1

Lévy processes

Summary Section 1.1 is a review of basic measure and probability theory. In Section 1.2, we meet the key concepts of the infinite divisibility of random variables and of probability distributions, which underly the whole subject. Important examples are the Gaussian, Poisson and stable distributions. The celebrated Lévy–Khintchine formula classifies the set of all infinitely divisible probability distributions by means of a canonical form for the characteristic function. Lévy processes are introduced in Section 1.3. These are essentially stochastic processes with stationary and independent increments. Each random variable within the process is infinitely divisible, and hence its distribution is determined by the Lévy–Khintchine formula. Important examples are Brownian motion, Poisson and compound Poisson processes, stable processes and subordinators. Section 1.4 clarifies the relationship between Lévy processes, infinite divisibility and weakly continuous convolution semigroups of probability measures. Finally, in Section 1.5, we briefly survey recurrence and transience, Wiener–Hopf factorisation, local times for Lévy processes, regular variation and subexponentiality.

1.1 Review of measure and probability

The aim of this section is to give a brief resumé of key notions of measure theory and probability that will be used extensively throughout the book and to fix some notation and terminology once and for all. I emphasise that reading this section is no substitute for a systematic study of the fundamentals from books, such as Billingsley [48], Itô [177], Ash and Doléans-Dade [17], Rosenthal [311], Dudley [98] or, for measure theory without probability, Cohn [80]. Knowledgeable readers are encouraged to skip this section altogether or to use it as a quick reference when the need arises.

1.1.1 Measure and probability spaces

Let S be a non-empty set and $\mathcal{F}$ a collection of subsets of S. We call $\mathcal{F}$ a *σ-algebra* if the following hold:

(1) $S \in \mathcal{F}$.
(2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.
(3) If $(A_n, n \in \mathbb{N})$ is a sequence of subsets in $\mathcal{F}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The pair $(S, \mathcal{F})$ is called a *measurable space*. A *measure* on $(S, \mathcal{F})$ is a mapping $\mu : \mathcal{F} \to [0, \infty]$ that satisfies

(1) $\mu(\emptyset) = 0$,
(2)

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

for every sequence $(A_n, n \in \mathbb{N})$ of mutually disjoint sets in $\mathcal{F}$.

The triple $(S, \mathcal{F}, \mu)$ is called a *measure space*.

The quantity $\mu(S)$ is called the *total mass* of μ and μ is said to be *finite* if $\mu(S) < \infty$. More generally, a measure μ is *σ-finite* if we can find a sequence $(A_n, n \in \mathbb{N})$ in $\mathcal{F}$ such that $S = \bigcup_{n=1}^{\infty} A_n$ and each $\mu(A_n) < \infty$.

For the purposes of this book, there will be two cases of interest. The first comprises

- **Borel measures** The *Borel σ-algebra of* $\mathbb{R}^d$ is the smallest σ-algebra of subsets of $\mathbb{R}^d$ that contains all the open sets. We denote it by $\mathcal{B}(\mathbb{R}^d)$. If $S \in \mathcal{B}(\mathbb{R}^d)$ we define its Borel σ-algebra to be

$$\mathcal{B}(S) = \{E \cap S; E \in \mathcal{B}(\mathbb{R}^d)\}.$$

Equivalently, $\mathcal{B}(S)$ is the smallest σ-algebra of subsets of S that contains every open set in S when S is equipped with the relative topology induced from $\mathbb{R}^d$, so that $U \subseteq S$ is open in S if $U \cap S$ is open in $\mathbb{R}^d$. Elements of $\mathcal{B}(S)$ are called *Borel sets* and any measure on $(S, \mathcal{B}(S))$ is called a *Borel measure*.

One of the best known examples of a Borel measure is given by the *Lebesgue measure* on $S = \mathbb{R}^d$. This takes the following explicit form on sets in the shape of boxes $A = (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d)$ where each $-\infty < a_i < b_i < \infty$:

$$\mu(A) = \prod_{i=1}^{d} (b_i - a_i).$$

Lebesgue measure is clearly σ-finite but not finite.

