Ergodic Theory for Semigroups of Markov Kernels

Julian Newman

Original version 3rd Jan 2015, current version 5th July 2015

The evolution of a homogeneous Markov process may be described statistically by "transition probabilites" which form a semigroup of Markov kernels.[1] These notes concern the abstract theory of stationary distributions and invariant sets of such semigroups.

Feel free to contact me by e-mail if you have any questions or comments, or if you believe that you have found any errors (either typographical or mathematical).

Section 0: Preliminaries (#1–23)**.** We start with some general notation and terminology (including a useful convention for integrating any real-valued measurable function). We then look at several important preliminary topics.

Section 1: Markov kernels (#24–29)**.** We define the notion of a Markov kernel on a measurable space. We introduce stationary probability measures, and consider subsets of the state space that are "invariant (mod null sets)"; naturally, the term that we use to describe such sets is "almost invariant".[2] As well as "almost invariant sets", we also define "almost invariant functions". We consider Lebesgue decomposition of stationary probability measures (Proposition 29). We also introduce ergodic probability measures (which we define in terms of the triviality of all almost-invariant sets).

Section 2: Semigroups of kernels and ergodicity (#30–36)**.** We introduce a natural monoid structure on the space of Markov kernels, and thence define the notion of a "semigroup of Markov kernels". We define stationarity (of probability measures) and almost-invariance (of sets and functions) by considering all the Markov kernels comprising the semigroup. Once again, we then define ergodicity in terms of the triviality of all almost-invariant sets. We give further characterisations of ergodicity (Theorem 34), including its equivalence to being an extremal point of the convex set of stationary probability measures. We also show that the class of ergodic probabliity measures is mutually singular (Theorem 36).

Section 3: Ergodicity in measurable semigroups (#37–49)**.** Until this point, all our notions of invariance have been "modulo null sets". We now introduce forward-invariance, backward-invariance, strict forward-invariance and strict backward-invariance of sets, and super-invariance, sub-invariance and strict invariance of functions. We show that if a semigroup of kernels is "measurable" (i.e. jointly measurable in its spatial and temporal variables), then ergodicity can be characterised in terms of these new notions of

[1] The terms "transition function" and "family of transition probabilities" are often used to refer to either a Markov kernel or a semigroup of Markov kernels.

[2] However, the reader should be warned that the phrase "almost invariant" sometimes appears in other literature to refer to sets that differ from being invariant by a "small but strictly positive amount".

invariance that do not, in and of themselves, make reference to a measure (Theorem 39). (However, we make the interesting observation in Exercise 46(A) that, even though the triviality of all bounded strictly invariant *functions* is sufficient for ergodicity, the triviality of all *sets* whose indicator function is strictly invariant is *not* sufficient for ergodicity.) In Proposition 49 we relate ergodicity for continuous-time semigroups to ergodicity for kernels.

Section 4: Markov processes and pointwise ergodic theorems (#50–90)**.** We introduce the notion of a "Markov measure" (that is, the law of a homogeneous Markov process). We state and prove a form of the "Markov-processes version" of the pointwise ergodic theorem (Theorem 55, with the proof extending from Proposition 58 to Corollary 73). We also state (without full proof) an "ergodic theorem for semigroups of kernels" (Proposition 78). Converses of these theorems are also given (Proposition 81 and Corollary 85). We then give an appendix, introducing a special class of Markov processes, namely processes with stationary and independent increments. (As in Exercise 135(B), these are often the stochastic processes driving a random dynamical system.)

Section 5: Ergodic decomposition (#91–111)**.** We show that if the state space of a measurable semigroup is standard then every stationary probability measure has an integral representation via ergodic probability measures (Theorem 99/Corollary 100). Specifically, an integral representation is obtained by conditioning the stationary measure with respect to the σ-algebra of almost-invariant sets. (The pointwise ergodic theorem for semigroups of kernels is used to show that the resulting measures are indeed ergodic). An application of this is presented in Corollary 109.

Section 6: Feller-continuity (#112–123)**.** If the state space of a semigroup of kernels is a separable metric space, then one can ask about "continuity" of the semigroup. We present a natural way of defining continuity of a semigroup with respect to its spatial variable, known as "Feller-continuity". (We do not consider continuity in time in these notes.) We present the Krylov-Bogolyubov theorem for existence of stationary probability measures (Theorem 114). We also discuss strong-Feller-continuity.

Section 7: Random maps and random dynamical systems (#124–143)**.** We introduce random maps and their associated Markov kernels. We introduce filtered random dynamical systems (adapted to a one-parameter filtration), and focus on the case of "memoryless noise". In this case, we show that the transition probabilities of a RDS form a semigroup of Markov kernels (Proposition 140, which is essentially a corollary of Proposition 127). We describe the stationary and ergodic probability measures of this semigroup in terms of the skew-product dynamics induced by the RDS (Theorem 143).

Appendix: Markov operators (#144–149)**.** We explain the link between Markov kernels and "Markov operators".

To maintain the flow of the material, several supporting lemmas and relevant remarks are left as exercises.

0 Preliminaries

0.1 Some notational conventions

I. For a function $f : S \to T$ and a collection $\mathcal{C}$ of subsets of T, let $f^{-1}\mathcal{C} := \{f^{-1}(C) : C \in \mathcal{C}\}$. (Let us recall at this point the general fact that $f^{-1}\sigma(\mathcal{C}) = \sigma(f^{-1}\mathcal{C})$.)

II. For a set Ω, a collection of measurable spaces $\{(X_\alpha, \Sigma_\alpha)\}_{\alpha \in I}$ and a collection $\{f_\alpha\}_{\alpha \in I}$ of functions $f_\alpha : \Omega \to X_\alpha$, we use the notation $\sigma(f_\alpha : \alpha \in I)$ as a shorthand for $\sigma(\{f_\alpha^{-1}(A) : \alpha \in I, A \in \Sigma_\alpha\})$. In other words, $\sigma(f_\alpha : \alpha \in I)$ is the smallest σ-algebra on Ω with respect to which f_α is measurable for every $\alpha \in I$; it is easy to show that if I is infinite then

$$\sigma(f_\alpha : \alpha \in I) \ = \bigcup_{\substack{S \subset I \\ |S| = |\mathbb{N}|}} \sigma(f_\alpha : \alpha \in S).$$

Beware that whenever the notation "$\sigma(\ldots)$" is used, it is always defined with reference to some underlying set that does not explicitly appear within the notation. Nonetheless, we still have the following useful fact (which is very easy to prove): given sets $\Omega_1 \subset \Omega_2$ and a collection $\{f_\alpha\}_{\alpha \in I}$ of functions $f_\alpha : \Omega_2 \to X_\alpha$, the σ-algebra $\sigma(f_\alpha|_{\Omega_1} : \alpha \in I)$ on Ω_1 coincides with the σ-algebra on Ω_1 induced from the σ-algebra $\sigma(f_\alpha : \alpha \in I)$ on Ω_2.

III. For a collection of sets $\{X_\alpha\}_{\alpha \in I}$, the *Cartesian product* $\times_{\alpha \in I} X_\alpha$ denotes the set of all I-indexed families $(x_\alpha)_{\alpha \in I}$ with x_α being a member of X_α for all α. Obviously, if there exists $\alpha' \in I$ such that $X_{\alpha'} = \varnothing$ then $\times_{\alpha \in I} X_\alpha = \varnothing$. (The converse clearly holds if the number of distinct members of the collection $\{X_\alpha\}_{\alpha \in I}$ is finite. Mathematicians uncontroversially take as axiomatic that the converse holds whenever I is countably infinite; the controversial "axiom of choice" asserts that the converse *always* holds.) Given a set X and a set I, we write X^I as a shorthand for $\times_{\alpha \in I} X$.

If $\{(X_\alpha, \Sigma_\alpha)\}_{\alpha \in I}$ is a collection of measurable spaces with $\times_{\alpha \in I} X_\alpha \neq \varnothing$, we define the *product σ-algebra* $\bigotimes_{\alpha \in I} \Sigma_\alpha$ on $\times_{\alpha \in I} X_\alpha$ by

$$\bigotimes_{\alpha \in I} \Sigma_\alpha \ := \ \sigma(\, (x_\alpha)_{\alpha \in I} \mapsto x_{\alpha'} \ : \ \alpha' \in I \,).$$

Given a measurable space (X, Σ) and a set I, we write $\Sigma^{\otimes I}$ as a shorthand for $\bigotimes_{\alpha \in I} \Sigma$.

IV. Given a probability space (X, Σ, ρ) and a set I, it is known (e.g. as a special case of the *Ionescu-Tulcea extension theorem*[3]) that there exists a unique probability measure μ on $(X^I, \Sigma^{\otimes I})$ with the property that for any distinct $\alpha_1, \ldots, \alpha_n \in I$ and any $A_1, \ldots, A_n \in \Sigma$,

$$\mu\Big(\big\{\, (x_\alpha)_{\alpha \in I} \in X^I \ : \ x_{\alpha_i} \in A_i \ \forall\, 1 \leq i \leq n \,\big\}\Big) \ = \ \prod_{i=1}^{n} \rho(A_i).$$

We will denote the unique probability measure with this property by $\rho^{\otimes I}$.

V. By a "measure", we specifically mean a $[0, \infty]$-valued σ-additive function on some

[3] Theorem 1 of here.

given σ-algebra on a set. We write f_*m to denote the image measure of a measure m under a measurable map f — i.e. $f_*m := m(f^{-1}(\cdot))$. Note that under a given map f, the image measure of a Dirac mass at a point x is equal to the Dirac mass at the image point of x (i.e. $f_*\delta_x = \delta_{f(x)}$). Thus we may "recover" the original map f from the image-measure operation f_* by restricting to Dirac measures. Also note that the image-measure operation respects composition: that is, $(g \circ f)_*m = g_*(f_*m)$.

Given a measure space $(\Omega, \mathcal{F}, m)$, we say that a set $A \subset \Omega$ is *m-null* if there exists $A' \in \mathcal{F}$ such that $A \subset A'$ and $m(A') = 0$. (So a measurable set $A \in \mathcal{F}$ is m-null if and only if $m(A) = 0$.) We say that a set $A \subset \Omega$ is *m-full* if $\Omega \setminus A$ is m-null. (If m is a finite measure then this is equivalent to saying that there exists $A' \in \mathcal{F}$ such that $A' \subset A$ and $m(A') = m(\Omega)$.)

We may define an equivalence relation $\sim$ on 2^Ω by

$$A \sim B \iff A \vartriangle B \text{ is an } m\text{-null set.}$$

The *m-completion of $\mathcal{F}$*, denoted $\bar{\mathcal{F}}_m$, is the union of all equivalence classes of $\sim$ that intersect $\mathcal{F}$. It is easy to show that $\bar{\mathcal{F}}_m$ is a σ-algebra, and that there is a unique measure $\bar{m}$ on $\bar{\mathcal{F}}_m$ with the properties that $\bar{m}|_{\mathcal{F}} = m$ and $\bar{m}$ assigns the same value to all members of the same equivalence class of $\sim$. $(\Omega, \bar{\mathcal{F}}_m, \bar{m})$ is called the *completion* of $(\Omega, \mathcal{F}, m)$. It is easy to show that $(\Omega, \bar{\mathcal{F}}_m, \bar{m})$ is a complete probability space, and that for any complete probability space $(\Omega, \mathcal{J}, l)$ with $\mathcal{F} \subset \mathcal{J}$ and $l|_{\mathcal{F}} = m$, we have $\bar{\mathcal{F}}_m \subset \mathcal{J}$ and $l|_{\bar{\mathcal{F}}_m} = \bar{m}$. (Heuristically, $(\Omega, \bar{\mathcal{F}}_m, \bar{m})$ is the "smallest complete probability space containing $(\Omega, \mathcal{F}, m)$".) It is also easy to show that for any measurable space (X, Σ) and any measurable map $g : \Omega \to X$, $g^{-1}(\bar{\Sigma}_{g_*m}) \subset \bar{\mathcal{F}}_m$ and for all $A \in \bar{\Sigma}_{g_*m}$, $\overline{g_*m}(A) = \bar{m}(g^{-1}(A))$.

Given a second-countable topological space X and a measure m on X (where X is equipped with its Borel σ-algebra $\mathcal{B}(X)$), it is easy to show that there exists a largest open m-null set; the complement of this set is called the *support of m* (denoted $\operatorname{supp} m$). Note that a point $x \in X$ belongs to $\operatorname{supp} m$ if and only if m assigns positive measure to every neighbourhood of x.

VI. (a) We denote the extended real line by $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, and equip it with the obvious ordering and corresponding order topology (so the function $\arctan\colon \bar{\mathbb{R}} \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ is both an order-isomorphism and a homeomorphism), and the corresponding Borel σ-algebra. For any $a, b \in \bar{\mathbb{R}}$ with $a \leq b$, the notations $[a,b]$, $[a,b)$, $(a,b]$ and (a,b) have the obvious meaning. We work with the usual arithmetic on $\bar{\mathbb{R}}$:

(i) $0.\infty = \infty.0 = 0.(-\infty) = (-\infty).0 = 0$;

(ii) for all $c \in (0, \infty]$,

$$\begin{aligned} c.\infty = \infty.c = (-\infty).(-c) = (-c).(-\infty) &= \infty \\ \text{and} \quad (-c).\infty = \infty.(-c) = (-\infty).c = c.(-\infty) &= -\infty; \end{aligned}$$

(iii) for all $c \in \mathbb{R} \cup \{-\infty\}$, $(-\infty) + c = c + (-\infty) = -\infty$ and $\infty + (-c) = (-c) + \infty = \infty$;

(iv) $|\infty| = |-\infty| = \infty$.

(b) Certain theorems involving integrals can be rather cumbersome to state (and to prove) when one has to take into account integrals that are not well-defined. Accordingly, for the purpose of efficiently stating and proving several of our results (especially Theorem 12, Exercise 13, Corollary 14, and the ergodic theorems in Section 4), we introduce the following convention, that may initially seem unusual but will be of enormous help: Given a measure space $(\Omega, \mathcal{F}, m)$ and a measurable function $g : \Omega \to \bar{\mathbb{R}}$ with $\int_\Omega g^+(\omega)\, m(d\omega) = \int_\Omega g^-(\omega)\, m(d\omega) = \infty$, we will declare the value of $\int_\Omega g(x)\, m(d\omega)$ to be NaN ("not a number"). We will write $\mathbb{R}'$ to denote $\bar{\mathbb{R}} \cup \{\text{NaN}\}$, and will equip $\mathbb{R}'$ with the obvious σ-algebra (namely $\mathcal{B}(\bar{\mathbb{R}}) \cup \{A \cup \{\text{NaN}\} : A \in \mathcal{B}(\bar{\mathbb{R}})\}$). We define the following rules:

(i) $0.\text{NaN} = \text{NaN}.0 = 0$

(ii) for all $c \in \mathbb{R}' \setminus \{0\}$, $c.\text{NaN} = \text{NaN}.c = \text{NaN}$;

(iii) $\text{NaN} + c = c + \text{NaN} = \text{NaN}$ for all $c \in \mathbb{R}'$;

(iv) $|\text{NaN}| = \text{NaN}$;

(v) if $g : \Omega \to \mathbb{R}'$ is a measurable function with $m(g^{-1}(\{\text{NaN}\})) = 0$ then $\int_\Omega g(\omega)\, m(d\omega) = \int_{g^{-1}(\bar{\mathbb{R}})} g(\omega)\, m(d\omega)$;

(vi) if $g : \Omega \to \mathbb{R}'$ is a measurable function with $m(g^{-1}(\{\text{NaN}\})) > 0$ then $\int_\Omega g(\omega)\, m(d\omega) = \text{NaN}$.

(We emphasise rule (v), which is essentially what will make our convention so useful.) Where useful, given a measurable function $g : \Omega \to \mathbb{R}'$, we will write $m(g)$ as a shorthand for $\int_\Omega g(\omega)\, m(d\omega)$. (So $m(A) = m(\mathbb{1}_A)$ for all $A \in \mathcal{F}$.)

Given a topological space T, a point $a \in T$ and a function $g : T \setminus \{a\} \to \mathbb{R}'$, if $\text{NaN} \in g(U \setminus \{a\})$ for every neighbourhood U of a, then we automatically say that $\lim_{x \to a} g(x)$ does not exist. (In particular, NaN itself can never be obtained as a limit.)

