Appendix A

Useful combinatorial formulas

Recall the following facts about factorials and binomial coefficients:

$$\frac{n^n}{e^{n-1}} \leq n! \leq \frac{n^{n+1}}{e^{n-1}},$$

$$\frac{n^k}{k^k} \leq \binom{n}{k} \leq \frac{e^k n^k}{k^k},$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},$$

$$\sum_{k=0}^{d} \binom{n}{k} \leq \left(\frac{en}{d}\right)^d,$$

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

$$\binom{2n}{n} = (1+o(1))\frac{4^n}{\sqrt{\pi n}},$$

and

$$\log \binom{n}{k} = (1+o(1)) n H(k/n),$$

where $H(p) := -p\log p - (1-p)\log(1-p)$. The third one is the *binomial theorem*. The fifth one is *Stirling's formula*.

Version: December 20, 2023
Modern Discrete Probability: An Essential Toolkit

Appendix B

Measure-theoretic foundations

This appendix contains relevant background on measure-theoretic probability. We follow closely the highly recommended [Wil91]. Missing proofs (and a lot more details and examples) can be found there. Another excellent textbook on this topic is [Dur10].

B.1 Probability spaces

Let S be a set. In general it turns out that we cannot assign a probability to every subset of S. Here we discuss "well-behaved" collections of subsets. First an algebra on S is a collection of subsets stable under finitely many set operations.

Definition B.1.1 (Algebra on S)**.** *A collection Σ_0 of subsets of S is an* algebra on *S if the following conditions hold:*

(i) $S \in \Sigma_0$;

(ii) $F \in \Sigma_0$ *implies* $F^c \in \Sigma_0$;

(iii) $F, G \in \Sigma_0$ *implies* $F \cup G \in \Sigma_0$.

That, of course, implies that the empty set as well as all pairwise intersections are also in Σ_0. The collection Σ_0 is an actual algebra (i.e., a vector space with a bilinear product) with the symmetric difference as its "sum," the intersection as its "product" and the underlying field being the field with two elements.

Example B.1.2. On $\mathbb{R}$, sets of the form

$$\bigcup_{i=1}^{k}(a_i, b_i]$$

where the union is disjoint with $k < +\infty$ and $-\infty \leq a_i \leq b_i \leq +\infty$ form an algebra. ◀

Finite set operations are not enough for our purposes. For instance, we want to be able to take limits. A σ-algebra is stable under countably many set operations.

Definition B.1.3 (σ-algebra on S)**.** *A collection Σ of subsets of S is a* σ-algebra on S *(or* σ-field on S*) if*

(i) $S \in \Sigma$*;*

(ii) $F \in \Sigma$ *implies* $F^c \in \Sigma$*;*

(iii) $F_n \in \Sigma$, $\forall n$ *implies* $\cup_n F_n \in \Sigma$.

Example B.1.4. 2^S is a trivial example. ◀

To give a nontrivial example, we need the following definition. We begin with a lemma.

Lemma B.1.5 (Intersection of σ-algebras)**.** *Let $\mathcal{F}_i$, $i \in I$, be σ-algebras on S where I is arbitrary. Then $\cap_i \mathcal{F}_i$ is a σ-algebra.*

Proof. We prove only one of the conditions. The other ones are similar. Suppose $A \in \mathcal{F}_i$ for all i. Then A^c is in $\mathcal{F}_i$ for all i since each $\mathcal{F}_i$ is itself a σ-algebra. ■

Definition B.1.6 (σ-algebra generated by $\mathcal{C}$)**.** *Let $\mathcal{C}$ be a collection of subsets of S. Then we let $\sigma(\mathcal{C})$ be the smallest σ-algebra containing $\mathcal{C}$, defined as the intersection of all such σ-algebras (including in particular 2^S).*

Example B.1.7. The smallest σ-algebra containing all open sets in $\mathbb{R}$, denoted $\mathcal{B}(\mathbb{R})$, is called the *Borel σ-algebra*. This is a non-trivial σ-algebra in the sense that it can be proved that there exist subsets of $\mathbb{R}$ that are *not* in $\mathcal{B}$, but that any "reasonable" set is in $\mathcal{B}$. In particular, it contains the algebra in Example B.1.2. ◀

Example B.1.8. The σ-algebra generated by the algebra in Example B.1.2 is $\mathcal{B}(\mathbb{R})$. This follows from the fact that all open sets of $\mathbb{R}$ can be written as a countable union of open intervals. (Indeed, for $x \in O$ an open set, let I_x be the largest open interval contained in O and containing x. If $I_x \cap I_y \neq \emptyset$ then $I_x = I_y$ by maximality (i.e., take the union). Then $O = \cup_x I_x$ and there are only countably many disjoint ones because each one contains a rational.) ◀

We now define measures.

Definition B.1.9 (Additivity and σ-additivity)**.** *A non-negative set function on an algebra Σ_0*

$$\mu_0 : \Sigma_0 \to [0, +\infty],$$

is additive *if*

(i) $\mu_0(\emptyset) = 0$*;*

(ii) $F, G \in \Sigma_0$ *with* $F \cap G = \emptyset$ *implies* $\mu_0(F \cup G) = \mu_0(F) + \mu_0(G)$*.*

Moreover μ_0 is said to be σ-additive *if condition (ii) is true for any countable collection of disjoint sets whose union is in Σ_0, that is, if $F_n \in \Sigma_0$, $n \geq 0$, all pairwise disjoint with $\cup_n F_n \in \Sigma_0$, then $\mu_0(\cup_n F_n) = \sum_n \mu_0(F_n)$.*

Example B.1.10. For the algebra in the Example B.1.2, the set function

$$\lambda_0 \left(\bigcup_{i=1}^{k} (a_i, b_i] \right) = \sum_{i=1}^{k} (b_i - a_i)$$

is additive. (In fact, it is also σ-additive. We will show this later.) ◀

Definition B.1.11 (Measure space)**.** *Let Σ be a σ-algebra on S. Then (S, Σ) is a* measurable space*. A σ-additive function μ on Σ is called a* measure *and (S, Σ, μ) is called a* measure space*.*

Definition B.1.12 (Probability space)**.** *If $(\Omega, \mathcal{F}, \mathbb{P})$ is a measure space with $\mathbb{P}(\Omega) = 1$ then $\mathbb{P}$ is called a* probability measure *and $(\Omega, \mathcal{F}, \mathbb{P})$ is called a* probability space *(or* probability triple*).*

probability space

The sets in $\mathcal{F}$ are referred to as *events*.

events

To define a measure on $\mathcal{B}(\mathbb{R})$ we need the following tools from abstract measure theory.

Theorem B.1.13 (Caratheodory's extension theorem)**.** *Let Σ_0 be an algebra on S and let $\Sigma = \sigma(\Sigma_0)$. If μ_0 is σ-additive on Σ_0 then there exists a measure μ on Σ that agrees with μ_0 on Σ_0.*

If in addition μ_0 is finite, the next lemma implies that the extension is unique.

Lemma B.1.14 (Uniqueness of extensions)**.** *Let $\mathcal{I}$ be a* π-system *on S, that is, a family of subsets closed under finite intersections, and let $\Sigma = \sigma(\mathcal{I})$. If μ_1, μ_2 are finite measures on (S, Σ) that agree on $\mathcal{I}$, then they agree on Σ.*

Example B.1.15. The sets $(-\infty, x]$ for $x \in \mathbb{R}$ form a π-system generating $\mathcal{B}(\mathbb{R})$. That is, $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing that π-system. ◀

Finally we can define Lebesgue measure. We start with $(0, 1]$ and extend to $\mathbb{R}$ in the obvious way. We need the following lemma.

Lemma B.1.16 (σ-additivity of λ_0)**.** *Let λ_0 be the set function defined above in Example B.1.10, restricted to $(0, 1]$. Then λ_0 is σ-additive.*

Definition B.1.17 (Lebesgue measure on unit interval)**.** *The unique extension of λ_0 (see Example B.1.10) to $(0, 1]$ is denoted λ and is called* Lebesgue measure*.*

B.2 Random variables

Let (S, Σ, μ) be a measure space and let $\mathcal{B} = \mathcal{B}(\mathbb{R})$.

Definition B.2.1 (Measurable function)**.** *Suppose $h : S \to \mathbb{R}$ and define*

$$h^{-1}(A) = \{s \in S : h(s) \in A\}.$$

The function h is Σ-measurable *if $h^{-1}(B) \in \Sigma$ for all $B \in \mathcal{B}$. We denote by $m\Sigma$ (resp., $(m\Sigma)^+$, $b\Sigma$) the Σ-measurable functions (resp., that are non-negative, bounded).*

In the probabilistic case:

Definition B.2.2. *A* random variable *is a measurable function on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.* *random variable*

The behavior of a random variable is characterized by its distribution function.

Definition B.2.3 (Distribution function)**.** *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The* law of X *is*

$$\mathcal{L}_X = \mathbb{P} \circ X^{-1},$$

which is a probability measure on $(\mathbb{R}, \mathcal{B})$. By Lemma B.1.14, $\mathcal{L}_X$ is determined by the distribution function (DF) of X

distribution function

$$F_X(x) = \mathbb{P}[X \leq x], \quad x \in \mathbb{R}.$$

Example B.2.4. The distribution function of a constant random variable is a jump of size 1 at the value it takes almost surely. The distribution function of a random variable with law equal to Lebesgue measure on $(0, 1]$ is

$$F_X(x) = \begin{cases} x & x \in (0,1], \\ 0 & x \le 0, \\ 1 & x > 1. \end{cases}$$

We refer to such as random variable as a *uniform random variable* over $(0, 1]$. ◀

Distribution functions are characterized by a few simple properties.

Proposition B.2.5. *Suppose $F = F_X$ is the distribution function of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the following hold:*

(i) F is non-decreasing;

(ii) $\lim_{x\to+\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$;

(iii) F is right-continuous.

Proof. The first property follows from the monotonicity of probability measure (which itself follows immediately from σ-additivity).

For the second property, note that the limit exists by the first property. The value of the limit follows from the following important lemma.

Lemma B.2.6 (Monotone convergence properties of measures)**.** *Let (S, Σ, μ) be a measure space.*

(i) If $F_n \in \Sigma$, $n \ge 1$, with $F_n \uparrow F$, then $\mu(F_n) \uparrow \mu(F)$.

(ii) If $G_n \in \Sigma$, $n \ge 1$, with $G_n \downarrow G$ and $\mu(G_k) < +\infty$ for some k, then $\mu(G_n) \downarrow \mu(G)$.

Proof. Clearly $F = \cup_n F_n \in \Sigma$. For $n \ge 1$, write $H_n = F_n \backslash F_{n-1}$ (with $F_0 = \emptyset$). Then by disjointness

$$\mu(F_n) = \sum_{k \le n} \mu(H_k) \uparrow \sum_{k < +\infty} \mu(H_k) = \mu(F).$$

The second statement is similar. ■

Similarly, for the third property, by Lemma B.2.6 again

$$\mathbb{P}[X \le x_n] \downarrow \mathbb{P}[X \le x],$$

if $x_n \downarrow x$. ■

It turns out that the properties above characterize distribution functions in the following sense.