Of course, Borel measures make sense in arbitrary topological spaces, but we will not have need of this degree of generality here.

The second case comprises

- **Probability measures** Here we usually write $S = \Omega$ and take Ω to represent the set of outcomes of some random experiment. Elements of $\mathcal{F}$ are called *events* and any measure on $(\Omega, \mathcal{F})$ of total mass 1 is called a *probability measure* and denoted P. The triple $(\Omega, \mathcal{F}, P)$ is then called a *probability space*.

Occasionally we will also need *counting measures*, which are those that take values in $\mathbb{N} \cup \{0\}$.

A proposition p about the elements of S is said to hold *almost everywhere* (usually shortened to *a.e.*) with respect to a measure μ if $\mathcal{N} = \{s \in S; p(s) \text{is false}\} \in \mathcal{F}$ and $\mu(\mathcal{N}) = 0$. In the case of probability measures, we use the terminology '*almost surely*' (shortened to *a.s.*) instead of 'almost everywhere', or alternatively '*with probability* 1'. Similarly, we say that '*almost all*' the elements of a set A have a certain property if the subset of A for which the property fails has measure zero.

Continuity of measures Let $(A(n), n \in \mathbb{N})$ be a sequence of sets in $\mathcal{F}$ with $A(n) \subseteq A(n+1)$ for each $n \in \mathbb{N}$. We then write $A(n) \uparrow A$ where $A = \bigcup_{n=1}^{\infty} A(n)$, and we have

$$\mu(A) = \lim_{n \to \infty} \mu(A(n)).$$

When μ is a probability measure, this is usually called *continuity of probability*.

Let G be a group whose members act as measurable transformations of $(S, \mathcal{F})$, so that $g : S \to S$ for each $g \in G$ and $gA \in \mathcal{F}$ for all $A \in \mathcal{F}$, $g \in G$, where $gA = \{ga, a \in A\}$. We say that a measure μ on $(S, \mathcal{F})$ is G-*invariant* if

$$\mu(gA) = \mu(A)$$

for each $g \in G, A \in \mathcal{F}$.

A (finite) *measurable partition* of a set $A \in \mathcal{F}$ is a family of sets $B_1, B_2, \ldots, B_n \in \mathcal{F}$ for which $B_i \cap B_j = \emptyset$ whenever $i \neq j$ and $\bigcup_{i=1}^{n} B_i = A$. We use the term *Borel partition* when $\mathcal{F}$ is a Borel σ-algebra.

We say that a σ-algebra $\mathcal{G}$ is a *sub-σ-algebra* of $\mathcal{F}$ if $\mathcal{G} \subseteq \mathcal{F}$, i.e. $A \subseteq \mathcal{G} \Rightarrow A \subseteq \mathcal{F}$. If $\{\mathcal{G}_i, i \in I\}$ is a (not necessarily countable) family of sub-σ-algebras of $\mathcal{F}$ then $\bigcap_{i \in I} \mathcal{G}_i$ is the largest sub-σ-algebra contained in each $\mathcal{G}_i$ and $\bigvee_{i \in I} \mathcal{G}_i$ denotes the smallest sub-σ-algebra that contains each $\mathcal{G}_i$.

If P is a probability measure and $A, B \in \mathcal{F}$, it is sometimes notationally convenient to write $P(A, B) = P(A \cap B)$.

Completion of a measure Let $(S, \mathcal{F}, \mu)$ be a measure space. Define

$$\mathcal{N} = \{A \subseteq S; \exists N \in \mathcal{F} \text{ with } \mu(N) = 0 \text{ and } A \subseteq N\}$$

and

$$\overline{\mathcal{F}} = \{A \cup B; A \in \mathcal{F}, B \in \mathcal{N}\}.$$

Then $\overline{\mathcal{F}}$ is a σ-algebra and the *completion* of the measure μ on $(S, \mathcal{F})$ is the measure $\overline{\mu}$ on $(S, \overline{\mathcal{F}})$ defined by

$$\overline{\mu}(A \cup B) = \mu(A), \qquad A \in \mathcal{F}, \quad B \in \mathcal{N}.$$

In particular, $\overline{\mathcal{B}(S)}$ is called the σ-algebra of *Lebesgue measurable sets* in S.