Given a function $g : \Omega \to \mathbb{R}'$, we write $g^+ : \Omega \to [0, \infty]$ and $g^- : \Omega \to [0, \infty]$ to denote respectively the positive and negative parts of the function $\mathbb{1}_{\bar{\mathbb{R}}}(g(\cdot))g(\cdot)$ on Ω. We say that g is nonnegative if $g(\Omega) \subset [0, \infty]$, i.e. if $g = g^+$. We say that g is bounded below (resp. above) if $g(\Omega)$ is a subset of $\mathbb{R} \cup \{\infty\}$ (resp. of $\mathbb{R} \cup \{-\infty\}$) that is bounded below (resp. above); and we will say that g is bounded if g is both bounded below and bounded above. We will say that g is *finite* if $g(\Omega) \subset \mathbb{R}$. We will say that g is ($\mathcal{F}$-)*simple* if g is measurable and $g(\Omega)$ is a finite subset of $\mathbb{R}$. We will say that g is *integrable with respect to* m (or m-*integrable*) if g is measurable and $\int_\Omega |g(\omega)|\, m(d\omega) \in [0, \infty)$ (which is equivalent to saying that $m(g) \in \mathbb{R}$). So then, $m(g) \neq \text{NaN}$ if and only if both $m(g^{-1}(\{\text{NaN}\})) = 0$ and at least one of the integrals $m(g^+)$ and $m(g^-)$ is finite; and g is m-integrable if and only if both $m(g^{-1}(\{\text{NaN}\}) = 0$ and the integrals $m(g^+)$ and $m(g^-)$ are both finite.

Define $sub : \mathbb{R}' \times \mathbb{R}' \to \mathbb{R}'$ by $sub(x, y) = x - y$ for $(x, y) \in (\bar{\mathbb{R}} \times \bar{\mathbb{R}}) \setminus \{(-\infty, -\infty), (\infty, \infty)\}$ and $sub(x, y) = \text{NaN}$ otherwise. Note that sub is measurable.

VII. (a) Let m_1 and m_2 be measures on a measurable space $(\Omega, \mathcal{F})$, and suppose we have a measurable function $g : \Omega \to [0, \infty)$ such that $m_2(A) = \int_A g(\omega)\, m_1(d\omega)$ for all $A \in \mathcal{F}$. Then we will say that g is "a version of $\frac{dm_2}{dm_1}$" [to be read: "a version of the density of m_2 with respect to m_1"]. The *Radon-Nikodym theorem* states that provided m_1 and m_2 are σ-finite, a version of $\frac{dm_2}{dm_1}$ exists if and only if m_2 is absolutely continuous with respect to m_1, and in this case any two versions of $\frac{dm_2}{dm_1}$ agree m_1-almost everywhere.

(b) Let (X, Σ, ρ) be a probability space, and let $\mathcal{E}$ be a sub-σ-algebra of Σ. Given any $A \in \Sigma$ we will say that a function $h : X \to [0,1]$ is "a version of $\rho(A|\mathcal{E})$" [to be read: "a version of the conditional probability under ρ of A given $\mathcal{E}$"] if h is $\mathcal{E}$-measurable and for all $E \in \mathcal{E}$

$$\rho(A \cap E) \;=\; \int_E h(x)\, \rho(dx).$$

Likewise, given a function $g : X \to \mathbb{R}'$ that is integrable with respect to ρ, we will say that a function $h : X \to \mathbb{R}'$ is "a version of $\rho(g|\mathcal{E})$" [to be read: "a version of the conditional expectation under ρ of g given $\mathcal{E}$"] if h is $\mathcal{E}$-measurable and for all $E \in \mathcal{E}$

$$\int_E g(x)\, \rho(dx) \;=\; \int_E h(x)\, \rho(dx).$$

Since we will be taking conditional expectations under various different probability measures in the course of these notes, we will dispense of any $\mathbb{E}$ notation, and just use the notations introduced above. We will assume knowledge of the most basic properties of conditional expectations.

Finally, in this document $\mathbb{N}$ denotes the set of positive integers (i.e. $0 \notin \mathbb{N}$).

Exercise 1. Let $(\Omega, \mathcal{F}, m)$ be a measure space. (A) Show that for any measurable $g : \Omega \to [0, \infty]$ and any $c \in \mathbb{R}'$, $m(cg) = cm(g)$. (B) Show that if $g_1, g_2 : \Omega \to \mathbb{R}'$ are measurable functions and g_2 is integrable with respect to m, then $m(sub(g_1, g_2)) = m(g_1) - m(g_2)$. (C) Show that if $g_1, g_2 : \Omega \to \mathbb{R}'$ are measurable functions with $m(g_1) \in \mathbb{R} \cup \{\infty\}$ and $m(g_2) \in \mathbb{R} \cup \{-\infty\}$ (or vice versa), then $m(sub(g_1, g_2)) = m(g_1) - m(g_2)$. (D) Let $g : \Omega \to \mathbb{R}'$ be an m-integrable function, let $E := g^{-1}(\mathbb{R}' \setminus \{0\})$, and let $\mathcal{F}_E$ be the set of $\mathcal{F}$-measurable subsets of E. Show that the measure space $(E, \mathcal{F}_E, m|_{\mathcal{F}_E})$ is σ-finite. (E) Suppose m is a finite measure. Then one may be tempted to "improve" the definition of an m-full set (by allowing it to cover more cases) as follows: "We say that a set $E \subset \Omega$ is m-full if there exists a measure m' on E (equipped with the induced σ-algebra of $\mathcal{F}$ from Ω onto E) such that $m(A) = m'(A \cap E)$ for all $A \in \mathcal{F}$." Identify a problem with this approach to defining full-measure sets.

0.2 Measurability of extrema and limits

A topology or a topological space is said to be *Polish* if it is separable and completely metrisable. A σ-algebra or a measurable space is said to be *standard* if it is generated by a Polish topology. It turns out (e.g. Proposition 424G in Chapter 42 of here) that for any standard measurable space $(I, \mathcal{I})$ and any non-empty $J \in \mathcal{I}$, the set of $\mathcal{I}$-measurable subsets of J is standard (as a σ-algebra on J).

Let $(\Omega, \mathcal{F})$ be a measurable space. The *universal completion* $\bar{\mathcal{F}}$ *of* $\mathcal{F}$ is defined to be the intersection, over all probability measures $\mathbb{P}$ on $(\Omega, \mathcal{F})$, of the $\mathbb{P}$-completion of $\mathcal{F}$. It is easy to check that $\bar{\bar{\mathcal{F}}} = \bar{\mathcal{F}}$. A set or a function is said to be *universally measurable with respect to* $\mathcal{F}$ if it is measurable with respect to $\bar{\mathcal{F}}$.

Working in any given metric space, the notations $B_\varepsilon(x)$ and $B_\varepsilon(A)$ denote respectively the ball of radius ε about a point x and the ε-neighbourhood of a set A.

The proof of the following is left as an exercise to the reader. (Part (B) is based on the measurable projection theorem.)

Lemma 2. *(A) Let* $(g_n)_{n\in\mathbb{N}}$ *be a sequence of measurable functions* $g_n : \Omega \to \bar{\mathbb{R}}$. *The functions* $g_s : \Omega \to \bar{\mathbb{R}}$ *and* $g_i : \Omega \to \bar{\mathbb{R}}$ *given by*

$$\begin{aligned} g_s(\omega) &= \sup_{n\in\mathbb{N}} g_n(\omega) \\ g_i(\omega) &= \inf_{n\in\mathbb{N}} g_n(\omega) \end{aligned}$$

are measurable. (B) Let $(I, \mathcal{I})$ *be a standard measurable space, and let* $g : \Omega \times I \to \bar{\mathbb{R}}$ *be a measurable function. The functions* $g_s : \Omega \to \bar{\mathbb{R}}$ *and* $g_i : \Omega \to \bar{\mathbb{R}}$ *given by*

$$\begin{aligned} g_s(\omega) &= \sup_{\alpha\in I} g(\omega, \alpha) \\ g_i(\omega) &= \inf_{\alpha\in I} g(\omega, \alpha) \end{aligned}$$

are universally measurable. (C) Let I *be a separable metric space, and let* $g : \Omega \times I \to \bar{\mathbb{R}}$ *be a measurable function such that the map* $\alpha \mapsto g(\omega, \alpha)$ *is continuous for each* $\omega \in \Omega$. *Then the functions* g_s *and* g_i *(as defined in part (B)) are measurable.*

Note that as a special case of Lemma 2(B), for any $A \in \mathcal{F} \otimes \mathcal{B}(\bar{\mathbb{R}})$ the maps $\omega \mapsto \sup A_\omega$ and $\omega \mapsto \inf A_\omega$ are universally measurable, where $A_\omega := \{x \in \bar{\mathbb{R}} : (\omega, x) \in A\}$. (To see this: setting

$$g(\omega, x) = x\,\mathbb{1}_A(\omega, x) + k\,\mathbb{1}_{(\Omega\times\bar{\mathbb{R}})\setminus A}(\omega, x),$$

we have that $g_i(\omega) = \inf A_\omega$ if $k = \infty$, and $g_s(\omega) = \sup A_\omega$ if $k = -\infty$.)

Exercise 3. Let I be a metric space, fix any $a \in I$ and let $g : \Omega \times (I \setminus \{a\}) \to \bar{\mathbb{R}}$ be a measurable function. Define the functions $g_{ls,a} : \Omega \to \bar{\mathbb{R}}$ and $g_{li,a} : \Omega \to \bar{\mathbb{R}}$ by

$$\begin{aligned} g_{ls,a}(\omega) &= \limsup_{\alpha\to a} g(\omega, \alpha) \\ g_{li,a}(\omega) &= \liminf_{\alpha\to a} g(\omega, \alpha). \end{aligned}$$

(A) Show that if $\mathcal{B}(I)$ is standard then $g_{ls,a}$ and $g_{li,a}$ are universally measurable. (B) Show that if I is separable and the map $\alpha \mapsto g(\omega, \alpha)$ from $I \setminus \{a\}$ to $\bar{\mathbb{R}}$ is continuous then $g_{ls,a}$ and $g_{li,a}$ are measurable. (C) [*Extended Fatou's lemma*] Suppose we have a measure m on Ω, a measurable function $\tilde{g} : \Omega \to \bar{\mathbb{R}}$ agreeing with $g_{li,a}$ m-almost everywhere, and a measurable function $l : \Omega \to \bar{\mathbb{R}}$ such that $m(l^-) < \infty$ and $g(\omega, \alpha) \geq l(\omega)$ for all ω and α. Show that

$$\int_\Omega \tilde{g}(\omega)\, m(d\omega) \leq \liminf_{\alpha\to a} \int_\Omega g(\omega, \alpha)\, m(d\omega).$$

Lemma 4. *(A) Let Y be a metric space, and let $(g_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions $g_n : \Omega \to Y$ such that $\lim_{n\to\infty} g_n(\omega) =: g(\omega)$ exists for all $\omega \in \Omega$. Then $g : \Omega \to Y$ is measurable. (B) Let Y be a Polish space, and let $(g_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions $g_n : \Omega \to Y$. Then the set $\Omega' := \{\omega \in \Omega : \lim_{n\to\infty} g_n(\omega)$ exists$\}$ is measurable, and the function $g : \Omega' \to Y$ given by $g(\omega) = \lim_{n\to\infty} g_n(\omega)$ is measurable.*

Proof. (A) For any closed $G \subset Y$,

$$g^{-1}(G) = \bigcap_{n=1}^{\infty}\bigcup_{i=1}^{\infty}\bigcap_{j=i}^{\infty} g_j^{-1}\Big(B_{\frac{1}{n}}(G)\Big).$$

So g is measurable. (B) Fix a separable complete metrisation d of Y. Then

$$\Omega' = \bigcap_{n=1}^{\infty}\bigcup_{i=1}^{\infty}\bigcap_{j,k=i}^{\infty} \{\omega : d(g_j(\omega), g_k(\omega)) < \tfrac{1}{n}\}$$

and so $\Omega' \in \mathcal{F}$. Part (A) gives the rest. □

Lemma 5. *Let I be a metric space, and fix a point $a \in I$ such that a is an accumulation point of I (i.e. $a \in \overline{I \setminus \{a\}}$). (A) Let Y be a metric space, and let $g : \Omega \times (I \setminus \{a\}) \to Y$ be a measurable function such that $\lim_{\alpha\to a} g(\omega,\alpha) =: g_{l,a}(\omega)$ exists for all $\omega \in \Omega$. Then the function $g_{l,a} : \Omega \to Y$ is measurable. (B) Assume I is separable. Let Y be a Polish space, let $g : \Omega \times (I \setminus \{a\}) \to Y$ be a measurable function, and let $\Omega' := \{\omega \in \Omega : \lim_{\alpha\to a} g(\omega,\alpha)$ exists$\}$. If $\mathcal{B}(I)$ is standard then Ω' is universally measurable. If the map $\alpha \mapsto g(\omega,\alpha)$ from $I \setminus \{a\}$ to Y is continuous for each $\omega \in \Omega$, then Ω' is measurable. In any case, the function $g_{l,a} : \Omega' \to Y$ given by $g_{l,a}(\omega) = \lim_{\alpha\to a} g(\omega,\alpha)$ is measurable with respect to the induced σ-algebra of $\mathcal{F}$ onto Ω'.*

Proof. (A) Let (a_n) be a sequence in $I \setminus \{a\}$ converging to a. Then for all $\omega \in \Omega$, $g_{l,a}(\omega) = \lim_{n\to\infty} g(\omega, a_n)$. So by Lemma 4(A), $g_{l,a}$ is measurable. (B) Fix a separable complete metrisation d of Y. Then

$$\Omega' = \bigcap_{n=1}^{\infty}\bigcup_{m=1}^{\infty} \{\omega : \sup_{\alpha,\beta\in B_{\frac{1}{m}}(a)\setminus\{a\}} d(g(\omega,\alpha), g(\omega,\beta)) < \tfrac{1}{n}\}.$$

Parts (B) and (C) of Lemma 2 (applied to the map $(\omega,(\alpha,\beta)) \mapsto d(g(\omega,\alpha), g(\omega,\beta))$) and part (A) of this lemma then give the rest. □

0.3 Integration preserves measurability

A *π-system* is a collection of sets that is closed under pairwise intersections. A *λ-system* on a set Ω is a collection of subsets of Ω that includes Ω itself and is closed under both complements in Ω and countable disjoint unions. Now in order to show that all the members of some σ-algebra $\mathcal{F}$ have some particular property, a common approach is to show that the set of *all* sets with the desired property is a σ-algebra, and that there is a generator $\mathcal{C}$ of $\mathcal{F}$ all of whose members have the desired property. However, sometimes we are not quite able to show that the set of all sets with the desired property is a σ-algebra, but only a λ-system. (An important example is the set of measurable sets on which two given probability measures agree.) The *π-λ theorem* says that this is still fine, *if* our generator $\mathcal{C}$ is a π-system. More precisely: the theorem states that if $\mathcal{D}$ is a λ-system on a set Ω and $\mathcal{C} \subset \mathcal{D}$ is a π-system, then the σ-algebra on Ω generated by $\mathcal{C}$ is contained in $\mathcal{D}$.

Exercise 6. Let μ and ν be measures on a measurable space $(\Omega, \mathcal{F})$. Let $\mathcal{C}$ be a π-system generating $\mathcal{F}$, containing an increasing sequence $E_1 \subset E_2 \subset E_3 \subset \ldots$ with $\mu(E_n) < \infty$ for all n and $\bigcup_{n=1}^{\infty} E_n = \Omega$ (so μ is σ-finite). Show that if $\mu(E) = \nu(E) \ \forall\, E \in \mathcal{C}$, then $\mu = \nu$.

Let $(\Omega, \mathcal{F})$ be a measurable space.

Lemma 7. *Fix $a \in [1, \infty]$. Let $\mathcal{H}$ be a set of functions from Ω to $[0, \infty]$ such that:*

(a) there exists a π-system $\mathcal{C}$ generating $\mathcal{F}$, with $\Omega \in \mathcal{C}$, such that $\mathbb{1}_E \in \mathcal{H}$ for all $E \in \mathcal{C}$;

(b) for any $E \in \mathcal{F}$, if $\mathbb{1}_E \in \mathcal{H}$ then $\mathbb{1}_{\Omega \setminus E} \in \mathcal{H}$;

(c) for any $c_1, c_2 \in [0, \infty)$ and $g_1, g_2 \in \mathcal{H}$, $c_1 g_1 + c_2 g_2 \in \mathcal{H}$;

(d) for any increasing sequence (g_n) in $\mathcal{H}$ with $g_n(\Omega) \subset [0, a]$ for all n, the pointwise limit of (g_n) is in $\mathcal{H}$.

Then $\mathcal{H}$ includes all measurable functions $g : \Omega \to [0, \infty]$ with $g(\Omega) \subset [0, a]$.

Proof. Let $\mathcal{A} := \{E \in \mathcal{F} : \mathbb{1}_E \in \mathcal{H}\}$. For any sequence $(E_n)_{n \in \mathbb{N}}$ of mutually disjoint members of $\mathcal{A}$, properties (c) and (d) together yield that $\bigcup_{n=1}^{\infty} E_n \in \mathcal{A}$. Combining this with property (b) and the fact that $\Omega \in \mathcal{A}$ (from property (a)), we have that $\mathcal{A}$ is a λ-system. Hence, by property (a) and the π-λ theorem, $\mathcal{A} = \mathcal{F}$. Property (c) then gives that $\mathcal{H}$ includes all the nonnegative simple functions, and hence property (d) completes the result. □

Throughout these notes, (X, Σ) is a measurable space, and $\mathcal{M}$ is the set of measures on (X, Σ), equipped with its natural σ-algebra, namely $\sigma(\rho \mapsto \rho(A) : A \in \Sigma)$. So a map $\omega \mapsto \mu_\omega$ from Ω to $\mathcal{M}$ is measurable if and only if the map $\omega \mapsto \mu_\omega(A)$ is measurable for every $A \in \Sigma$. This is, in turn, equivalent to saying that the map $\omega \mapsto \mu_\omega(g)$ is measurable for every measurable $g : X \to \mathbb{R}$ (see Exercise 9).

