

5

Ergodic decomposition

For convex subsets of vector spaces with finite dimension, it is clear that every element of the convex set may be written as a convex combination of the extremal elements. For example, every point in a triangle may be written as a convex combination of the vertices of the triangle. In view of the results in Section 4.3, it is natural to ask whether a similar property holds in the space of invariant probability measures, that is, whether every invariant measure is a convex combination of ergodic measures.

The ergodic decomposition theorem, which we prove in this chapter (Theorem 5.1.3), asserts that the answer is positive, except that the number of “terms” in this combination is not necessarily finite, not even countable. This theorem has several important applications; in particular, it permits the reduction of the proof of many results to the case when the system is ergodic.

We are going to deduce the ergodic decomposition theorem from another important result from measure theory, the Rokhlin disintegration theorem. The simplest instance of this theorem holds when we have a partition of a probability space (M, μ) into finitely many measurable subsets P_1, \dots, P_N with positive measure. Then, obviously, we may write μ as a linear combination

$$\mu = \mu(P_1)\mu_1 + \dots + \mu(P_N)\mu_N$$

of its normalized restrictions $\mu_i(E) = \mu(E \cap P_i) / \mu(P_i)$ to each of the partition elements. The Rokhlin disintegration theorem (Theorem 5.1.11) states that this type of disintegration of the probability measure is possible for any partition \mathcal{P} (possibly uncountable!) that can be obtained as the limit of an increasing sequence of finite partitions.

5.1 Ergodic decomposition theorem

Before stating the ergodic decomposition theorem, let us analyze a couple of examples that help motivate and clarify its content:

Example 5.1.1. Let $f : [0, 1] \rightarrow [0, 1]$ be given by $f(x) = x^2$. The Dirac measures δ_0 and δ_1 are invariant and ergodic for f . It is also clear that $x = 0$ and $x = 1$ are the unique recurrent points for f and so every invariant probability measure μ must satisfy $\mu(\{0, 1\}) = 1$. Then, $\mu = \mu(\{0\})\delta_0 + \mu(\{1\})\delta_1$ is a (finite) convex combination of the ergodic measures.

Example 5.1.2. Let $f : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ be given by $f(x, y) = (x + y, y)$. The Lebesgue measure m on the torus is preserved by f . Observe that every horizontal circle $H_y = S^1 \times \{y\}$ is invariant under f and the restriction $f : H_y \rightarrow H_y$ is the rotation R_y . Let m_y be the Lebesgue measure on H_y . Observe that m_y is also invariant under f . Moreover, m_y is ergodic whenever y is irrational. On the other hand, by the Fubini theorem,

$$m(E) = \int m_y(E) dy \quad \text{for every measurable set } E. \tag{5.1.1}$$

The identity is not affected if we consider the integral restricted to the subset of irrational values of y . Then (5.1.1) presents m as an (uncountable) convex combination of ergodic measures.

5.1.1 Statement of the theorem

Let us start by introducing some useful terminology. In what follows, (M, \mathcal{B}, μ) is a probability space and \mathcal{P} is a partition of M into measurable subsets. We denote by $\pi : M \rightarrow \mathcal{P}$ the canonical projection that assigns to each point $x \in M$ the element $\mathcal{P}(x)$ of the partition that contains it. This projection map endows \mathcal{P} with the structure of a probability space, as follows. Firstly, by definition, a subset \mathcal{Q} of \mathcal{P} is measurable if and only if its pre-image

$$\pi^{-1}(\mathcal{Q}) = \text{union of all } P \in \mathcal{P} \text{ that belong to } \mathcal{Q}$$

is a measurable subset of M . It is easy to check that this definition is consistent: the family $\hat{\mathcal{B}}$ of measurable subsets is a σ -algebra in \mathcal{P} . Then, we define the *quotient measure* $\hat{\mu}$ by

$$\hat{\mu}(\mathcal{Q}) = \mu(\pi^{-1}(\mathcal{Q})) \quad \text{for every } \mathcal{Q} \in \hat{\mathcal{B}}.$$

Theorem 5.1.3 (Ergodic decomposition). *Let M be a complete separable metric space, $f : M \rightarrow M$ be a measurable transformation and μ be an invariant probability measure. Then there exist a measurable set $M_0 \subset M$ with $\mu(M_0) = 1$, a partition \mathcal{P} of M_0 into measurable subsets and a family $\{\mu_P : P \in \mathcal{P}\}$ of probability measures on M , satisfying*

- (i) $\mu_P(P) = 1$ for $\hat{\mu}$ -almost every $P \in \mathcal{P}$;
- (ii) $P \mapsto \mu_P(E)$ is measurable, for every measurable set $E \subset M$;
- (iii) μ_P is invariant and ergodic for $\hat{\mu}$ -almost every $P \in \mathcal{P}$;
- (iv) $\mu(E) = \int \mu_P(E) d\hat{\mu}(P)$, for every measurable set $E \subset M$.

Part (iv) of the theorem means that μ is a convex combination of the ergodic probability measures μ_P , where the “weight” of each μ_P is determined by the probability measure $\hat{\mu}$. Part (ii) ensures that the integral in (iv) is well defined. Moreover (see Exercise 5.1.3), it implies that the map $\mathcal{P} \rightarrow \mathcal{M}_1(M)$ given by $P \mapsto \mu_P$ is measurable.

5.1.2 Disintegration of a measure

We are going to deduce Theorem 5.1.3 from an important result in measure theory, the Rokhlin disintegration theorem, which has many other applications. To state this theorem we need the following notion.

Definition 5.1.4. A *disintegration* of μ with respect to a partition \mathcal{P} is a family $\{\mu_P : P \in \mathcal{P}\}$ of probability measures on M such that, for every measurable set $E \subset M$:

- (i) $\mu_P(P) = 1$ for $\hat{\mu}$ -almost every $P \in \mathcal{P}$;
- (ii) the map $\mathcal{P} \rightarrow \mathbb{R}$, defined by $P \mapsto \mu_P(E)$ is measurable;
- (iii) $\mu(E) = \int \mu_P(E) d\hat{\mu}(P)$.