π-systems and d-systems Let $\mathcal{C}$ be an arbitrary collection of subsets of S. We denote the smallest σ-algebra containing $\mathcal{C}$ by $\sigma(\mathcal{C})$, so $\sigma(\mathcal{C})$ is the intersection of all the σ-algebras which contain $\mathcal{C}$.

Sometimes we have to deal with collections of sets which do not form a σ-algebra but which still have enough structure to be useful. To this end we introduce π- and d-systems. A collection $\mathcal{H}$ of subsets of S is called a π-system if $A \cap B \in \mathcal{H}$ for all $A, B \in \mathcal{H}$.

A collection $\mathcal{D}$ of subsets of S is called a d-system if

(i) $S \in \mathcal{D}$,
(ii) If $A, B \in \mathcal{D}$ with $B \subseteq A$ then the set theoretic difference $A - B \in \mathcal{D}$,
(ii) If $(A_n, n \in \mathbb{N})$ is a sequence of subsets wherein $A_n \in \mathcal{D}$ and $A_n \subseteq A_{n+1}$ for each $n \in \mathbb{N}$, then $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}$.

If $\mathcal{C}$ is an arbitrary collection of subsets of S then we denote the smallest d-system containing $\mathcal{C}$ by $d(\mathcal{C})$, so $d(\mathcal{C})$ is the intersection of all the d-systems which contain $\mathcal{C}$.

The key result that we will need about π-systems and d-systems is the following.

Lemma 1.1.1 (Dynkin's lemma) *If $\mathcal{H}$ is a π-system then $d(\mathcal{H}) = \sigma(\mathcal{H})$.*

1.1.2 Random variables, integration and expectation

For $i = 1, 2$, let $(S_i, \mathcal{F}_i)$ be measurable spaces. A mapping $f : S_1 \to S_2$ is said to be $(\mathcal{F}_1, \mathcal{F}_2)$-*measurable* if $f^{-1}(A) \in \mathcal{F}_1$ for all $A \in \mathcal{F}_2$. If each $S_1 \subseteq \mathbb{R}^d$, $S_2 \subseteq \mathbb{R}^m$ and $\mathcal{F}_i = \mathcal{B}(S_i)$, f is said to be *Borel measurable*. In the case $d = 1$, we sometimes find it useful to write each Borel measurable f as $f^+ - f^-$ where, for each $x \in S_1$, $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = -\min\{f(x), 0\}$. If $f = (f_1, f_2, \ldots, f_d)$ is a measurable mapping from S_1 to $\mathbb{R}^d$, we write $f^+ = (f_1^+, f_2^+, \ldots, f_d^+)$ and $f^- = (f_1^-, f_2^-, \ldots, f_d^-)$.

In what follows, whenever we speak of measurable mappings taking values in a subset of $\mathbb{R}^d$, we always take it for granted that the latter is equipped with its Borel σ-algebra.

When we are given a probability space $(\Omega, \mathcal{F}, P)$ then measurable mappings from Ω into $\mathbb{R}^d$ are called *random variables*. Random variables are usually denoted $X, Y, \ldots$. Their values should be thought of as the results of quantitative observations on the set Ω. Note that if X is a random variable then so is $f(X) = f \circ X$, where f is a Borel measurable mapping from $\mathbb{R}^d$ to $\mathbb{R}^m$. A measurable mapping $Z = X + iY$ from Ω into $\mathbb{C}$ (equipped with the natural Borel structure inherited from $\mathbb{R}^2$) is called a *complex random variable*. Note that Z is measurable if and only if both X and Y are measurable.

