Teoria Ergódica Diferenciável

lecture 19: Disintegration of measures and unique ergodicity

Instituto Nacional de Matemática Pura e Aplicada

Misha Verbitsky, November 22, 2017

Choquet theorem (reminder)

THEOREM: (Choquet theorem) Let $K \subset V$ be a compact, convex subset in a locally convex topological vector space, R the closure of the set $E(K)$ of its extreme points, and P the space of all probabilistic Borel measures on R. Consider the map $\Phi: P \longrightarrow K$ putting μ to $\int_{x\in R} x\mu$. **Then Φ is surjective.**

Proof: By weak-$*$ compactness of the space of measures, P is compact. The image of Φ is convex and contains all points of R which correspond to atomic measures. On the other hand, an image of a compact set under a continuous map is compact, hence $\Phi(P)$ is compact and complete. Finally, K is a completion of a convex hull of R, hence $K = \Phi(P)$. ■

REMARK: The measure μ associated with a point $k \in K$ is not necessarily unique. If $\Phi: P \longrightarrow K$ is bijective, the set K is called **a simplex**.

Ergodic decomposition of a measure (reminder)

THEOREM: Let Γ be a group (or a semigroup) acting on a topological space M and preserving the Borel σ-algebra, P the space of all Γ-invariant probabilistic measures on M, and R the space of ergodic probabilistic measures. Then, for each $\mu \in P$, **there exists a probability measure ρ_μ on R, such that $\mu = \int_{x \in R} x \rho_\mu$.** Moreover, if Γ is countable, **the measure ρ_μ is uniquely determined by μ.**

REMARK: Such a form ρ_μ is called **ergodic decomposition** of a form μ.

Existence of ergodic decomposition follows from Choquet theorem. Uniqueness follows from the disintegration, see the next slides.

Probability kernels and disintegragion of measures

DEFINITION: Let X, Y be spaces with σ-algebras, P the space of probability measures on X, and $y \stackrel{\varphi}{\mapsto} \mu_y$ a map from Y to P. We say that φ is **probability kernel** if the map $y \longrightarrow \int_X f\mu_y$ gives a measurable function on Y for any bounded, measurable function f on X.

EXAMPLE: Let (A, μ) and (B, ν) be probability spaces, and $A \times B \xrightarrow{\pi} B$ the projection. By Fubini theorem, for any measurable, bounded function f on $A \times B$, the restriction of f to $\pi^{-1}(b)$ is integrable almost everywhere, and $\int_{A\times B} f = \int_{b\in B} \nu \int_{A\times \{b\}} f\mu$. Then $b \longrightarrow \mu\Big|_{X\times\{b\}}$ is a probability kernel.

DEFINITION: Let μ, μ' be measures, with μ absolutely continuous with respect to μ'. Radon-Nikodym tell us that $\mu = f\mu'$, for some non-negative measurable function f. Then f is called **Radon-Nikodym derivative** and denoted by $f = \frac{\mu}{\mu'}$.

Disintegragion of measures

THEOREM: (disintegration of measures) Let (X, μ), (Y, ν) be spaces with probability measures, and $\pi: X \longrightarrow Y$ measurable map such that $\pi_*(\mu) = \nu$. Denote the space of probability measures on X by P. Assume that X is a metrizable topological space with Borel σ-algebra. Then $\pi_*(f\mu)$ is absolutely continuous with respect to ν. Moreover, **there exists a probability kernel $Y \longrightarrow P$ mapping $y \in Y$ to μ_y, such that**

$$\frac{\pi_*(f\mu)}{\nu}(y) = \int_{\pi^{-1}(y)} f\mu_y. \quad (*)$$

Proof. Step 1: Absolute continuity of $\pi_*(f\mu)$ is clear, because a preimage of measure zero subset in Y has measure zero in X, hence it has measure zero in the measure $f\mu$. **It remains to check that $\mu_y(f) := \frac{\pi_*(f\mu)}{\nu}(y)$ defines a probability measure.**

Step 2: This functional is a measure by Riesz representation theorem. Indeed, it is non-negative and continuous on $C^0(M)$. Since $\pi_*\mu = \nu$, one has $\mu_y(1) = 1$, and this measure is probabilistic. ■