Recall that the partition \mathcal{P} inherits from M a natural structure of probability space, with a σ -algebra $\hat{\mathcal{B}}$ and a probability measure $\hat{\mu}$. The measures μ_P are called *conditional probabilities* of μ with respect to \mathcal{P} .

Example 5.1.5. Let $\mathcal{P} = \{P_1, \dots, P_n\}$ be a finite partition of M into measurable subsets with $\mu(P_i) > 0$ for every i . The quotient measure $\hat{\mu}$ is given by $\hat{\mu}(\{P_i\}) = \mu(P_i)$. Consider the normalized restriction μ_i of μ to each P_i :

$$\mu_i(E) = \frac{\mu(E \cap P_i)}{\mu(P_i)} \quad \text{for every measurable set } E \subset M.$$

Then $\{\mu_1, \dots, \mu_n\}$ is a disintegration of μ with respect to \mathcal{P} : it is clear that $\mu(E) = \sum_{i=1}^n \hat{\mu}(\{P_i\})\mu_i(E)$ for every measurable set $E \subset M$.

This construction extends immediately to countable partitions. In the next example we treat an uncountable case:

Example 5.1.6. Let $M = \mathbb{T}^2$ and \mathcal{P} be the partition of M into horizontal circles $S^1 \times \{y\}$, $y \in S^1$. Let m be the Lebesgue measure on \mathbb{T}^2 and \hat{m} be the Lebesgue measure on S^1 . Denote by m_y the Lebesgue measure (arc-length) on each horizontal circle $S^1 \times \{y\}$. By the Fubini theorem,

$$m(E) = \int m_y(E) d\hat{m}(y) \quad \text{for every measurable set } E \subset \mathbb{T}^2.$$

Hence, $\{m_y : y \in S^1\}$ is a disintegration of m with respect to \mathcal{P} .

The next proposition asserts that disintegrations are essentially unique, when they exist. The hypothesis is very general: it holds, for example, if M is a

topological space with a countable basis of open sets and \mathcal{B} is the Borel σ -algebra:

Proposition 5.1.7. *Assume that the σ -algebra \mathcal{B} admits some countable generator. If $\{\mu_P : P \in \mathcal{P}\}$ and $\{\mu'_P : P \in \mathcal{P}\}$ are disintegrations of μ with respect to \mathcal{P} , then $\mu_P = \mu'_P$ for $\hat{\mu}$ -almost every $P \in \mathcal{P}$.*

Proof. Let Γ be a countable generator of the σ -algebra \mathcal{B} and \mathcal{A} be the algebra generated by Γ . Note that \mathcal{A} is countable, since it coincides with the union of the (finite) algebras generated by the finite subsets of Γ . For each $A \in \mathcal{A}$, consider the sets

$$\mathcal{Q}_A = \{P \in \mathcal{P} : \mu_P(A) > \mu'_P(A)\} \quad \text{and} \quad \mathcal{R}_A = \{P \in \mathcal{P} : \mu_P(A) < \mu'_P(A)\}.$$

If $P \in \mathcal{Q}_A$ then P is contained in $\pi^{-1}(\mathcal{Q}_A)$ and, using property (i) in the definition of disintegration, $\mu_P(A \cap \pi^{-1}(\mathcal{Q}_A)) = \mu_P(A)$. Otherwise, P is disjoint from $\pi^{-1}(\mathcal{Q}_A)$ and, hence, $\mu_P(A \cap \pi^{-1}(\mathcal{Q}_A)) = 0$. Moreover, these conclusions remain valid when one takes μ'_P in the place of μ_P . Hence, using property (iii) in the definition of disintegration,

$$\mu(A \cap \pi^{-1}(\mathcal{Q}_A)) = \begin{cases} \int_{\mathcal{P}} \mu_P(A \cap \pi^{-1}(\mathcal{Q}_A)) d\hat{\mu}(P) = \int_{\mathcal{Q}_A} \mu_P(A) d\hat{\mu}(P) \\ \int_{\mathcal{P}} \mu'_P(A \cap \pi^{-1}(\mathcal{Q}_A)) d\hat{\mu}(P) = \int_{\mathcal{Q}_A} \mu'_P(A) d\hat{\mu}(P). \end{cases}$$

Since $\mu_P(A) > \mu'_P(A)$ for every $P \in \mathcal{Q}_A$, this implies that $\hat{\mu}(\mathcal{Q}_A) = 0$ for every $A \in \mathcal{A}$. A similar argument shows that $\hat{\mu}(\mathcal{R}_A) = 0$ for every $A \in \mathcal{A}$. So,

$$\bigcup_{A \in \mathcal{A}} \mathcal{Q}_A \cup \mathcal{R}_A$$

is also a subset of \mathcal{P} with measure zero. For every P in the complement of this subset, the measures μ_P and μ'_P coincide on the generating algebra \mathcal{A} and, consequently, they coincide on the whole σ -algebra \mathcal{B} .

On the other hand, disintegrations may fail to exist:

Example 5.1.8. Let $f : S^1 \rightarrow S^1$ be an irrational rotation and \mathcal{P} be the partition of S^1 whose elements are the orbits $\{f^n(x) : n \in \mathbb{Z}\}$ of f . Assume that there exists a disintegration $\{\mu_P : P \in \mathcal{P}\}$ of the Lebesgue measure μ with respect to \mathcal{P} . Consider the iterates $\{f_* \mu_P : P \in \mathcal{P}\}$ of the conditional probabilities. Since the partition elements are invariant sets, $f_* \mu_P(P) = \mu_P(P) = 1$ for $\hat{\mu}$ -almost every P . It is clear that, given any measurable set $E \subset M$,

$$P \mapsto f_* \mu_P(E) = \mu_P(f^{-1}(E))$$

is a measurable function. Moreover, since μ is an invariant measure,

$$\mu(E) = \mu(f^{-1}(E)) = \int \mu_P(f^{-1}(E)) d\hat{\mu}(P) = \int f_* \mu_P(E) d\hat{\mu}(P).$$

These observations show that $\{f_*\mu_P : P \in \mathcal{P}\}$ is a disintegration of μ with respect to the partition \mathcal{P} . By uniqueness (Proposition 5.1.7), it follows that $f_*\mu_P = \mu_P$ for $\hat{\mu}$ -almost every P . That is, almost every conditional probability μ_P is invariant. That is a contradiction, because $P = \{f^n(x) : n \in \mathbb{Z}\}$ is an infinite countable set and so there can be no invariant probability measure giving P a positive weight.