If X is a random variable, its *law* (or *distribution*) is the Borel probability measure p_X on $\mathbb{R}^d$ defined by

$$p_X = P \circ X^{-1}.$$

We say that X is *symmetric* if $p_X(A) = p_X(-A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$.

Two random variables X and Y that have the same probability law are said to be *identically distributed*, and we sometimes denote this as $X \overset{\mathrm{d}}{=} Y$. For a one-dimensional random variable X, its *distribution function* is the right-continuous increasing function defined by $F_X(x) = p_X((-\infty, x])$ for each $x \in \mathbb{R}$.

If $W = (X, Y)$ is a random variable taking values in $\mathbb{R}^{2d}$, the probability law of W is sometimes called the *joint distribution* of X and Y. The quantities p_X and p_Y are then called the *marginal distributions* of W, where $p_X(A) = p_W(A, \mathbb{R}^d)$ and $p_Y(A) = p_W(\mathbb{R}^d, A)$ for each $A \in \mathcal{B}(\mathbb{R}^d)$.

Suppose that we are given a collection of random variables $(X_i, i \in I)$ in a fixed probability space; then we denote by $\sigma(X_i, i \in I)$ the smallest σ-algebra contained in $\mathcal{F}$ with respect to which all the X_i are measurable. When there is only a single random variable X in the collection, we denote this σ-algebra as $\sigma(X)$.

The *Doob–Dynkin lemma* states that a random variable Y is measurable with respect to $\sigma(X_1, \ldots, X_n)$ if and only if there is a Borel measurable function $g : \mathbb{R}^{dn} \to \mathbb{R}^d$ such that $Y = g(X_1, \ldots, X_n)$.

Let S be a Borel subset of $\mathbb{R}^d$ that is locally compact in the relative topology. We denote as $B_b(S)$ the linear space of all bounded Borel measurable functions from S to $\mathbb{R}$Banach space) with respect to $||f|| = \sup_{x \in S} |f(x)|$ for each $f \in B_b(S)$. Let $C_b(S)$ be the subspace of $B_b(S)$ comprising continuous functions, $C_0(S)$ be the subspace comprising continuous functions that vanish at infinity and $C_c(S)$ be the subspace comprising functions with compact support, so that

$$C_c(S) \subseteq C_0(S) \subseteq C_b(S).$$

$C_b(S)$ and $C_0(S)$ are both Banach spaces under $|| \cdot ||$ and $C_c(S)$ is norm dense in $C_0(S)$. When S is compact, all three spaces coincide. For each $n \in \mathbb{N}$, $C_b^n(\mathbb{R}^d)$ is the space of all $f \in C_b(\mathbb{R}^d) \cap C^n(\mathbb{R}^d)$ such that all the partial derivatives of f, of order up to and including n, are in $C_b(\mathbb{R}^d)$. We further define $C_b^\infty(\mathbb{R}^d) = \bigcap_{n \in \mathbb{N}} C_b^n(\mathbb{R}^d)$. We define $C_c^n(\mathbb{R}^d)$ and $C_0^n(\mathbb{R}^d)$ analogously, for each $1 \leq n \leq \infty$.

Let $(S, \mathcal{F})$ be a measurable space. A measurable function, $f : S \to \mathbb{R}^d$, is said to be *simple* if

$$f = \sum_{j=1}^{n} c_j \chi_{A_j}$$

for some $n \in \mathbb{N}$, where $c_j \in \mathbb{R}^d$ and $A_j \in \mathcal{F}$ for $1 \leq j \leq n$. We call χ_A the *indicator function*, defined for any $A \in \mathcal{F}$ by

$$\chi_A(x) = 1 \quad \text{whenever } x \in A; \qquad \chi_A(x) = 0 \quad \text{whenever } x \notin A.$$