Theorem B.2.7 (Skorokhod representation)**.** *Let F satisfy the three properties above in Proposition B.2.5. Then there is a random variable X on*

$$(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1], \mathcal{B}(0,1], \lambda),$$

with distribution function F. The law of X is called the Lebesgue-Stieltjes measure associated to F.

The result says that all real random variables can be generated from uniform random variables over $(0,1]$.

Proof. Assume first that F is continuous and strictly increasing. Define $X(\omega) = F^{-1}(\omega)$ for all $\omega \in \Omega$. Then, $\forall x \in \mathbb{R}$,

$$\mathbb{P}[X \leq x] = \mathbb{P}[\{\omega \,:\, F^{-1}(\omega) \leq x\}] = \mathbb{P}[\{\omega \,:\, \omega \leq F(x)\}] = F(x).$$

In general, let

$$X(\omega) = \inf\{x \,:\, F(x) \geq \omega\}.$$

It suffices to prove that

$$X(\omega) \leq x \quad \iff \quad \omega \leq F(x).$$

The $\Leftarrow$ direction is clear by definition of X. On the other hand, by the right-continuity of F, we have that $\omega \leq F(X(\omega))$. Therefore, by monotonicity of F,

$$X(\omega) \leq x \quad \Rightarrow \quad \omega \leq F(X(\omega)) \leq F(x).$$

That proves the claim. ■

Turning measurability on its head, we get the following important definition.

Definition B.2.8. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let Y_γ, $\gamma \in \Gamma$, be a collection of maps from Ω to $\mathbb{R}$. We let*

$$\sigma(Y_\gamma, \gamma \in \Gamma),$$

be the smallest σ-algebra on which the Y_γ's are measurable.

In a sense, the above σ-algebra corresponds to "the partial information available when the Y_γs are observed."

Example B.2.9. Suppose we flip two unbiased coins and let X be the number of heads observed. Then, denoting heads by H and tails by T,

$$\sigma(X) = \sigma(\{\{\texttt{HH}\}, \{\texttt{HT}, \texttt{TH}\}, \{\texttt{TT}\}\}),$$

which is coarser than the full σ-algebra 2^Ω. ◀

Note that h^{-1} preserves all set operations. For example, $h^{-1}(A \cup B) = h^{-1}(A) \cup h^{-1}(B)$. This gives the following important lemma.

Lemma B.2.10 (Sufficient condition for measurability)**.** *Suppose $\mathcal{C} \subseteq \mathcal{B}$ with $\sigma(\mathcal{C}) = \mathcal{B}$. Then $h^{-1} : \mathcal{C} \to \Sigma$ implies $h \in m\Sigma$. That is, it suffices to check measurability on a collection generating $\mathcal{B}$.*

Proof. Let $\mathcal{E}$ be the sets such that $h^{-1}(B) \in \Sigma$. By the observation before the statement, $\mathcal{E}$ is a σ-algebra. But $\mathcal{C} \subseteq \mathcal{E}$ which implies $\sigma(\mathcal{C}) \subseteq \mathcal{E}$ by minimality. ■

As a consequence we get the following properties of measurable functions.

Proposition B.2.11 (Properties of measurable functions)**.** *Let h, h_n, $n \geq 1$, be in $m\Sigma$ and $f \in m\mathcal{B}$.*

(i) $f \circ h \in m\Sigma$.

(ii) If S is a topological space and h is continuous, then h is $\mathcal{B}(S)$-measurable, where $\mathcal{B}(S)$ is generated by the open sets of S.

(iii) The function $g : S \to \mathbb{R}$ is in $m\Sigma$ if for all $c \in \mathbb{R}$,

$$\{g \leq c\} \in \Sigma.$$

(iv) $\forall \alpha \in \mathbb{R}$, $h_1 + h_2, h_1 h_2, \alpha h \in m\Sigma$.

(v) $\inf h_n$, $\sup h_n$, $\liminf h_n$, $\limsup h_n$ *are in* $m\Sigma$.

(vi) The set

$$\{s : \lim h_n(s) \text{ exists in } \mathbb{R}\},$$

is measurable.

Proof. We sketch the proof of a few of them.

(ii) This follows from Lemma B.2.10 by taking $\mathcal{C}$ as the open sets of $\mathbb{R}$.

(iii) Similarly, take $\mathcal{C}$ to be the sets of the form $(-\infty, c]$.

(iv) This follows from (iii). For example note that, writing the left-hand side as $h_1 > c - h_2$,

$$\{h_1 + h_2 > c\} = \cup_{q \in \mathbb{Q}} [\{h_1 > q\} \cap \{q > c - h_2\}],$$

which is a countable union of measurable sets by assumption.

(v) Note that

$$\{\sup h_n \leq c\} = \cap_n \{h_n \leq c\}.$$

Further, note that $\liminf$ is the $\sup$ of an $\inf$. ■

B.3 Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition B.3.1 (Independence)**.** *Sub-σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots$ of $\mathcal{F}$ are* independent *if: for all $G_i \in \mathcal{G}_i$, $i \geq 1$, and distinct $i_1, \ldots, i_n$ we have* *independence*

$$\mathbb{P}[G_{i_1} \cap \cdots \cap G_{i_n}] = \prod_{j=1}^{n} \mathbb{P}[G_{i_j}].$$

Specializing to events and random variables:

Definition B.3.2 (Independent random variables)**.** *Random variables $X_1, X_2, \ldots$ are* independent *if the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots$ are independent.*

Definition B.3.3 (Independent events)**.** *Events $E_1, E_2, \ldots$ are* independent *if the σ-algebras*

$$\mathcal{E}_i = \{\emptyset, E_i, E_i^c, \Omega\}, \quad i \geq 1,$$

are independent.

Recall the more familiar definitions.

Theorem B.3.4 (Independent random variables: familiar definition)**.** *Random variables X, Y are independent if and only if for all $x, y \in \mathbb{R}$*

$$\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x]\, \mathbb{P}[Y \leq y].$$

Theorem B.3.5 (Independent events: familiar definition)**.** *Events E_1, E_2 are independent if and only if*

$$\mathbb{P}[E_1 \cap E_2] = \mathbb{P}[E_1]\, \mathbb{P}[E_2].$$

The proofs of these characterizations follow immediately from the following lemma.

Lemma B.3.6 (Independence and π-systems)**.** *Suppose that $\mathcal{G}$ and $\mathcal{H}$ are sub-σ-algebras and that $\mathcal{I}$ and $\mathcal{J}$ are π-systems such that*

$$\sigma(\mathcal{I}) = \mathcal{G}, \quad \sigma(\mathcal{J}) = \mathcal{H}.$$

Then $\mathcal{G}$ and $\mathcal{H}$ are independent if and only if $\mathcal{I}$ and $\mathcal{J}$ are as well, that is,

$$\mathbb{P}[I \cap J] = \mathbb{P}[I]\,\mathbb{P}[J], \quad \forall I \in \mathcal{I}, J \in \mathcal{J}.$$

Proof. Suppose $\mathcal{I}$ and $\mathcal{J}$ are independent. For fixed $I \in \mathcal{I}$, the measures $\mathbb{P}[I \cap H]$ and $\mathbb{P}[I]\,\mathbb{P}[H]$ are equal for $H \in \mathcal{J}$ and have total mass $\mathbb{P}[I] < +\infty$. By the Uniqueness of Extensions Lemma (Lemma B.1.14) the above measures agree on $\sigma(\mathcal{J}) = \mathcal{H}$.

Repeat the argument. Fix $H \in \mathcal{H}$. Then the measures $\mathbb{P}[G \cap H]$ and $\mathbb{P}[G]\,\mathbb{P}[H]$ agree on $\mathcal{I}$ and have total mass $\mathbb{P}[H] < +\infty$. Therefore they must agree on $\sigma(\mathcal{I}) = \mathcal{G}$. ■

We give a standard construction of an infinite sequence of independent random variables with prescribed distributions.

Let $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1], \mathcal{B}(0,1], \lambda)$ and for $\omega \in \Omega$ consider the binary expansion

$$\omega = 0.\omega_1\omega_2 \ldots.$$

(For dyadic rationals, use the all-1 ending and note that the dyadic rationals have measure 0 by countability.) This construction produces a sequence of independent so-called *Bernoulli trials*. That is, under λ, each bit is Bernoulli$(1/2)$ and any finite collection is independent.

Bernoulli trials

To get two independent uniform random variables, consider the following construction:

$$\begin{aligned} U_1 &= 0.\omega_1\omega_3\omega_5 \ldots \\ U_2 &= 0.\omega_2\omega_4\omega_6 \ldots \end{aligned}$$

Let $\mathcal{A}_1$ (resp. $\mathcal{A}_2$) be the π-system consisting of all finite intersections of events of the form $\{\omega_i \in H_i\}$ for odd i (resp. even i). By Lemma B.3.6, the σ-fields $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent.

More generally, let

$$\begin{aligned} V_1 &= 0.\omega_1\omega_3\omega_6 \ldots \\ V_2 &= 0.\omega_2\omega_5\omega_9 \ldots \\ V_3 &= 0.\omega_4\omega_8\omega_{13} \ldots \\ \vdots &= \ddots \end{aligned}$$

that is, fill up the array diagonally. By the argument above, the V_i's are independent and Bernoulli($1/2$).

Finally let μ_n, $n \geq 1$, be a sequence of probability measures with distribution functions F_n, $n \geq 1$. For each n, define

$$X_n(\omega) = \inf\{x \,:\, F_n(x) \geq V_n(\omega)\}$$

By the (proof of the) Skorokhod Representation (Theorem B.2.7), X_n has distribution function F_n.

Definition B.3.7 (I.i.d. random variables)**.** *A sequence of independent random variables (X_n) as above is independent and identically distributed (i.i.d.) if $F_n = F$ for some n.*

Alternatively, we have the following more general result.

Theorem B.3.8 (Kolmogorov's extension theorem)**.** *Suppose we are given probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ that are* consistent*, that is,*

$$\mu_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Then there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ with

$$\mathbb{P}[\omega \,:\, \omega_i \in (a_i, b_i], 1 \leq i \leq n] = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Here $\mathcal{R}^{\mathbb{N}}$ is the product σ-algebra*, that is, the σ-algebra generated by finite-dimensional rectangles.*

Next, we discuss a first non-trivial result about independent sequences.

Definition B.3.9 (Tail σ-algebra)**.** *Let $X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define*

$$\mathcal{T} = \bigcap_{n \geq 1} \mathcal{T}_n,$$

where

$$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots).$$

As an intersection of σ-algebras, $\mathcal{T}$ is a σ-algebra. It is called the tail σ-algebra *of the sequence (X_n).*

Intuitively, an event is in the tail if changing a finite number of values does not affect its occurence.

