|| \int_a^b h(t)\, dt|| \leq \int_a^b ||h(t)||\, dt$.

(e) The function

$$t \to \int_a^t h(s)\, ds$$

is continuously differentiable, and

$$\frac{d}{dt} \int_a^t h(s)\, ds = h(t).$$

Theorem A.9. *If $h : [a, b] \to V$ is continuously differentiable, then*

$$\int_a^b h'(t)\, dt = h(b) - h(a).$$

Improper integrals are defined in the natural way:

$$\int_a^\infty h(t)\, dt = \lim_{b \to \infty} \int_a^b h(t)\, dt,$$

provided the limit exists strongly. A sufficient condition for this is that h be continuous on $[a, \infty)$ and

$$\int_0^\infty ||h(t)||\, dt < \infty.$$

A.4. The Poisson distribution

A nonnegative integer-valued random variable ξ is said to be Poisson distributed with parameter $\lambda > 0$ if

$$P(\xi = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

The Poisson distribution has a number of very special properties. Here are some of the most important:

(a) If $\xi_1, \ldots, \xi_m$ are independent Poisson distributed random variables with parameters $\lambda_1, \ldots, \lambda_m$, and $m_0 = 0 < m_1 < \cdots < m_l = m$, then

$$\sum_{i=1}^{m_1} \xi_i, \ldots, \sum_{i=m_{l-1}+1}^{m_l} \xi_i$$

are independent Poisson distributed random variables.

(b) If $\xi_1, \ldots, \xi_m$ are independent Poisson distributed random variables with parameters $\lambda_1, \ldots, \lambda_m$ and $\xi = \sum_i \xi_i$, then conditionally on $\xi = k$, $\xi_1, \ldots, \xi_m$ have a multinomial distribution with parameters k and $p_1, \ldots, p_m$, where $p_i = \lambda_i / \sum_j \lambda_j$.

(c) If ξ is Poisson distributed with parameter λ and conditionally on $\xi = k$, $\xi_1, \ldots, \xi_m$ have a multinomial distribution with parameters k and $p_1, \ldots, p_m$, then unconditionally, $\xi_1, \ldots, \xi_m$ are independent Poisson distributed random variables with parameters $p_1\lambda, \ldots, p_m\lambda$. They are called thinnings of ξ.

A.5. Random series and laws of large numbers

An important tool in analysis of sums of independent random variables is the Borel-Cantelli lemma. Given a sequence of events A_n, define

$$\{A_n \ i.o.\} = \bigcap_n \bigcup_{k \geq n} A_k.$$

Lemma A.10 (Borel-Cantelli). *Suppose A_n is a sequence of events.*

(a) If $\sum_n P(A_n) < \infty$, then $P(A_n \ i.o.) = 0$. [**18**, page 46]

(b) If the events are pairwise independent and $\sum_n P(A_n) = \infty$, then $P(A_n \ i.o.) = 1$. [**18**, page 50]

This is but one example of a result that says that the probability of a certain type of event must be zero or one. Here are two others.

Definition A.11. If $\xi_1, \xi_2, \ldots$ are random variables, the tail σ-algebra is

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(\xi_k, k \geq n).$$

Theorem A.12 (Kolmogorov $0-1$ law). *If $\xi_1, \xi_2, \ldots$ are independent random variables, then $P(A) = 0$ or 1 for every $A \in \mathcal{T}$.* [**18**, page 61]

Definition A.13. If $\xi_1, \xi_2, \ldots$ are random variables, the exchangeable σ-algebra $\mathcal{E}$ consists of those events defined in terms of this sequence that are not changed when finitely many of the ξ_i's are permuted.

Theorem A.14 (Hewitt-Savage $0-1$ law). *If $\xi_1, \xi_2, \ldots$ are independent and identically distributed random variables, then $P(A) = 0$ or 1 for every $A \in \mathcal{E}$.* [**18**, page 172]

One consequence of Theorem A.12 is that a series of independent random variables either converges a.s. or diverges a.s. The following theorem is useful in showing that such a series converges a.s.

Theorem A.15 (Lévy). *If $\xi_1, \xi_2, \ldots$ are independent random variables, then $\sum_k \xi_k$ converges a.s. if and only if it converges in probability.* [**23**, page 201]

Here is an example of its usefulness. Convergence in L_2 is easy to check directly, and implies convergence in probability; a.s. convergence is harder to check.

Corollary A.16. *If $\xi_1, \xi_2, \ldots$ are independent random variables with mean 0 and $\sum_n E\xi_n^2 < \infty$, then $\sum_n \xi_n$ converges a.s.*

Laws of large numbers are often obtained as consequences of convergence of random series. Here is the main statement.

Theorem A.17 (Strong law of large numbers). *Suppose $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with a finite absolute first moment. Let $S_n = \xi_1 + \cdots + \xi_n$ be their partial sums. Then*

$$\lim_{n \to \infty} \frac{S_n}{n} = E\xi_1 \ a.s.$$

[**18**, page 64]

A.6. The central limit theorem and related topics

Many theorems in probability theory are distributional, rather than pointwise. The central limit theorem is, of course, the most important of these.