REMARK: Disintegration of measures **is unique by construction.**

Disintegration and orthogonal projection

CLAIM: Let (X, μ), (Y, ν) be spaces with probability measure, and $\pi: X \longrightarrow Y$ measurable map such that $\pi_*(\mu) = \nu$. Consider the pullback map $L^2(Y, \nu) \longrightarrow L^2(X, \mu)$, which is by construction an isometry, and let Π be the orthogonal projection from $L^2(X, \mu)$ to the image of $L^2(Y, \nu)$. **Then $\Pi(f)(y) = \int_X f\mu_y$, where $y \mapsto \mu_y$ is the disintegration probability kernel constructed above.**

Proof: Let $g \in L^2(Y)$. Then $\int_X f\pi^* g\mu = \int_Y \pi_*(f\mu)g$. This gives

$$\left\langle \frac{\pi(f)\mu}{\nu}, g \right\rangle = \langle f, \pi^* g\rangle = \langle \Pi(f), g\rangle.$$

We obtained that $\frac{\pi(f)\mu}{\nu} = \Pi(f)$, giving $\int_X f\mu_y = \frac{\pi(f)\mu}{\nu}(y) = \Pi(f)(y)$. ■

Disintegration and conditional expectation

DEFINITION: Probability space is the set M, elements of which are called **outcomes**, equipped with a σ-algebra of subsets, called **events**, and a probability measure μ. In this interpretation, the measure of an event $U \subset M$ is its probability. A **random variable** is a measurable map $f: M \longrightarrow \mathbb{R}$. Its **expected value** is $E(f) := \int_M f\mu$.

DEFINITION: Let $A \subset M$ be an event with $\mu(A) > 0$. **Conditional expectation** of the random variable f is $E_A(f) := \frac{\int_A f\mu}{\mu(A)}$. This is an expectation of f under the condition that the event A happened. The conditional expectation $E_A(\chi_B) := \frac{\mu(A\cap B)}{\mu(A)}$ is probability that B happens under the condition that A happened.

REMARK: Consider now the map $(X,\mu) \xrightarrow{\pi} (Y,\nu)$, and let

$$\frac{\pi_*(f\mu)}{\nu}(y) = \int_{\pi^{-1}(y)} f\mu_y,$$

define the probability kernel μ_y. **The conditional expectation $E_{\pi^{-1}(y)}(f)$ (expectation of f on the set $\pi^{-1}(y)$) is equal to $\int_M f\mu_y$.**

Disintegration and ergodic decomposition

THEOREM: Let X be a metrizable topological space, A its Borel σ-algebra, $T: X \longrightarrow X$ a measurable map, and μ a T-invariant measure. Consider the σ-algebra A^T of T-invariant Borel sets, and let $\pi: (X, A) \longrightarrow (X, A^T)$ be the identity map. Consider the corresponding disintegration $y \longrightarrow \mu_y$ of μ. **Then μ_y are ergodic for a. e. y.**

REMARK: By definition of disintegration, $\int_X f\mu = \int_{y\in X} \int_X f\mu_y$. Therefore, this theorem gives another construction of ergodic decomposition. **Uniqueness of ergodic decomposition is immediately implied by uniqueness of disintegration.**

Proof. Step 1: Notice that all measures μ_y are T-invariant. Indeed, $\pi_* f\mu = \pi_* Tf\mu$. Also, all measurable functions on (X, A^T) are T-invariant, hence $L^2(X, A^T)$ is the space of all L^2-integrable T-invariant functions. This implies that $\int_X f\mu_y = \Pi(f)(y)$ where $\Pi: L^2(X) \longrightarrow L^2(X, A^T)$ is orthogonal projection.

Step 2: To prove that μ_y is ergodic, we need to show that for any bounded L^2-measurable function f, the sequence $C_n(f) := \frac{1}{n}\sum_{i=0}^{n-1} T^i f$ converges to constant a.e. in μ_y for y a.e.

Step 3: The sequence $C_n(f)$ converges to $\Pi(f)$ a.e. in μ. However, $\Pi(f)$ is constant a.e. with respect to μ_y, because $\int g\Pi(f)\mu_y = \Pi(g)\Pi(f)(y)$ and this indegral depends only on $\int_M g\mu_y$. ■

Unique ergodicity

DEFINITION: From now on in this lecture we consider dynamical systems (M, μ, T), where M is a compact space, μ a probability Borel measure, and $T: M \longrightarrow M$ continuous. We say that μ is **uniquely ergodic** if μ is a unique T-invariant probability measure on M.