The theorem of Rokhlin states that disintegrations always exist if the partition \mathcal{P} is the limit of an increasing sequence of countable partitions and the space M is reasonably well behaved. The precise statement is given in the section that follows.

5.1.3 Measurable partitions

We say that \mathcal{P} is a *measurable partition* if there exists some measurable set $M_0 \subset M$ with full measure such that, restricted to M_0 ,

$$\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$$

for some increasing sequence $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \dots \prec \mathcal{P}_n \prec \dots$ of countable partitions (see also Exercise 5.1.1). By $\mathcal{P}_i \prec \mathcal{P}_{i+1}$ we mean that every element of \mathcal{P}_{i+1} is contained in some element of \mathcal{P}_i or, equivalently, every element of \mathcal{P}_i coincides with a union of elements of \mathcal{P}_{i+1} . Then we say that \mathcal{P}_i is *coarser* than \mathcal{P}_{i+1} or, equivalently, that \mathcal{P}_{i+1} is *finer* than \mathcal{P}_i .

Represent by $\bigvee_{n=1}^{\infty} \mathcal{P}_n$ the partition whose elements are the non-empty intersections of the form $\bigcap_{n=1}^{\infty} P_n$ with $P_n \in \mathcal{P}_n$ for every n . Equivalently, this is the coarser partition such that

$$\mathcal{P}_n \prec \bigvee_{n=1}^{\infty} \mathcal{P}_n \quad \text{for every } n.$$

It follows immediately from the definition that every countable partition is measurable. It is easy to find examples of uncountable measurable partitions:

Example 5.1.9. Let $M = \mathbb{T}^2$, endowed with the Lebesgue measure m , and let \mathcal{P} be the partition of M into horizontal circles $S^1 \times \{y\}$. Then \mathcal{P} is a measurable partition. To see that, consider

$$\mathcal{P}_n = \{S^1 \times I(i, n) : i = 1, \dots, 2^n\},$$

where $I(i, n)$, $1 \leq i \leq 2^n$ is the segment of $S^1 = \mathbb{R}/\mathbb{Z}$ corresponding to the interval $[(i - 1)/2^n, i/2^n) \subset \mathbb{R}$. The sequence $(\mathcal{P}_n)_n$ is increasing and $\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$.

On the other hand, not all partitions are measurable:

Example 5.1.10. Let $f : M \rightarrow M$ be a measurable transformation and μ be an ergodic probability measure. Let \mathcal{P} be the partition of M whose elements are the orbits of f . Then \mathcal{P} is *not* measurable, unless f exhibits an orbit with full measure. Indeed, suppose that there exists a non-decreasing sequence $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \dots \prec \mathcal{P}_n \prec \dots$ of countable partitions such that $\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$ restricted to some full measure subset. This last condition implies that almost every orbit of f is contained in some element P_n of the partition \mathcal{P}_n . In other words, up to measure zero, every element of \mathcal{P}_n is invariant under f . By ergodicity, it follows that for every n there exists exactly one $P_n \in \mathcal{P}_n$ such that $\mu(P_n) = 1$. Denote $P = \bigcap_{n=1}^{\infty} P_n$. Then P is an element of the partition $\bigvee_{n=1}^{\infty} \mathcal{P}_n = \mathcal{P}$, that is, P is an orbit of f , and it has $\mu(P) = 1$.

Theorem 5.1.11 (Rokhlin disintegration). *Assume that M is a complete separable metric space and \mathcal{P} is a measurable partition. Then the probability measure μ admits some disintegration with respect to \mathcal{P} .*

Theorem 5.1.11 is proven in Section 5.2. The hypothesis that \mathcal{P} is measurable is, actually, also necessary for the conclusion of the theorem (see Exercise 5.2.2).

5.1.4 Proof of the ergodic decomposition theorem

At this point we are going to use Theorem 5.1.11 to prove the ergodic decomposition theorem. Let \mathcal{U} be a countable basis of open sets of M and \mathcal{A} be the algebra generated by \mathcal{U} . Note that \mathcal{A} is countable and that it generates the Borel σ -algebra of M . By the ergodic theorem of Birkhoff, for every $A \in \mathcal{A}$ there exists a set $M_A \subset M$ with $\mu(M_A) = 1$ such that the mean sojourn time $\tau(A, x)$ is well defined for every $x \in M_A$. Let $M_0 = \bigcap_{A \in \mathcal{A}} M_A$. Note that $\mu(M_0) = 1$, since the intersection is countable.

Now consider the partition \mathcal{P} of M_0 defined as follows: two points $x, y \in M_0$ are in the same element of \mathcal{P} if and only if $\tau(A, x) = \tau(A, y)$ for every $A \in \mathcal{A}$. We claim that this partition is measurable. To prove that it is so, consider any enumeration $\{A_k : k \in \mathbb{N}\}$ of the elements of the algebra \mathcal{A} and let $\{q_k : k \in \mathbb{N}\}$ be an enumeration of the rational numbers. For each $n \in \mathbb{N}$, consider the partition \mathcal{P}_n of M_0 defined as follows: two points $x, y \in M_0$ are in the same element of \mathcal{P}_n if and only if, given any $i, j \in \{1, \dots, n\}$,

$$\begin{aligned} &\text{either } \tau(A_i, x) \leq q_j \text{ and } \tau(A_i, y) \leq q_j \\ &\text{or } \tau(A_i, x) > q_j \text{ and } \tau(A_i, y) > q_j. \end{aligned}$$

It is clear that every \mathcal{P}_n is a finite partition, with no more than 2^{n^2} elements. It follows immediately from the definition that x and y are in the same element

of $\bigvee_{n=1}^{\infty} \mathcal{P}_n$ if and only if $\tau(A_i, x) = \tau(A_i, y)$ for every i . This means that

$$\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n,$$

which implies our claim.