Example B.3.10. If $S_n = \sum_{k \leq n} X_k$, then

$$\{\lim_n S_n \text{ exists}\} \in \mathcal{T},$$

$$\{\limsup_n n^{-1} S_n > 0\} \in \mathcal{T},$$

but

$$\{\limsup_n S_n > 0\} \notin \mathcal{T}.$$

◀

Theorem B.3.11 (Kolmogorov's 0-1 law)**.** *Let* (X_n) *be a sequence of independent random variables with tail* σ*-algebra* $\mathcal{T}$*. Then* $\mathcal{T}$ *is* $\mathbb{P}$*-trivial, that is, for all* $A \in \mathcal{T}$ *we have either* $\mathbb{P}[A] = 0$ *or* 1*.*

Proof. Let $\mathcal{X}_n = \sigma(X_1, \ldots, X_n)$. Note that $\mathcal{X}_n$ and $\mathcal{T}_n$ are independent. Moreover, since $\mathcal{T} \subseteq \mathcal{T}_n$ we have that $\mathcal{X}_n$ is independent of $\mathcal{T}$. Now let

$$\mathcal{X}_\infty = \sigma(X_n, n \geq 1).$$

Note that

$$\mathcal{K}_\infty = \bigcup_{n \geq 1} \mathcal{X}_n,$$

is a π-system generating $\mathcal{X}_\infty$. Therefore, by Lemma B.3.6, $\mathcal{X}_\infty$ is independent of $\mathcal{T}$. But $\mathcal{T} \subseteq \mathcal{X}_\infty$ and therefore $\mathcal{T}$ is independent of itself! Hence if $A \in \mathcal{T}$,

$$\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]^2,$$

which can occur only if $\mathbb{P}[A] \in \{0, 1\}$. ■

B.4 Expectation

Let (S, Σ, μ) be a measure space. We denote by $\mathbf{1}_A$ the indicator of a set A, that is,

$$\mathbf{1}_A(s) = \begin{cases} 1, & \text{if } s \in A \\ 0, & \text{o.w.} \end{cases}$$

Definition B.4.1 (Simple functions)**.** *A* simple function *is a function of the form*

$$f = \sum_{k=1}^{m} a_k \mathbf{1}_{A_k},$$

where $a_k \in [0, +\infty]$ *and* $A_k \in \Sigma$ *for all* k. *We denote the set of all such functions by* SF^+. *We define the* integral of f *by*

$$\mu(f) := \sum_{k=1}^{m} a_k \mu(A_k) \leq +\infty.$$

We also write $\mu f = \mu(f)$.

The following is left as a (somewhat tedious but) immediate exercise.

Proposition B.4.2. *Let* $f, g \in \mathrm{SF}^+$.

(i) If $\mu(f \neq g) = 0$, *then* $\mu f = \mu g$. *[Hint: Rewrite* f *and* g *over the same disjoint sets.]*

(ii) For all $c \geq 0$, $f + g, cf \in \mathrm{SF}^+$ *and*

$$\mu(f+g) = \mu f + \mu g, \quad \mu(cf) = c\mu f.$$

[Hint: This one is obvious by definition.]

(iii) If $f \leq g$ *then* $\mu f \leq \mu g$. *[Hint: Show that* $g - f \in \mathrm{SF}^+$ *and use linearity.]*

The main definition and theorem of integration theory follows.

Definition B.4.3 (Nonnegative functions)**.** *Let* $f \in (\mathrm{m}\Sigma)^+$. *Then the* integral of f *is defined by*

$$\mu(f) = \sup\{\mu(h) \,:\, h \in \mathrm{SF}^+,\ h \leq f\}.$$

Again we also write $\mu f = \mu(f)$.

Theorem B.4.4 (Monotone convergence theorem)**.** *If* $f_n, f \in (\mathrm{m}\Sigma)^+$, $n \geq 1$, *with* $f_n \uparrow f$, *then*

$$\mu f_n \uparrow \mu f.$$

Many theorems in integration follow from the monotone convergence theorem. In that context, the following approximation is useful.

Definition B.4.5 (Staircase function)**.** *For* $f \in (\mathrm{m}\Sigma)^+$ *and* $r \geq 1$, *the* r-th staircase function $\alpha^{(r)}$ *is*

$$\alpha^{(r)}(x) = \begin{cases} 0, & \text{if } x = 0, \\ (i-1)2^{-r}, & \text{if } (i-1)2^{-r} < x \leq i2^{-r} \leq r, \\ r, & \text{if } x > r, \end{cases}$$

We let $f^{(r)} = \alpha^{(r)}(f)$. *Note that* $f^{(r)} \in \mathrm{SF}^+$ *and* $f^{(r)} \uparrow f$ *as* $r \to +\infty$.

Using the previous definition, we get for example the following properties.

Proposition B.4.6. *Let $f, g \in (\mathrm{m}\Sigma)^+$.*

(i) If $\mu(f \neq g) = 0$, then $\mu(f) = \mu(g)$.

(ii) For all $c \geq 0$, $f + g, cf \in (\mathrm{m}\Sigma)^+$ and

$$\mu(f+g) = \mu f + \mu g, \quad \mu(cf) = c\mu f.$$

(iii) If $f \leq g$ then $\mu f \leq \mu g$.

For a function f, let f^+ and f^- be the positive and negative parts of f, that is,

$$f^+(s) = f(s) \vee 0, \quad f^-(s) = (-f(s)) \vee 0.$$

Note that $|f| = f^+ + f^-$. Finally we define

$$\mu(f) := \mu(f^+) - \mu(f^-),$$

provided $\mu(f^+) + \mu(f^-) < +\infty$, in which case we write $f \in L^1(S, \Sigma, \mu)$. Proposition B.4.6 can be generalized naturally to this definition. Moreover we have the following.

Theorem B.4.7 (Dominated convergence theorem)**.** *If $f_n, f \in \mathrm{m}\Sigma$, $n \geq 1$, with $f_n(s) \to f(s)$ for all $s \in S$, and there is a nonnegative function $g \in L^1(S, \Sigma, \mu)$ such that $|f_n| \leq g$, then*

$$\mu(|f_n - f|) \to 0,$$

and in particular

$$\mu f_n \to \mu f,$$

as $n \to \infty$.

More generally, for $0 < p < +\infty$, the space $L^p(S, \Sigma, \mu)$ contains all functions $f : S \to \mathbb{R}$ such that $\|f\|_p < +\infty$, where

$$\|f\|_p := \mu(|f|^p)^{1/p},$$

up to equality almost everywhere. We state the following results without proof.

Theorem B.4.8 (Hölder's inequality)**.** *Let $1 < p, q < +\infty$ such that $p^{-1} + q^{-1} = 1$. Then, for any $f \in L^p(S, \Sigma, \mu)$ and $g \in L^q(S, \Sigma, \mu)$, it holds that $fg \in L^1(S, \Sigma, \mu)$ and further*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

The case $p = q = 2$ is known as the Cauchy-Schwarz inequality *(or* Schwarz inequality*).*

Cauchy-Schwarz inequality

Theorem B.4.9 (Minkowski's inequality). *Let $1 < p < +\infty$. Then, for any $f, g \in L^p(S, \Sigma, \mu)$, it holds that $f + g \in L^p(S, \Sigma, \mu)$ and further*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Theorem B.4.10 (L^p completeness). *Let $1 \leq p < +\infty$. If $(f_n)_n$ in $L^p(S, \Sigma, \mu)$ is Cauchy, that is,*

$$\sup_{n,m \geq k} \|f_n - f_m\|_p \to 0,$$

as $k \to +\infty$, then there exists $f \in L^p(S, \Sigma, \mu)$ such that

$$\|f_n - f\|_p \to 0,$$

as $n \to +\infty$.

We can now define the expectation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition B.4.11 (Expectation). *If $X \geq 0$ is a random variable then we define the* expectation of X, *denoted by $\mathbb{E}[X]$, as the integral of X over $\mathbb{P}$. More generally (i.e., not assuming non-negativity), if*

$$\mathbb{E}|X| = \mathbb{E}[X^+] + \mathbb{E}[X^-] < +\infty,$$

we let

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

We denote the set of all such integrable random variables *(up to equality almost surely) by $L^1(\Omega, \mathcal{F}, \mathbb{P})$.* *integrable*

The properties of the integral for nonnegative functions (see Proposition B.4.6) extend to the expectation.

Proposition B.4.12. *Let X, X_1, X_2 be random variables in $L^1(\Omega, \mathcal{F}, \mathbb{P})$.*

(LIN) *If $a_1, a_2 \in \mathbb{R}$, then $\mathbb{E}[a_1 X_1 + a_2 X_2] = a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2]$.*

(POS) *If $X \geq 0$, then $\mathbb{E}[X] \geq 0$.*

One useful implication of (POS) is that $|X| - X \geq 0$ so that $\mathbb{E}[X] \leq \mathbb{E}|X|$ and, by applying the same argument to $-X$, we have further $|\mathbb{E}[X]| \leq \mathbb{E}|X|$.

The monotone convergence theorem (Theorem B.4.4) implies the following results. We first need a definition.

Definition B.4.13 (Convergence almost sure). *We say that $X_n \to X$ almost surely (a.s.) if*

$$\mathbb{P}[X_n \to X] = 1.$$

Proposition B.4.14. *Let X, Y, X_n, $n \geq 1$, be random variables in $L^1(\Omega, \mathcal{F}, \mathbb{P})$.*

(MON) *If $0 \leq X_n \uparrow X$, then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X] \leq +\infty$.*

(FATOU) *If $X_n \geq 0$, then $\mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n]$.*

(DOM) *If $|X_n| \leq Y$, $n \geq 1$, with $\mathbb{E}[Y] < +\infty$ and $X_n \to X$ a.s., then*

$$\mathbb{E}|X_n - X| \to 0,$$

and, hence,

$$\mathbb{E}[X_n] \to \mathbb{E}[X].$$

(Indeed,

$$\begin{aligned} |\mathbb{E}[X_n] - \mathbb{E}[X]| &= |\mathbb{E}[X_n - X]| \\ &\leq \mathbb{E}|X_n - X|.) \end{aligned}$$

(SCHEFFE) *If $X_n \to X$ a.s. and $\mathbb{E}|X_n| \to \mathbb{E}|X|$ then*

$$\mathbb{E}|X_n - X| \to 0.$$

(BDD) *If $X_n \to X$ a.s. and $|X_n| \leq K < +\infty$ for all n then*

$$\mathbb{E}|X_n - X| \to 0.$$

Proof. We only prove (FATOU). To use (MON) we write the $\liminf$ as an increasing limit. Letting $Z_k = \inf_{n \geq k} X_n$, we have

$$\liminf_n X_n = \uparrow \lim_k Z_k,$$

so that by (MON)

$$\mathbb{E}[\liminf_n X_n] = \uparrow \lim_k \mathbb{E}[Z_k].$$

For $n \geq k$ we have $X_n \geq Z_k$ so that $\mathbb{E}[X_n] \geq \mathbb{E}[Z_k]$ hence

$$\mathbb{E}[Z_k] \leq \inf_{n \geq k} \mathbb{E}[X_n].$$

Finally, we get

$$\mathbb{E}[\liminf_n X_n] \leq \uparrow \lim_k \inf_{n \geq k} \mathbb{E}[X_n].$$

■

The following inequality is often useful. We give an example below.