REMARK: Clearly, **uniquely ergodic measures are ergodic.** Indeed, any T-invariant non-negative measurable function is constant a.e. in μ.

THEOREM: Let (M, μ, T) be as above, and μ uniquely ergodic. **Then the closure of any orbit of T contains the support of μ.**

Proof: Let $x \in M$ and $x_i = T^i(x)$. Consider the atomic measure δ_{x_i}, and let $C_i := \frac{1}{n}\sum_{i=0}^{n-1} \delta_{x_i}$. As shown in Lecture 5, any limit point C of the sequence $\{C_i\}$ is a T-invariant measure; the limit points exist by weak-$*$ compactness. However, **C is supported on the closure $\overline{\{x_i\}}$ of $\{x_i\}$,** because all δ_i vanish on continuous functions which vanish on $\{x_i\}$, and for any point $z \notin \overline{\{x_i\}}$, there exists a continuous function vanishing on $\overline{\{x_i\}}$ and positive in z. ■

EXERCISE: Find a map $T: M \longrightarrow M$ such that **μ is uniquely ergodic, but its support is not the whole M.**

REMARK: Density of all orbits **does not** imply unique ergodicity.

Unique ergodicity and uniform convergence

THEOREM: Let (M,μ,T) be a dynamical system, with M a compact metric space. Denote by $C_n(f)$ the sum $\frac{1}{n}\sum_{i=0}^{n-1} T^i(f)$. **Then the following are equivalent.**

(i) (M,μ,T) is uniquely ergodic.

(ii) For any continuous function f, the sequence $C_n(f)$ converges everywhere to a constant.

(iii) For any continuous function f, the sequence $C_n(f)$ converges uniformly to a constant.

(iv) For any Lipschitz function f, the sequence $C_n(f)$ converges uniformly to a constant.

Proof: Equivalence of (iii) and (iv) is clear, because Lipschitz functions are dense in uniform topology by Stone-Weierstrass. The implications (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) are also clear. It remains to show that (i) implies (iii). Suppose that $C_n(f)$ does not convegre uniformly to $\int_M f\mu$. Then there exists a sequence x_{j_n} such that $C_{j_n}(f)(x_{j_n}) \geqslant \int_M f\mu + \varepsilon$ for some $\varepsilon > 0$. Consider the sequence of measures $\rho_n := \frac{1}{j_n}\sum_{i=0}^{j_n-1} T^i(\delta_{x_{j_n}})$. Then $\int_M f\rho_n = C_{j_n}(f)(x_{j_n}) \geqslant \int_M f\mu + \varepsilon$. Then the same is true for any limit point ρ of $\{\rho_n\}$: $\int_M f\rho > \int_M f\mu + \varepsilon$. However, any such ρ is T-invariant, as shown in Lecture 5. **Then μ and ρ are non-equal T-invariant probability measures.** We obtained a contradiction.

■

Unique ergodicity for isometries

THEOREM: Let (M, μ, T) be a dynamical system, with M a compact metric space, and T an ergodic isometry. **Then it is uniquely ergodic.**

Proof. Step 1: It would suffice to show that $C_n(f) := \frac{1}{n}\sum_{i=0}^{n-1} T^i(f)$ uniformly converges for any Lipschitz f. **Then by ergodicity of T it converges to a constant.**

Step 2: If F is C-Lipschitz, then $C_n(f)$ is also C-Lipschitz. However, $C_n(f)$ converges to f in $L^2(M)$, hence **it converges pointwise on a dense subset of M.**

Step 3: In Lecture 4 it was shown that **a sequence of C-Lipschitz functions converging pointwise in a dense subset of M converges uniformly.** ■

COROLLARY: Irrational circle rotations are uniquely ergodic.

DEFINITION: A sequence $\{x_i\}$ in a measured space (M, μ) is **equidistributed** if the sequence $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{x_i}$ converges to μ.

COROLLARY: Let R be an irrational circle rotation. **Then the sequence $\{R^i(x)\}$ is equidistributed.** ■