A.6.1. Weak convergence.

Weak convergence is an important tool in proving distributional limit theorems. The most natural setting for this theory is a metric space. An excellent reference for this material is [**3**].

Definition A.18. Suppose S is a metric space.

(a) If μ_k and μ are probability measures on the Borel sets of S, then μ_k is said to converge weakly to μ (written $\mu_k \Rightarrow \mu$) if

$$\lim_{k\to\infty} \int f d\mu_k = \int f d\mu$$

for all bounded continuous real-valued functions f on S.

(b) If $X_1, X_2, \ldots$ and X are S-valued random variables, then X_k converges weakly, or in distribution, to X (written $X_k \Rightarrow X$) if their distributions converge weakly, which means that

$$\lim_{k\to\infty} Ef(X_k) = Ef(X)$$

for all bounded continuous real-valued functions f on S.

Proposition A.19. *Suppose that μ_k and μ are probability measures on the Borel sets of the metric space S. Then $\mu_k \Rightarrow \mu$ if and only if*

$$\lim_{k\to\infty} \mu_k(A) = \mu(A)$$

for every Borel set $A \subset S$ such that $\mu(\partial A) = 0$. (Here ∂A denotes the boundary of A.) [**3**, page 11]; [**20**, page 108]

The following two concepts play an important role is proving weak convergence of probability measures.

Definition A.20. (a) A family Π of probability measures is said to be relatively compact if every sequence in Π contains a weakly convergent subsequence.

(b) A family Π of probability measures is said to be tight if for every $\epsilon > 0$ there exists a compact $K \subset S$ so that $\mu(K) > 1 - \epsilon$ for every $\mu \in \Pi$.

It turns out that these concepts are equivalent in commonly occurring metric spaces:

Theorem A.21 (Prohorov). *If S is complete and separable, then Π is relatively compact if and only if it is tight.* [**3**, page 37]; [**20**, page 104]

A property that makes weak convergence particularly useful, is the fact that the weak convergence of one sequence implies the weak convergence of many others.

Proposition A.22. *Suppose S and T are metric spaces and X_k and X are S-valued random variables. If $\phi : S \to T$ is measurable and satisfies $P(X \in A) = 1$ for some measurable set A on which ϕ is continuous, then $\phi(X_k) \Rightarrow \phi(X)$ whenever $X_k \Rightarrow X$.* [**3**, page 30]

A.6.2. Characteristic functions. The main tool for proving central limit theorems is the characteristic function.

Definition A.23. The characteristic function of a random vector $\xi = (\xi_1, \ldots, \xi_n)$ in R^n is the function

$$\phi(t) = \phi(t_1, \ldots, t_n) = Ee^{i\langle t,\xi\rangle} = E\exp\left\{i\sum_{j=1}^{n} t_j\xi_j\right\}.$$

Proposition A.24. *If the characteristic functions of two random vectors agree, then so do their distributions.* [**18**, page 167]

Proposition A.25. *If ϕ_k is the characteristic function of the random vector X_k, $\phi_k \to \phi$ pointwise, and ϕ is continuous at the origin, then ϕ is the characteristic function of a random vector X, and $X_k \Rightarrow X$.* [**11**, page 161]

Theorem A.26 (The Cramér-Wold device)**.** *For random vectors X_k and X in R^n, $X_k \Rightarrow X$ if and only if $\langle t, X_k\rangle \Rightarrow \langle t, X\rangle$ for every $t \in R^n$.* [**18**, page 168]

A.6.3. The central limit problem. The main case of the central limit problem is given by the theorem below. More generally, one considers arbitrary normalizing sequences and/or triangular arrays of random variables, rather than sequences.

Theorem A.27 (Central limit theorem)**.** *Suppose $\xi_1, \xi_2, \ldots$ are i.i.d. random vectors with finite second moments, and let $S_n = \xi_1 + \cdots + \xi_n$ be their partial sums. Then*

$$\frac{S_n - nm}{\sqrt{n}} \Rightarrow N(0, \Sigma),$$

where $m = E\xi_i$ and Σ is the covariance matrix of ξ_i. [**18**, page 168]

The following result follows from the central limit theorem and the Kolmogorov $0-1$ law, Theorem A.12.

Corollary A.28. *If $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with mean 0 and finite variance, and $S_n = \xi_1 + \cdots + \xi_n$ are their partial sums, then*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = +\infty \quad \textit{and} \quad \liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty.$$

Stable laws play an important role in the treatment of the general central limit problem. They are exactly the possible distributional limits of sequences of the form

$$\frac{S_n - a_n}{b_n},$$

where S_n is the nth partial sum of a sequence of i.i.d. random variables, and $b_n > 0$ and a_n are normalizing sequences. Here is the definition:

Definition A.29. A probability measure μ on R^1 is said to be stable if for every $a_1, a_2 > 0$ there exist $a > 0$ and b so that $a_1\xi_1 + a_2\xi_2$ and $a\xi + b$ have the same distribution, where ξ_1, ξ_2, and ξ are independent random variables with distribution μ.