$\mathcal{M}_{<\infty} \subset \mathcal{M}$ will denote the set of finite measures on X, and $\mathcal{M}_1 \subset \mathcal{M}_{<\infty}$ the set of probability measures on X; note that $\mathcal{M}_{<\infty}$ and $\mathcal{M}_1$ are measurable subsets of $\mathcal{M}$.

Lemma 8. *(A) Let $(I, \mathcal{I})$ be a measurable space, and suppose we have a measurable mapping $\alpha \mapsto \rho_\alpha$ from I to $\mathcal{M}_{<\infty}$. For any measurable function $g : \Omega \times I \times X \to \mathbb{R}'$, the function $\overline{g} : \Omega \times I \to \mathbb{R}'$ given by*

$$\overline{g}(\omega, \alpha) = \int_X g(\omega, \alpha, x)\, \rho_\alpha(dx)$$

is measurable. (B) Suppose we have a σ-finite measure ρ on X. For any measurable function $g : \Omega \times X \to \mathbb{R}'$, the function $\overline{g} : \Omega \to \mathbb{R}'$ given by

$$\overline{g}(\omega) = \int_X g(\omega, x)\, \rho(dx)$$

is measurable.

Proof. (A) Let $\mathcal{H}$ be the set of measurable functions $g : \Omega \times I \times X \to \bar{\mathbb{R}}$ for which $\overline{g}$ is measurable, and let $\mathcal{C} := \{E \times B \times A : E \in \mathcal{F}, B \in \mathcal{I}, A \in \Sigma\}$. Note that for any member $S = E \times B \times A$ of $\mathcal{C}$, $\overline{\mathbb{1}_S}(\omega, \alpha) = \mathbb{1}_{E\times B}(\omega, \alpha)\rho_\alpha(A)$ and so $\mathbb{1}_S \in \mathcal{H}$. Also note that for any $S \in \mathcal{F} \otimes \mathcal{I} \otimes \Sigma$, $\overline{\mathbb{1}_{X\setminus S}}(\omega, \alpha) = \rho_\alpha(X) - \overline{\mathbb{1}_S}(\omega, \alpha)$, and so if $\mathbb{1}_S \in \mathcal{H}$ then $\mathbb{1}_{X\setminus S} \in \mathcal{H}$. By the monotone convergence theorem and Lemma 7 (with $a = \infty$), it follows that $\mathcal{H}$ consists of all the nonnegative measurable functions on $\Omega \times I \times X$; in other words, the desired statement is true whenever g is nonnegative. Now for a general measurable function $g : \Omega \times I \times X \to \mathbb{R}'$, we can apply our previous statement to the function $\mathbb{1}_{\{\mathrm{NaN}\}} \circ g$ to obtain that the set

$$S \;:=\; \{\, (\omega, \alpha) \in \Omega \times I : \rho_\alpha(x \in X : g(\omega, \alpha, x) = \mathrm{NaN}) > 0 \,\}$$

is measurable. We have that

$$\overline{g}(\omega, \alpha) \;=\; \begin{cases} \mathrm{NaN} & (\omega, \alpha) \in S \\ sub(\overline{g^+}(\omega, \alpha), \overline{g^-}(\omega, \alpha)) & (\omega, \alpha) \in (\Omega \times I) \setminus S. \end{cases}$$

Now $\overline{g^+}$ and $\overline{g^-}$ are measurable (since g^+ and g^- are nonnegative); so $\overline{g}$ is measurable. (B) If $\rho(X) < \infty$ then the result follows immediately from (A); so assume that $\rho(X) = \infty$. Let $(X_n)_{n\in\mathbb{N}}$ be an increasing sequence of sets belonging to Σ, with $\bigcup_{n=1}^\infty X_n = X$ and $0 < \rho(X_n) < \infty$ for all n. For any nonnegative measurable function g on $\Omega \times X$, we have

$$\overline{g}(\omega) \;=\; \lim_{n\to\infty} \int_{X_n} g(\omega, x)\, \rho(dx)$$

and so, by Lemma 4(A) and part (A) of this lemma, $\overline{g}$ is measurable. To extend to general g, argue as in part (A). $\square$

Exercise 9. Show that the σ-algebra $\sigma(\,\rho \mapsto \rho(g) \;:\; \text{measurable } g : X \to \mathbb{R}'\,)$ on $\mathcal{M}$ precisely coincides with the σ-algebra on $\mathcal{M}$ that we introduced earlier. (So if $\omega \mapsto \mu_\omega$ is a measurable mapping from Ω to $\mathcal{M}$ then the mapping $\omega \mapsto \mu_\omega(g)$ is measurable for any measurable $g : X \to \mathbb{R}'$.)

Exercise 10. Show that if the diagonal in $X \times X$ is $(\Sigma \otimes \Sigma)$-measurable (e.g. if Σ is the Borel σ-algebra of a second-countable Hausdorff topology), then Σ includes all the singletons in X, and the map $(\rho, x) \mapsto \rho(\{x\})$ from $\mathcal{M} \times X$ to $[0, 1]$ is measurable.

Exercise 11. Recall that for any measure w on $\bar{\mathbb{R}}$, given a value $a \in \bar{\mathbb{R}}$ and a set $A \subset \bar{\mathbb{R}}$, a is called an an *essential upper bound (under w) of A* if the set $\{x \in A : x > a\}$ is a w-null set; and for any $A \subset \bar{\mathbb{R}}$, the set of essential upper bounds under w of A has a least element, which is called the *essential supremum (under w) of A*. The notion of an *essential infimum* can be defined similarly. Given a measure space $(I, \mathcal{I}, m)$ and a measurable function $g : I \to \bar{\mathbb{R}}$, for any set $J \subset I$ we write

$$m\text{-}\operatorname*{ess\,sup}_{\alpha\in J} g(\alpha) \quad \text{and} \quad m\text{-}\operatorname*{ess\,inf}_{\alpha\in J} g(\alpha)$$

to denote respectively the essential supremum under g_*m of $g(J)$ and the essential infimum under g_*m of $g(J)$. (A) Let $(I, \mathcal{I}, m)$ be a σ-finite measure space, and let

$g : \Omega \times I \to \bar{\mathbb{R}}$ be a measurable function. Show that the functions $g_{es} : \Omega \to \bar{\mathbb{R}}$ and $g_{ei} : \Omega \to \bar{\mathbb{R}}$ given by

$$\begin{aligned} g_{es}(\omega) &= m\text{-}\operatorname*{ess\,sup}_{\alpha \in I} g(\omega, \alpha) \\ g_{ei}(\omega) &= m\text{-}\operatorname*{ess\,inf}_{\alpha \in I} g(\omega, \alpha) \end{aligned}$$

are measurable. (B) Let I be a metric space, fix $a \in I$, and let m be a σ-finite measure on $I \setminus \{a\}$ (equipped with its Borel σ-algebra). Given a function $h : I \setminus \{a\} \to \bar{\mathbb{R}}$, we define the "essential superior limit" and "essential inferior limit" of h at a by

$$\begin{aligned} m\text{-}\lim\operatorname*{ess\,sup}_{\alpha \to a} h(\alpha) &:= \lim_{\varepsilon \to 0} m\text{-}\operatorname*{ess\,sup}_{\alpha \in B_\varepsilon(a) \setminus \{a\}} h(\alpha) \\ m\text{-}\lim\operatorname*{ess\,inf}_{\alpha \to a} h(\alpha) &:= \lim_{\varepsilon \to 0} m\text{-}\operatorname*{ess\,inf}_{\alpha \in B_\varepsilon(a) \setminus \{a\}} h(\alpha). \end{aligned}$$

Now let $g : \Omega \times (I \setminus \{a\}) \to \bar{\mathbb{R}}$ be a measurable function. Show that the functions $g_{les,a} : \Omega \to \bar{\mathbb{R}}$ and $g_{lei,a} : \Omega \to \bar{\mathbb{R}}$ given by

$$\begin{aligned} g_{les,a}(\omega) &= m\text{-}\lim\operatorname*{ess\,sup}_{\alpha \to a} g(\omega, \alpha) \\ g_{lei,a}(\omega) &= m\text{-}\lim\operatorname*{ess\,inf}_{\alpha \to a} g(\omega, \alpha) \end{aligned}$$

are measurable.

0.4 Shifting integral signs

Note that by the monotone convergence theorem, given a measurable mapping $\omega \mapsto \mu_\omega$ from Ω to $\mathcal{M}$ and a measure m on $(\Omega, \mathcal{F})$, $A \mapsto \int_\Omega \mu_\omega(A)\, m(d\omega)$ is a measure on (X, Σ). The integral with respect to this measure of a measurable function $g : X \to \mathbb{R}'$ may be denoted

$$\int_X g(x) \int_\Omega \mu_\omega(dx)\, m(d\omega).$$

Theorem 12. *Suppose we have a measurable mapping $\omega \mapsto \mu_\omega$ from Ω to $\mathcal{M}$ and a measure m on Ω. Let $g : X \to \mathbb{R}'$ be a measurable function. If*

$$\int_X g(x) \int_\Omega \mu_\omega(dx)\, m(d\omega) \neq \mathrm{NaN}$$

then

$$\int_X g(x) \int_\Omega \mu_\omega(dx)\, m(d\omega) = \int_\Omega \int_X g(x)\, \mu_\omega(dx)\, m(d\omega).$$

Note that, using Theorem 12 itself,

$$\int_X g(x) \int_\Omega \mu_\omega(dx)\, m(d\omega) \neq \mathrm{NaN} \quad [\text{resp. } \in \mathbb{R}]$$

if and only if the following two statements hold:

- $g^{-1}(\{\mathrm{NaN}\})$ is a μ_ω-null set for m almost all $\omega \in \Omega$;

- at least one of the integrals

$$\int_\Omega \int_X g^+(x)\,\mu_\omega(dx)\,m(d\omega) \quad \text{and} \quad \int_\Omega \int_X g^-(x)\,\mu_\omega(dx)\,m(d\omega)$$

[resp. the integral

$$\int_\Omega \int_{g^{-1}(\bar{\mathbb{R}})} |g(x)|\,\mu_\omega(dx)\,m(d\omega)\ ]$$

is finite.

Proof of Theorem 12. If $g = \mathbb{1}_A$ for some $A \in \Sigma$ then LHS $= \int_\Omega \mu_\omega(A)\,m(d\omega) =$ RHS. Hence the monotone convergence theorem and Lemma 7 (with $a = \infty$) yield that the desired statement is true whenever g is nonnegative. Now, for convenience, let $\bar{\mu}$ denote the measure $\int_\Omega \mu_\omega(\cdot)\,m(d\omega)$. Let $g : X \to \mathbb{R}'$ be any measurable function such that $\bar{\mu}(g) \neq \mathrm{NaN}$. Firstly, $\bar{\mu}(g^{-1}(\mathrm{NaN})) = 0$, and so for m-almost every $\omega \in \Omega$, for μ_ω-almost all $x \in X$, $g(x) \neq \mathrm{NaN}$. We also know that either $\bar{\mu}(g^+) < \infty$ or $\bar{\mu}(g^-) < \infty$. Assume the latter case (the former case is similar); hence

$$\int_\Omega \mu_\omega(g^-)\,m(d\omega) \ < \ \infty,$$

from which it follows that for m-almost every $\omega \in \Omega$, $\mu_\omega(g^-) < \infty$. So then, for m-almost every $\omega \in \Omega$, $\mu_\omega(g) = \mu_\omega(g^+) - \mu_\omega(g^-)$. With all this, we have:

$$\begin{aligned} \bar{\mu}(g) &= \bar{\mu}(g^+) - \bar{\mu}(g^-) \\ &= \int_\Omega \mu_\omega(g^+)\,m(d\omega) - \int_\Omega \mu_\omega(g^-)\,m(d\omega) \\ &= \int_{\tilde{\Omega}} \mu_\omega(g^+) - \mu_\omega(g^-)\,m(d\omega) \\ &= \int_\Omega \mu_\omega(g)\,m(d\omega) \end{aligned}$$

where $\tilde{\Omega}$ is an m-full set on which $\mu_\omega(g^-) < \infty$. So we are done. □

Exercise 13. (A) Let $(\Omega, \mathcal{F}, m)$ be a measure space. Let $f : \Omega \to X$ and $h : \Omega \to [0, \infty]$ be measurable functions, and define the measure ν on X by

$$\nu(A) \ = \ \int_{f^{-1}(A)} h(\omega)\,m(d\omega).$$

Show that for *any* measurable $g : X \to \mathbb{R}'$,

$$\int_X g(x)\,\nu(dx) \ = \ \int_\Omega g(f(\omega))h(\omega)\,m(d\omega).$$

Note that if $h \equiv 1$ then we recover the "transformation-of-integrals formula", namely $m(g \circ f) = f_* m(g)$; and if $\Omega = X$ with $f = \mathrm{id}_\Omega$, then we recover the well-known formula that "$\int_\Omega g\,d\nu = \int_\Omega g \cdot \frac{d\nu}{dm}\,dm$". (B) Let $(A, \mathcal{A})$ and $(B, \mathcal{B})$ be measurable spaces, fix some $a \in A$, let p be a measure on B, and define the measure q_a on $A \times B$ by

$$q_a(S) \ = \ \int_B \mathbb{1}_S(a,b)\,p(db) \ = \ p(b \in B : (a,b) \in S).$$

Show that

$$q_a(g) \;=\; \int_B g(a,b)\, p(db).$$

for any measurable $g : A \times B \to \mathbb{R}'$. (C) [*Conditional transformation-of-integrals formula*] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(E, \mathcal{E})$ be a measurable space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{E}$, let $f : \Omega \to E$ be a measurable function, and let $g : E \to \mathbb{R}$ be an $(f_*\mathbb{P})$-integrable function. Show that if $\tilde{g} : E \to \mathbb{R}$ is a version of $f_*\mathbb{P}(g|\mathcal{G})$, then $\tilde{g} \circ f$ is a version of $\mathbb{P}(g \circ f | f^{-1}\mathcal{G})$.

As an important case of Theorem 12, we have the following:

Corollary 14 (Fubini-Tonelli theorem for σ-finite spaces)**.** *Let $(A, \mathcal{A}, \mu)$ and $(B, \mathcal{B}, \nu)$ be σ-finite measure spaces. For any $S \in \mathcal{A} \otimes \mathcal{B}$, let*

$$\mu \otimes \nu(S) \;:=\; \int_A \int_B \mathbb{1}_S(a,b)\, \nu(db)\, \mu(da).$$

(This is well-defined by Lemma 8(B).) $\mu \otimes \nu$ is a σ-finite measure on $A \times B$, and is the only measure on $A \times B$ assigning the value $\mu(U)\nu(V)$ to $U \times V$ for all $U \in \mathcal{A}$ and $V \in \mathcal{B}$. For any measurable $g : A \times B \to \mathbb{R}'$, if

$$\int_{A \times B} g(a,b)\, \mu \otimes \nu(d(a,b)) \;\neq\; \mathrm{NaN}$$

then

$$\begin{aligned}\int_{A \times B} g(a,b)\, \mu \otimes \nu(d(a,b)) \;&=\; \int_A \int_B g(a,b)\, \nu(db)\, \mu(da) \\ &=\; \int_B \int_A g(a,b)\, \mu(da)\, \nu(db).\end{aligned}$$

Note once again that for any measurable $g : A \times B \to \mathbb{R}'$, $\mu \otimes \nu(g) \neq \mathrm{NaN}$ [resp. $\in \mathbb{R}$] if and only if the following two statements hold:

- $\mu \otimes \nu(g^{-1}(\{\mathrm{NaN}\})) = 0$;
- at least one of the four integrals

  $$\begin{array}{ll} \int_A \int_B g^+(a,b)\, \nu(db)\, \mu(da) & \int_A \int_B g^-(a,b)\, \nu(db)\, \mu(da) \\ \int_B \int_A g^+(a,b)\, \mu(da)\, \nu(db) & \int_B \int_A g^-(a,b)\, \mu(da)\, \nu(db) \end{array}$$

  [resp. at least one of the two integrals

  $$\int_A \int_B |g_{\bar{\mathbb{R}}}(a,b)|\, \nu(db)\, \mu(da) \quad \text{and} \quad \int_B \int_A |g_{\bar{\mathbb{R}}}(a,b)|\, \mu(da)\, \nu(db)$$

  where $g_{\bar{\mathbb{R}}}(a,b) := g(a,b)\mathbb{1}_{\bar{\mathbb{R}}}(g(a,b))$] is finite.