So, by Theorem 5.1.11, there exists some disintegration $\{\mu_P : P \in \mathcal{P}\}$ of μ with respect to \mathcal{P} . Parts (i), (ii) and (iv) of Theorem 5.1.3 are contained in the definition of disintegration. To prove part (iii) it suffices to show that μ_P is invariant and ergodic for $\hat{\mu}$ -almost every P , which is what we do now.

Consider the family of probability measures $\{f_*\mu_P : P \in \mathcal{P}\}$. Observe that every $P \in \mathcal{P}$ is an invariant set, since mean sojourn times are constant on orbits. It follows that

$$f_*\mu_P(P) = \mu_P(f^{-1}(P)) = \mu_P(P) = 1.$$

Moreover, given any measurable set $E \subset M$, the function

$$P \mapsto f_*\mu_P(E) = \mu_P(f^{-1}(E))$$

is measurable and, using the fact that μ is invariant under f ,

$$\mu(E) = \mu(f^{-1}(E)) = \int \mu_P(f^{-1}(E)) d\hat{\mu}(P) = \int f_*\mu_P(E) d\hat{\mu}(P).$$

This shows that $\{f_*\mu_P : P \in \mathcal{P}\}$ is a disintegration of μ with respect to \mathcal{P} . By uniqueness (Proposition 5.1.7), it follows that $f_*\mu_P = \mu_P$ for almost every P .

We are left to prove that μ_P is ergodic for almost every P . Since $\mu(M_0) = 1$, we have that $\mu_P(M_0 \cap P) = 1$ for almost every P . Hence, it is enough to prove that, given any $P \in \mathcal{P}$ and any measurable set $E \subset M$, the mean sojourn time $\tau(E, x)$ is well defined for every $x \in M_0 \cap P$ and is constant on that set. Fix P and denote by \mathcal{C} the class of all measurable sets E for which this holds. By construction, \mathcal{C} contains the generating algebra \mathcal{A} . Observe that if $E_1, E_2 \in \mathcal{C}$ with $E_1 \supset E_2$ then $E_1 \setminus E_2 \in \mathcal{C}$:

$$\tau(E_1 \setminus E_2, x) = \tau(E_1, x) - \tau(E_2, x)$$

is well defined and it is constant on $M_0 \cap P$. In particular, if $E \in \mathcal{C}$ then E^c is also in \mathcal{C} . Analogously, \mathcal{C} is closed under countable pairwise disjoint unions: if $E_j \in \mathcal{C}$ are pairwise disjoint then

$$\tau\left(\bigcup_j E_j, x\right) = \sum_j \tau(E_j, x)$$

is well defined and it is constant on $M_0 \cap P$. It is easy to deduce that \mathcal{C} is a monotone class: given any sequences $A_n, B_n \in \mathcal{C}$, $n \geq 1$ with $A_n \subset A_{n+1}$ and $B_n \supset B_{n+1}$ for every n , the two previous observations yield

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup \bigcup_{n=1}^{\infty} (A_{n+1} \setminus A_n) \in \mathcal{C} \quad \text{and} \quad \bigcap_{n=1}^{\infty} B_n = \left(\bigcup_{n=1}^{\infty} B_n^c\right)^c \in \mathcal{C}.$$

By Theorem A.1.18, it follows that \mathcal{C} contains the Borel σ -algebra of M .

This concludes the proof of Theorem 5.1.3 from Theorem 5.1.11.

5.1.5 Exercises

5.1.1. Show that a partition \mathcal{P} is measurable if and only if there exist measurable subsets $M_0, E_1, E_2, \dots, E_n, \dots$ such that $\mu(M_0) = 1$ and, restricted to M_0 ,

$$\mathcal{P} = \bigvee_{n=1}^{\infty} \{E_n, M \setminus E_n\}.$$

5.1.2. Let μ be an ergodic probability measure for a transformation f . Then μ is also invariant under f^k for any $k \geq 2$. Describe the ergodic decomposition of μ for the iterate f^k .

5.1.3. Let M be a metric space and X be a measurable space. Prove that the following conditions are all equivalent:

- (a) the map $\nu : X \rightarrow \mathcal{M}_1(M)$, $x \mapsto \nu_x$ is measurable;
- (b) the map $X \rightarrow \mathbb{R}$, $x \mapsto \int \varphi d\nu_x$ is measurable, for every bounded continuous function $\varphi : M \rightarrow \mathbb{R}$;
- (c) the map $X \rightarrow \mathbb{R}$, $x \mapsto \int \psi d\nu_x$ is measurable, for every bounded measurable function $\psi : M \rightarrow \mathbb{R}$;
- (d) the map $X \rightarrow \mathbb{R}$, $x \mapsto \nu_x(E)$ is measurable, for every measurable set $E \subset M$.

5.1.4. Prove that if $\{\mu_P : P \in \mathcal{P}\}$ is a disintegration of μ then

$$\int \psi d\mu = \int \left(\int \psi d\mu_P \right) d\hat{\mu}(P)$$

for every bounded measurable function $\psi : M \rightarrow \mathbb{R}$.

5.1.5. Let μ be a probability measure invariant under a measurable transformation $f : M \rightarrow M$. Let $\hat{f} : \hat{M} \rightarrow \hat{M}$ be the natural extension of f and $\hat{\mu}$ be the lift of μ (Section 2.4.2). Relate the ergodic decomposition of μ to the ergodic decomposition of $\hat{\mu}$.