The class of stable laws is known explicitly. Here is a the description of the most general stable characteristic function.

Theorem A.30. *A function $\phi(t)$ is the characteristic function of a stable distribution if and only if it is of the form*

$$\phi(t) = \exp\left\{i\gamma t - c|t|^\alpha\left(1 + i\beta\frac{t}{|t|}\omega(t,\alpha)\right)\right\},$$

where $\gamma \in R^1$, $c \geq 0$, $0 < \alpha \leq 2$, $|\beta| \leq 1$, and

$$\omega(t,\alpha) = \begin{cases} \tan\frac{\pi\alpha}{2} & \text{if } \alpha \neq 1; \\ \frac{2}{\pi}\log|t| & \text{if } \alpha = 1. \end{cases}$$

[**18**, page 153]

The parameter α is called the index of the stable law. The stable law is said to be one sided if $\beta = \pm 1$, and symmetric if $\beta = \gamma = 0$. If $\alpha = 2$, the stable law is normal; if $\alpha = 1, \beta = 0$, it is Cauchy.

A second tool that is used in proving convergence in distribution is the method of moments. Here is the main statement.

Theorem A.31. *Suppose the sequence ξ_n satisfies the following:*

$$c_k = \lim_{n\to\infty} E\xi_n^k$$

exists and is finite for $k = 1, 2, \ldots$. Then there exists a random variable ξ with moments given by $E\xi^k = c_k$. If, in addition, the distribution of ξ is uniquely determined by its moments, then $\xi_n \Rightarrow \xi$. [**18**, page 105]

Most potential limiting distributions, such as the Poisson, normal, and absolute value of a normal, as well as all distributions with bounded support, are uniquely determined by their moments.

A.6.4. The moment problem. The classical moment problem asks for conditions on a sequence of numbers that guarantee that it is the sequence of moments for some distribution. The cleanest solution to this problem is for distributions on $[0, 1]$. Suppose $0 \leq \xi \leq 1$, and put $c_n = E\xi^n$. Then for nonnegative integers m, n,

$$\sum_{k=0}^{n}\binom{n}{k}(-1)^k c_{k+m} = \sum_{k=0}^{n}\binom{n}{k}(-1)^k E\xi^{k+m} = E\xi^m(1-\xi)^n \geq 0.$$

It turns out that this condition on the sequence c_k is both necessary and sufficient for the existence of a distribution on $[0,1]$ with these moments.

Theorem A.32. *The sequence c_k is the sequence of moments of a distribution on $[0,1]$ if and only if $c_0 = 1$, and the sequence satisfies*

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^k c_{k+m} \geq 0$$

for every $m, n \geq 0$.

Proof. One direction was proved above. For the other direction, consider random variables ξ_n with distributions given by

$$P\left(\xi_n = \frac{j}{n}\right) = \binom{n}{j} \sum_{k=0}^{n-j} \binom{n-j}{k} (-1)^k c_{k+j}, \quad j = 0, 1, \ldots, n. \tag{A.1}$$

These probabilities are nonnegative by assumption. We will see shortly that they sum to 1. Take $l \geq 0$, and compute as follows for $n \geq l$, using the identity

$$\binom{j}{l}\binom{n}{j}\binom{n-j}{k} = \binom{n}{l}\binom{n-l}{n-k-j}\binom{k+j-l}{k},$$

and making the change of variables $m = k + j$ in the sum:

$$\begin{aligned} E\binom{n\xi_n}{l} &= \sum_{j=l}^{n} \binom{j}{l}\binom{n}{j} \sum_{k=0}^{n-j} \binom{n-j}{k} (-1)^k c_{k+j} \\ &= \binom{n}{l} \sum_{m=l}^{n} \binom{n-l}{n-m} c_m \sum_{k=0}^{m-l} (-1)^k \binom{m-l}{k} = \binom{n}{l} c_l. \end{aligned}$$

Setting $l = 0$ in this identity shows that the probabilities in (A.1) sum to 1. Then dividing both sides of the identity by n^l and passing to the limit shows that

$$\lim_{n\to\infty} E\xi_n^l = c_l, \quad l \geq 0.$$

Therefore $\xi_n \Rightarrow \xi$ for some random ξ satisfying $E\xi^l = c_l$ for $l \geq 0$ by Theorem A.31. □

An important application of this result is the following characterization of infinite exchangeable sequences of Bernoulli random variables. A sequence of random variables is said to be exchangeable if the joint distributions are invariant under permutations of finitely many indexes. For Bernoulli random variables ξ_n, this means that

$$P(\xi_{n_1} = 1, \ldots, \xi_{n_k} = 1) = P(\xi_1 = 1, \ldots, \xi_k = 1)$$

for any $k \geq 1$ and any choice of indexes $n_1 < n_2 < \cdots < n_k$.