Jensen's inequality

Theorem B.4.15 (Jensen's inequality)**.** *Let $h : G \to \mathbb{R}$ be a convex function on an open interval G such that $\mathbb{P}[X \in G] = 1$ and $X, h(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then*

$$\mathbb{E}[h(X)] \geq h(\mathbb{E}[X]).$$

The L^p norm defined earlier applies to random variables as well. That is, for $p \geq 1$, we let $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$ and denote by $L^p(\Omega, \mathcal{F}, \mathbb{P})$ the collection of random variables X (up to almost sure equality) such that $\|X\|_p < +\infty$. Jensen's inequality (Theorem B.4.15) implies the following relationship.

Lemma B.4.16 (Monotonicity of norms)**.** *For $1 \leq p \leq r < +\infty$, we have $\|X\|_p \leq \|X\|_r$.*

Proof. For $n \geq 0$, let

$$X_n = (|X| \wedge n)^p.$$

Take $h(x) = x^{r/p}$which is convex on $(0, +\infty)$. Then, by Jensen's inequality,

$$(\mathbb{E}[X_n])^{r/p} \leq \mathbb{E}[(X_n)^{r/p}] = \mathbb{E}[(|X| \wedge n)^r] \leq \mathbb{E}[|X|^r].$$

Take $n \to \infty$ and use (MON). ■

This latter inequality is useful among other things to argue about the convergence of expectations. We say that X_n converges to X_∞ in L^p if $\|X_n - X_\infty\|_p \to 0$. By the previous lemma, convergence on L^r implies convergence in L^p for $r \geq p \geq 1$. Further we have:

Lemma B.4.17 (Convergence of expectations)**.** *Assume $X_n, X_\infty \in L^1$. Then*

$$\|X_n - X_\infty\|_1 \to 0,$$

implies

$$\mathbb{E}[X_n] \to \mathbb{E}[X_\infty].$$

Proof. Note that

$$|\mathbb{E}[X_n] - \mathbb{E}[X_\infty]| \leq \mathbb{E}|X_n - X_\infty| \to 0.$$

■

So, a fortiori, convergence in L^p, $p \geq 1$, implies convergence of expectations.

Square integrable random variables have a nice geometry by virtue of forming a Hilbert space.

Definition B.4.18 (Square integrable variables)**.** *Recall that $L^2(\Omega, \mathcal{F}, \mathbb{P})$ denotes the set of all* square integrable random variables *(up to equality almost surely), that is, those X with $\mathbb{E}[X^2] < +\infty$ For $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, define the inner product $\langle X, Y\rangle := \mathbb{E}[XY]$. Then the L^2 norm is $\|X\|_2 = \sqrt{\langle X, X\rangle}$.* *square integrable*

Theorem B.4.19 (Cauchy-Schwarz inequality)**.** *If $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and*

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]},$$

or put differently

$$|\langle X, Y\rangle| \leq \|X\|_2 \|Y\|_2.$$

Theorem B.4.20 (Parallelogram law)**.** *If $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, then* *parallelogram law*

$$\|X + Y\|_2^2 + \|X - Y\|_2^2 = 2\|X\|_2^2 + 2\|Y\|_2^2.$$

B.5 Fubini's theorem

We now define product measures and state (without proof) Fubini's Theorem.

Definition B.5.1 (Product σ-algebra)**.** *Let (S_1, Σ_1) and (S_2, Σ_2) be measure spaces. Let $S = S_1 \times S_2$ be the Cartesian product of S_1 and S_2. For $i = 1, 2$, let $\pi_i : S \to S_i$ be the projection on the i-th coordinate, that is,*

$$\pi_i(s_1, s_2) = s_i.$$

The product σ-algebra $\Sigma = \Sigma_1 \times \Sigma_2$ *is defined as*

$$\Sigma = \sigma(\pi_1, \pi_2).$$

In words, it is the smallest σ-algebra that makes coordinate maps measurable. It is generated by sets of the form

$$\pi_1^{-1}(B_1) = B_1 \times S_2, \quad \pi_2^{-1}(B_2) = S_1 \times B_2, \quad B_1 \in \Sigma_1, B_2 \in \Sigma_2.$$

Theorem B.5.2 (Fubini's Theorem)**.** *For $F \in \Sigma$, let $f = \mathbf{1}_F$ and define* *Fubini's Theorem*

$$\mu(F) := \int_{S_1} I_1^f(s_1)\mu_1(\mathrm{d}s_1) = \int_{S_2} I_2^f(s_2)\mu_2(\mathrm{d}s_2),$$

where

$$I_1^f(s_1) := \int_{S_2} f(s_1, s_2)\mu_2(\mathrm{d}s_2) \in \mathrm{b}\Sigma_1,$$
$$I_2^f(s_2) := \int_{S_1} f(s_1, s_2)\mu_1(\mathrm{d}s_1) \in \mathrm{b}\Sigma_2.$$

(The equality and inclusions above are part of the statement.) The set function μ is a measure on (S, Σ) called the product measure *of μ_1 and μ_2 and we write $\mu = \mu_1 \times \mu_2$ and*

$$(S, \Sigma, \mu) = (S_1, \Sigma_1, \mu_1) \times (S_2, \Sigma_2, \mu_2).$$

Moreover μ is the unique measure on (S, Σ) for which

$$\mu(A_1 \times A_2) = \mu(A_1)\mu(A_2), \quad A_i \in \Sigma_i.$$

If $f \in (\mathrm{m}\Sigma)^+$ then

$$\mu(f) = \int_{S_1} I_1^f(s_1)\mu_1(\mathrm{d}s_1) = \int_{S_2} I_2^f(s_2)\mu_2(\mathrm{d}s_2),$$

where I_1^f, I_2^f are defined as before (i.e., as the sup *over bounded functions from below). The same is valid if $f \in \mathrm{m}\Sigma$ and $\mu(|f|) < +\infty$.*

Some applications of Fubini's Theorem (Theorem B.5.2) follow. We first recall the following useful formula.

Theorem B.5.3 (Change-of-variables formula)**.** *Let X be a random variable with law $\mathcal{L}$. If $f : \mathbb{R} \to \mathbb{R}$ is such that either $f \geq 0$ or $\mathbb{E}|f(X)| < +\infty$ then*

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(y)\mathcal{L}(\mathrm{d}y).$$

Proof. We use the standard machinery.

1. For $f = \mathbf{1}_B$ with $B \in \mathcal{B}$,

$$\mathbb{E}[\mathbf{1}_B(X)] = \mathcal{L}(B) = \int_{\mathbb{R}} \mathbf{1}_B(y)\mathcal{L}(\mathrm{d}y).$$

2. If $f = \sum_{k=1}^m a_k \mathbf{1}_{A_k}$ is a simple function, then by (LIN)

$$\mathbb{E}[f(X)] = \sum_{k=1}^m a_k \mathbb{E}[\mathbf{1}_{A_k}(X)] = \sum_{k=1}^m a_k \int_{\mathbb{R}} \mathbf{1}_{A_k}(y)\mathcal{L}(\mathrm{d}y) = \int_{\mathbb{R}} f(y)\mathcal{L}(\mathrm{d}y).$$

3. Let $f \geq 0$ and approximate f by a sequence $\{f_n\}$ of increasing simple functions. By (MON)

$$\mathbb{E}[f(X)] = \lim_n \mathbb{E}[f_n(X)] = \lim_n \int_{\mathbb{R}} f_n(y)\mathcal{L}(\mathrm{d}y) = \int_{\mathbb{R}} f(y)\mathcal{L}(\mathrm{d}y).$$

4. Finally, assume that f is such that $\mathbb{E}|f(X)| < +\infty$. Then by (LIN)

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[f^+(X)] - \mathbb{E}[f^-(X)] \\
&= \int_{\mathbb{R}} f^+(y)\mathcal{L}(\mathrm{d}y) - \int_{\mathbb{R}} f^-(y)\mathcal{L}(\mathrm{d}y) \\
&= \int_{\mathbb{R}} f(y)\mathcal{L}(\mathrm{d}y).
\end{aligned}$$

■

Theorem B.5.4. *Let X and Y be independent random variables with respective laws μ and ν. Let f and g be measurable functions such that either $f, g \geq 0$ or $\mathbb{E}|f(X)|, \mathbb{E}|g(Y)| < +\infty$. Then*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

Proof. From the change-of-variables formula (Theorem B.5.3) and Fubini's Theorem (Theorem B.5.2), we get

$$\begin{aligned}
\mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^2} f(x)g(y)(\mu \times \nu)(\mathrm{d}x \times \mathrm{d}y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x)g(y)\mu(\mathrm{d}x) \right) \nu(\mathrm{d}y) \\
&= \int_{\mathbb{R}} \left(g(y)\mathbb{E}[f(X)] \right) \nu(\mathrm{d}y) \\
&= \mathbb{E}[f(X)]\mathbb{E}[g(Y)].
\end{aligned}$$

■

Definition B.5.5 (Density)**.** *Let X be a random variable with law μ. We say that X has density f_X if for all $B \in \mathcal{B}(\mathbb{R})$*

$$\mu(B) = \mathbb{P}[X \in B] = \int_B f_X(x)\lambda(\mathrm{d}x).$$

Theorem B.5.6 (Convolution)**.** *Let X and Y be independent random variables with distribution functions F and G respectively. Then the distribution function, H, of $X + Y$ is*

$$H(z) = \int F(z - y)\mathrm{d}G(y).$$

This is called the convolution *of F and G. Moreover, if X and Y have densities f and g respectively, then $X + Y$ has density*

$$h(z) = \int f(z - y)g(y)\mathrm{d}y.$$

Proof. From Fubini's Theorem (Theorem B.5.3), denoting the laws of X and Y by μ and ν respectively,

$$\begin{aligned}
\mathbb{P}[X+Y\leq z] &= \int\int \mathbf{1}_{\{x+y\leq z\}}\mu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&= \int F(z-y)\nu(\mathrm{d}y)\\
&= \int F(z-y)\mathrm{d}G(y)\\
&= \int \left(\int_{-\infty}^{z} f(x-y)\mathrm{d}x\right)\mathrm{d}G(y)\\
&= \int_{-\infty}^{z} \left(\int f(x-y)\mathrm{d}G(y)\right)\mathrm{d}x\\
&= \int_{-\infty}^{z} \left(\int f(x-y)g(y)\mathrm{d}y\right)\mathrm{d}x.
\end{aligned}$$

■

See Exercise 2.1 for a proof of the following standard formula.

Theorem B.5.7 (Moments of nonegative random variables)**.** *For any nonnegative random variable X and positive integer k,*

$$\mathbb{E}[X^k] = \int_0^{+\infty} kx^{k-1}\mathbb{P}[X>x]\,\mathrm{d}x. \tag{B.5.1}$$

B.6 Conditional expectation

Before defining the conditional expectation, we recall some elementary concepts. For two events A, B, the conditional probability of A given B is defined as

$$\mathbb{P}[A\,|\,B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]},$$

where we assume $\mathbb{P}[B] > 0$.