Let $\Sigma(S)$ denote the linear space of all simple functions on S and let μ be a measure on $(S, \mathcal{F})$. The *integral* with respect to μ is the linear mapping from $\Sigma(S)$ into $\mathbb{R}^d$ defined by

$$I_\mu(f) = \sum_{j=1}^{n} c_j \mu(A_j)$$

for each $f \in \Sigma(S)$. The integral is extended to measurable functions $f = (f_1, f_2, \ldots, f_d)$, where each $f_i \geq 0$, by the prescription for $1 \leq i \leq d$

$$I_\mu(f_i) = \sup\{I_\mu(g_i), \quad g = (g_1, \ldots, g_d) \in \Sigma(S), g_i \leq f_i\}$$

and to arbitrary measurable functions f by

$$I_\mu(f) = I_\mu(f^+) - I_\mu(f^-).$$

We write $I_\mu(f) = \int f(x)\mu(dx)$ or, alternatively, $I_\mu(f) = \int f d\mu$. Note that at this stage there is no guarantee that any of the $I_\mu(f_i)$ is finite.

We say that f is *integrable* if $|I_\mu(f^+)| < \infty$ and $|I_\mu(f^-)| < \infty$. For arbitrary $A \in \mathcal{F}$, we define

$$\int_A f(x)\mu(dx) = I_\mu(f\chi_A).$$

It is worth pointing out that the key estimate

$$\left|\int_A f(x)\mu(dx)\right| \leq \int_A |f(x)|\mu(dx)$$

holds in this vector-valued framework (see e.g. Cohn [80], pp. 352–3).

In the case where we have a probability space $(\Omega, \mathcal{F}, P)$, the linear mapping I_P is called the *expectation* and written simply as $\mathbb{E}$ so, for a random variable X and Borel measurable mapping $f : \mathbb{R}^d \to \mathbb{R}^m$, we have

$$\mathbb{E}(f(X)) = \int_\Omega f(X(\omega))P(d\omega) = \int_{\mathbb{R}^m} f(x)p_X(dx),$$

if $f \circ X$ is integrable. If $A \in \mathcal{F}$, we sometimes write $\mathbb{E}(X; A) = \mathbb{E}(X\chi_A)$.

In the case $d = m = 1$ we have *Jensen's inequality*,

$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X)),$$

whenever $f : \mathbb{R} \to \mathbb{R}$ is a convex function and X and $f(X)$ are both integrable.

The *mean* of X is the vector $\mathbb{E}(X)$ (when it exists) and this is sometimes denoted μ (if there is no measure called μ already in the vicinity) or μ_X, if we want to emphasise the underlying random variable. If $X = (X_1, X_2, \ldots, X_d)$ and $Y = (Y_1, Y_2, \ldots, Y_d)$ are two random variables then the $d \times d$ matrix with (i, j)th entry $\mathbb{E}[(X_i - \mu_{X_i})(Y_j - \mu_{Y_j})]$ is called the *covariance* of X and Y (when it exists) and denoted $\mathrm{Cov}(X, Y)$. In the case $X = Y$ and $d = 1$, we write $\mathrm{Var}(X) = \mathrm{Cov}(X, Y)$ and call this quantity the *variance* of X. It is sometimes denoted σ^2 or σ_X^2. When $d = 1$ the quantity $\mathbb{E}(X^n)$, where $n \in \mathbb{N}$, is called the *nth moment* of X, when it exists. X is said to have *moments to all orders* if $\mathbb{E}(|X|^n) < \infty$, for all $n \in \mathbb{N}$. A sufficient condition for this is that X has an *exponential moment*, i.e. $\mathbb{E}(e^{\alpha|X|}) < \infty$ for some $\alpha > 0$.

For an arbitrary $\mathbb{R}^d$-valued random variable X, we can easily verify the following for all $p > 0$:

- $\mathbb{E}(|X|^p) < \infty$ if and only if $\mathbb{E}(|X_j|^p) < \infty$, for all $1 \leq j \leq d$.
- If $\mathbb{E}(|X|^p) < \infty$ then $\mathbb{E}(|X|^q) < \infty$, for all $0 < q < p$.