Theorem A.33 (De Finetti). *If ξ_n is an infinite exchangeable sequence of Bernoulli random variables, then there is a random variable $0 \le \xi \le 1$ so that*

$$P(\xi_1 = 1, \dots, \xi_k = 1) = E\xi^k$$

for $k = 1, 2, \dots$.

Proof. Let

$$c_{m,n} = P(\xi_1 = 1, \dots, \xi_m = 1, \xi_{m+1} = 0, \dots, \xi_{m+n} = 0).$$

Then

$$c_{m,n} = c_{m+1,n} + c_{m,n+1}.$$

By induction on n, using the identity

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1},$$

we then see that

$$c_{m,n} = \sum_{k=0}^{n} \binom{n}{k} (-1)^k c_{k+m,0}.$$

By Theorem A.32, there is a random variable $0 \le \xi \le 1$ so that $E\xi^k = c_{k,0}$, $k \ge 1$. $\square$

An alternate statement of Theorem A.33 is the following. If S is countably infinite, then every exchangeable probability measure μ on $\{0, 1\}^S$ can be expressed as a mixture

$$\mu = \int_0^1 \nu_\rho \gamma(d\rho),$$

where ν_ρ is the homogeneous product measure with density ρ and γ is a probability measure on $[0, 1]$. (The measure γ is the distribution of the random variable ξ appearing in the statement of Theorem A.33.)

A.7. Discrete time martingales

A filtration is an increasing sequence $\{\mathcal{F}_n\}$ of σ-algebras. A sequence X_n of random variables is said to be adapted to the filtration if X_n is $\mathcal{F}_n$-measurable for each n.

Given an integrable random variable X and σ-algebra $\mathcal{G}$, the conditional expectation $E(X \mid \mathcal{G})$ is the a.s. unique random variable that has the following two properties:

(a) $E(X \mid \mathcal{G})$ is $\mathcal{G}$ measurable.

(b) $E[E(X \mid \mathcal{G}), A] = E(X, A)$ for every $A \in \mathcal{G}$.

Conditional expectations satisfy a number of inequalities. Perhaps the most useful is the following:

Proposition A.34 (Jensen's inequality). *If ϕ is a convex function and both X and $\phi(X)$ are integrable, then*

$$\phi(E(X \mid \mathcal{G})) \leq E(\phi(X) \mid \mathcal{G}) \quad a.s.$$

[**18**, page 223]

The following property is often useful. The idea behind it is that when conditioning on $\mathcal{G}$, any $\mathcal{G}$ measurable random variable can be treated as a constant.

Proposition A.35. *Suppose that $f(x, y)$ is a bounded measurable function, X is $\mathcal{G}$ measurable, and Y is independent of $\mathcal{G}$. Then*

$$E(f(X, Y) \mid \mathcal{G}) = g(X) \quad a.s.,$$

where $g(x) = Ef(x, Y)$.

Definition A.36. Given a filtration $\{\mathcal{F}_n\}$, a sequence of integrable adapted random variables $\{M_n\}$ is said to be a martingale if

$$E(M_{n+1} \mid \mathcal{F}_n) = M_n \quad \text{for each } n.$$

It is said to be a submartingale if

$$E(M_{n+1} \mid \mathcal{F}_n) \geq M_n \quad \text{for each } n,$$

and a supermartingale if

$$E(M_{n+1} \mid \mathcal{F}_n) \leq M_n \quad \text{for each } n.$$

An application of Jensen's inequality gives the following result, which is useful in constructing submartingales from martingales.

Proposition A.37. *If $\{M_n\}$ is a martingale and ϕ is a convex function for which $\phi(M_n)$ is integrable for each n, then $\{\phi(M_n)\}$ is a submartingale.*

The two main substantive results for martingales are the stopping time theorem and the convergence theorem. In order to state them, we need two definitions:

(a) A random variable τ with values in $\{0, 1, \ldots, \infty\}$ is said to be a stopping time if $\{\omega : \tau(\omega) = n\} \in \mathcal{F}_n$ for each $0 \leq n < \infty$.

(b) A family $\{X_\alpha\}$ of random variables is said to be uniformly integrable if

$$\lim_{N\to\infty} \sup_\alpha E(|X_\alpha|, |X_\alpha| \geq N) = 0.$$

The following is a useful sufficient condition for uniform integrability.