5.1.6. When M is a compact metric space, we may obtain the ergodic decomposition of an invariant probability measure μ by taking for M_0 the subset of points $x \in M$ such that

$$\mu_x = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \delta^{f^j(x)}$$

exists in the weak* topology and taking for \mathcal{P} the partition of M_0 defined by $\mathcal{P}(x) = \mathcal{P}(y) \Leftrightarrow \mu_x = \mu_y$. Check the details of this alternative proof of Theorem 5.1.3 for compact metric spaces.

5.1.7. Let $\sigma : \Sigma \rightarrow \Sigma$ be the shift map in $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$. Consider the partition \mathcal{W}^s of Σ into “stable sets”

$$\mathcal{W}^s((a_n)_n) = \{(x_n)_n : x_n = a_n \text{ for every } n \geq 0\}.$$

Given any probability measure μ invariant under σ , let $\{\mu_P : P \in \mathcal{P}\}$ be an ergodic decomposition of μ . Check that $\mathcal{P} \prec \mathcal{W}^s$, restricted to a full measure subset of M .

5.2 Rokhlin disintegration theorem

Now we prove Theorem 5.1.11. Fix any increasing $\mathcal{P}_1 < \mathcal{P}_2 < \dots < \mathcal{P}_n < \dots$ of countable partitions such that $\mathcal{P} = \bigvee_{n=1}^\infty \mathcal{P}_n$ restricted to some full measure set $M_0 \subset M$. As before, we use $\mathcal{P}_n(x)$ to denote the element of \mathcal{P}_n that contains a given point $x \in M$.

5.2.1 Conditional expectations

Let $\psi : M \rightarrow \mathbb{R}$ be any bounded measurable function. For each $n \geq 1$, define $e_n(\psi) : M \rightarrow \mathbb{R}$ as follows:

$$e_n(\psi, x) = \begin{cases} \frac{1}{\mu(\mathcal{P}_n(x))} \int_{\mathcal{P}_n(x)} \psi \, d\mu & \text{if } \mu(\mathcal{P}_n(x)) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5.2.1}$$

Since the partitions \mathcal{P}_n are countable, the second case of the definition corresponds to a subset of points with total measure zero. Observe also that $e_n(\psi)$ is constant on each $P_n \in \mathcal{P}_n$; let us denote by $E_n(\psi, P_n)$ the value of that constant. Then,

$$\int \psi \, d\mu = \sum_{P_n} \int_{P_n} \psi \, d\mu = \sum_{P_n} \mu(P_n) E_n(\psi, P_n) = \int e_n(\psi) \, d\mu \tag{5.2.2}$$

for every $n \in \mathbb{N}$ (the sums involve only partition elements $P_n \in \mathcal{P}_n$ with positive measure).

Lemma 5.2.1. *Given any bounded measurable function $\psi : M \rightarrow \mathbb{R}$, there exists a subset M_ψ of M with $\mu(M_\psi) = 1$ such that*

- (i) $e(\psi, x) = \lim_n e_n(\psi, x)$ exists for every $x \in M_\psi$;
- (ii) $e(\psi) : M_\psi \rightarrow \mathbb{R}$ is measurable and constant on each $P \in \mathcal{P}$;
- (iii) $\int \psi \, d\mu = \int e(\psi) \, d\mu$.

Proof. Initially, suppose that $\psi \geq 0$. For each $\alpha < \beta$, let $S(\alpha, \beta)$ be the set of points $x \in M$ such that

$$\liminf_n e_n(\psi, x) < \alpha < \beta < \limsup_n e_n(\psi, x).$$

It is clear that the sequence $e_n(\psi, x)$ diverges if and only if $x \in S(\alpha, \beta)$ for some pair of rational numbers $\alpha < \beta$. In other words, the limit $e(\psi, x)$ exists if and only if x belongs to the intersection M_ψ of all $S(\alpha, \beta)^c$ with rational $\alpha < \beta$. As this is a countable intersection, in order to prove that $\mu(M_\psi) = 1$ it suffices to show that $\mu(S(\alpha, \beta)) = 0$ for every $\alpha < \beta$. We do this next.

Let α and β be fixed and denote $S = S(\alpha, \beta)$. Given $x \in S$, fix any sequence of integers $1 \leq a_1^x < b_1^x < \dots < a_i^x < b_i^x < \dots$ such that

$$e_{a_i^x}(\psi, x) < \alpha \quad \text{and} \quad e_{b_i^x}(\psi, x) > \beta \quad \text{for every } i \geq 1.$$

Define A_i to be the union of the partition elements $A_i(x) = \mathcal{P}_{a_i^x}(x)$ and B_i to be the union of the partition elements $B_i(x) = \mathcal{P}_{b_i^x}(x)$ obtained in this way, for all points $x \in S$. By construction, $S \subset A_{i+1} \subset B_i \subset A_i$ for every $i \geq 1$. In particular, S is contained in the set

$$\tilde{S} = \bigcap_{i=1}^{\infty} B_i = \bigcap_{i=1}^{\infty} A_i.$$

Since the sequence $\mathcal{P}_n, n \geq 1$, is monotone increasing, given any two of the sets $A_i(x) = \mathcal{P}_{a_i^x}(x)$ that form A_i , either they are disjoint or one is contained in the other. It follows that the maximal sets $A_i(x)$ are pairwise disjoint and, hence, they constitute a partition of A_i . Hence, adding only over such maximal sets with positive measure,

$$\int_{A_i} \psi \, d\mu = \sum_{A_i(x)} \int_{A_i(x)} \psi \, d\mu \leq \sum_{A_i(x)} \alpha \mu(A_i(x)) = \alpha \mu(A_i),$$

for every $i \geq 1$. Analogously,

$$\int_{B_i} \psi \, d\mu = \sum_{B_i(x)} \int_{B_i(x)} \psi \, d\mu \geq \sum_{B_i(x)} \beta \mu(B_i(x)) = \beta \mu(B_i).$$