Now let X and Z be random variables taking values $x_1,\ldots,x_m$ and $z_1,\ldots,z_n$ respectively. The conditional expectation of X given $Z=z_j$ is defined as

$$y_j = \mathbb{E}[X\,|\,Z=z_j] = \sum_i x_i\mathbb{P}[X=x_i\,|\,Z=z_j],$$

where we assume $\mathbb{P}[Z=z_j] > 0$ for all j. As motivation for the general definition, we make the following observations.

- We can think of the conditional expectation as a *random variable* $Y = \mathbb{E}[X \mid Z]$ defined as follows

$$Y(\omega) = y_j \text{ on } G_j = \{\omega : Z(\omega) = z_j\}.$$

- Then Y is $\mathcal{G}$-measurable where $\mathcal{G} = \sigma(Z)$.

- On sets in $\mathcal{G}$, the expectation of Y agrees with the expectation of X. Indeed, note first that

$$\begin{aligned}
\mathbb{E}[Y; G_j] &= y_j \mathbb{P}[G_j] \\
&= \sum_i x_i \mathbb{P}[X = x_i \mid Z = z_j] \mathbb{P}[Z = z_j] \\
&= \sum_i x_i \mathbb{P}[X = x_i, Z = z_j] \\
&= \mathbb{E}[X; G_j].
\end{aligned}$$

  This is also true for all $G \in \mathcal{G}$ by summation over j.

We are ready to state the general definition of the conditional expectation. Its existence and uniqueness follow from the next theorem.

Theorem B.6.1 (Conditional expectation)**.** *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra. Then:*

(i) (Existence) *There exists a random variable $Y \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ such that*

$$\mathbb{E}[Y; G] = \mathbb{E}[X; G], \ \forall G \in \mathcal{G}. \tag{B.6.1}$$

Such a Y is called a version of the conditional expectation of X given $\mathcal{G}$ *and is denoted by $\mathbb{E}[X \mid \mathcal{G}]$.* *conditional expectation*

(ii) (Uniqueness) *It is unique in the sense that, if Y and Y' are two versions of the conditional expectation, then $Y = Y'$ almost surely.*

When $\mathcal{G} = \sigma(Z)$, we sometimes use the notation $\mathbb{E}[X \mid Z] := \mathbb{E}[X \mid \mathcal{G}]$. A similar convention applies to collections of random variables, for example, $\mathbb{E}[X \mid Z_1, Z_2] := \mathbb{E}[X \mid \sigma(Z_1, Z_2)]$ and so on.

We first prove uniqueness. Existence is proved below after some more concepts are introduced.

Proof of Theorem B.6.1 (ii). By way of contradiction, let Y, Y' be two versions of $\mathbb{E}[X \mid G]$ such that without loss of generality $\mathbb{P}[Y > Y'] > 0$. By monotonicity, there is $n \geq 1$ with $G = \{Y > Y' + n^{-1}\} \in \mathcal{G}$ such that $\mathbb{P}[G] > 0$. Then, by definition,

$$0 = \mathbb{E}[Y - Y'; G] > n^{-1}\mathbb{P}[G] > 0,$$

which gives a contradiction. ■

To prove existence, we use the L^2 method. In $L^2(\Omega, \mathcal{F}, \mathbb{P})$, the conditional expectation reduces to an orthogonal projection.

Theorem B.6.2 (Conditional expectation: L^2 case)**.** *Let $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra. Then there exists an (almost surely) unique $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ such that*

$$\|X - Y\|_2 = \Delta := \inf\{\|X - W\|_2 \,:\, W \in L^2(\Omega, \mathcal{G}, \mathbb{P})\},$$

and, moreover, $\langle Z, X - Y\rangle = 0$, $\forall Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. In particular, it satisfies (B.6.1)*. Such a Y is called* the orthogonal projection of X on $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

Proof. Take (Y_n) such that $\|X - Y_n\|_2 \to \Delta$. We use the fact that $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is complete (Theorem B.4.10) and first seek to prove that (Y_n) is Cauchy. Using the parallelogram law (Theorem B.4.20), note that

$$\|X - Y_r\|_2^2 + \|X - Y_s\|_2^2 = 2\left\|X - \frac{1}{2}(Y_r + Y_s)\right\|_2^2 + 2\left\|\frac{1}{2}(Y_r - Y_s)\right\|_2^2.$$

The first term on the right-hand side is $\geq 2\Delta^2$ by definition of Δ, so taking limits $r, s \to +\infty$ we have what we need, that is, that (Y_n) is indeed Cauchy.

Let Y be the limit of (Y_n) in $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Note that by the triangle inequality

$$\Delta \leq \|X - Y\|_2 \leq \|X - Y_n\|_2 + \|Y_n - Y\|_2 \to \Delta,$$

as $n \to +\infty$. As a result, for any $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ and $t \in \mathbb{R}$,

$$\|X - Y - tZ\|_2^2 \geq \Delta^2 = \|X - Y\|_2^2,$$

so that, expanding and rearranging, we have

$$-2t\langle Z, X - Y\rangle + t^2\|Z\|_2^2 \geq 0,$$

which is only possible *for every* $t \in \mathbb{R}$ if the first term is 0.

Uniqueness follows from the parallelogram law and the definition of Δ. ■

We return to the proof of existence of the conditional expectation. We use the standard machinery.

Proof of Theorem B.6.1 (i). The previous theorem implies that conditional expectations exist for indicators and simple functions. Now take $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and write $X = X^+ - X^-$, so we can assume X is in fact nonnegative without loss of generality. Using the staircase function

$$X^{(r)} = \begin{cases} 0, & \text{if } X = 0 \\ (i-1)2^{-r}, & \text{if } (i-1)2^{-r} < X \leq i2^{-r} \leq r \\ r, & \text{if } X > r, \end{cases}$$

we have $0 \leq X^{(r)} \uparrow X$. Let $Y^{(r)} = \mathbb{E}[X^{(r)} \,|\, \mathcal{G}]$. Using an argument similar to the proof of uniqueness, it follows that $U \geq 0$ implies $\mathbb{E}[U \,|\, \mathcal{G}] \geq 0$ for a simple function U. Using linearity (which is immediate from the definition), we then have $Y^{(r)} \uparrow Y := \limsup Y^{(r)}$ which is measurable in $\mathcal{G}$. By (MON),

$$\mathbb{E}[Y; G] = \mathbb{E}[X; G], \ \forall G \in \mathcal{G}.$$

That concludes the proof. ■

Before deriving some properties, we give a few examples.

Example B.6.3. If $X \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ then $\mathbb{E}[X \,|\, \mathcal{G}] = X$ almost surely trivially. ◀

Example B.6.4. If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}[X]$. ◀

Example B.6.5. Let $A, B \in \mathcal{F}$ with $0 < \mathbb{P}[B] < 1$. If $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$ and $X = \mathbf{1}_A$, then

$$\mathbb{P}[A \,|\, \mathcal{G}] = \begin{cases} \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}, & \text{on } \omega \in B, \\ \frac{\mathbb{P}[A \cap B^c]}{\mathbb{P}[B^c]}, & \text{on } \omega \in B^c. \end{cases}$$

◀

Intuition about the conditional expectation sometimes breaks down.

Example B.6.6. On $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1], \mathcal{B}(0,1], \lambda)$, let $\mathcal{G}$ be the σ-algebra of all countable and co-countable (i.e., whose complement in $(0,1]$ is countable) subsets of $(0,1]$. Then $\mathbb{P}[G] \in \{0,1\}$ for all $G \in \mathcal{G}$ and

$$\mathbb{E}[X; G] = \mathbb{E}[\mathbb{E}[X]; G] = \mathbb{E}[X]\mathbb{P}[G],$$

so that $\mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}[X]$. Yet, $\mathcal{G}$ contains all singletons and we seemingly have "full information," which would lead to the wrong guess $\mathbb{E}[X \,|\, \mathcal{G}] = X$. ◀

We show that the conditional expectation behaves similarly to the ordinary expectation. Below all X and X_is are in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$.

Lemma B.6.7 (cLIN)**.** *If* $a_1, a_2 \in \mathbb{R}$*, then* $\mathbb{E}[a_1 X_1 + a_2 X_2 \,|\, \mathcal{G}] = a_1 \mathbb{E}[X_1 \,|\, \mathcal{G}] + a_2 \mathbb{E}[X_2 \,|\, \mathcal{G}]$ *a.s.*

Proof. Use the linearity of expectation and the fact that a linear combination of random variables in $\mathcal{G}$ is also in $\mathcal{G}$. ■

Lemma B.6.8 (cPOS)**.** *If* $X \geq 0$ *then* $\mathbb{E}[X \,|\, \mathcal{G}] \geq 0$ *a.s.*

Proof. Let $Y = \mathbb{E}[X \,|\, \mathcal{G}]$ and assume for contradiction that $\mathbb{P}[Y < 0] > 0$. There is $n \geq 1$ such that $\mathbb{P}[Y < -n^{-1}] > 0$. But that implies, for $G = \{Y < -n^{-1}\}$,

$$\mathbb{E}[X; G] = \mathbb{E}[Y; G] < -n^{-1}\mathbb{P}[G] < 0,$$

a contradiction. ■

Lemma B.6.9 (cMON)**.** *If* $0 \leq X_n \uparrow X$ *then* $\mathbb{E}[X_n \,|\, \mathcal{G}] \uparrow \mathbb{E}[X \,|\, \mathcal{G}]$ *a.s.*

Proof. Let $Y_n = \mathbb{E}[X_n \,|\, \mathcal{G}]$. By (cLIN) and (cPOS), $0 \leq Y_n \uparrow$. Then letting $Y = \limsup Y_n$, by (MON),

$$\mathbb{E}[X; G] = \mathbb{E}[Y; G],$$

for all $G \in \mathcal{G}$. ■

Lemma B.6.10 (cFATOU)**.** *If* $X_n \geq 0$ *then* $\mathbb{E}[\liminf X_n \,|\, \mathcal{G}] \leq \liminf \mathbb{E}[X_n \,|\, \mathcal{G}]$ *a.s.*

Proof. Note that, for $n \geq m$,

$$X_n \geq Z_m := \inf_{k \geq m} X_k \uparrow \in \mathcal{G},$$

so that $\inf_{n \geq m} \mathbb{E}[X_n \,|\, \mathcal{G}] \geq \mathbb{E}[Z_m \,|\, \mathcal{G}]$. Applying (cMON)

$$\mathbb{E}[\lim Z_m \,|\, \mathcal{G}] = \lim \mathbb{E}[Z_m \,|\, \mathcal{G}] \leq \lim \inf_{n \geq m} \mathbb{E}[X_n \,|\, \mathcal{G}].$$

■

Lemma B.6.11 (cDOM)**.** *If* $X_n \leq V \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *and* $X_n \to X$ *a.s., then*

$$\mathbb{E}[X_n \,|\, \mathcal{G}] \to \mathbb{E}[X \,|\, \mathcal{G}] \; a.s.$$

Proof. Applying (cFATOU) to $W_n := 2V - |X_n - X| \geq 0$,

$$\begin{aligned}\mathbb{E}[2V \,|\, \mathcal{G}] &= \mathbb{E}[\liminf_n W_n \,|\, \mathcal{G}] \\ &\leq \liminf_n \mathbb{E}[W_n \,|\, \mathcal{G}] \\ &= \mathbb{E}[2V \,|\, \mathcal{G}] - \liminf_n \mathbb{E}[|X_n - X| \,|\, \mathcal{G}],\end{aligned}$$

so we must have

$$\liminf_n \mathbb{E}[|X_n - X| \,|\, \mathcal{G}] = 0.$$

Now use that $|\mathbb{E}[X_n - X \,|\, \mathcal{G}]| \leq \mathbb{E}[|X_n - X| \,|\, \mathcal{G}]$ (which follows from (cPOS)). ■

Lemma B.6.12 (cJENSEN)**.** *If f is convex and $\mathbb{E}[|f(X)|] < +\infty$ then*

$$f(\mathbb{E}[X \,|\, \mathcal{G}]) \leq \mathbb{E}[f(X) \,|\, \mathcal{G}].$$

In addition, we highlight (without proof) the following important properties of the conditional expectation.