If μ and ν are probability measures, then the probability measure $\mu \otimes \nu$ represents the probability distribution for a random selection of a pair $(a,b) \in A \times B$ in which a and b are selected *independently of each other*, with a selected from A with probability distribution μ and b selected from B with probability distribution ν.

Proof of Corollary 14. It is clear (by the monotone convergence theorem) that $\mu\otimes\nu$ is a measure. For any $U\in\mathcal{A}$ and $V\in\mathcal{B}$,

$$\mu\otimes\nu(U\times V) \;=\; \int_A\int_B \mathbb{1}_U(a)\mathbb{1}_V(b)\,\nu(db)\,\mu(da) \;=\; \int_A \nu(V)\mathbb{1}_U(a)\,\mu(da) \;=\; \mu(U)\nu(V).$$

Now given increasing sequences $(U_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ and $\mathcal{B}$ respectively such that $\bigcup_{n=1}^\infty U_n = A$, $\bigcup_{n=1}^\infty V_n = B$ and $\mu(U_n),\nu(V_n) < \infty$ for all $n\in\mathbb{N}$, it is clear that $(U_n\times V_n)_{n\in\mathbb{N}}$ is an increasing sequence in $\mathcal{A}\otimes\mathcal{B}$ with $\mu\otimes\nu(U_n\times V_n)<\infty$ for all n and $\bigcup_{n=1}^\infty(U_n\times V_n) = A\times B$. So $\mu\otimes\nu$ is σ-finite. The fact that $\mu\otimes\nu$ is the only measure assigning the value $\mu(U)\nu(V)$ to $U\times V$ for all $U\in\mathcal{A}$ and $V\in\mathcal{B}$ then follows from Exercise 6, with $\mathcal{C}=\{U\times V : U\in\mathcal{A}, V\in\mathcal{B}\}$. Now applying Theorem 12 with $\Omega = A$, $X = A\times B$ and $\mu_a(S) = \int_B \mathbb{1}_S(a,b)\,\nu(db)$, we obtain (using Exercise 13(B)) that

$$\int_{A\times B} g(a,b)\,\mu\otimes\nu(d(a,b)) \;=\; \int_A\int_B g(a,b)\,\nu(db)\,\mu(da)$$

provided $\mu\otimes\nu(g)\neq \mathrm{NaN}$. Finally, if we define

$$\mu\,\tilde{\otimes}\,\nu(S) \;:=\; \int_B\int_A \mathbb{1}_S(a,b)\,\mu(da)\,\nu(db)$$

for all $S\in\mathcal{A}\otimes\mathcal{B}$, then it is easy to check once again that $\mu\,\tilde{\otimes}\,\nu$ is a measure on $A\times B$ assigning the value $\mu(U)\nu(V)$ to $U\times V$ for all $U\in\mathcal{A}$ and $V\in\mathcal{B}$. So $\mu\,\tilde{\otimes}\,\nu = \mu\otimes\nu$, and therefore, by Theorem 12 again,

$$\int_{A\times B} g(a,b)\,\mu\otimes\nu(d(a,b)) \;=\; \int_B\int_A g(a,b)\,\mu(da)\,\nu(db)$$

provided $\mu\otimes\nu(g)\neq\mathrm{NaN}$. □

Exercise 15 (*Conditional Fubini theorem*)**.** Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, with $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$, and let $(I,\mathcal{I},\nu)$ be a σ-finite measure space. Let $g:\Omega\times I\to\mathbb{R}'$ be a function that is integrable with respect to $\mathbb{P}\otimes\nu$, and let $\tilde{g}:\Omega\times I\to\mathbb{R}'$ be a measurable function such that for ν-almost all $\alpha\in I$, the map $\omega\mapsto\tilde{g}(\omega,\alpha)$ is a version of $\mathbb{P}(\omega\mapsto g(\omega,\alpha)|\mathcal{G})$. Show that the map $\omega\mapsto\int_I\tilde{g}(\omega,\alpha)\,\nu(d\alpha)$ is a version of $\mathbb{P}\big(\omega\mapsto\int_I g(\omega,\alpha)\,\nu(d\alpha)\big|\,\mathcal{G}\big)$.

0.5 A lemma on joint measurability

Given sets $I\subset J\subset\bar{\mathbb{R}}$, we will say that I is *right-dense in* J if for every $t\in J$ and $\varepsilon>0$, $I\cap[t,t+\varepsilon)\neq\varnothing$. (This is equivalent to saying that I is both dense in J and contains every point $t\in J$ with the property that $J\cap(t,t+\delta)=\varnothing$ for some $\delta>0$.)

Lemma 16. *Let $(\Omega,\mathcal{F})$ be a measurable space, let Z be a metric space, let J be a separable metric space, and suppose we have a function $g:J\times\Omega\to Z$.*

(A) Suppose that (i) the map $x\mapsto g(x,\omega)$ is continuous for each $\omega\in\Omega$, and (ii) J admits a countable dense subset S such that the map $\omega\mapsto g(x,\omega)$ is measurable for each $x\in S$. Then g is a measurable function.

(B) In the case that J is a subspace of $\bar{\mathbb{R}}$: suppose that (i) the map $t\mapsto g(t,\omega)$ is right-continuous for each $\omega\in\Omega$, and (ii) J admits a countable right-dense subset S such that the map $\omega\mapsto g(t,\omega)$ is measurable for each $t\in S$. Then g is a measurable function.

To prove this: (A) (Following Lemma 4.51 of here) we leave it as an exercise to show that for any closed $G \subset Z$,

$$g^{-1}(G) = \bigcap_{n=1}^{\infty} \bigcup_{x \in S} \Big(B_{\frac{1}{n}}(x) \times g(x,\cdot)^{-1}\Big(B_{\frac{1}{n}}(G)\Big) \Big).$$

(B) We leave it as an exercise to show that for any closed $G \subset Z$,

$$g^{-1}(G) = \bigcap_{n=1}^{\infty} \bigcup_{t \in S} \Big(\big(J \cap (t - \tfrac{1}{n}, t]\big) \times g(t,\cdot)^{-1}\Big(B_{\frac{1}{n}}(G)\Big) \Big).$$

(Lemma 16 can be further generalised, but such generalisations have harder proofs and will not be needed.)

0.6 The narrow topology for separable metric spaces

Theorem 17. *For any separable metrisable topology $\mathcal{T}$ on X generating Σ, there is a corresponding separable metrisable topology $\mathcal{N}_{\mathcal{T}}$ on $\mathcal{M}_1$ (generating the natural σ-algebra on $\mathcal{M}_1$), in which a sequence (μ_n) of probability measures on X converges to a probability measure μ on X if and only if the following equivalent statements hold:*

(i) $\mu_n(g) \to \mu(g)$ *for every bounded $d_{\mathcal{T}}$-Lipschitz $g : X \to \mathbb{R}$;*

(ii) $\mu(g) \leq \liminf_{n\to\infty} \mu_n(g)$ *for every $\mathcal{T}$-lower-semicontinuous $g : X \to \mathbb{R} \cup \{\infty\}$ that is bounded below;*

(iii) $\mu(g) \geq \limsup_{n\to\infty} \mu_n(g)$ *for every $\mathcal{T}$-upper-semicontinuous $g : X \to \mathbb{R} \cup \{-\infty\}$ that is bounded above;*

(iv) $\mu(U) \leq \liminf_{n\to\infty} \mu_n(U)$ *for every $\mathcal{T}$-open $U \subset X$;*

(v) $\mu(G) \geq \limsup_{n\to\infty} \mu_n(G)$ *for every $\mathcal{T}$-closed $G \subset X$;*

where, in (i), $d_{\mathcal{T}}$ may be any separable metrisation of $\mathcal{T}$. $\mathcal{N}_{\mathcal{T}}$ is compact if and only if $\mathcal{T}$ is compact, and $\mathcal{N}_{\mathcal{T}}$ is Polish if and only if $\mathcal{T}$ is Polish.

$\mathcal{N}_{\mathcal{T}}$ is called the *narrow topology* (or *topology of weak convergence*) associated to $\mathcal{T}$. Observe that $\mu_n \to \mu$ in the narrow topology if and only if $\mu_n(g) \to \mu(g)$ for every bounded continuous $g : X \to \mathbb{R}$. (In fact, this is probably the most commonly given definition of the narrow topology.)

It is also worth saying that the equivalence of (i)–(v) does not actually rely on the metrisable topology $\mathcal{T}$ being separable. In addition to our above characterisations of the narrow topology, a further well-known characterisation is that $\mu_n \to \mu$ if and only $\mu_n(A) \to \mu(A)$ for every $A \in \Sigma$ with $\mu(\partial A) = 0$; this is proved in many textbooks on probability theory, but we will not need it here.

Also note that, by the above theorem, if (X, Σ) is standard then $\mathcal{M}_1$ (equipped with its natural σ-algebra) is standard.

Although the above theorem is (at least in most of its details) well-known, we will write out a proof of all but the last sentence[4]. We start with the following exercise:

Exercise 18. Fix a metric d on X. Show that for any $A \subset X$, the sequence $(g_n^A)_{n\in\mathbb{N}}$ of functions $g_n^A : X \to [0,1]$ given by $g_n^A(x) = \min(1, nd(x, X \setminus A))$ is an increasing sequence of Lipschitz functions converging pointwise to $\mathbb{1}_{A^\circ}$. (So g_n^A converges pointwise to $\mathbb{1}_A$ if and only if A is open.)

Proof of the equivalence of (i)–(v). Fix a metric d on X whose Borel σ-algebra is Σ. We will show (i)$\Rightarrow$(iv)$\Rightarrow$(ii); the rest is then clear.

Suppose we have $(\mu_n)_{n\in\mathbb{N}}$ and μ satisfying (i). Let $U \subset X$ be any open set. Using the monotone convergence theorem and Exercise 18,

$$\liminf_{n\to\infty} \mu_n(U) \;=\; \liminf_{n\to\infty} \sup_{m\in\mathbb{N}} \mu_n(g_m^U) \;\geq\; \sup_{m\in\mathbb{N}} \liminf_{n\to\infty} \mu_n(g_m^U) \;=\; \sup_{m\in\mathbb{N}} \mu(g_m^U) \;=\; \mu(U)$$

where g_m^U is as in Exercise 18. So (i)$\Rightarrow$(iv).

Now suppose $(\mu_n)_{n\in\mathbb{N}}$ and μ satisfy (iv). For each $m \in \mathbb{N}$, let $R_m := \{\frac{k}{2^m}\}_{k\in\mathbb{Z}\cup\{\infty\}}$ and define the function $h_m : \mathbb{R}\cup\{\infty\} \to \mathbb{R}\cup\{\infty\}$ by $h_m(y) = \sup(R_m \cap (-\infty, y))$. (So for $y < \infty$, h_m rounds y down to the nearest $\frac{1}{2^m}$-division that is strictly less than y.) Note that h_m increases pointwise to the identity function $\mathrm{id}_{\mathbb{R}\cup\{\infty\}}$ as $m \to \infty$, and that for any $m \in \mathbb{N}$, given any $c \in R_m$

$$h_m(y) \;=\; c + \frac{1}{2^m}\sum_{k=1}^{\infty} \mathbb{1}_{(c+\frac{k}{2^m},\infty]}(y) \quad \forall\; y \in (c,\infty].$$

Let $g : X \to \mathbb{R}\cup\{\infty\}$ be any lower-semicontinuous that is bounded below, with $c \in \mathbb{Z}$ being a strict lower bound of g. So for each $m \in \mathbb{N}$,

$$h_m(g(x)) \;=\; c + \frac{1}{2^m}\sum_{k=1}^{\infty} \mathbb{1}_{U_{k,m}}(x) \quad \forall\; x \in X$$

where $U_{k,m} = g^{-1}((c + \frac{k}{2^m}, \infty])$ for each $k, m \in \mathbb{N}$. So then, since the sequence $(h_m \circ g)$ is

[4] see e.g. Theorem 9.4 of here, and Exercise 22(B). Further facts along the same lines can be found in Theorem III.60 of here.

uniformly bounded below and increases pointwise to g, we have

$$\begin{aligned}
\liminf_{n\to\infty} \mu_n(g) &= \liminf_{n\to\infty} \sup_{m\in\mathbb{N}} \mu_n(h_m \circ g) \quad \text{(by MCT)} \\
&\geq \sup_{m\in\mathbb{N}} \liminf_{n\to\infty} \mu_n(h_m \circ g) \\
&= \sup_{m\in\mathbb{N}} \liminf_{n\to\infty} c + \frac{1}{2^m} \sum_{k=1}^{\infty} \mu_n(U_{k,m}) \quad \text{(by MCT)} \\
&\geq \sup_{m\in\mathbb{N}} c + \frac{1}{2^m} \sum_{k=1}^{\infty} \liminf_{n\to\infty} \mu_n(U_{k,m}) \\
&\qquad \text{(by Fatou's lemma, applied to the counting measure on } \mathbb{N}) \\
&\geq \sup_{m\in\mathbb{N}} c + \frac{1}{2^m} \sum_{k=1}^{\infty} \mu(U_{k,m}) \quad \text{(by (iv))} \\
&= \sup_{m\in\mathbb{N}} \mu(h_m \circ g) \quad \text{(by MCT)} \\
&= \mu(g) \quad \text{(by MCT).}
\end{aligned}$$

("MCT" stands for "the monotone convergence theorem".) This proves that (iv)⇒(ii). So we are done. □

We now consider metrisability of the convergence described in (i)–(v) above (which we refer to as "narrow convergence" or "weak convergence"). Recall that $[0,1]^{\mathbb{N}}$ equipped with the infinite product topology (the "Hilbert cube") is a compact metrisable space, with an exemplary metric being $d_\infty((x_n),(y_n)) = \sum_{n=1}^{\infty} \frac{1}{2^n}|x_n - y_n|$.

Lemma 19. *Fix a separable metric d on X whose Borel σ-algebra is Σ. Then there exists a countable set $\{g_k\}_{k\in\mathbb{N}}$ of d-Lipschitz functions $g_n : X \to [0,1]$ such that for any sequence (μ_n) in $\mathcal{M}_1$ and any $\mu \in \mathcal{M}_1$, μ_n converges narrowly to μ if and only if $\mu_n(g_k) \to \mu(g_k)$ for each $k \in \mathbb{N}$. Hence, if we let d_∞ be a metrisation of the topology of $[0,1]^{\mathbb{N}}$, then the function $d_{\mathcal{M}_1} : \mathcal{M}_1 \times \mathcal{M}_1 \to [0,\infty)$ given by*

$$d_{\mathcal{M}_1}(\mu_1,\mu_2) = d_\infty\big(\,(\mu_1(g_k))_{k\in\mathbb{N}}\,,\,(\mu_2(g_k))_{k\in\mathbb{N}}\,\big)$$

is a separable metric on $\mathcal{M}_1$ whose convergence is precisely narrow convergence.

Another metrisation of the narrow topology is the "Lévy-Prokhorov metric"; however, the above metrisation will be useful for us.

Proof of Lemma 19. Let $\tilde{\mathcal{C}}$ be a countable base for the topology induced by d, and let $\mathcal{C}$ be the collection of all finite unions of members of $\tilde{\mathcal{C}}$. Since $\mathcal{C}$ is countable, we can write $\mathcal{C} = \{U_r\}_{r\in\mathbb{N}}$. So for every open $U \subset X$ there exists a sequence $(r_m)_{m\in\mathbb{N}}$ of positive integers such that $\mathbb{1}_{U_{r_m}}$ increases pointwise to $\mathbb{1}_U$ as $m \to \infty$. For each $r \in \mathbb{N}$, let $(g_m^r)_{m\in\mathbb{N}}$ be an increasing sequence of d-Lipschitz functions $g_m^r : X \to [0,1]$ converging pointwise to $\mathbb{1}_{U_r}$ (e.g. as in Exercise 18). (The set $\{g_m^r\}_{r,m\in\mathbb{N}}$ will be precisely the countable set $\{g_k\}_{k\in\mathbb{N}}$ referred to in the statement of the lemma.)

Suppose we have $(\mu_n)_{n\in\mathbb{N}}$ and μ such that $\mu_n(g_m^r) \to \mu(g_m^r)$ for every r and m. As in the proof that (i)⇒(iv), we have that $\mu(U_r) \leq \liminf_{n\to\infty} \mu_n(U_r)$ for every r. If we then fix any open $U \subset X$ and let $(r_m)_{m\in\mathbb{N}}$ be a sequence of positive integers such that $\mathbb{1}_{U_{r_m}}$ increases pointwise to $\mathbb{1}_U$ as $m \to \infty$, we have (using the monotone convergence theorem)

$$\liminf_{n\to\infty} \mu_n(U) \;=\; \liminf_{n\to\infty} \sup_{m\in\mathbb{N}} \mu_n(U_{r_m}) \;\geq\; \sup_{m\in\mathbb{N}} \liminf_{n\to\infty} \mu_n(U_{r_m}) \;\geq\; \sup_{m\in\mathbb{N}} \mu(U_{r_m}) \;=\; \mu(U).$$

So μ_n converges narrowly to μ. □

Exercise 20. (A) Show that if X is equipped with a second-countable topology generating Σ and $\mathcal{C}$ is a countable subbase for this topology, then $\Sigma = \sigma(\mathcal{C})$. (B) Show that for any topological space $(T, \mathcal{T})$ and any function $f : X \to T$, if $\mathcal{D}$ is a subbase for $\mathcal{T}$ then $\{f^{-1}(U) : U \in \mathcal{D}\}$ is a subbase for the topology $\{f^{-1}(U) : U \in \mathcal{T}\}$ on X. (C) Show that for any topological space $(S, \mathcal{S})$, if $\mathcal{U}$ is a subbase for $\mathcal{S}$ then

$$\big\{ \{ (x_n) \in S^{\mathbb{N}} : x_k \in V \} : k \in \mathbb{N}, V \in \mathcal{U} \big\}$$

is a subbase for the product topology on $S^{\mathbb{N}}$.