The *Chebyshev–Markov inequality* for a random variable X is

$$P(|X - \alpha\mu| \geq C) \leq \frac{\mathbb{E}(|X - \alpha\mu|^n)}{C^n},$$

where $C > 0$, $\alpha \in \mathbb{R}$, $n \in \mathbb{N}$. The commonest forms of this are the Chebyshev inequality ($n = 2$, $\alpha = 1$) and the Markov inequality ($n = 1$, $\alpha = 0$).

We return to a general measure space $(S, \mathcal{F}, \mu)$ and list some key theorems for establishing the integrability of functions from S to $\mathbb{R}^d$. For the first two of these we require $d = 1$.

Theorem 1.1.2 (Monotone convergence theorem) *If $(f_n, n \in \mathbb{N})$ is a sequence of non-negative measurable functions on S that is (a.e.) monotone increasing and converging pointwise to f (a.e.), then*

$$\lim_{n\to\infty} \int_S f_n(x)\mu(dx) = \int_S f(x)\mu(dx).$$

From this we easily deduce the following corollary.

Corollary 1.1.3 (Fatou's lemma) *If $(f_n, n \in \mathbb{N})$ is a sequence of non-negative measurable functions on S, then*

$$\liminf_{n\to\infty} \int_S f_n(x)\mu(dx) \geq \int_S \liminf_{n\to\infty} f_n(x)\mu(dx),$$

which is itself then applied to establish the following theorem.

Theorem 1.1.4 (Lebesgue's dominated convergence theorem) *If $(f_n, n \in \mathbb{N})$ is a sequence of measurable functions from S to $\mathbb{R}^d$ converging pointwise to f (a.e.) and $g \geq 0$ is an integrable function such that $|f_n(x)| \leq g(x)$ (a.e.) for all $n \in \mathbb{N}$, then*

$$\lim_{n\to\infty} \int_S f_n(x)\mu(dx) = \int_S f(x)\mu(dx).$$

We close this section by recalling function spaces of integrable mappings. Let $1 \leq p < \infty$ and denote by $L^p(S, \mathcal{F}, \mu; \mathbb{R}^d)$ the Banach space of all equivalence classes of mappings $f : S \to \mathbb{R}^d$ which agree a.e. (with respect to μ) and for which $||f||_p < \infty$, where $|| \cdot ||_p$ denotes the norm

$$||f||_p = \left[\int_S |f(x)|^p \mu(dx) \right]^{1/p}.$$

In particular, when $p = 2$ we obtain a Hilbert space with respect to the inner product

$$\langle f, g \rangle = \int_S (f(x), g(x)) \mu(dx),$$

for each $f, g \in L^2(S, \mathcal{F}, \mu; \mathbb{R}^d)$. If $\langle f, g \rangle = 0$, we say that f and g are *orthogonal*. A linear subspace V of $L^2(S, \mathcal{F}, \mu; \mathbb{R}^d)$ is called a *closed subspace* if it is closed with respect to the topology induced by $|| \cdot ||_2$, i.e. if $(f_n; n \in \mathbb{N})$ is a sequence in V that converges to f in $L^2(S, \mathcal{F}, \mu; \mathbb{R}^d)$ then $f \in V$.

When there can be no room for doubt, we will use the notation $L^p(S)$ or $L^p(S, \mu)$ for $L^p(S, \mathcal{F}, \mu; \mathbb{R}^d)$.

Hölder's inequality is extremely useful. Let $p, q > 1$ be such that

$$1/p + 1/q = 1.$$

Let $f \in L^p(S)$ and $g \in L^q(S)$ and define $(f, g) : S \to \mathbb{R}$ by $(f, g)(x) = (f(x), g(x))$ for all $x \in S$. Then $(f, g) \in L^1(S)$ and we have

$$||(f, g)||_1 \leq ||f||_p ||g||_q.$$

When $p = 2$, this is called the *Cauchy–Schwarz inequality*.