Lemma B.6.13 (Taking out what is known)**.** *If $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $Z \in \mathrm{m}\mathcal{G}$ is bounded or if X is bounded and $Z \in L^1(\Omega, \mathcal{G}, \mathbb{P})$, then $\mathbb{E}[ZX \,|\, \mathcal{G}] = Z\,\mathbb{E}[X \,|\, \mathcal{G}]$. This is also true if $X, Z \geq 0$, $\mathbb{E}[X] < +\infty$ and $\mathbb{E}[ZX] < +\infty$, or $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$.*

Lemma B.6.14 (Role of independence)**.** *If $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ is independent of $\mathcal{H}$ then $\mathbb{E}[X \,|\, \mathcal{H}] = \mathbb{E}[X]$. In fact, if $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then $\mathbb{E}[X \,|\, \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \,|\, \mathcal{G}]$.*

Lemma B.6.15 (Conditioning on an independent random variable)**.** *Suppose X, Y are independent. Let ϕ be a function with $\mathbb{E}|\phi(X, Y)| < +\infty$ and let $g(x) = \mathbb{E}(\phi(x, Y))$. Then,*

$$\mathbb{E}(\phi(X, Y)|X) = g(X).$$

Lemma B.6.16 (Tower property)**.** *If $\mathcal{H} \subseteq \mathcal{G}$ is a σ-algebra and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$* *tower property*

$$\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{H}] = \mathbb{E}[\mathbb{E}[X \,|\, \mathcal{H}] \,|\, \mathcal{G}] = \mathbb{E}[X \,|\, \mathcal{H}].$$

That is, the "smallest σ-algebra wins."

An important special case of the latter, also known as the law of total probability or the law of total expectation, is $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]] = \mathbb{E}[X]$.

One last useful property:

Lemma B.6.17. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $Y_1 = Y_2$ a.s. on $B \in \mathcal{F}$ then $\mathbb{E}[Y_1 \,|\, \mathcal{F}] = \mathbb{E}[Y_2 \,|\, \mathcal{F}]$ a.s. on B.*

B.7 Filtered spaces

Finally we define stochastic processes. Let E be a set and let $\mathcal{E}$ be a σ-algebra defined over E.

Definition B.7.1. *A* stochastic process *(or* process*) is a collection* $\{X_t\}_{t\in\mathscr{T}}$ *of* $(E,\mathcal{E})$*-valued random variables on a probability space* $(\Omega,\mathcal{F},\mathbb{P})$*, where* $\mathscr{T}$ *is an arbitrary* index set*.* *process*

Here is a typical example.

Example B.7.2. When $\mathscr{T} = \mathbb{Z}_+$ (or $\mathscr{T} = \mathbb{N}$ or $\mathscr{T} = \mathbb{Z}$) we have a *discrete-time process*, in which case we often write the process as a sequence $(X_t)_{t\geq 0}$. For instance:

- $X_0, X_1, X_2, \ldots$ i.i.d. random variables;
- $(S_t)_{t\geq 0}$ where $S_t = \sum_{i\leq t} X_i$ with X_i as above.

We let

$$\mathcal{F}_t = \sigma(X_0, X_1, \ldots, X_t),$$

which can be thought of as "the information known up to time t." For a fixed $\omega \in \Omega$, $(X_t(\omega) : t \in \mathscr{T})$ is called a *sample path.* ◀ *sample path*

Definition B.7.3. *A* random walk on $\mathbb{R}^d$ *is a process of the form:*

$$S_t = S_0 + \sum_{i=1}^{t} X_i, \quad t \geq 1$$

where the X_i*s are i.i.d. in* $\mathbb{R}^d$*, independent of* S_0*. The case* X_i *uniform in* $\{-1,+1\}$ *is called* simple random walk on $\mathbb{Z}$*.*

Filtered spaces provide a formal framework for time-indexed processes. We restrict ourselves to discrete time. (We will not discuss continuous-time processes in this book.)

Definition B.7.4. *A* filtered space *is a tuple* $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in\mathbb{Z}_+},\mathbb{P})$ *where:*

- $(\Omega,\mathcal{F},\mathbb{P})$ *is a probability space;*
- $(\mathcal{F}_t)_{t\in\mathbb{Z}_+}$ *is a* filtration*, that is,* *filtration*

  $$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_\infty := \sigma(\cup_t \mathcal{F}_t) \subseteq \mathcal{F}.$$

  where each $\mathcal{F}_i$ *is a* σ*-algebra.*

Definition B.7.5. *Fix* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{Z}_+}, \mathbb{P})$. *A process* $(W_t)_{t \geq 0}$ *is* adapted *if* $W_t \in \mathcal{F}_t$ *for all* t. *adapted*

Intuitively, in the previous definition, the value of W_t is "known at time t."

Definition B.7.6. *A process* $(C_t)_{t \geq 1}$ *is* predictable *if* $C_t \in \mathcal{F}_{t-1}$ *for all* $t \geq 1$. *predictable*

Example B.7.7. Continuing Example B.7.2. The collection $(\mathcal{F}_t)_{t \geq 0}$ forms a filtration. The process $(S_t)_{t \geq 0}$ is adapted. On the other hand, the process $C_t = \mathbf{1}\{S_{t-1} \leq k\}$ is predictable. ◀

Bibliography

[Abb18] Emmanuel Abbe. Community Detection and Stochastic Block Models. *Foundations and Trends® in Communications and Information Theory*, 14(1-2):1–162, June 2018. Publisher: Now Publishers, Inc.

[ABH16] Emmanuel Abbe, Afonso S. Bandeira, and Georgina Hall. Exact Recovery in the Stochastic Block Model. *IEEE Transactions on Information Theory*, 62(1):471–487, January 2016. Conference Name: IEEE Transactions on Information Theory.

[Ach03] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, 2003.

[AD78] Rudolf Ahlswede and David E. Daykin. An inequality for the weights of two families of sets, their unions and intersections. *Z. Wahrsch. Verw. Gebiete*, 43(3):183–185, 1978.

[AF] David Aldous and James Allen Fill. Reversible Markov chains and random walks on graphs. http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[AGG89] R. Arratia, L. Goldstein, and L. Gordon. Two Moments Suffice for Poisson Approximations: The Chen-Stein Method. *Annals of Probability*, 17(1):9–25, January 1989. Publisher: Institute of Mathematical Statistics.

[AJKS22] Alekh Agarwal, Nan Jiang, Sham M. Kakade, and Wen Sun. Reinforcement learning: Theory and algorithms. Available at https://rltheorybook.github.io/rltheorybook_AJKS.pdf, 2022.

[AK97] Noga Alon and Michael Krivelevich. The concentration of the chromatic number of random graphs. *Combinatorica*, 17(3):303–313, 1997.

[Ald83] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.

[Ald90] David J. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.*, 3(4):450–465, 1990.

[Ald97] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997.

[Alo03] Noga Alon. Problems and results in extremal combinatorics. I. *Discrete Math.*, 273(1-3):31–53, 2003. EuroComb'01 (Barcelona).

[AMS09] Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902, April 2009.

[AN04] K. B. Athreya and P. E. Ney. *Branching processes*. Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040].

[ANP05] Dimitris Achlioptas, Assaf Naor, and Yuval Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435:759–764, 2005.

[AS11] N. Alon and J.H. Spencer. *The Probabilistic Method*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011.

[AS15] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 670–688. IEEE Computer Society, 2015.

[Axl15] Sheldon Axler. *Linear algebra done right*. Undergraduate Texts in Mathematics. Springer, Cham, third edition, 2015.

[AZ18] Martin Aigner and Günter M. Ziegler. *Proofs from The Book*. Springer, Berlin, sixth edition, 2018. See corrected reprint of the 1998 original [MR1723092], Including illustrations by Karl H. Hofmann.

[Azu67] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, 19:357–367, 1967.

[BA99] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[BBFdlV00] D. Barraez, S. Boucheron, and W. Fernandez de la Vega. On the fluctuations of the giant component. *Combin. Probab. Comput.*, 9(4):287–304, 2000.

[BC03] Bo Brinkman and Moses Charikar. On the Impossibility of Dimension Reduction in L1. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[BCB12] Sébastien Bubeck and Nicolò Cesa-Bianchi. *Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems*. Now Publishers, 2012. Google-Books-ID: Rl2skwEACAAJ.

[BCMR06] Christian Borgs, Jennifer T. Chayes, Elchanan Mossel, and Sébastien Roch. The Kesten-Stigum reconstruction bound is tight for roughly symmetric binary channels. In *FOCS*, pages 518–530, 2006.

[BD97] Russ Bubley and Martin E. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997*, pages 223–231. IEEE Computer Society, 1997.

[BDDW08] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[BDJ99] Jinho Baik, Percy Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

[Ber46] S.N. Bernstein. *Probability Theory (in russian)*. M.-L. Gostechizdat, 1946.

[Ber14] Nathanaël Berestycki. Lectures on mixing times: A crossroad between probability, analysis and geometry. Available at https://homepage.univie.ac.at/nathanael.berestycki/wp-content/uploads/2022/05/mixing3.pdf, 2014.

[BH57] S. R. Broadbent and J. M. Hammersley. Percolation processes. I. Crystals and mazes. *Proc. Cambridge Philos. Soc.*, 53:629–641, 1957.

[BH16] Afonso S. Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, July 2016. Publisher: Institute of Mathematical Statistics.

[Bil12] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 2012.

[BKW14] Itai Benjamini, Gady Kozma, and Nicholas Wormald. The mixing time of the giant component of a random graph. *Random Structures Algorithms*, 45(3):383–407, 2014.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[Bol81] Béla Bollobás. Random graphs. In *Combinatorics (Swansea, 1981)*, volume 52 of *London Math. Soc. Lecture Note Ser.*, pages 80–102. Cambridge Univ. Press, Cambridge-New York, 1981.