Proof that $\mathcal{N}_{\mathcal{T}}$ generates the natural σ-algebra of $\mathcal{M}_1$.[5] Fix a separable metrisable topology $\mathcal{T}$ on X generating Σ, and let $\mathcal{B}(\mathcal{M}_1) = \sigma(\mathcal{N}_{\mathcal{T}})$ denote the Borel σ-algebra of the corresponding narrow topology $\mathcal{N}_{\mathcal{T}}$. Let $\mathfrak{K}$ be the natural σ-algebra of $\mathcal{M}_1$. As in Exercise 9, $\mathfrak{K}$ is the smallest σ-algebra on $\mathcal{M}_1$ with respect to which the map $\mu \mapsto \mu(g)$ is measurable for every measurable $g : X \to [0,1]$, i.e.

$$\mathfrak{K} \;=\; \sigma(\, \{\mu \in \mathcal{M}_1 : \mu(g) \in B\} \,:\, \text{measurable } g : X \to [0,1],\, B \in \mathcal{B}([0,1]) \,).$$

Letting $\mathcal{U}$ be a countable base (or subbase) for the topology on $[0,1]$ and letting $\{g_k\}_{k\in\mathbb{N}}$ be as in Lemma 19, we have (by Exercise 20) that

$$\mathcal{B}(\mathcal{M}_1) \;=\; \sigma(\, \{\mu \in \mathcal{M}_1 : \mu(g_k) \in V\} \,:\, k \in \mathbb{N},\, V \in \mathcal{U} \,).$$

Hence it is clear that $\mathcal{B}(\mathcal{M}_1) \subset \mathfrak{K}$.

Conversely, for any bounded continuous $g : X \to \mathbb{R}$, the map $\mu \mapsto \mu(g)$ is continuous and hence $\mathcal{B}(\mathcal{M}_1)$-measurable; and therefore, by Exercise 18, the map $\mu \mapsto \mu(U)$ is $\mathcal{B}(\mathcal{M}_1)$-measurable for every open $U \subset X$. Now the collection of all sets $A \in \Sigma$ for which the map $\mu \mapsto \mu(A)$ is $\mathcal{B}(\mathcal{M}_1)$-measurable is a λ-system. Hence, by the π-λ theorem, the map $\mu \mapsto \mu(A)$ is $\mathcal{B}(\mathcal{M}_1)$-measurable for every $A \in \Sigma$; in other words, $\mathfrak{K} \subset \mathcal{B}(\mathcal{M}_1)$. □

This completes our proof of Theorem 17.

Corollary 21. *Fix a separable metrisable topology on X generating Σ, and let $K \subset X$ be a non-empty compact set. Then the set $\mathcal{K}_K := \{\rho \in \mathcal{M}_1 : \rho(K) = 1\}$ is a compact subset of $\mathcal{M}_1$ (equipped with the narrow topology).*

Proof. Let $\mathcal{M}_1^K$ denote the set of Borel probability measures on K, equipped with the narrow topology associated to the topology on K induced from X. We know from Theorem 17 that $\mathcal{M}_1^K$ is compact, and it is clear that the map $\varphi : \mu \mapsto \mu(\,\cdot \cap K)$ from $\mathcal{M}_1^K$ to $\mathcal{K}_K$ is surjective (and in fact bijective). Hence it is sufficient to show that φ is continuous; but this is clear, since for any bounded continuous $g : X \to \mathbb{R}$, $g|_K$ is bounded and continuous. □

[5] I am grateful to Prof Dan Crisan for his help in the construction of this proof.

Exercise 22. Fix a separable metrisable topology on X (with Σ being the Borel σ-algebra). (A) Give an elementary proof (i.e. without citing Theorem 17 or Lemma 19) that if μ_1 and μ_2 are probability measures on X such that $\mu_1(g) = \mu_2(g)$ for every bounded continuous $g : X \to \mathbb{R}$, then $\mu_1 = \mu_2$. (B) Recall that, given two topological spaces T_1 and T_2, a *closed embedding* of T_1 into T_2 is a function $f : T_1 \to T_2$ such that $f(T_1)$ is a closed subset of T_2 and f maps T_1 homeomorphically into its image $f(T_1)$. (In the case that T_1 and T_2 are metric spaces, this is equivalent to saying that f is a continuous function under which divergent sequences in T_1 are mapped into divergent sequences in T_2.) Show that the map $x \mapsto \delta_x$ serves as a closed embedding of X into $\mathcal{M}_1$ (equipped with the narrow topology). Hence prove the "only if" statements in the last sentence of Theorem 17. (C) [*a.s. convergence implies convergence in distribution*] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let (g_n) be a sequence of measurable functions $g_n : \Omega \to X$ and let $g : \Omega \to X$ be a measurable function such that for $\mathbb{P}$-almost all $\omega \in \Omega$, $g_n(\omega) \to g(\omega)$ as $n \to \infty$. Show that $g_{n*}\mathbb{P}$ converges in the narrow topology to $g_*\mathbb{P}$.

Exercise 23. Find a sequence of probability measures (μ_n) on $\mathbb{N}$ converging in the narrow topology to a probability measure μ on $\mathbb{N}$, such that $\mu(\mathrm{id}_\mathbb{N}) < \liminf_{n\to\infty} \mu_n(\mathrm{id}_\mathbb{N})$.

1 Markov kernels

Recall that throughout this document, (X, Σ) is a measurable space, with $\mathcal{M}$ and $\mathcal{M}_1$ denoting respectively the space of measures and the space of probability measures on X.

A *Markov kernel on X* (which we will sometimes just call a "*kernel*") is an X-indexed family $(\mu_x)_{x\in X}$ of probability measures μ_x on X such that the map $x \mapsto \mu_x$ from X to $\mathcal{M}_1$ is measurable (which is equivalent to saying that the map $x \mapsto \mu_x(A)$ is measurable for all $A \in \Sigma$). Heuristically, one can regard a kernel $(\mu_x)_{x\in X}$ as a probabilistic description of a "random relocation procedure" for a particle in X: if the particle is at position x prior to relocation, then μ_x denotes the probability distribution for where the particle will be relocated to.

Note that the map $x \mapsto \delta_x$ is measurable; hence we may associate to each measurable function $f : X \to X$ the corresponding Markov kernel $(\delta_{f(x)})_{x\in X}$.

Given a kernel (μ_x) on X, we define an associated map $\mu^* : \mathcal{M}_1 \to \mathcal{M}_1$ by

$$\mu^*\rho(A) \;=\; \int_X \mu_x(A)\,\rho(dx)$$

for all $\rho \in \mathcal{M}_1$ and $A \in \Sigma$. By Lemma 8(A), this map is measurable. Observe that in the case that $(\mu_x) = (\delta_{f(x)})$ for some measurable $f : X \to X$, $\mu^*\rho$ is simply equal to $f_*\rho$. A heuristic interpretation of $\mu^*\rho$ is as follows: Suppose we have a random relocation procedure for a particle in X, whose transition probabilities are given by $(\mu_x)_{x\in X}$; and suppose the position of the particle *prior* to relocation is itself selected randomly (independently of the relocation procedure) with probability distribution ρ. Then, prior to the selection of the initial position of the particle, the probability distribution for where the particle will be subsequent to relocation is given by $\mu^*\rho$.[6]

Note that, given a probability measure ρ on X and a measurable function $g : X \to \mathbb{R}'$ with $\rho(g) \neq \text{NaN}$, Theorem 12 yeids that

$$\mu^*\rho(g) \;=\; \int_X \mu_x(g)\,\rho(dx).$$

(As in Exercise 13(A), the condition that $\rho(g) \neq \text{NaN}$ can be dropped if $(\mu_x) = (\delta_{f(x)})$ for some measurable $f : X \to X$.)

We now introduce the most fundamental object of study in ergodic theory: A probability measure ρ on X is said to be *stationary* (with respect to $(\mu_x)_{x\in X}$) if $\mu^*\rho = \rho$. In the particular case that $(\mu_x) = (\delta_{f(x)})$ for some measurable $f : X \to X$, this reduces to the following: a probability measure ρ on X is said to be *invariant with respect to* f if $f_*\rho = \rho$. (In this case, we also say that f is ρ*-preserving*.)

Exercise 24. Given a stationary probability measure ρ, show that for any measurable functions $g_1, g_2 : X \to \mathbb{R}'$ with $g_1(x) = g_2(x)$ for ρ-almost all x, $\mu_x(g_1) = \mu_x(g_2)$ for ρ-almost all x.

Exercise 25. For any probability measure ρ on X, define the probability measure $\boldsymbol{\mu}_\rho$ on $X \times X$ by

$$\boldsymbol{\mu}_\rho(B) \;=\; \int_X \int_X \mathbb{1}_B(x,y)\,\mu_x(dy)\,\rho(dx)$$

for all $B \in \Sigma \otimes \Sigma$. (So ρ is stationary if and only if $\rho(A) = \boldsymbol{\mu}_\rho(X \times A)$ for all $A \in \Sigma$.) Note that, combining Theorem 12 and Exercise 13(B),

$$\boldsymbol{\mu}_\rho(g) \;=\; \int_X \int_X g(x,y)\,\mu_x(dy)\,\rho(dx)$$

for any measurable $g : X \times X \to \mathbb{R}'$ with $\boldsymbol{\mu}_\rho(g) \neq \text{NaN}$. (A) Show that ρ is stationary if and only if for every $A \in \Sigma$, $\boldsymbol{\mu}_\rho(A \times (X \setminus A)) = \boldsymbol{\mu}_\rho((X \setminus A) \times A)$. (B) Show that for *any* measurable function $g : X \to \mathbb{R}'$,

$$\mu^*\rho(g) \;=\; \int_{X\times X} g(y)\,\boldsymbol{\mu}_\rho(d(x,y)).$$

Now let ρ be a stationary probability measure. Then we will say that a set $A \in \Sigma$ is ρ*-almost invariant* (with respect to (μ_x)) if $\mu_x(A) = 1$ for ρ-almost all $x \in A$. In the case that $(\mu_x) = (\delta_{f(x)})$ for some ρ-preserving measurable function $f : X \to X$, this reduces to the following: we will say that a set $A \in \Sigma$ is ρ*-almost invariant with respect to* f if $f(x) \in A$ for ρ-almost all $x \in A$ (i.e. if $\rho(A \setminus f^{-1}(A)) = 0$).

[6] One way to make this rigorous is as follows: Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable function $f : \Omega \times X \to X$ such that for each $x \in X$ and $A \in \Sigma$, $\mathbb{P}(\omega \in \Omega : f(\omega,x) \in A) = \mu_x(A)$. Given a probability measure ρ on X, Corollary 14 yields that $\mathbb{P}\otimes\rho(\,(\omega,x) \in \Omega\times X \;:\; f(\omega,x) \in A\,) = \mu^*\rho(A)$ for all $A \in \Sigma$. (Here, we treat the "random relocation procedure" as a *random self-mapping* of X. For more on this, see section 7.)

Exercise 26. (A) Show that the set of ρ-almost invariant sets forms a σ-algebra on X, and that a set $A \in \Sigma$ is ρ-almost invariant if and only if $\mu_x(A) = \mathbb{1}_A(x)$ for ρ-almost all $x \in X$. (B) Show that $A \in \Sigma$ is ρ-almost invariant if and only if $\boldsymbol{\mu}_\rho(A \times (X \setminus A)) = 0$. (C) Show that $A \in \Sigma$ is ρ-almost invariant if and only if either (i) $\rho(A) = 0$, or (ii) $\rho(A) > 0$ and the probability measure $C \mapsto \frac{\rho(A \cap C)}{\rho(A)}$ is stationary with respect to (μ_x). (D) In the case that $(\mu_x) = (\delta_{f(x)})$ for some ρ-preserving measurable function $f : X \to X$, show that for any $A \in \Sigma$ the following are equivalent:

- A is ρ-almost invariant with respect to f;
- $\rho(f^{-1}(A) \setminus A) = 0$;
- $\rho(A \triangle f^{-1}(A)) = 0$;

and show that a *sufficient* condition for these statements to be true is the following: $f(A) \setminus A$ is a ρ-null set.

Now let us still assume that ρ is a stationary probability measure. We will say that a function $g : X \to \mathbb{R}'$ is *ρ-almost invariant* (with respect to (μ_x)) if g is measurable with respect to the σ-algebra of ρ-almost invariant sets. If $(\mu_x) = (\delta_{f(x)})$ for some ρ-preserving measurable function $f : X \to X$, then we will say in this case that g is *ρ-almost invariant with respect to f*.

Since every ρ-null $A \in \Sigma$ is ρ-almost invariant, it follows that if $g_1, g_2 : X \to \mathbb{R}'$ are measurable functions with $g_1(x) = g_2(x)$ for ρ-almost all x, then g_1 is ρ-almost invariant if and only if g_2 is ρ-almost invariant.

Theorem 27. *Let ρ be a stationary probability measure. (A) A measurable function $g : X \to \mathbb{R}'$ is ρ-almost invariant if and only if for ρ-almost every $x \in X$,*

$$\mu_x(\, y \in X \,:\, g(y) = g(x)\,) \;=\; 1.$$

(So in the case that $(\mu_x) = (\delta_{f(x)})$, g is ρ-almost invariant if and only if for ρ-almost every $x \in X$, $g(f(x)) = g(x)$.) (B) For any measurable $g : X \to \bar{\mathbb{R}}$, if either

(a) $\rho(g^+) < \infty$ and $\mu_x(g) \geq g(x)$ for ρ-almost all $x \in X$; or

(b) $\rho(g^-) < \infty$ and $\mu_x(g) \leq g(x)$ for ρ-almost all $x \in X$;

then g is ρ-almost invariant. In the case that $(\mu_x) = (\delta_{f(x)})$ for some measurable map $f : X \to X$ (so $f_*\rho = \rho$), the conditions on $\rho(g^+)$ and $\rho(g^-)$ can be dropped—that is: for any measurable $g : X \to \bar{\mathbb{R}}$, if either $g(f(x)) \geq g(x)$ for ρ-almost all $x \in X$ or $g(f(x)) \leq g(x)$ for ρ-almost all $x \in X$, then g is ρ-almost invariant.

Proof. (A) Suppose g is ρ-almost invariant. Let $X_1 := g^{-1}(\bar{\mathbb{R}})$, $X_2 := g^{-1}(\{\mathrm{NaN}\}) = X \setminus X_1$, and let $\mathcal{I}_\rho^{X_1}$ be the set of ρ-almost invariant subsets of X_1. Let $(g_n)_{n \in \mathbb{N}}$ be a sequence of functions $g_n : X \to \mathbb{R}'$ such that

- $g_n(x) = \mathrm{NaN}$ for all $x \in X_2$ and $n \in \mathbb{N}$;
- $g_n|_{X_1}$ is $\mathcal{I}_\rho^{X_1}$-simple for all $n \in \mathbb{N}$;

- $g_n|_{X_1}$ converges pointwise to $g|_{X_1}$ as $n \to \infty$.

It is clear that g_n is ρ-almost invariant for each n. For each n, since g_n only takes finitely many different values and the preimage of each of these values is a ρ-almost invariant set, we have that for ρ-almost all $x \in X$,

$$\mu_x(\, y \in X \,:\, g_n(y) = g_n(x)\,) \;=\; 1.$$

Obviously, therefore, ρ-almost every $x \in X$ has the property that

$$\mu_x(\, y \in X \,:\, g_n(y) = g_n(x)\,) \;=\; 1 \;\text{ for all } n \in \mathbb{N}$$

and therefore

$$\mu_x(\, y \in X \,:\, g_n(y) = g_n(x) \text{ for all } n \in \mathbb{N}) \;=\; 1$$

and therefore

$$\mu_x(\, y \in X \,:\, g(y) = g(x)\,) \;=\; 1.$$

as required. Now on the other hand, suppose we have a ρ-full set $X' \in \Sigma$ such that $\mu_x(\, y \in X \,:\, g(y) = g(x)\,) = 1$ for all $x \in X'$. Given any measurable $B \subset \mathbb{R}'$, for any $x \in g^{-1}(B) \cap X'$,

$$\mu_x(g^{-1}(B)) \;\geq\; \mu_x(g^{-1}(\,\{g(x)\}\,)) \;=\; 1$$

and so $g^{-1}(B)$ is ρ-almost invariant. Hence g is ρ-almost invariant.