Since $A_i \supset B_i$ and we are assuming that $\psi \geq 0$, it follows that

$$\alpha \mu(A_i) \geq \int_{A_i} \psi \, d\mu \geq \int_{B_i} \psi \, d\mu \geq \beta \mu(B_i),$$

for every $i \geq 1$. Taking the limit as $i \rightarrow \infty$, we find that $\alpha \mu(\tilde{S}) \geq \beta \mu(\tilde{S})$. This implies that $\mu(\tilde{S}) = 0$ and, hence, $\mu(S) = 0$. This proves the claim when ψ is non-negative. The general case follows immediately, since we may always write $\psi = \psi^+ - \psi^-$, where ψ^\pm are measurable, non-negative and bounded. Note that $e_n(\psi) = e_n(\psi^+) - e_n(\psi^-)$ for every $n \geq 1$ and, hence, the conclusion of the lemma holds for ψ if it holds for ψ^+ and ψ^- . This ends the proof of claim (i).

The other claims are simple consequences of the definition. The fact that $e(\psi)$ is measurable follows directly from Proposition A.1.31. Since \mathcal{P}_n is coarser than \mathcal{P} , it is clear that $e_n(\psi)$ is constant on each $P \in \mathcal{P}$, restricted to a subset of M with full measure. Hence, the same is true for $e(\psi)$. This proves part (ii). Observe also that $|e_n(\psi)| \leq \sup |\psi|$ for every $n \geq 1$. Hence, we may use the dominated convergence theorem to pass to the limit in (5.2.2). In this way, we get part (iii).

We are especially interested in the case when ψ is a characteristic function: $\psi = \chi_A$ for some measurable set $A \subset M$. In this case, the definition means that

$$e(\psi, x) = \lim_n \frac{\mu(\mathcal{P}_n(x) \cap A)}{\mu(\mathcal{P}_n(x))}. \tag{5.2.3}$$

We denote by \mathcal{P}_A the subset of elements P of the partition \mathcal{P} that intersect M_ψ . Observe that $\hat{\mu}(\mathcal{P}_A) = 1$. Moreover, we define $E(A) : \mathcal{P}_A \rightarrow \mathbb{R}$ by setting

$E(A, P) = e(\psi, x)$ for any $x \in M_\psi \cap P$. Note that $e(\psi) = E(A) \circ \pi$. Hence, the function $E(A)$ is measurable and satisfies:

$$\int \psi \, d\mu = \int e(\psi) \, d\mu = \int E(A) \, d\hat{\mu}. \tag{5.2.4}$$

5.2.2 Criterion for σ -additivity

The hypothesis that the ambient space M is complete separable metric space is used in the proof of the important criterion for σ -additivity that we now state and prove:

Proposition 5.2.2. *Let M be a complete separable metric space and \mathcal{A} be an algebra generated by a countable basis $\mathcal{U} = \{U_k : k \in \mathbb{N}\}$ of open sets of M . Let $\mu : \mathcal{A} \rightarrow [0, 1]$ be an additive function with $\mu(\emptyset) = 0$. Then μ extends to a probability measure on the Borel σ -algebra of M .*

First, let us outline the proof. We consider the product space $\Sigma = \{0, 1\}^{\mathbb{N}}$, endowed with the topology generated by the cylinders

$$[0; a_0, \dots, a_s] = \{(i_k)_{k \in \mathbb{N}} : i_0 = a_0, \dots, i_s = a_s\}, \quad s \geq 0.$$

Note that Σ is compact (Exercise A.1.11). Using the fact that M is a complete metric space, we will show that the map

$$\gamma : M \rightarrow \Sigma, \quad \gamma(x) = (\mathcal{X}_{U_k}(x))_{k \in \mathbb{N}}$$

is a measurable embedding of M inside Σ . Moreover, the function μ yields an additive function ν defined on the algebra \mathcal{A}_Σ generated by the cylinders of Σ . This algebra is compact (Definition A.1.15), since every element is compact. Hence, ν extends to a probability measure on the Borel σ -algebra of Σ ; we still represent this extension by ν . We will show that the image $\gamma(M)$ has full measure for ν . Then, the image $\gamma_*^{-1}\nu$ is a probability measure on the Borel σ -algebra of M . Finally, we will check that this probability measure is an extension of the function μ .

Now let us detail these arguments. In what follows, given any set $A \subset M$, we denote $A^1 = A$ and $A^0 = A^c$.

Lemma 5.2.3. *The image $\gamma(M)$ is a Borel subset of Σ .*

Proof. Let $x \in M$ and $(i_k)_k = \gamma(x)$. It is clear that

$$(A) \quad \bigcap_{j=0}^k U_j^{i_j} \neq \emptyset \text{ for every } k \in \mathbb{N},$$

since x is in the intersection. Moreover, since \mathcal{U} is a basis of open sets of M ,

$$(B) \quad \text{there exists } k \in \mathbb{N} \text{ such that } i_k = 1 \text{ and } \text{diam } U_k \leq 1, \text{ and}$$

$$(C) \quad \text{for every } k \in \mathbb{N} \text{ such that } i_k = 1 \text{ there exists } l(k) > k \text{ such that } i_{l(k)} = 1 \text{ and } \bar{U}_{l(k)} \subset U_k \text{ and } \text{diam } U_{l(k)} \leq \text{diam } U_k / 2.$$

Conversely, suppose that $(i_k)_k \in \Sigma$ satisfies conditions (A), (B) and (C). We are going to show that there exists $x \in M$ such that $\gamma(x) = (i_k)_k$. For that, define

$$F_n = \bigcap_{k=0}^n V_k,$$

where $V_k = U_k^c$ if $i_k = 0$ and $V_k = \bar{U}_{l(k)}$ if $i_k = 1$. Then $(F_n)_n$ is a decreasing sequence of closed sets. Condition (A) assures that $F_n \neq \emptyset$ for every $n \geq 1$. Conditions (B) and (C) imply that the diameter of F_n converges to zero as $n \rightarrow \infty$. Then, since M is a complete metric space, the intersection $\bigcap_n F_n$ contains some point x . By construction, F_n is contained in $\bigcap_{k=0}^n U_k^{i_k}$ for every n . It follows that

$$x \in \bigcap_{k=0}^{\infty} U_k^{i_k},$$

that is, $\gamma(x) = (i_k)_k$. In this way, we have shown that the image of γ is perfectly characterized by the conditions (A), (B) and (C).