Another useful fact is that for each $1 \leq p < \infty$ if we define $\Sigma^p(S) = \Sigma(S) \cap L^p(S)$, then $\Sigma^p(S)$ is dense in $L^p(S)$, i.e. given any $f \in L^p(S)$ we can find a sequence $(f_n, n \in \mathbb{N})$ in $\Sigma^p(S)$ such that $\lim_{n \to \infty} ||f - f_n||_p = 0$.

The space $L^p(S, \mathcal{F}, \mu)$ is said to be *separable* if it has a countable dense subset. A sufficient condition for this is that the σ-algebra $\mathcal{F}$ is *countably generated*, i.e. there exists a countable set $\mathcal{C}$ such that $\mathcal{F}$ is the smallest σ-algebra containing $\mathcal{C}$. If $S \in \mathcal{B}(\mathbb{R}^d)$ then $\mathcal{B}(S)$ is countably generated.

1.1.3 Conditional expectation

Let $(S, \mathcal{F}, \mu)$ be an arbitrary measure space. A measure ν on $(S, \mathcal{F})$ is said to be *absolutely continuous* with respect to μ if $A \in \mathcal{F}$ and $\mu(A) = 0 \Rightarrow \nu(A) = 0$. We then write $\nu \ll \mu$. Two measures μ and ν are said to be *equivalent* if they are mutually absolutely continuous. The key result on absolutely continuous measures is

Theorem 1.1.5 (Radon–Nikodým) *If μ is σ-finite and ν is finite with $\nu \ll \mu$, then there exists a measurable function $g : S \to \mathbb{R}^+$ such that, for each $A \in \mathcal{F}$,*

$$\nu(A) = \int_A g(x)\mu(dx).$$

The function g is unique up to μ-almost-everywhere equality.

The functions g appearing in this theorem are sometimes denoted $d\nu/d\mu$ and called (versions of) the *Radon–Nikodým derivative* of ν with respect to μ. For example, if X is a random variable with law p_X that is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, we usually write $f_X = dp_X/dx$ and call f_X a *probability density function* (or sometimes a density or a pdf for short).

Now let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G}$ be a *sub-σ-algebra* of $\mathcal{F}$. Let X be an $\mathbb{R}$-valued random variable with $\mathbb{E}(|X|) < \infty$, and for now assume that $X \geq 0$. We define a finite measure Q_X on $(\Omega, \mathcal{G})$ by the prescription $Q_X(A) = \mathbb{E}(X\chi_A)$ for $A \in \mathcal{G}$; then $Q_X \ll P$, and we write

$$\mathbb{E}(X|\mathcal{G}) = \frac{dQ_X}{dP}.$$

We call $\mathbb{E}(X|\mathcal{G})$ the *conditional expectation* of X with respect to $\mathcal{G}$. It is a random variable on $(\Omega, \mathcal{G}, P)$ and is uniquely defined up to sets of P-measure zero. For arbitrary real-valued X with $\mathbb{E}(|X|) < \infty$, we define

$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X^+|\mathcal{G}) - \mathbb{E}(X^-|\mathcal{G}).$$

When $X = (X_1, X_2, \ldots, X_d)$ takes values in $\mathbb{R}^d$ with $\mathbb{E}(|X|) < \infty$, we define

$$\mathbb{E}(X|\mathcal{G}) = (\mathbb{E}(X_1|\mathcal{G}), E(X_2|\mathcal{G}), \ldots, \mathbb{E}(X_d|\mathcal{G})).$$

We sometimes write $\mathbb{E}_{\mathcal{G}}(\cdot) = \mathbb{E}(\cdot|\mathcal{G})$.

We now list a number of key properties of the conditional expectation:

- $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$.
- $|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X||\mathcal{G})$ a.s.
- If Y is a $\mathcal{G}$-measurable random variable and $\mathbb{E}(|(X, Y)|) < \infty$ then

$$\mathbb{E}((X, Y)|\mathcal{G}) = (\mathbb{E}(X|\mathcal{G}), Y) \quad \text{a.s.}$$