(B) Let us consider case (a) (case (b) is similar). Fix any $a \in \mathbb{R}$; we will show that the set $A := g^{-1}([a, \infty])$ is ρ-almost invariant. On the basis of Exercise 25(B), since $\rho(g^+) < \infty$ we know that

$$\int_{A \times X} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \;\neq\; \text{NaN}.$$

Since $a > -\infty$ and g^+ is integrable with respect to ρ, the map $x \mapsto \mathbb{1}_A(x) g(x)$ is integrable with respect to ρ. Hence, again using Exercise 25(B), we have that

$$\int_{X \times A} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \;\in\; \mathbb{R}$$

and so

$$\int_{A \times A} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \;\in\; \mathbb{R}.$$

Consequently, we have the following:

$$\begin{aligned}
\int_{A \times (X \setminus A)} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) &= \int_{A \times X} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \;-\; \int_{A \times A} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \\
&= \int_A \int_X g(y)\, \mu_x(dy)\, \rho(dx) \;-\; \int_{A \times A} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \\
&\geq \int_A g(x)\, \rho(dx) \;-\; \int_{A \times A} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \\
&= \int_{X \times A} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \;-\; \int_{A \times A} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)) \\
&= \int_{(X \setminus A) \times A} g(y)\, \boldsymbol{\mu}_\rho(d(x,y)).
\end{aligned}$$

Given that (by Exercise 25(A)) $\boldsymbol{\mu}_\rho(A\times(X\smallsetminus A)) = \boldsymbol{\mu}_\rho((X\smallsetminus A)\times A)$, and yet the range of g on $X\smallsetminus A$ is strictly lower than the range of g on A, it must follow that $\boldsymbol{\mu}_\rho(A\times(X\smallsetminus A)) = 0$. So, by Exercise 26(B), A is ρ-almost invariant.

Now consider the case that $(\mu_x) = (\delta_{f(x)})$ for some measurable map $f : X \to X$, and suppose that $g(f(x)) \geq g(x)$ for ρ-almost all $x \in X$. Fix any $a \in \mathbb{R}$. For ρ-almost every x with $g(x) \geq a$, $g(f(x)) \geq a$; so $g^{-1}([a,\infty])$ is ρ-almost invariant. Hence g is ρ-almost invariant. (The case that $g(f(x)) \leq g(x)$ ρ-a.e. is similar.) □

Exercise 28. Let ρ be a stationary probability measure. Show by elementary means (i.e. without citing Theorem 27) that for any ρ-integrable $g : X \to \bar{\mathbb{R}}$ the following are equivalent:

- $\mu_x(g) = g(x)$ for ρ-almost all $x \in X$;
- $\mu_x(g) \geq g(x)$ for ρ-almost all $x \in X$;
- $\mu_x(g) \leq g(x)$ for ρ-almost all $x \in X$.

The following proposition will be our main tool in proving important results concerning the "structure" of the space of stationary probability measures (Theorem 34(i)⇔(iv), and Theorem 36).

Proposition 29. *(A) Let ρ_1 and ρ_2 be stationary probability measures on X, let*

$$\rho_2 \;=\; \rho_{sing} + \rho_{cont}$$

denote the Radon-Nikodym decomposition of ρ_2 into its singular and absolutely continuous parts with respect to ρ_1, and suppose that $\rho_{cont}(X) > 0$. Then the probability measure $\rho_c = \frac{\rho_{cont}(\cdot)}{\rho_{cont}(X)}$ is stationary. (B) Let ρ_1 and ρ_2 be stationary probability measures, with ρ_2 being absolutely continuous with respect to ρ_1, and let $g : X \to [0,\infty)$ be a version of $\frac{d\rho_2}{d\rho_1}$. Then g is ρ_1-almost invariant.

Proof. (A) Take a set $A \in \Sigma$ with $\rho_1(A) = 1$ and $\rho_{sing}(A) = 0$. Then $\rho_2(A) = \rho_{cont}(A) = \rho_{cont}(X)$, and for any Σ-measurable $B \subset A$, $\rho_2(B) = \rho_{cont}(B)$. Hence

$$\rho_c(C) \;=\; \frac{\rho_2(A\cap C)}{\rho_2(A)}$$

for any $C \in \Sigma$, and so (by Exercise 26(C)) it is sufficient to show that A is ρ_2-almost invariant. Since ρ_1 is stationary, we have that $\rho_1(x \in A : \mu_x(A) < 1) = 0$, and so

$$\rho_2(x \in A : \mu_x(A) < 1) \;=\; \rho_{cont}(x \in A : \mu_x(A) < 1) \;=\; 0.$$

Hence A is ρ_2-invariant. (B) Fix any $a \in [0,\infty)$, and let $A = g^{-1}([a,\infty])$. We will show that A is ρ_1-almost invariant. We have

$$\begin{aligned}
\int_{A\times(X\setminus A)} g(x)\,\boldsymbol{\mu}_{\rho_1}(d(x,y)) &= \int_A \mu_x(X\setminus A)g(x)\,\rho_1(dx)\\
&= \int_A \mu_x(X\setminus A)\,\rho_2(dx)\\
&= \boldsymbol{\mu}_{\rho_2}(A\times(X\setminus A))\\
&= \boldsymbol{\mu}_{\rho_2}((X\setminus A)\times A)\\
&= \int_{(X\setminus A)\times A} g(x)\,\boldsymbol{\mu}_{\rho_1}(d(x,y)).
\end{aligned}$$

As in the proof of Theorem 27(B), we have that the range of g on $X\setminus A$ is strictly lower than the range of g on A and therefore A is ρ_1-almost invariant. □

A probability measure ρ is said to be *ergodic* (with respect to (μ_x)) if it is stationary and assigns trivial measure (i.e. either 0 or 1) to every ρ-almost invariant set. (Other equivalent formulations will be given in Theorem 34 and in Section 3.) If $(\mu_x)=(\delta_{f(x)})$ for some measurable $f:X\to X$, then we will say in this case that ρ is *ergodic with respect to f*.

2 Semigroups of kernels and ergodicity

From now on, $\mathbb{T}^+$ denotes either $\mathbb{N}\cup\{0\}$ or $[0,\infty)$. We equip $\mathbb{T}^+$ with its standard topology and the corresponding Borel σ-algebra. λ denotes the counting measure on $\mathbb{T}^+$ if $\mathbb{T}^+=\mathbb{N}\cup\{0\}$, and the Lebesgue measure on $\mathbb{T}^+$ if $\mathbb{T}^+=[0,\infty)$. For any set $S\subset[0,\infty)$, we write $\mathbb{T}_S$ to mean $\mathbb{T}^+\cap S$.

Given two Markov kernels $(\mu_x)_{x\in X}$ and $(\nu_x)_{x\in X}$ on X, we refer to the Markov kernel $(\nu^*\mu_x)_{x\in X}$ as the *composition* of (ν_x) with (μ_x). We refer to $(\delta_x)_{x\in X}$ as the *identity kernel* on X. One can easily check the set of Markov kernels on X forms a monoid under composition. We will say that two Markov kernels (μ_x) and (ν_x) *commute* if $\nu^*\mu_x=\mu^*\nu_x$ for all $x\in X$.

Note that for any two measurable functions $f_1,f_2:X\to X$, $(\delta_{f_2\circ f_1(x)})_{x\in X}$ is precisely the composition of $(\delta_{f_2(x)})_{x\in X}$ with $(\delta_{f_1(x)})_{x\in X}$.

Exercise 30. Let (μ_x) and (ν_x) be kernels on X. (A) Given any probability measure ρ on X, show that $(\nu^*\mu)^*\rho=\nu^*(\mu^*\rho)$ and $\delta^*\rho=\rho$. (Here, $(\nu^*\mu)^*$ denotes the map on $\mathcal{M}_1$ associated with the Markov kernel $(\nu^*\mu_x)_{x\in X}$, and δ^* denotes the map on $\mathcal{M}_1$ associated with the identity kernel.) (B) Given any bounded measurable $g:X\to\mathbb{R}$, show that for all $x\in X$, $\mu_x(\nu_.(g))=\nu^*\mu_x(g)$ and $\delta_x(g)=g(x)$. (Here, $\nu_.(g)$ denotes the function from X to $\mathbb{R}$ sending $y\mapsto\nu_y(g)$.)

Exercise 30 can be summarised by saying that the map $((\mu_x)_{x\in X},\rho)\mapsto\mu^*\rho$ is a left monoid action of the space of Markov kernels upon the set $\mathcal{M}_1$, and the map $((\mu_x)_{x\in X},g)\mapsto\mu_.(g)$ is a right monoid action of the space of Markov kernels upon the set of bounded measurable functions $g:X\to\mathbb{R}$.

A *semigroup*[7] *of Markov kernels on* X (which we will sometimes just call a "*semigroup*") is a $\mathbb{T}^+$-indexed family $(\mu_x^t)_{x\in X,\, t\in\mathbb{T}^+}$ of Markov kernels $(\mu_x^t)_{x\in X}$ on X such that (μ_x^0) is the identity kernel and (μ_x^{s+t}) is the composition of (μ_x^t) with (μ_x^s) for all $s,t\in\mathbb{T}^+$ (i.e., such that the map $t\mapsto(\mu_x^t)_{x\in X}$ serves as a monoid homomorphism from $(\mathbb{T}^+,+)$ to the space of Markov kernels on X). Writing this in full:

$$\begin{aligned}\mu_x^0(A) &= \mathbb{1}_A(x),\\ \mu_x^{s+t}(A) &= \int_X \mu_y^t(A)\,\mu_x^s(dy)\end{aligned}$$

for all $x\in X$, $A\in\Sigma$ and $s,t\in\mathbb{T}^+$. (The latter equation is sometimes referred to as the "Chapman-Kolmogorov equation".) Obviously, this implies that for any $s,t\in\mathbb{T}^+$ the kernels $(\mu_x^s)_{x\in X}$ and $(\mu_x^t)_{x\in X}$ commute. Note that for any kernel (μ_x) on X there is a unique discrete-time semigroup $(\mu_x^n)_{x\in X,\, n\in\mathbb{N}\cup\{0\}}$ on X such that $(\mu_x^1)=(\mu_x)$.

Semigroups of Markov kernels typically appear as the family of transition probabilities associated to a Markov process (Section 4) or stochastic flow / random dynamical system (Section 7, in particular Proposition 140, which may be regarded as a version of the "Chapman-Kolmogorov theorem").

Recall that an *autonomous dynamical system* $(f^t)_{t\in\mathbb{T}^+}$ on X is a family of measurable functions $f^t: X\to X$ such that f^0 is the identity map and $f^{s+t}=f^t\circ f^s$ for all $s,t\in\mathbb{T}^+$. (For convenience, we will write $f^{-t}(A)$ as a shorthand for $(f^t)^{-1}(A)$.) We may associate with each autonomous dynamical system $(f^t)_{t\in\mathbb{T}^+}$ on X a corresponding semigroup of kernels $(\delta_{f^t(x)})_{x\in X,\, t\in\mathbb{T}^+}$. Also note that by Exercise 30(A), for any semigroup (μ_x^t) on X, $(\mu^{t*})_{t\in\mathbb{T}^+}$ defines an autonomous dynamical system on $\mathcal{M}_1$.

Given a semigroup (μ_x^t) on X, we will say that a probability measure ρ is *stationary* with respect to the semigroup (μ_x^t) if it is stationary with respect to the kernel $(\mu_x^t)_{x\in X}$ for every $t\in\mathbb{T}^+$; if $(\mu_x^t)=(\delta_{f^t(x)})$ for some autonomous dynamical system (f^t), then we will say in this case that ρ is *invariant with respect to* (f^t) (or that (f^t) is ρ*-preserving*).

Given a stationary probability measure ρ of the semigroup (μ_x^t), we will say that a set $A\in\Sigma$ or a measurable function $g: X\to\mathbb{R}'$ is ρ*-almost invariant* with respect to the semigroup (μ_x^t) if it is ρ-almost invariant with respect to the kernel $(\mu_x^t)_{x\in X}$ for every $t\in\mathbb{T}^+$. Obviously, Proposition 29 holds for semigroups of kernels just as it does for individual kernels.

Now it is essentially a tautology that a function is measurable with respect to the intersection of a collection of σ-algebras if and only if it is measurable with respect to each member of the collection. Hence, given a semigroup of kernels (μ_x^t) on X and a stationary probability measure ρ, a measurable function $g: X\to\mathbb{R}'$ is ρ-almost invariant if and only if it is measurable with respect to the σ-algebra of ρ-almost invariant sets.

[7] Generally, a "semigroup" simply means a set equipped with an associative binary operator; however, the term is often used specifically in connection with the image of $\mathbb{T}^+$ under a monoid homomorphism.

We will say that a probability measure ρ is *ergodic* (with respect to the semigroup (μ_x^t)) if it is stationary and assigns trivial measure (i.e. either 0 or 1) to every ρ-almost invariant set. Observe that if a stationary probability measure ρ is ergodic with respect to the kernel $(\mu_x^{t'})_{x\in X}$ for some $t' \in \mathbb{T}^+$ then it is ergodic with respect to the semigroup (μ_x^t).

We define ρ-almost invariance and ergodicity with respect to an autonomous dynamical system (f^t) as being the same as ρ-almost invariance and ergodicity with respect to the semigroup of kernels $(\delta_{f^t(x)})$.

Proposition 31. *Suppose $\mathbb{T}^+ = \mathbb{N} \cup \{0\}$; then the notions of stationarity, almost-invariance (of sets or functions) and ergodicity are the same for a semigroup of kernels (μ_x^n) as they are for the kernel (μ_x^1).*

Proof. By Exercise 30(A), it is clear that a probability measure ρ is stationary with respect to the semigroup (μ_x^n) if and only if it is stationary with respect to the kernel (μ_x^1). Now fix a stationary probability measure ρ and let $A \in \Sigma$ be a set that is ρ-almost invariant under the kernel (μ_x^1). Obviously A is ρ-almost invariant under the kernel (μ_x^0) (since every measurable set is ρ-almost invariant under the identity kernel). Now fix any $k \in \mathbb{N}$ such that A is ρ-almost invariant under the kernel (μ_x^k); we will show that A is ρ-almost invariant under the kernel (μ_x^{k+1}). Let $\tilde{A} := \{y \in A : \mu_y^1(A) = 1\}$. Since $\rho(A \smallsetminus \tilde{A}) = 0$ and ρ is stationary with respect to (μ_x^k), it follows that for ρ-almost all $x \in X$, $\mu_x^k(A \smallsetminus \tilde{A}) = 0$. But we also know that for ρ-almost all $x \in A$, $\mu_x^k(A) = 1$. Combining these two facts, we have that for ρ-almost all $x \in A$, $\mu_x^k(\tilde{A}) = 1$ and therefore

$$\mu_x^{k+1}(A) \;=\; \int_X \mu_y^1(A)\,\mu_x^k(dy) \;\geq\; \int_{\tilde{A}} \mu_y^1(A)\,\mu_x^k(dy) \;=\; 1.$$

Hence A is ρ-almost invariant under the kernel (μ_x^{k+1}). It follows by induction that A is ρ-almost invariant under the whole semigroup (μ_x^n). The rest is then immediate. $\square$

Exercise 32. Suppose $\mathbb{T}^+ = [0, \infty)$, and let (μ_x^t) be a semigroup of kernels on X. (A) Given a probability measure ρ on X, show that if there exists a Lebesgue-positive measure set $R \subset [0, \infty)$ such that ρ is stationary with respect to the kernel $(\mu_x^s)_{x\in X}$ for all $s \in R$, then ρ is stationary with respect to the semigroup (μ_x^t). (Hint: reduce the problem to a problem about fixed points of autonomous dynamical systems; the Lebesgue density theorem may also be useful.) (B) Hence, given a stationary probability measure ρ of the semigroup (μ_x^t) and a set $A \in \Sigma$, show that if there exists a Lebesgue-positive measure set $R \subset [0, \infty)$ such that A is ρ-almost invariant under the kernel $(\mu_x^s)_{x\in X}$ for all $s \in R$, then A is ρ-almost invariant under the semigroup (μ_x^t).

We now come to our first important theorem characterising ergodicity. We start with the following exercise:

Exercise 33. Given a probability measure ρ on X, we will say that a measurable function $g : X \to \mathbb{R}'$ is *ρ-almost constant* if there exists $c \in \mathbb{R}'$ such that $g(x) = c$ for ρ-almost all $x \in X$. Show that a measurable function $g : X \to \mathbb{R}'$ is ρ-almost constant if and only if it is measurable with respect to the σ-algebra of ρ-trivial measure sets.

Now let (μ_x^t) be a semigroup of kernels on X. Note that any convex combination of stationary probability measures is stationary; i.e. the set of stationary probability measures is a convex set (within the vector space of functions from Σ to $\mathbb{R}$).