[Bol84] Béla Bollobás. The evolution of random graphs. *Trans. Amer. Math. Soc.*, 286(1):257–274, 1984.

[Bol98] Béla Bollobás. *Modern graph theory*, volume 184 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

[Bol01] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[BR06a] Béla Bollobás and Oliver Riordan. *Percolation*. Cambridge University Press, New York, 2006.

[BR06b] Béla Bollobás and Oliver Riordan. A short proof of the Harris-Kesten theorem. *Bull. London Math. Soc.*, 38(3):470–484, 2006.

[Bre17] Pierre Bremaud. *Discrete probability models and methods*, volume 78 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017. Probability on graphs and trees, Markov chains and random fields, entropy and coding.

[Bre20] Pierre Bremaud. *Markov chains—Gibbs fields, Monte Carlo simulation and queues*, volume 31 of *Texts in Applied Mathematics*. Springer, Cham, 2020. Second edition [of 1689633].

[Bro89] Andrei Z. Broder. Generating random spanning trees. In *FOCS*, pages 442–447. IEEE Computer Society, 1989.

[BRST01] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001.

[BS89] Ravi Boppona and Joel Spencer. A useful elementary correlation inequality. *J. Combin. Theory Ser. A*, 50(2):305–307, 1989.

[Bub10] Sébastien Bubeck. *Bandits Games and Clustering Foundations*. phdthesis, Université des Sciences et Technologie de Lille - Lille I, June 2010.

[BV04] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Berichte über verteilte messysteme. Cambridge University Press, 2004.

[Car85] Thomas Keith Carne. A transmutation formula for Markov chains. *Bull. Sci. Math. (2)*, 109(4):399–405, 1985.

[CDL$^+$12] P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly. Glauber dynamics for the mean-field Potts model. *J. Stat. Phys.*, 149(3):432–477, 2012.

[CF07] Colin Cooper and Alan Frieze. The cover time of sparse random graphs. *Random Structures Algorithms*, 30(1-2):1–16, 2007.

[Che52] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952.

[Che70] Jeff Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Sympos. in honor of Salomon Bochner, Princeton Univ., Princeton, N.J., 1969)*, pages 195–199. Princeton Univ. Press, Princeton, N.J., 1970.

[Che75] Louis H. Y. Chen. Poisson Approximation for Dependent Trials. *The Annals of Probability*, 3(3):534–545, 1975. Publisher: Institute of Mathematical Statistics.

[Chu97] Fan R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1997.

[CL06] Fan Chung and Linyuan Lu. *Complex graphs and networks*, volume 107 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006.

[CR92] V. Chvatal and B. Reed. Mick gets some (the odds are on his side) [satisfiability]. In *Foundations of Computer Science, 1992. Proceedings., 33rd Annual Symposium on*, pages 620–627, Oct 1992.

[Cra38] H. Cramér. Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles*, 736:5–23, 1938.

[CRR+89] Ashok K. Chandra, Prabhakar Raghavan, Walter L. Ruzzo, Roman Smolensky, and Prasoon Tiwari. The electrical resistance of a graph captures its commute and cover times (detailed abstract). In David S. Johnson, editor, *STOC*, pages 574–586. ACM, 1989.

[CRT06a] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[CRT06b] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[CT05] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[CW08] E.J. Candès and M.B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, March 2008.

[Dev98] Luc Devroye. Branching processes and their applications in the analysis of tree structures and tree algorithms. In Michel Habib, Colin McDiarmid, Jorge Ramirez-Alfonsin, and Bruce Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16 of *Algorithms and Combinatorics*, pages 249–314. Springer Berlin Heidelberg, 1998.

[Dey] Partha Dey. Lecture notes on "Stein-Chen method for Poisson approximation". https://faculty.math.illinois.edu/~psdey/414CourseNotes.pdf.

[DGG+00] Martin Dyer, Leslie Ann Goldberg, Catherine Greenhill, Mark Jerrum, and Michael Mitzenmacher. An extension of path coupling and its application to the Glauber dynamics for graph colourings (extended abstract). In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms (San Francisco, CA, 2000)*, pages 616–624. ACM, New York, 2000.

[dH] Frank den Hollander. Probability theory: The coupling method, 2012. http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf.

[Dia88] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988.

[Dia09] Persi Diaconis. The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):179–205, 2009.

[Die10] Reinhard Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, Heidelberg, fourth edition, 2010.

[DKLP11] Jian Ding, Jeong Han Kim, Eyal Lubetzky, and Yuval Peres. Anatomy of a young giant component in the random graph. *Random Structures Algorithms*, 39(2):139–178, 2011.

[DMR11] Constantinos Daskalakis, Elchanan Mossel, and Sébastien Roch. Evolutionary trees and the ising model on the bethe lattice: a proof of steel's conjecture. *Probability Theory and Related Fields*, 149:149–189, 2011. 10.1007/s00440-009-0246-2.

[DMS00] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Phys. Rev. Lett.*, 85:4633–4636, Nov 2000.

[Doe38] Wolfgang Doeblin. Exposé de la théorie des chaînes simples constantes de markoff à un nombre fini d'états. *Rev. Math. Union Interbalkan*, 2:77–105, 1938.

[Don06] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[Doo01] J.L. Doob. *Classical Potential Theory and Its Probabilistic Counterpart*. Classics in Mathematics. Springer Berlin Heidelberg, 2001.

[DP11] Jian Ding and Yuval Peres. Mixing time for the Ising model: a uniform lower bound for all graphs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):1020–1028, 2011.

[DS84] P.G. Doyle and J.L. Snell. *Random walks and electric networks*. Carus mathematical monographs. Mathematical Association of America, 1984.

[DS91] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[DSC93a] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.*, 21(4):2131–2156, 1993.

[DSC93b] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3(3):696–730, 1993.

[DSS22] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. *Ann. of Math. (2)*, 196(1):1–388, 2022.

[Dur85] Richard Durrett. Some general results concerning the critical exponents of percolation processes. *Z. Wahrsch. Verw. Gebiete*, 69(3):421–437, 1985.

[Dur06] R. Durrett. *Random Graph Dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2006.

[Dur10] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010.

[Dur12] Richard Durrett. *Essentials of stochastic processes*. Springer Texts in Statistics. Springer, New York, second edition, 2012.

[DZ10] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[Ebe] Andreas Eberle. Markov Processes. 2021. https://uni-bonn.sciebo.de/s/kzTUFff5FrWGAay.

[EKPS00] W. S. Evans, C. Kenyon, Y. Peres, and L. J. Schulman. Broadcasting on trees and the Ising model. *Ann. Appl. Probab.*, 10(2):410–433, 2000.

[ER59] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[ER60] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[Fel71] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[FK16] Alan Frieze and MichałKaroński. *Introduction to random graphs*. Cambridge University Press, Cambridge, 2016.

[FKG71] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[FM88] P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44(3):355–362, 1988.

[Fos53] F. G. Foster. On the stochastic matrices associated with certain queuing processes. *Ann. Math. Statistics*, 24:355–360, 1953.

[FR98] AlanM. Frieze and Bruce Reed. Probabilistic analysis of algorithms. In Michel Habib, Colin McDiarmid, Jorge Ramirez-Alfonsin, and Bruce Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16 of *Algorithms and Combinatorics*, pages 36–92. Springer Berlin Heidelberg, 1998.

[FR08] N. Fountoulakis and B. A. Reed. The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures Algorithms*, 33(1):68–86, 2008.

[FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser Basel, 2013.

[FV18] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems*. Cambridge University Press, Cambridge, 2018. A concrete mathematical introduction.

[GC11] Aurélien Garivier and Olivier Cappé. The KL-UCB Algorithm for Bounded Stochastic Bandits and Beyond. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 359–376. JMLR Workshop and Conference Proceedings, December 2011. ISSN: 1938-7228.

[GCS+14] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian data analysis*. Texts in Statistical Science Series. CRC Press, Boca Raton, FL, third edition, 2014.

[Gil59] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30:1141–1144, 1959.

[GL06] Dani Gamerman and Hedibert Freitas Lopes. *Markov chain Monte Carlo*. Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2006. Stochastic simulation for Bayesian inference.

[Gri97] Geoffrey Grimmett. Percolation and disordered systems. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 153–300. Springer, Berlin, 1997.

[Gri10a] Geoffrey Grimmett. *Probability on graphs*, volume 1 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, Cambridge, 2010. Random processes on graphs and lattices.

[Gri10b] G.R. Grimmett. *Percolation*. Grundlehren der mathematischen Wissenschaften. Springer, 2010.

[Gri75] David Griffeath. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:95–106, 1974/75.

[GS20] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and random processes*. Oxford University Press, Oxford, 2020. Fourth edition [of 0667520].

[Ham57] J. M. Hammersley. Percolation processes. II. The connective constant. *Proc. Cambridge Philos. Soc.*, 53:642–645, 1957.

[Har] Nicholas Harvey. Lecture notes for CPSC 536N: Randomized Algorithms. http://www.cs.ubc.ca/~nickhar/W12/.

[Har60] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[Har63] Theodore E. Harris. *The theory of branching processes*. Die Grundlehren der mathematischen Wissenschaften, Band 119. Springer-Verlag, Berlin; Prentice Hall, Inc., Englewood Cliffs, N.J., 1963.

[Haz16] Elad Hazan. Introduction to Online Convex Optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, August 2016. Publisher: Now Publishers, Inc.

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.

[HMRAR98] M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors. *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1998.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[HS07] Thomas P. Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *Ann. Appl. Probab.*, 17(3):931–952, 2007.

[IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In Jeffrey Scott Vitter, editor, *STOC*, pages 604–613. ACM, 1998.

[Jan90] Svante Janson. Poisson approximation for large deviations. *Random Structures Algorithms*, 1(2):221–229, 1990.

[JH01] Galin L. Jones and James P. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.*, 16(4):312–334, 2001.

[JL84] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[JLR11] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011.

[JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.

[Kan86] Masahiko Kanai. Rough isometries and the parabolicity of Riemannian manifolds. *J. Math. Soc. Japan*, 38(2):227–238, 1986.

[Kar90] Richard M. Karp. The transitive closure of a random digraph. *Random Structures Algorithms*, 1(1):73–93, 1990.

[Kes80] Harry Kesten. The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Comm. Math. Phys.*, 74(1):41–59, 1980.

[Kes82] Harry Kesten. *Percolation theory for mathematicians*, volume 2 of *Progress in Probability and Statistics*. Birkhäuser, Boston, Mass., 1982.

[KP] Júlia Komjáthy and Yuval Peres. Lecture notes for Markov chains: mixing times, hitting times, and cover times, 2012. Saint-Petersburg Summer School. http://www.win.tue.nl/~jkomjath/SPBlecturenotes.pdf.

[KRS] Michael J. Kozdron, Larissa M. Richards, and Daniel W. Stroock. Determinants, their applications to markov processes, and a random walk proof of kirchhoff's matrix tree theorem, 2013. Preprint available at http://arxiv.org/abs/1306.2059.