To conclude the proof it suffices to show that the subset described by each of these conditions may be constructed from cylinders through countable unions and intersections. Given $k \in \mathbb{N}$, let $N(k)$ be the set of $(k + 1)$ -tuples (a_0, \dots, a_k) in $\{0, 1\}$ such that $U_0^{a_0} \cap \dots \cap U_k^{a_k} \neq \emptyset$. Condition (A) corresponds to the subset

$$\bigcap_{k=0}^{\infty} \bigcup_{(a_0, \dots, a_k) \in N(k)} [0; a_0, \dots, a_k].$$

Let $D = \{k \in \mathbb{N} : \text{diam } U_k \leq 1\}$. Then, condition (B) corresponds to

$$\bigcup_{k \in D} \bigcup_{(a_0, \dots, a_{k-1})} [0; a_0, \dots, a_{k-1}, 1].$$

Finally, given any $k \in \mathbb{N}$, let $L(k)$ be the set of all $l > k$ such that $\bar{U}_l \subset U_k$ and $\text{diam } U_l \leq \text{diam } U_k / 2$. Condition (C) corresponds to the subset

$$\bigcap_{k=0}^{\infty} \bigcup_{a_0, \dots, a_{k-1}} \left([0; a_0, \dots, a_{k-1}, 0] \cup \bigcup_{l \in L(k)} \bigcup_{a_{k+1}, \dots, a_{l-1}} [0; a_0, \dots, a_{k-1}, 1, a_{k+1}, \dots, a_{l-1}, 1] \right).$$

This completes the proof of the lemma.

Corollary 5.2.4. *The map $\gamma : M \rightarrow \gamma(M)$ is a measurable bijection whose inverse is also measurable.*

Proof. Given any points $x \neq y$ in M , there exists $k \in \mathbb{N}$ such that U_k contains one of the points but not the other. This ensures that γ is injective. For any $s \geq 0$ and $a_0, \dots, a_s \in \{0, 1\}$,

$$\gamma^{-1}([0; a_0, \dots, a_s]) = U_0^{a_0} \cap \dots \cap U_s^{a_s}. \tag{5.2.5}$$

This implies that γ is measurable, because the cylinders generate the Borel σ -algebra of Σ . Next, observe that

$$\gamma(U_0^{a_0} \cap \dots \cap U_s^{a_s}) = [0; a_0, \dots, a_s] \cap \gamma(M) \tag{5.2.6}$$

for every s, a_0, \dots, a_s . Using Lemma 5.2.3, it follows that $\gamma(U_0^{a_0} \cap \dots \cap U_s^{a_s})$ is a Borel subset of Σ for every s, a_0, \dots, a_s . This proves that the inverse transformation γ^{-1} is measurable.

Now we are ready to prove that μ extends to a probability measure on the Borel σ -algebra of M , as claimed in Proposition 5.2.2. For that, let us consider the algebra \mathcal{A}_Σ generated by the cylinders of Σ . Note that the elements of \mathcal{A} are the finite pairwise disjoint unions of cylinders. In particular, every element of \mathcal{A}_Σ is compact and, consequently, \mathcal{A}_Σ is a compact algebra (Definition A.1.15). Define:

$$\nu([0; a_0, \dots, a_s]) = \mu(U_0^{a_0} \cap \dots \cap U_s^{a_s}), \tag{5.2.7}$$

for every $s \geq 0$ and a_0, \dots, a_s in $\{0, 1\}$. Then ν is an additive function in the set of all cylinders, with values in $[0, 1]$. It extends in a natural way to an additive function defined on the algebra \mathcal{A}_Σ , which we still denote as ν .

It is clear that $\nu(\Sigma) = 1$. Moreover, since the algebra \mathcal{A}_Σ is compact, we may use Theorem A.1.14 to conclude that the function $\nu : \mathcal{A}_\Sigma \rightarrow [0, 1]$ is σ -additive. Hence, by Theorem A.1.13, the function ν extends to a probability measure defined on the Borel σ -algebra of Σ . Given any cover \mathcal{C} of $\gamma(M)$ by cylinders, it follows from the definition (5.2.7) that

$$\nu\left(\bigcup_{C \in \mathcal{C}} C\right) = \mu\left(\bigcup_{C \in \mathcal{C}} \gamma^{-1}(C)\right) = \mu(M) = 1.$$

Taking the infimum over all covers, we conclude that $\nu(\gamma(M)) = 1$.

By Corollary 5.2.4, the image $\gamma_*^{-1}\nu$ is a Borel probability measure on M . By definition, and using the relation (5.2.6),

$$\begin{aligned} \gamma_*^{-1}\nu(U_0^{a_0} \cap \dots \cap U_s^{a_s}) &= \nu\left(\gamma(U_0^{a_0} \cap \dots \cap U_s^{a_s})\right) = \nu([0; a_0, \dots, a_s] \cap \gamma(M)) \\ &= \nu([0; a_0, \dots, a_s]) = \mu(U_0^{a_0} \cap \dots \cap U_s^{a_s}) \end{aligned}$$

for any s, a_0, \dots, a_s . This implies that $\gamma_*^{-1}\nu$ is an extension of the function $\mu : \mathcal{A} \rightarrow [0, 1]$. Therefore, the proof of Proposition 5.2.2 is complete.