Proposition A.38. *If $\sup_\alpha EX_\alpha^2 < \infty$, then $\{X_\alpha\}$ is uniformly integrable.*

Proof. Use

$$E(|X_\alpha|, |X_\alpha| \geq N) \leq \frac{EX_\alpha^2}{N}. \qquad \square$$

Proposition A.39. *If $E|X| < \infty$, then the collection $\{X_{\mathcal{G}} := E(X \mid \mathcal{G})\}$, where $\mathcal{G}$ runs over all σ-algebras, is uniformly integrable.*

Proof. Without loss of generality, assume that $X \geq 0$. Since $\{X_{\mathcal{G}} \geq N\} \in \mathcal{G}$,

$$E(X_{\mathcal{G}}, X_{\mathcal{G}} \geq N) = E(X, X_{\mathcal{G}} \geq N).$$

Now use

$$P(X_{\mathcal{G}} \geq N) \leq \frac{EX_{\mathcal{G}}}{N} = \frac{EX}{N}. \qquad \square$$

Theorem A.40 (Stopping time theorem). *Suppose $M = \{M_n, n \geq 0\}$ is a submartingale, and $\sigma \leq \tau$ are two finite stopping times. If either τ is a.s. bounded or M is uniformly integrable, then*

$$E(M_\tau \mid \mathcal{F}_\sigma) \geq M_\sigma \ \text{a.s.}$$

[**18**, page 270]

Closely related is the following:

Theorem A.41 (Wald's identity). *If $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with $E|\xi_1| < \infty$, $S_n = \xi_1 + \cdots + \xi_n$, and τ is a stopping time relative to the filtration $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ that satisfies $E\tau < \infty$, then $E|S_\tau| < \infty$ and $ES_\tau = E\xi_1 E\tau$.* [**18**, page 178]

The stopping time theorem can be used to prove the next two important inequalities.

Theorem A.42 (Doob's inequality). *If $\{M_k, 0 \leq k \leq n\}$ is a nonnegative submartingale, then*

$$\lambda P(\max_{0 \leq k \leq n} M_k \geq \lambda) \leq EM_n, \quad \lambda > 0,$$

and as a consequence,

$$E(\max_{0 \leq k \leq n} M_k)^2 \leq 4EM_n^2.$$

[**18**, pages 247–248]

Given a real sequence $\{x_0, \ldots, x_n\}$ and $a < b$, the number of upcrossings from below a to above b is the maximum value of k so that there exist indexes $0 \leq j_1 < j_2 < \cdots < j_{2k} \leq n$ satisfying $x_{j_i} \leq a$ for odd i and $x_{j_i} \geq b$ for even i. The martingale convergence theorem is proved using the bound on upcrossings given next.

Theorem A.43 (Upcrossing inequality). *If $M = \{M_k, 0 \le k \le n\}$ is a submartingale, then*

$$E\#\{\textit{upcrossings by } M \textit{ from below } a \textit{ to above } b\} \le \frac{EM_n^+ + |a|}{b-a}.$$

[**18**, page 232]

Theorem A.44 (Convergence theorem). *If $M_0, M_1, \dots$ is a submartingale that satisfies $\sup_n E|M_n| < \infty$, then $\lim_{n\to\infty} M_n$ exists and is finite a.s. If, in addition, the submartingale is uniformly integrable, then the convergence also occurs in L_1.* [**18**, page 233]

When applied to the martingale $M_n = E(X \mid \mathcal{F}_n)$ for some integrable random variable X, we have the following statement:

Corollary A.45. *If X is integrable, then $E(X \mid \mathcal{F}_n)$ converges a.s. and in L_1 to $E(X \mid \mathcal{F})$, where $\mathcal{F}$ is the smallest σ-algebra containing all the $\mathcal{F}_n$'s.*

Occasionally, reversed martingales will be needed. The difference in the definition is that now the $\mathcal{F}_n$'s are decreasing rather than increasing, and the martingale property becomes

$$E(M_n \mid \mathcal{F}_{n+1}) = M_{n+1} \quad \text{for each } n.$$

The convergence theorem is even simpler in this case:

Theorem A.46. *Every reversed martingale converges a.s. and in L_1.* [**18**, page 263]

A.8. Discrete time Markov chains

A discrete time Markov chain on a countable set S is determined by a matrix $p(x,y)$ indexed by $x, y \in S$ that satisfies

$$p(x,y) \ge 0 \text{ and } \sum_y p(x,y) = 1.$$

These are the one-step transition probabilities. The k-step transition probabilities $p_k(x,y)$ are defined recursively by $p_1(x,y) = p(x,y)$ and

$$p_{k+1}(x,y) = \sum_z p_k(x,z)p(z,y)$$

for $k \ge 1$. These satisfy the Chapman-Kolmogorov equations

$$p_{j+k}(x,y) = \sum_z p_j(x,z)p_k(z,y).$$

The corresponding chain with initial state $X_0 = x_0$ is the sequence of random variables X_k with finite-dimensional distributions given by

$$P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = p(x_0,x_1)p(x_1,x_2)\cdots p(x_{n-1},x_n).$$

Probabilities and expectations for the chain with initial state x are denoted by P^x and E^x respectively.

Definition A.47. The Markov chain is said to be irreducible if for every $x, y \in S$ there is a k so that $p_k(x, y) > 0$.

Definition A.48. A state $x \in S$ is said to be recurrent if

$$P^x(X_n = x \text{ for some } n \geq 1) = 1.$$

Otherwise, it is said to be transient.