Theorem 34. *For any stationary probability measure ρ, the following are equivalent:*

(i) *ρ is ergodic;*

(ii) *any measurable function $g : X \to \mathbb{R}'$ that is ρ-almost invariant is ρ-almost constant;*

(iii) *the only stationary probability measure that is absolutely continuous with respect to ρ is ρ itself;*

(iv) *ρ is an extreme point of the convex set of stationary probability measures.*

Proof. (i)⇒(ii) follows from Exercise 33. (ii)⇒(iii) follows from Proposition 29(B). Now given two probability measures ρ_1 and ρ_2 on X, it is clear that ρ_1 and ρ_2 are absolutely continuous with respect to any strict convex combination of ρ_1 and ρ_2; hence (iii)⇒(iv). Finally, if $A \in \Sigma$ is a ρ-almost invariant set that is not of ρ-trivial measure, then ρ can be expressed as a convex combination of $\frac{\rho(A \cap \cdot)}{\rho(A)}$ and $\frac{\rho(\cdot \setminus A)}{\rho(X \setminus A)}$, which are stationary probability measures by parts (A) and (C) of Exercise 26; hence (iv)⇒(i). □

Recall that two measures m_1 and m_2 on X are said to be *mutually singular* if there exists $A \in \Sigma$ such that $m_1(A) = 0$ and $m_2(X \setminus A) = 0$. In the proof that (iv)⇒(i) in Theorem 34, we actually see that any non-stationary probability measure can be expressed as a strict convex combination of two mutually singular probability measures.

A set $\mathcal{S} \subset \mathcal{M}$ of measures on X is said to be mutually singular if for all $m_1, m_2 \in \mathcal{S}$, m_1 and m_2 are mutually singular.

Exercise 35. Show that if $\mathcal{S}$ is a mutually singular set of measures on X, then for any finite or countable subset $\mathcal{C}$ of $\mathcal{S}$ one can associate to each $m \in \mathcal{S}$ an m-full measure set $A_m \in \Sigma$ in such a way that for any distinct $m_1, m_2 \in \mathcal{S}$, $m_1(A_{m_2}) = 0$.

Theorem 36. *The set of ergodic probability measures is mutually singular.*

Proof. This follows immediately from Proposition 29(A) and characterisation (iii) of ergodicity in Theorem 34. □

3 Ergodicity in measurable semigroups

Let (μ_x) be a Markov kernel on X. We will say that a set $A \in \Sigma$ is *strictly forward-invariant* under (μ_x) if for every $x \in X$, $\mu_x(A) = 1$ if and only if $x \in A$. We will say that $A \in \Sigma$ is *strictly backward-invariant* under (μ_x) if for every $x \in X$, $\mu_x(A) = 0$ if and only $x \in X \setminus A$; this is equivalent to saying that $X \setminus A$ is strictly forward-invariant. We will say that a bounded measurable function $g : X \to \mathbb{R}$ is *strictly invariant* under (μ_x) if for every $x \in X$, $g(x) = \mu_x^t(g)$.

Now (and throughout the rest of this section) let (μ_x^t) be a semigroup of Markov kernels on X. We will say that a set $A \in \Sigma$ is strictly forward-invariant (resp. strictly backward-invariant) under the semigroup (μ_x^t) if and only if it is strictly forward-invariant (resp. strictly backward-invariant) under the kernel $(\mu_x^t)_{x \in X}$ for all $t \in \mathbb{T}^+$. We will say that a bounded measurable function $g : X \to \mathbb{R}$ is strictly invariant under the semigroup (μ_x^t) if and only if it is strictly invariant under the kernel $(\mu_x^t)_{x \in X}$ for all $t \in \mathbb{T}^+$.

Proposition 37. *Suppose $\mathbb{T}^+ = \mathbb{N} \cup \{0\}$; then the notions of strict forward-invariance (of sets), strict backward-invariance (of sets) and strict invariance (of functions) are the same for the semigroup (μ_x^n) as they are for the kernel (μ_x^1).*

Proof. We first claim that, given a set $A \in \Sigma$ with $\mu_x^1(A) = 1$ for all $x \in A$, we have that $\mu_x^n(A) = 1$ for all $x \in A$ and $n \in \mathbb{N} \cup \{0\}$. The case that $n = 0$ is clear, since (μ_x^0) is the identity kernel. Now fix any $k \in \mathbb{N}$ such that $\mu_x^k(A) = 1$ for all $x \in A$. Then for any $x \in A$,

$$\mu_x^{k+1}(A) \;=\; \int_X \mu_y^1(A)\,\mu_x^k(dy) \;\geq\; \int_A \mu_y^1(A)\,\mu_x^k(dy) \;=\; 1.$$

Hence, our claim is proved by induction. We now claim that, given a set $A \in \Sigma$ with $\mu_x^1(A) < 1$ for all $x \in X \setminus A$, we have that $\mu_x^n(A) < 1$ for all $x \in X \setminus A$ and $n \in \mathbb{N} \cup \{0\}$. The case that $n = 0$ is clear, since (μ_x^0) is the identity kernel. Now fix any $k \in \mathbb{N}$ such that $\mu_x^k(A) < 1$ for all $x \in X \setminus A$. Then for any $x \in X \setminus A$,

$$\mu_x^{k+1}(X \setminus A) \;=\; \int_X \mu_y^1(X \setminus A)\,\mu_x^k(dy) \;\geq\; \int_{X \setminus A} \mu_y^1(X \setminus A)\,\mu_x^k(dy) \;>\; 0.$$

Hence, our claim is proved by induction. Combining these two claims gives that if a set $A \in \Sigma$ is strictly forward-invariant under the kernel (μ_x^1) then it is strictly forward-invariant under the whole semigroup (μ_x^n). The same is then also true for the notion of strict backward-invariance, since (both for a kernel and for a semigroup) a strictly backward-invariant set is precisely a set whose complement is strictly forward-invariant.

Now let $g : X \to \mathbb{R}$ be a bounded measurable function that is strictly invariant under the kernel (μ_x^1). Again, obviously g is strictly invariant under the kernel (μ_x^0). Now fix any $k \in \mathbb{N}$ such that g is strictly invariant under the kernel (μ_x^k). For any $x \in X$, using Theorem 12 we have that

$$\mu_x^{k+1}(g) \;=\; \int_X \mu_y^1(g)\,\mu_x^k(dy) \;=\; \int_X g(y)\,\mu_x(dy) \;=\; g(x).$$

Hence, by induction, g is strictly invariant under the semigroup (μ_x^n). □

Exercise 38. Given an autonomous dynamical system $(f^t)_{t \in \mathbb{T}^+}$ on X, we will say that a set $A \in \Sigma$ is *strictly invariant under* $(f^t)_{t \in \mathbb{T}^+}$ if $f^{-t}(A) = A$ for all $t \in \mathbb{T}^+$, and we will say that a measurable function $g : X \to \mathbb{R}'$ is *strictly invariant under* $(f^t)_{t \in \mathbb{T}^+}$ if $g \circ f^t = g$ for all $t \in \mathbb{T}^+$. (So a set A is strictly invariant if and only if its indicator function $\mathbb{1}_A$ is strictly invariant; and a bounded measurable function $g : X \to \mathbb{R}$ is strictly invariant under (f^t) if and only if it is strictly invariant under the semigroup $(\delta_{f^t(x)})$.) (A) Show that for any $A \in \Sigma$, the following are equivalent:

- A is strictly invariant under (f^t);
- A is strictly forward-invariant under $(\delta_{f^t(x)})$;
- A is strictly backward-invariant under $(\delta_{f^t(x)})$.

(B) Show that the set of strictly invariant sets under (f^t) is a σ-algebra, and that a measurable function $g : X \to \mathbb{R}'$ is strictly invariant under (f^t) if and only if it is measurable with respect to this σ-algebra.

Now we will say that the semigroup (μ_x^t) is *measurable* if the map $(t,x) \mapsto \mu_x^t$ from $\mathbb{T}^+ \times X$ to $\mathcal{M}_1$ is measurable. (So in the case that $\mathbb{T}^+ = \mathbb{N} \cup \{0\}$, *every* semigroup of Markov kernels is measurable.)

Theorem 39. *Suppose that (μ_x^t) is measurable, and let ρ be a stationary probability measure. Then the following are equivalent:*

(i) ρ is ergodic;

(ii) ρ assigns trivial measure to every strictly forward-invariant set;

(iii) ρ assigns trivial measure to every strictly backward-invariant set;

(iv) every bounded measurable function $g : X \to \mathbb{R}$ that is strictly invariant is ρ-almost constant.

Note that by Propositions 31 and 37, Theorem 39 (which we have stated for measurable semigroups of kernels) also holds for individual Markov kernels.

Remark 40. An autonomous dynamical system (f^t) on X is said to be *measurable* if the map $(t,x) \mapsto f^t(x)$ is measurable. It is clear that if (f^t) is measurable then $(\delta_{f^t(x)})$ is measurable. So as a special case of Theorem 39 (together with Exercise 38(A)), an invariant probability measure of a measurable autonomous dynamical system is ergodic if and only if it assigns trivial measure to every strictly invariant set.

Before proving Theorem 39, let us introduce some further concepts for which Theorem 39 has important implications.

We say that a set $A \in \Sigma$ is *forward-invariant* (under the semigroup (μ_x^t)) if $\mu_x^t(A) = 1$ for all $x \in A$ and $t \in \mathbb{T}^+$; in the case that $(\mu_x^t) = (\delta_{f^t(x)})$ for some autonomous dynamical system (f^t), this reduces to the following: we will say that a set $A \in \Sigma$ is *forward-invariant under* (f^t) if $f^t(A) \subset A$ for all $t \in \mathbb{T}^+$. We say that $A \in \Sigma$ is *backward-invariant* if $\mu_x^t(A) = 0$ for all $x \in X \setminus A$ and $t \in \mathbb{T}^+$; so A is backward-invariant if and only if $X \setminus A$ is forward-invariant. In the case that $(\mu_x^t) = (\delta_{f^t(x)})$ for some autonomous dynamical system (f^t), this reduces to the following: we will say that a set $A \in \Sigma$ is *backward-invariant under* (f^t) if $f^{-t}(A) \subset A$ for all $t \in \mathbb{T}^+$.

It is easy to show that the set of forward-invariant sets is closed under countable intersections and relatively closed in Σ under arbitrary unions (and so, in particular, is closed under countable unions).

Now we say that a measurable function $g : X \to \mathbb{R} \cup \{-\infty\}$ which is bounded above is *super-invariant* if $\mu_x^t(g) \geq g(x)$ for all $x \in X$ and $t \in \mathbb{T}^+$; so a set $A \in \Sigma$ is forward-invariant if and only if $\mathbb{1}_A$ is super-invariant. We say that a measurable function $g : X \to \mathbb{R} \cup \{\infty\}$ which is bounded below is *sub-invariant* if $\mu_x^t(g) \leq g(x)$ for all $x \in X$ and $t \in \mathbb{T}^+$; so a set

$A \in \Sigma$ is backward-invariant if and only if $\mathbb{1}_A$ is sub-invariant. Obviously, a function g is super-invariant if and only if $-g$ is sub-invariant.

All these notions (which we have just introduced for semigroups of kernels) can also be defined for individual kernels, and (following the proof of Proposition 37) it is easy to show that if $\mathbb{T}^+ = \mathbb{N} \cup \{0\}$ then these notions are the same for the semigroup (μ_x^n) as they are for the kernel (μ_x^1). (In fact, we already proved this for forward-invariance, in the first "claim" within the proof of Proposition 37.)

Note that, given a stationary probability measure ρ, the notion of forward-invariance "lies between" ρ-almost invariance and strict forward-invariance; so as a "special case" of Theorem 39, if (μ_x^t) is measurable then ρ is ergodic if and only if it assigns trivial measure to every forward-invariant set. (And obviously, the same statement holds for backward-invariant sets.) Moreover, using Theorem 27, super-invariance (of bounded-above measurable functions) "lies between" ρ-almost invariance (of measurable functions) and strict invariance (of bounded measurable functions). So once again, as a "special case" of Theorem 39, if (μ_x^t) is measurable then ρ is ergodic if and only if all super-invariant functions are ρ-almost constant; and likewise for sub-invariant functions.

Let us now start the proof of Theorem 39. It is clear that (i) implies (ii) and (iii), and by Theorem 34 we know that (i) implies (iv). So it remains to show that (ii) implies (i) (from which it immediately follows that (iii) implies (i)) and that (iv) implies (ii).

Lemma 41. *Suppose (μ_x^t) is measurable. For any $A \in \Sigma$, the set*

$$\mu_+(A) \;:=\; \{\, x \in X \,:\, \text{for } \lambda\text{-almost all } t \in \mathbb{T}^+,\ \mu_x^t(A) = 1 \}$$

is forward-invariant, and the set

$$\mu_-(A) \;:=\; \{\, x \in X \,:\, \lambda(t \in \mathbb{T}^+ : \mu_x^t(A) > 0) \;>\; 0 \,\}$$

is backward-invariant.

Proof. By Lemma 8(B), $\mu_+(A) \in \Sigma$. Now fix any $x \in \mu_+(A)$ and $t \in \mathbb{T}^+$. Then (using Corollary 14)

$$\begin{aligned}
\int_X \int_{\mathbb{T}^+} \mu_y^s(X \setminus A)\, \lambda(ds)\, \mu_x^t(dy) \;&=\; \int_{\mathbb{T}^+} \int_X \mu_y^s(X \setminus A)\, \mu_x^t(dy)\, \lambda(ds) \\
&=\; \int_{\mathbb{T}^+} \mu_x^{t+s}(X \setminus A)\, \lambda(ds) \\
&=\; 0
\end{aligned}$$

and hence for μ_x^t-almost all $y \in X$, for λ-almost all $s \in \mathbb{T}^+$, $\mu_y^s(X \setminus A) = 0$. In other words, for μ_x^t-almost all $y \in X$, y belongs to $\mu_+(A)$. Hence $\mu_+(A)$ is forward-invariant. Now notice that

$$\mu_-(A) \;=\; X \setminus \mu_+(X \setminus A).$$

So $\mu_-(A)$ is backward-invariant. $\square$

Lemma 42. *Suppose (μ_x^t) is measurable. For any $A \in \Sigma$, the set $A' := \mu_+(\mu_-(A))$ is strictly forward-invariant.*

Proof. By Lemma 41, A' is forward-invariant. Now fix any $x \in X$ and $t \in \mathbb{T}^+$ such that $\mu_x^t(A') = 1$; if we can show that $x \in A'$ then we are done. Since $\mu_-(A)$ is backward-invariant (by Lemma 41), we have that $A' \subset \mu_-(A)$. So for any $s \in \mathbb{T}_{[0,t]}$,

$$\int_X \mu_y^{t-s}(\mu_-(A))\, \mu_x^s(dy) \;=\; \mu_x^t(\mu_-(A)) \;=\; 1$$

and so $\mu_y^{t-s}(\mu_-(A)) = 1$ for μ_x^s-almost all $y \in Y$; since $\mu_-(A)$ is backward-invariant, it follows that $\mu_x^s(\mu_-(A)) = 1$. Now for any $s \in \mathbb{T}_{[t,\infty)}$,

$$\mu_x^s(A') \;=\; \int_X \mu_y^{s-t}(A')\, \mu_x^t(dy) \;\geq\; \int_{A'} \mu_y^{s-t}(A')\, \mu_x^t(dy) \;=\; 1$$

since $\mu_x^t(A') = 1$ and A' is forward-invariant. Since $A' \subset \mu_-(A)$, we have, in particular, that $\mu_x^s(\mu_-(A)) = 1$ for all $s \in \mathbb{T}_{[t,\infty)}$. Thus overall, we have seen that $\mu_x^s(\mu_-(A)) = 1$ for all $s \in \mathbb{T}^+$. Hence $x \in A'$. $\square$

Lemma 43. *Suppose (μ_x^t) is measurable. Let ρ be a stationary probability measure, and let $A \in \Sigma$ be any ρ-almost invariant set. Then*

$$\rho(A \triangle \mu_+(A)) \;=\; \rho(A \triangle \mu_-(A)) \;=\; 0.$$

Proof. Note that

$$\{x \in A : \text{for } \lambda\text{-almost all } t \in \mathbb{T}^+,\ \mu_x^t(A) = 1\} \;=\; A \cap \mu_+(A),$$
$$\{x \in X \setminus A : \text{for } \lambda\text{-almost all } t \in \mathbb{T}^+,\ \mu_x^t(A) = 0\} \;\subset\; (X \setminus A) \cap (X \setminus \mu_+(A)),$$

and therefore the complement of $A \triangle \mu_+(A)$ contains the set

$$X_A \;:=\; \{x \in X : \text{for } \lambda\text{-almost all } t \in \mathbb{T}^+,\ \mu_x^t(A) = \mathbb{1}_A(x)\}.$$

By Exercise 26(A) we know that for every $t \in \mathbb{T}^+$, for ρ-almost all $x \in X$, $\mu_x^t(A) = \mathbb{1}_A(x)$. Hence, by Corollary 14, X_A is a ρ-full set, and therefore $\rho(A \triangle \mu_+(A)) = 0$. Finally,

$$\rho(A \triangle \mu_-(A)) \;=\; \rho((X \setminus A) \triangle (X \setminus \mu_-(A))) \;=\; \rho((X \setminus A) \triangle \mu_+(X \setminus A)) \;=\; 0$$

since $X \setminus A$ is ρ-almost invariant. So we are done. $\square$

Corollary 44. *Suppose (μ_x^t) is measurable, and let ρ be a stationary probability measure. Then every ρ-almost invariant set differs from some strictly forward-invariant set by a ρ-null set.*

Proof. Let A be a ρ-almost invariant set. By Lemma 43, $\rho(A \triangle \mu_-(A)) = 0$; and since $\mu_-(A)$ is itself backward-invariant (by Lemma 41) and therefore ρ-almost invariant, $\rho(\mu_-(A) \triangle A') = 0$ by Lemma 43. We already said (in Lemma 42) that A' is strictly forward-invariant. So we are done. $\square$

It immediately follows that (ii)$\Rightarrow$(i) in Theorem 39.