[KS66a] H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Statist.*, 37:1463–1481, 1966.

[KS66b] H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. *Ann. Math. Statist.*, 37:1211–1223, 1966.

[KS67] H. Kesten and B. P. Stigum. Limit theorems for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.*, 17:309–338, 1967.

[KS05] Gil Kalai and Shmuel Safra. Threshold Phenomena and Influence: Perspectives from Mathematics, Computer Science, and Economics. In *Computational Complexity and Statistical Physics*. Oxford University Press, December 2005. _eprint: https://academic.oup.com/book/0/chapter/354512033/chapter-pdf/43716844/isbn-9780195177374-book-part-8.pdf.

[KSK76] John G. Kemeny, J. Laurie Snell, and Anthony W. Knapp. *Denumerable Markov chains*. Springer-Verlag, New York-Heidelberg-Berlin, second edition, 1976. With a chapter on Markov random fields, by David Griffeath, Graduate Texts in Mathematics, No. 40.

[Law05] Gregory F. Lawler. *Conformally invariant processes in the plane*, volume 114 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2005.

[Law06] Gregory F. Lawler. *Introduction to stochastic processes*. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2006.

[Led01] M. Ledoux. *The Concentration of Measure Phenomenon*. Mathematical surveys and monographs. American Mathematical Society, 2001.

[Lin02] Torgny Lindvall. *Lectures on the coupling method*. Dover Publications, Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

[LL10] G.F. Lawler and V. Limic. *Random Walk: A Modern Introduction*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010.

[Lov83] L. Lovász. Submodular functions and convexity. In *Mathematical programming: the state of the art (Bonn, 1982)*, pages 235–257. Springer, Berlin, 1983.

[Lov12] László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[LP16] Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016.

[LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[LPW06] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

[LR85] T.L Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, March 1985.

[LS88] Gregory F. Lawler and Alan D. Sokal. Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, 309(2):557–580, 1988.

[LS12] Eyal Lubetzky and Allan Sly. Critical Ising on the square lattice mixes in polynomial time. *Comm. Math. Phys.*, 313(3):815–836, 2012.

[LS20] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[Lu90] Tomasz Ł uczak. Component behavior near the critical point of the random graph process. *Random Structures Algorithms*, 1(3):287–310, 1990.

[Lug] Gabor Lugosi. Concentration-of-measure inequalities, 2004. Available at http://www.econ.upf.edu/~lugosi/anu.pdf.

[LuPW94] Tomasz Ł uczak, Boris Pittel, and John C. Wierman. The structure of a random graph at the point of the phase transition. *Trans. Amer. Math. Soc.*, 341(2):721–748, 1994.

[Lyo83] Terry Lyons. A simple criterion for transience of a reversible Markov chain. *Ann. Probab.*, 11(2):393–402, 1983.

[Lyo90] Russell Lyons. Random walks and percolation on trees. *Ann. Probab.*, 18(3):931–958, 1990.

[Mat88] Peter Matthews. Covering problems for Markov chains. *Ann. Probab.*, 16(3):1215–1228, 1988.

[Mau79] Bernard Maurey. Construction de suites symétriques. *C. R. Acad. Sci. Paris Sér. A-B*, 288(14):A679–A681, 1979.

[McD89] Colin McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[ML86] Anders Martin-Löf. Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J. Appl. Probab.*, 23(2):265–282, 1986.

[ML98] Anders Martin-Löf. The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *J. Appl. Probab.*, 35(3):671–682, 1998.

[MNS15a] Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for the planted bisection model. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 69–75. ACM, 2015.

[MNS15b] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3-4):431–461, 2015.

[Mor05] Frank Morgan. *Real analysis*. American Mathematical Society, Providence, RI, 2005.

[Mos01] E. Mossel. Reconstruction on trees: beating the second eigenvalue. *Ann. Appl. Probab.*, 11(1):285–300, 2001.

[Mos03] E. Mossel. On the impossibility of reconstructing ancestral data and phylogenies. *J. Comput. Biol.*, 10(5):669–678, 2003.

[Mos04] E. Mossel. Phase transitions in phylogeny. *Trans. Amer. Math. Soc.*, 356(6):2379–2404, 2004.

[MP03] E. Mossel and Y. Peres. Information flow on trees. *Ann. Appl. Probab.*, 13(3):817–844, 2003.

[MR95] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, 1995.

[MS86] Vitali D. Milman and Gideon Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[MT06] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006.

[MT09] Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, New York, NY, USA, 2005.

[Nic18] Bogdan Nica. *A brief introduction to spectral graph theory*. EMS Textbooks in Mathematics. European Mathematical Society (EMS), Zürich, 2018.

[Nor98] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.

[NP10] Asaf Nachmias and Yuval Peres. The critical random graph, with martingales. *Israel J. Math.*, 176:29–41, 2010.

[NW59] C. St. J. A. Nash-Williams. Random walk and electric currents in networks. *Proc. Cambridge Philos. Soc.*, 55:181–194, 1959.

[Ott49] Richard Otter. The multiplicative process. *Ann. Math. Statistics*, 20:206–224, 1949.

[Pem91] Robin Pemantle. Choosing a spanning tree for the integer lattice uniformly. *Ann. Probab.*, 19(4):1559–1574, 1991.

[Pem00] Robin Pemantle. Towards a theory of negative dependence. *J. Math. Phys.*, 41(3):1371–1390, 2000. Probabilistic techniques in equilibrium and nonequilibrium statistical physics.

[Per] Yuval Peres. Course notes on Probability on trees and networks, 2004. http://stat-www.berkeley.edu/~peres/notes1.pdf.

[Per09] Yuval Peres. The unreasonable effectiveness of martingales. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, pages 997–1000, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

[Pet] Gábor Pete. Probability and geometry on groups. Lecture notes for a graduate course. http://www.math.bme.hu/~gabor/PGG.html.

[Pey08] Rémi Peyre. A probabilistic approach to Carne's bound. *Potential Anal.*, 29(1):17–36, 2008.

[Pit76] J. W. Pitman. On coupling of Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(4):315–322, 1976.

[Pit90] Boris Pittel. On tree census and the giant component in sparse random graphs. *Random Structures Algorithms*, 1(3):311–342, 1990.

[RAS15] Firas Rassoul-Agha and Timo Seppäläinen. *A course on large deviations with an introduction to Gibbs measures*, volume 162 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015.

[RC04] Christian P. Robert and George Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.

[Res92] Sidney Resnick. *Adventures in stochastic processes*. Birkhäuser Boston, Inc., Boston, MA, 1992.

[Roc10] Sebastien Roch. Toward Extracting All Phylogenetic Information from Matrices of Evolutionary Distances. *Science*, 327(5971):1376–1379, 2010.

[Rom15] Dan Romik. *The surprising mathematics of longest increasing subsequences*, volume 4 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, New York, 2015.

[RS17] Sebastien Roch and Allan Sly. Phase transition in the sample complexity of likelihood-based phylogeny inference. *Probability Theory and Related Fields*, 169(1):3–62, Oct 2017.

[RT87] WanSoo T. Rhee and Michel Talagrand. Martingale inequalities and NP-complete problems. *Math. Oper. Res.*, 12(1):177–181, 1987.

[Rud73] Walter Rudin. *Functional analysis*. McGraw-Hill Series in Higher Mathematics. McGraw-Hill Book Co., New York-Düsseldorf-Johannesburg, 1973.

[Rus78] Lucio Russo. A note on percolation. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 43(1):39–48, 1978.

[SE64] M. F. Sykes and J. W. Essam. Exact critical percolation probabilities for site and bond problems in two dimensions. *J. Mathematical Phys.*, 5:1117–1127, 1964.

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.*, 82(1):93–133, 1989.

[Spi56] Frank Spitzer. A combinatorial lemma and its application to probability theory. *Trans. Amer. Math. Soc.*, 82:323–339, 1956.

[Spi12] Daniel A. Spielman. Lecture notes on spectral graph theory. `https://www.cs.yale.edu/homes/spielman/561/2012/index.html`, 2012.

[SS87] E. Shamir and J. Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, 7(1):121–129, 1987.

[SS03] Elias M. Stein and Rami Shakarchi. *Fourier analysis*, volume 1 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2003. An introduction.

[SS05] Elias M. Stein and Rami Shakarchi. *Real analysis*, volume 3 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2005. Measure theory, integration, and Hilbert spaces.

[SSBD14] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014.

[Ste] J. E. Steif. A mini course on percolation theory, 2009. http://www.math.chalmers.se/~steif/perc.pdf.

[Ste72] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, 1972.

[Ste97] J. Michael Steele. *Probability theory and combinatorial optimization*, volume 69 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[Ste98] G. W. Stewart. *Matrix algorithms. Vol. I*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1998. Basic decompositions.

[Str65] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.

[Str14] Daniel W. Stroock. Doeblin's Theory for Markov Chains. In *An Introduction to Markov Processes*, pages 25–47. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[SW78] P. D. Seymour and D. J. A. Welsh. Percolation probabilities on the square lattice. *Ann. Discrete Math.*, 3:227–245, 1978. Advances in graph theory (Cambridge Combinatorial Conf., Trinity College, Cambridge, 1977).

[Tao] Terence Tao. Open question: deterministic UUP matrices. https://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices/.

[Var85] Nicholas Th. Varopoulos. Long range estimates for Markov chains. *Bull. Sci. Math. (2)*, 109(3):225–252, 1985.

[vdH10] Remco van der Hofstad. Percolation and random graphs. In *New perspectives in stochastic geometry*, pages 173–247. Oxford Univ. Press, Oxford, 2010.

[vdH17] Remco van der Hofstad. *Random graphs and complex networks. Vol. 1*. Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017.

[vdHK08] Remco van der Hofstad and Michael Keane. An elementary proof of the hitting time theorem. *Amer. Math. Monthly*, 115(8):753–756, 2008.

[Vem04] Santosh S. Vempala. *The random projection method*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 65. American Mathematical Society, Providence, RI, 2004. With a foreword by Christos H. Papadimitriou.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, September 2018. Google-Books-ID: TahxDwAAQBAJ.

[vH16] Ramon van Handel. Probability in high dimension. http://www.princeton.edu/~rvan/APC550.pdf, 2016.

[Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[Wen75] J. G. Wendel. Left-continuous random walk and the Lagrange expansion. *Amer. Math. Monthly*, 82:494–499, 1975.

[Whi32] Hassler Whitney. Non-separable and planar graphs. *Trans. Amer. Math. Soc.*, 34(2):339–362, 1932.

[Wil91] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

[Wil96] David Bruce Wilson. Generating random spanning trees more quickly than the cover time. In Gary L. Miller, editor, *STOC*, pages 296–303. ACM, 1996.

[YP14] Se-Young Yun and Alexandre Proutière. Community detection via random and adaptive sampling. In Maria-Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 138–175. JMLR.org, 2014.

[Yur76] V. V. Yurinskiĭ. Exponential inequalities for sums of random vectors. *J. Multivariate Anal.*, 6(4):473–499, 1976.

Index