Proposition A.49. *For an irreducible chain, either all states are recurrent, or all states are transient.*

Definition A.50. An irreducible Markov chain is said to be recurrent if all states are recurrent, and transient if all states are transient.

The Green function $G(x, y)$ for a transient Markov chain is the expected amount of time spent at y by the chain starting at x:

$$G(x, y) = E^x \sum_{k=0}^{\infty} 1_{\{X_k = y\}} = \sum_{k=0}^{\infty} p_k(x, y).$$

It is always finite.

Next come definitions that are relevant to the convergence theorem.

Definition A.51. (a) A measure π on S is stationary for the chain with transition probabilities $p(x, y)$ if

$$\sum_x \pi(x) p(x, y) = \pi(y), \quad y \in S.$$

A stationary measure satisfying $\sum_x \pi(x) = 1$, it is called a stationary distribution.

(b) An irreducible Markov chain is positive recurrent if it has a (necessarily unique) stationary distribution.

(c) An irreducible recurrent Markov chain that is not positive recurrent is null recurrent.

Definition A.52. An irreducible Markov chain is aperiodic if the greatest common divisor of $\{n \geq 1 : p_n(x, x) > 0\}$ is 1 for every (equivalently, for some) $x \in S$.

Theorem A.53. *For an irreducible, aperiodic, positive recurrent chain,*

$$\lim_{n \to \infty} p_n(x, y) = \pi(y),$$

where π is the stationary distribution. [**18**, page 310]

If $S = Z^d$ and $p(x, y) = p(0, y - x)$, the corresponding Markov chain is called a random walk. In this case, the chain starting at x can be realized as a sum of independent, identically distributed random vectors,

$$X_n = x + \sum_{k=1}^{n} \xi_k, \tag{A.2}$$

where $P(\xi_k = y) = p(0, y)$. Here are some useful sufficient conditions for recurrence and transience of random walks.

Theorem A.54. *Consider an irreducible random walk X_n of the form (A.2) and assume that the moments occurring below are well-defined.*

(a) If $d \geq 3$, then X_n is transient. [**18**, page 193]

(b) If $E\xi_k \neq 0$, then X_n is transient.

(c) If $d = 2$, $E\xi_k = 0$, and $E|\xi_k|^2 < \infty$, then X_n is recurrent. [**18**, page 188]

(d) If $d = 1$ and $E\xi_k = 0$, then X_n is recurrent. [**18**, page 188]

A.9. The renewal theorem

Let $S_n = \xi_1 + \cdots + \xi_n$ be a strictly increasing random walk on Z^1. The renewal sequence $u(k)$ associated with this random walk is defined by

$$u(k) = P(S_n = k \text{ for some } n \geq 0). \tag{A.3}$$

It satisfies the renewal equation

$$u(k) = \sum_{j=1}^{k} P(\xi = j)u(k - j), \quad k \geq 1.$$

Theorem A.55. *Suppose that the possible values of the random walk are not contained in a proper subgroup of Z^1. Then*

$$\lim_{k \to \infty} u(k) = \frac{1}{E\xi},$$

where the limit is interpreted as 0 if $E\xi = \infty$. [**9**, Chapter 10]

A.10. Harmonic functions for discrete time Markov chains

A function α on S is said to be harmonic for the Markov chain with transition probabilities $p(x, y)$ if it satisfies the mean value property

$$\sum_{y} p(x, y)\alpha(y) = \alpha(x) \tag{A.4}$$

for all x. Typically, these arise in the following way: If $C \subset S$, the Markov property implies that

$$\alpha(x) = P^x(X_n \in C \ i.o.) \tag{A.5}$$

is harmonic.

Of course, constants are always harmonic. An important problem is to determine whether all harmonic functions that are either bounded or positive are constant. This is certainly not always the case. For example, if S is a tree in which every vertex has exactly three neighbors, and $p(x,y) = 1/3$ whenever x and y are neighbors, it is easy to construct many nonconstant bounded harmonic functions explicitly. A simple example is of the form (A.5) where C is "half" the tree, i.e., all vertices to one side of a particular edge — say the one joining vertices $u \notin C$ and $v \in C$. If $d(x,y)$ is the length of the shortest path joining x and y, then

$$\alpha(x) = \begin{cases} \frac{1}{3}2^{-d(x,u)} & \text{if } x \notin C; \\ 1 - \frac{1}{3}2^{-d(x,v)} & \text{if } x \in C. \end{cases}$$

See Figure 4.