5.2.3 Construction of conditional measures

Let $\mathcal{U} = \{U_k : k \in \mathbb{N}\}$ be a basis of open sets of M and \mathcal{A} be the algebra generated by \mathcal{U} . It is clear that \mathcal{A} generates the Borel σ -algebra of M . Observe also that \mathcal{A} is countable: it coincides with the union of the (finite) algebras generated by the subsets $\{U_k : 0 \leq k \leq n\}$, for every $n \geq 1$. Define

$$\mathcal{P}_* = \bigcap_{A \in \mathcal{A}} \mathcal{P}_A.$$

Then $\hat{\mu}(\mathcal{P}_*) = 1$, since the intersection is countable. For each $P \in \mathcal{P}_*$, define

$$\mu_P : \mathcal{A} \rightarrow [0, 1], \quad \mu_P(A) = E(A, P). \tag{5.2.8}$$

In particular, $\mu_P(M) = E(M, P) = 1$. It is clear that μ_P is an additive function: the definition (5.2.3) gives that

$$A \cap B = \emptyset \implies E(A \cup B, P) = E(A, P) + E(B, P) \quad \text{for every } P \in \mathcal{P}_*.$$

By Proposition 5.2.2, it follows that this function extends to a probability measure defined on the Borel σ -algebra of M , which we still denote as μ_P . We are left to check that this family of measures $\{\mu_P : P \in \mathcal{P}_*\}$ satisfies all the conditions in the definition of disintegration (Definition 5.1.4).

Let us start with condition (i). Let $P \in \mathcal{P}_*$ and, for every $n \geq 1$, let P_n be the element of the partition \mathcal{P}_n that contains P . Observe that if $A \in \mathcal{A}$ is such that $A \cap P_n = \emptyset$ for some n then

$$\mu_P(A) = E(A, P) = \lim_m \frac{\mu(A \cap P_m)}{\mu(P_m)} = 0,$$

since $P_m \subset P_n$ for every $m \geq n$. Fix n . For each $s \geq 0$, let P_n^s be the union of all sets of the form $U_0^{a_0} \cap \dots \cap U_s^{a_s}$ that intersect P_n . By the previous observation, the cylinders of length $s + 1$ that are not in P_n^s have measure zero for μ_P . Therefore, $\mu_P(P_n^s) = 1$ for every $s \geq 0$. Passing to the limit when $s \rightarrow \infty$, we conclude that $\mu_P(U) = 1$ for every open set U that contains P_n . Since the measure μ_P is regular (Proposition A.3.2), it follows that $\mu_P(P_n) = 1$. Passing to the limit when $n \rightarrow \infty$, we find that $\mu_P(P) = 1$ for every $P \in \mathcal{P}_*$.

Finally, let \mathcal{C} denote the family of all measurable sets $E \subset M$ for which conditions (ii) and (iii) hold. By construction (recall Lemma 5.2.1), given any $A \in \mathcal{A}$, the function $P \mapsto \mu_P(A) = E(A, P)$ is measurable and satisfies

$$\mu(A) = \int E(A, P) d\hat{\mu}(P) = \int \mu_P(A) d\hat{\mu}(P).$$

This means that $\mathcal{A} \subset \mathcal{C}$. We claim that \mathcal{C} is a monotone class. Indeed, suppose that B is the union of an increasing sequence $(B_j)_j$ of sets in \mathcal{C} . Then, by Proposition A.1.31,

$$P \mapsto \mu_P(B) = \sup_j \mu_P(B_j) \quad \text{is a measurable function}$$

and, using the monotone convergence theorem,

$$\mu(B) = \lim_n \mu(B_n) = \lim_n \int \mu_P(B_n) d\hat{\mu} = \int \lim_n \mu_P(B_n) d\hat{\mu} = \int \mu_P(B) d\hat{\mu}.$$

This means that $B \in \mathcal{C}$. Analogously, if B is the intersection of a decreasing sequence of sets in \mathcal{C} then $P \mapsto \mu_P(B)$ is measurable and $\mu(B) = \int \mu_P(B) d\hat{\mu}(P)$. That is, $B \in \mathcal{C}$. This proves that \mathcal{C} is a monotone class, as we claimed. By Theorem A.1.18 it follows that \mathcal{C} coincides with the Borel σ -algebra of M .

The proof of Theorem 5.1.11 is complete.

5.2.4 Exercises

- 5.2.1. Let \mathcal{P} and \mathcal{Q} be measurable partitions of (M, \mathcal{B}, μ) such that $\mathcal{P} \prec \mathcal{Q}$ up to measure zero. Let $\{\mu_P : P \in \mathcal{P}\}$ be a disintegration of μ with respect to \mathcal{P} and, for every $P \in \mathcal{P}$, let $\{\mu_{P,Q} : Q \in \mathcal{Q}, Q \subset P\}$ be a disintegration of μ_P with respect to \mathcal{Q} . Let $\pi : \mathcal{Q} \rightarrow \mathcal{P}$ be the canonical projection, such that $Q \subset \pi(Q)$ for almost every $Q \in \mathcal{Q}$. Show that $\{\mu_{\pi(Q),Q} : Q \in \mathcal{Q}\}$ is a disintegration of μ with respect to \mathcal{Q} .
- 5.2.2. (Converse to the theorem of Rokhlin) Let M be a complete separable metric space. Show that if \mathcal{P} satisfies the conclusion of Theorem 5.1.11, that is, if μ admits a disintegration with respect to \mathcal{P} , then the partition \mathcal{P} is measurable.
- 5.2.3. Let $\mathcal{P}_1 \prec \dots \prec \mathcal{P}_n \prec \dots$ be an increasing sequence of countable partitions such that the union $\bigcup_n \mathcal{P}_n$ generates the σ -algebra \mathcal{B} of measurable sets, up to measure zero. Show that the conditional expectation $e(\psi) = \lim_n e_n(\psi)$ coincides with ψ at almost every point, for every bounded measurable function.
- 5.2.4. Prove Proposition 2.4.4, using Proposition 5.2.2.