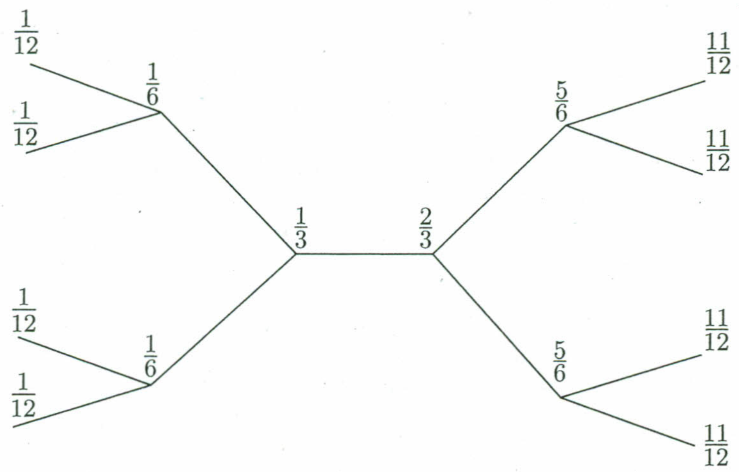

Figure 4: A nonconstant bounded harmonic function on the binary tree.

Here is one case in which all positive harmonic functions are constant.

Theorem A.56. *Suppose the chain is irreducible and recurrent. Then every positive harmonic function is constant.*

Proof. If α is positive and harmonic, then $\alpha(X_n)$ is a positive martingale. Therefore, by Theorem A.44, $\lim_n \alpha(X_n)$ exists a.s. Since X_n visits every state infinitely often, α is constant. $\square$

Corollary A.57. *An irreducible random walk on Z^d cannot be positive recurrent.*

Proof. If the random walk X_n is recurrent, so is the random walk $-X_n$. A stationary measure for one is a harmonic function for the other. $\square$

An important property of random walks is that all bounded harmonic functions are constant. There are various proofs of this result. For a coupling proof, see Section 1 of Chapter II in [**31**]. Here we give a proof based on the martingale convergence theorem.

Theorem A.58. *Suppose X_k is a random walk on Z^d that is weakly irreducible in the sense that for each $x, y \in Z^d$, there is a $k > 0$ so that*

$$p_k(x, y) + p_k(y, x) > 0.$$

If α is a bounded harmonic function for this random walk, then α is constant.

Proof. Suppose α is bounded and harmonic. Then $h(X_n)$ is a bounded martingale with respect to P^x for any x:

$$E^x[\alpha(X_{n+1}) \mid X_1, \ldots, X_n] = E^{X_n}\alpha(X_1) = \alpha(X_n).$$

The first equality uses the Markov property, while the second is just (A.4). Therefore, by Theorem A.44,

$$\lim_{n\to\infty} \alpha(X_n) \tag{A.6}$$

exists a.s. (P^x) and in L_1. This limit is exchangeable, so it is a constant, possibly depending on x, by Theorem A.14. Since $E^x\alpha(X_n) = \alpha(x)$ for every n, the limit in (A.6) is $\alpha(x)$. Writing this statement in the form

$$\alpha(x) = \lim_{n\to\infty} \alpha(x + \xi_1 + \cdots + \xi_n) = \lim_{n\to\infty} \alpha((x + \xi_1) + \xi_2 + \cdots + \xi_n),$$

we see that

$$\alpha(x) = \alpha(x + \xi_1) \text{ a.s.}$$

Iterating this leads to

$$\alpha(x) = \alpha(x + \xi_1 + \cdots + \xi_n) \text{ a.s.}$$

for every n. Therefore $\alpha(x) = \alpha(y)$ whenever $p_n(x, y) > 0$ for some $n \geq 1$, so α is constant by the irreducibility assumption. $\square$

A.11. Subadditive functions

Subadditive functions and processes arise in many probabilistic applications. The subadditive property is key in proving many limit theorems. Here is an example.

Theorem A.59. *Suppose that $f : [0, \infty) \to R^1$ is right continuous at 0, satisfies $f(0) = 0$, and is subadditive in the sense that $f(s+t) \le f(s) + f(t)$ for $s, t \ge 0$. Then*

$$c = \lim_{t \downarrow 0} \frac{f(t)}{t} = \sup_{t>0} \frac{f(t)}{t} \in (-\infty, \infty]$$

exists.

Proof. Let

$$c = \sup_{t>0} \frac{f(t)}{t}.$$

Fix $s > 0$ and for $0 < t \le s$, choose an integer $n \ge 0$ and $0 \le \epsilon < t$ so that $s = nt + \epsilon$. By subadditivity,

$$\frac{f(s)}{s} \le \frac{nt}{nt+\epsilon} \frac{f(t)}{t} + \frac{f(\epsilon)}{s}.$$

Pass to the limit as $t \downarrow 0$ along a sequence t_k for which

$$\lim_{k \to \infty} \frac{f(t_k)}{t_k} \to \liminf_{r \downarrow 0} \frac{f(r)}{r}.$$

Since the corresponding ϵ's and n's tend to 0 and ∞ respectively, it follows that

$$\frac{f(s)}{s} \le \liminf_{r \downarrow 0} \frac{f(r)}{r}.$$

Therefore

$$c \le \liminf_{r \downarrow 0} \frac{f(r)}{r}$$

as required. □

Bibliography

[1] J. Bertoin, *Lévy Processes*, Cambridge University Press, 1996.

[2] C. Bezuidenhout and G. Grimmett, *The critical contact process dies out*, Ann. Probab. **18** (1990), 1462–1482.

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley, 1968.

[4] R. M. Blumenthal and R. K. Getoor, *Markov Processes and Potential Theory*, Academic Press, 1968.

[5] J. Borcea, P. Brändén, and T. M. Liggett, *Negative dependence and the geometry of polynomials*, J. Amer. Math. Soc. **22** (2009), 521–567.

[6] A. N. Borodin and P. Salminen, *Handbook of Brownian Motion — Facts and Formulae*, 2nd edition, Birkhäuser, 2002.

[7] M. Bramson, T. M. Liggett, and T. Mountford, *Characterization of stationary measures for one-dimensional exclusion processes*, Ann. Probab. **30** (2002), 1539–1575.

[8] M. Bramson and T. M. Liggett, *Exclusion processes in higher dimensions: stationary measures and convergence*, Ann. Probab. **33** (2005), 2255–2313.

[9] L. Breiman, *Probability*, SIAM series Classics in Applied Mathematics, 1992.

[10] B. Bru and M. Yor, *Comments on the life and mathematical legacy of Wolfgang Doeblin*, Finance and Stochastics **6** (2002), 3–47.

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

[12] K. L. Chung, *Lectures from Markov Processes to Brownian Motion*, Springer, 1982.

[13] K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, 2nd edition, Birkhäuser, 1990.

[14] Z. Ciesielski and S. J. Taylor, *First passage times and sojourn times for Brownian motion in space and the exact Hausdorff measure of the sample path*, Trans. Amer. Math. Soc. **103** (1962), 434–450.

[15] P. Clifford and A. Sudbury, *A model for spatial conflict*, Biometrika **60** (1973), 581–588.

[16] R. L. Dobrushin, *Markov processes with a large number of locally interacting components: existence of a limit process and its ergodicity*, Prob. Inform. Trans. **7** (1971), 149–164.

[17] J. L. Doob, *Classical Potential Theory and its Probabilistic Counterpart*, Springer, 1984.

[18] R. Durrett, *Probability: Theory and Examples,* 3rd edition, Duxbury, 2005.

[19] R. Durrett, *Stochastic Calculus, A Practical Introduction*, CRC Press, 1996.

[20] S. N. Ethier and T. G. Kurtz, *Markov Processes — Characterization and Convergence*, Wiley, 1986.

[21] D. Freedman, *Brownian Motion and Diffusion*, Holden Day, 1971.

[22] D. Freedman, *Markov Chains*, Holden Day, 1971.

[23] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*, Birkhäuser, 1997.

[24] T. E. Harris, *Contact interactions on a lattice,* Ann. Probab. **2** (1974), 969–988.

[25] R. Holley and T. M. Liggett, *Ergodic theorems for weakly interacting systems and the voter model*, Ann. Probab. **3** (1975), 643–663.

[26] O. Kallenberg, *Foundations of Modern Probability Theory*, Springer, 1997.

[27] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus,* 2nd edition, Springer, 1991.

[28] J. W. Lamperti, *Probability,* 2nd edition, Holden Day, 1971.

[29] S. Lang, *Real Analysis*, Addison-Wesley, 1983.

[30] G. F. Lawler, *Conformally Invariant Processes in the Plane*, American Mathematical Society, 2005.

[31] T. M. Liggett, *Interacting Particle Systems*, Springer, 1985.

[32] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*, Springer, 1999.

[33] T. M. Liggett, *Distributional limits for the symmetric exclusion process*, Stoch. Proc. Appl. **119** (2009), 1–15.

[34] C. T. MacDonald, J. H. Gibbs, and A. C. Pipkin, *Kinetics of biopolymerization on nucleic acid templates,* Biopolymers **6** (1968), 1–25.

[35] P. Mörters and Y. Peres, *Brownian Motion*, Cambridge University Press, 2010.

[36] J. Norris, *Markov Chains*, Cambridge, 1997.

[37] J. Obłój, *The Skorokhod problem and its offspring,* Probab. Surv. **1** (2004), 321–390.

[38] J. Pitman, *Probabilistic bounds on the coefficients of polynomials with only real zeros*, J. Comb. Th. A **77** (1997), 279–303.

[39] S. C. Port and C. J. Stone, *Brownian Motion and Classical Potential Theory*, Academic Press, 1978.

[40] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion,* 2nd edition, Springer, 1994.

[41] H. L. Royden, *Real Analysis, 2nd edition*, Macmillan, 1968.

[42] F. Spitzer, *Interaction of Markov processes*, Adv. Math. **5** (1970), 246–290.

[43] J. M. Steele, *Stochastic Calculus and Financial Applications,* Springer, 2001.

[44] R. J. Williams, *Introduction to the Mathematics of Finance*, American Mathematical Society, 2006.