Lemma 45. *Suppose (μ_x^t) is measurable. Let $A \in \Sigma$ be a forward-invariant set. Then the map $t \mapsto \mu_x^t(A)$ is increasing for every $x \in X$, and the map $g_A : X \to [0,1]$ given by $g_A(x) = \lim_{t\to\infty} \mu_x^t(A)$ is strictly invariant. Given any stationary probability measure ρ, $g_A(x) = \mathbb{1}_A(x)$ for ρ-almost all $x \in X$.*

Proof. It is clear that for each $x \in X$, $\mu_x^t(A) \geq \mathbb{1}_A(x)$ for all $t \in \mathbb{T}^+$. Now for any $s, t \in \mathbb{T}^+$ and $x \in X$,

$$\mu_x^{s+t}(A) \;=\; \int_X \mu_y^t(A)\, \mu_x^s(dy) \;\geq\; \int_X \mathbb{1}_A(y)\, \mu_x^s(dy) \;=\; \mu_x^s(A).$$

So the map $t \mapsto \mu_x^t(A)$ is increasing for each x. Now (using the monotone or dominated convergence theorem), for any $x \in X$ and $t \in \mathbb{T}^+$,

$$\int_X g_A(y)\, \mu_x^t(dy) \;=\; \lim_{s\to\infty} \int_X \mu_y^s(A)\, \mu_x^t(dy) \;=\; \lim_{s\to\infty} \mu_x^{s+t}(A) \;=\; g_A(x).$$

So g_A is strictly invariant. Now fix a stationary probability measure ρ. Since A is ρ-almost invariant, there exists (by Exercise 26(A)) a ρ-full $Y \in \Sigma$ such that for all $x \in Y$, for all $n \in \mathbb{N}$, $\mu_x^n(A) = \mathbb{1}_A(x)$. Obviously $g_A(x) = \mathbb{1}_A(x)$ for all $x \in Y$. So we are done. $\square$

Proof that (iv)$\Rightarrow$(i). Suppose (iv) holds, and let A be a ρ-almost invariant set. By Lemmas 41 and 43, $\mu_+(A)$ is forward-invariant and $\rho(\mu_+(A)) = \rho(A)$. By Lemma 45, $g_{\mu_+(A)}$ is strictly invariant, and is therefore ρ-almost constant. But since, by Lemma 45, $g_{\mu_+(A)}(x) = \mathbb{1}_{\mu_+(A)}(x)$ for ρ-almost all x, it follows that $\mathbb{1}_{\mu_+(A)}(x)$ is ρ-almost constant. So $\rho(\mu_+(A))$ is equal to either 0 or 1, and hence $\rho(A)$ is equal to either 0 or 1. $\square$

This completes the proof of Theorem 39.

Exercise 46. Given a Markov kernel or a semigroup of Markov kernels on X, we will say that a set $A \in \Sigma$ is *strictly invariant* if $\mathbb{1}_A$ is strictly invariant; note that this is the same as saying that A is simultaneously forward-invariant and backward-invariant. (This is, in turn, equivalent to saying that A is simultaneously strictly forward-invariant and strictly backward-invariant.) (A) In contrast to Remark 40: give an example of a Markov kernel possessing a stationary probability measure ρ that assigns trivial measure to every strictly invariant set and yet is not ergodic. (Hint: this is possible on a discrete state space of just 3 elements!) (B) Show that for any kernel or semigroup of kernels, the set of strictly invariant sets forms a σ-algebra. (C) Is it necessarily the case that if a bounded measurable function $g : X \to \mathbb{R}$ is strictly invariant then it is measurable with respect to σ-algebra of strictly invariant sets?

Exercise 47. Suppose $\mathbb{T}^+ = [0, \infty)$. Let ρ be a stationary probability measure, and let $A \in \Sigma$ be a set such that for λ-almost all $t \geq 0$, A is ρ-almost invariant with respect to the kernel $(\mu_x^t)_{x\in X}$. We know from Exercise 32(B) that A must be ρ-almost invariant with respect to the whole semigroup (μ_x^t); but without using Exercise 32, find an alternative proof of this fact in the case that (μ_x^t) is measurable. (Hint: following the proof of Lemma 43, show that $\rho(A \vartriangle \mu_+(A)) = 0$.)

Exercise 48. Let $g : X \to \mathbb{R} \cup \{-\infty\}$ be a measurable function that is bounded above. (A) Show that if g can be expressed as the pointwise supremum of a collection of super-invariant functions then g is super-invariant. (B) Show that the following are equivalent:

(i) for any $a \in \mathbb{R} \cup \{-\infty\}$, $g^{-1}((a,\infty))$ is forward-invariant;

(ii) for any $a \in \mathbb{R} \cup \{-\infty\}$, $g^{-1}([a,\infty))$ is forward-invariant.

It is clear that (ii) implies that g is super-invariant, but does the converse hold? *For the rest of this exercise, assume that (μ_x^t) is measurable.* (C) Show that if g is bounded and super-invariant then Lemma 45 holds with A and $\mathbb{1}_A$ replaced by g. (D) [Warning: this part is long!] Recalling the notion of "essential superior limit" from Exercise 11(B), show that the function $\tilde{g} : X \to \mathbb{R} \cup \{-\infty\}$ given by $\tilde{g}(x) = \lambda\text{-}\lim\operatorname{ess\,sup}_{t\to\infty} \mu_x^t(g)$ is super-invariant. (Hint: first prove an "essential" version of Fatou's lemma.) Can a stronger conclusion be made in the case that $(\mu_x^t) = (\delta_{f^t(x)})$ for some measurable autonomous dynamical system (f^t)? Show moreover that if ρ is a stationary probability measure and g is ρ-almost invariant then $\tilde{g}(x) = g(x)$ for ρ-almost all $x \in X$.[8]

Proposition 49. *Suppose that $\mathbb{T}^+ = [0,\infty)$ and (μ_x^t) is measurable, and define the Markov kernel $(\bar{\mu}_x^1)_{x\in X}$ on X by*

$$\bar{\mu}_x^1(A) \;=\; \int_0^1 \mu_x^t(A)\,dt$$

for all $A \in \Sigma$ and $x \in X$. Let ρ be a stationary probability measure of the semigroup (μ_x^t). Then:

(A) ρ is stationary with respect to the kernel $(\bar{\mu}_x^1)$;

(B) a set $A \in \Sigma$ is ρ-almost invariant with respect to the semigroup (μ_x^t) if and only if it is ρ-almost invariant with respect to the kernel $(\bar{\mu}_x^1)$;

(C) ρ is ergodic with respect to the semigroup (μ_x^t) if and only if it is ergodic with respect to the kernel $(\bar{\mu}_x^1)$.

Proof. (A) For any $A \in \Sigma$, using Corollary 14 we have that

$$\int_X \bar{\mu}_x^1(A)\,\rho(dx) \;=\; \int_0^1 \int_X \mu_x^t(A)\,\rho(dx)\,dt \;=\; \rho(A).$$

(B) For any $A \in \Sigma$, again using Corollary 14 we have that

$$\int_A \bar{\mu}_x^1(A)\,\rho(dx) \;=\; \int_0^1 \int_A \mu_x^t(A)\,\rho(dx)\,dt.$$

The left-hand side is equal to $\rho(A)$ if and only if $\bar{\mu}_x^1(A) = 1$ for ρ-almost all $x \in A$, i.e. if and only if A is ρ-almost invariant with respect to $(\bar{\mu}_x^1)_{x\in X}$. Similarly, the right-hand side is equal to $\rho(A)$ if and only if for λ-almost all $t \in [0,1]$, A is ρ-almost invariant with respect to the kernel $(\mu_x^t)_{x\in X}$; but by Exercise 32(B), this is equivalent to saying that A is ρ-almost invariant with respect to the whole semigroup (μ_x^t). (C) Follows immediately from part (B). □

[8] The case of a measurable autonomous dynamical system is mentioned at the bottom of p537 / top of p538 of here.

4 Markov processes and pointwise ergodic theorems

Let (μ_x^t) be a semigroup of kernels, and let ρ be any probability measure on X. We define (by a recursive construction) a probability measure $\boldsymbol{\mu}_\rho^{t_1,\dots,t_n}$ on X^{n+1} for each list of times $t_1,\dots,t_n \in \mathbb{T}^+$ with $t_1 \le \dots \le t_n$, as follows:

$$\boldsymbol{\mu}_\rho^{t_1,\dots,t_n}(A) \;=\; \int_{X^n}\int_X \mathbb{1}_A(x_0,\dots,x_n)\,\mu_{x_{n-1}}^{t_n-t_{n-1}}(dx_n)\,\boldsymbol{\mu}_\rho^{t_1,\dots,t_{n-1}}(d(x_0,\dots,x_{n-1})) \qquad \forall\, A \in \Sigma^{\otimes(n+1)}$$

and

$$\boldsymbol{\mu}_\rho^{t}(A) \;=\; \int_X\int_X \mathbb{1}_A(x_0,x_1)\,\mu_{x_0}^{t}(dx_1)\,\rho(dx_0) \qquad \forall\, A \in \Sigma\otimes\Sigma.$$

A heuristic interpretation of $\boldsymbol{\mu}_\rho^{t_1,\dots,t_n}(A)$ is as follows: Imagine a particle in X which follows a random trajectory $(x(t) : t \ge 0)$; and imagine that at any given time s, if the particle is at location $y \in X$ then (independent of the history of the particle before time s) the probability distribution for where the particle will be at some later time $s+t$ is given by μ_y^t. (The trajectory of such a particle is called a *Markov process*.) Imagine moreover that the initial position $x(0)$ of the particle is itself random, having probability distribution ρ. Then, prior to the selection of the initial point $x(0)$, the probability that the sequence of positions $(x(0),x(t_1),\dots,x(t_n))$ will belong to the set $A \subset X^{n+1}$ is given by $\boldsymbol{\mu}_\rho^{t_1,\dots,t_n}(A)$.

Lemma 50. *Given any $t_1 \le \dots \le t_n$ in $\mathbb{T}^+$ and any measurable $g : X^{n+1} \to \mathbb{R}'$, if $\boldsymbol{\mu}_\rho^{t_1,\dots,t_n}(g) \ne \mathrm{NaN}$ then*

$$\boldsymbol{\mu}_\rho^{t_1,\dots,t_n}(g) \;=\; \int_X\int_X\int_X\dots\int_X g(x_0,\dots,x_n)\,\mu_{x_{n-1}}^{t_n-t_{n-1}}(dx_n)\,\dots\,\mu_{x_1}^{t_2-t_1}(dx_2)\,\mu_{x_0}^{t_1}(dx_1)\,\rho(dx_0).$$

Proof. First suppose $n=1$. As in Section 2, Theorem 12 and Exercise 13(B) give that if $\boldsymbol{\mu}_\rho^{t_1}(g) \ne \mathrm{NaN}$ then

$$\boldsymbol{\mu}_\rho^{t_1}(g) \;=\; \int_X\int_X g(x_0,x_1)\,\mu_{x_0}^{t_1}(dx_1)\,\rho(dx_0).$$

Now suppose $n = k+1$, for some k where the result is known to be true for $n=k$. Define the function $\tilde{g} : X^{k+1} \to \mathbb{R}'$ by

$$\tilde{g}(x_0,\dots,x_k) \;=\; \int_X g(x_0,\dots,x_{k+1})\,\mu_{x_k}^{t_{k+1}-t_k}(dx_{k+1}).$$

Then provided $\boldsymbol{\mu}_\rho^{t_1,\dots,t_{k+1}}(g) \ne \mathrm{NaN}$,

$$\begin{aligned}
&\boldsymbol{\mu}_\rho^{t_1,\dots,t_{k+1}}(g)\\
&= \int_{X^{k+1}}\int_X g(x_0,\dots,x_{k+1})\,\mu_{x_k}^{t_{k+1}-t_k}(dx_{k+1})\,\boldsymbol{\mu}_\rho^{t_1,\dots,t_k}(d(x_0,\dots,x_k))\\
&\qquad\qquad \text{(by Theorem 12 and Exercise 13(B))}\\
&= \boldsymbol{\mu}_\rho^{t_1,\dots,t_k}(\tilde{g})\\
&= \int_X\int_X\dots\int_X \tilde{g}(x_0,\dots,x_k)\mu_{x_{k-1}}^{t_k-t_{k-1}}(dx_k)\dots(dx_2)\mu_{x_0}^{t_1}(dx_1)\rho(dx_0)\\
&= \int_X\int_X\dots\int_X\int_X g(x_0,\dots,x_{k+1})\mu_{x_k}^{t_{k+1}-t_k}(dx_{k+1})\mu_{x_{k-1}}^{t_k-t_{k-1}}(dx_k)\dots\mu_{x_0}^{t_1}(dx_1)\rho(dx_0).
\end{aligned}$$

So the result is true for $n=k+1$. Hence, by induction, the result is true in general. □

Now let Y be a non-empty subset of $X^{\mathbb{T}^+}$ such that for any $(x_t)_{t\in\mathbb{T}^+} \in Y$ and $\tau \in \mathbb{T}^+$, $(x_{t+\tau})_{t\in\mathbb{T}^+} \in Y$. For each $t_1,\dots,t_n \in \mathbb{T}^+$, we define the projection $\pi_{t_1,\dots,t_n} : Y \to X^n$ by

$$\pi_{t_1,\dots,t_n}((x_t)_{t\in\mathbb{T}^+}) \;=\; (x_{t_1},\dots,x_{t_n}),$$

and we equip Y with the σ-algebra $\mathcal{Y} := \sigma(\pi_t : t \in \mathbb{T}^+)$. We will say that Y is *measurable in time* if the map $(t,\mathbf{x}) \mapsto \pi_t(\mathbf{x})$ from $\mathbb{T}^+ \times Y$ to X is measurable. Obviously if $\mathbb{T}^+ = \mathbb{N}\cup\{0\}$ then Y must be measurable in time.

Proposition 51. *Suppose $\mathbb{T}^+ = [0,\infty)$, and suppose we have a separable metric on X whose Borel σ-algebra coincides with Σ, such that for all $(x_t) \in Y$ the map $t \mapsto x_t$ is right-continuous. Then Y is measurable in time.*

Proof. Follows immediately from Lemma 16(B). □

Remark 52. Under the conditions of Proposition 51, we actually have that Y is "progressively" measurable in time, in the sense that for all $t \in \mathbb{T}^+$, if we equip Y with the σ-algebra $\sigma(\pi_s : s \in \mathbb{T}_{[0,t]})$ then the map $(s,\mathbf{x}) \mapsto \pi_s(\mathbf{x})$ from $\mathbb{T}_{[0,t]} \times Y$ to X is measurable.

Proposition 53. *There is at most one probability measure $\boldsymbol{\mu}_\rho^Y$ on Y such that for any $t_1 \le \dots \le t_n$ in $\mathbb{T}^+$, $\pi_{0,t_1,\dots,t_n*}\boldsymbol{\mu}_\rho^Y = \boldsymbol{\mu}_\rho^{t_1,\dots,t_n}$.*

Proof. It is clear that

$$\{\,\pi_{0,t_1,\dots,t_n}^{-1}(A) \,:\, n \in \mathbb{N}, A \in \Sigma^{\otimes(n+1)}, t_1 \le \dots \le t_n\}$$

is a π-system generating $\mathcal{Y}$. Hence the result follows from the π-λ theorem. □

A heuristic interpretation of $\boldsymbol{\mu}_\rho^Y$ is that it is the probability distribution for the entire random trajectory $(x(t) : t \ge 0)$ of the particle described earlier. (Y itself will represent some constraint on the motion of the particle, such as being continuous in time with respect to some topology on X.)

Note that if $\boldsymbol{\mu}_\rho^Y$ exists then for each $t \in \mathbb{T}^+$, $\pi_{t*}\boldsymbol{\mu}_\rho^Y = \mu^{t*}\rho$ (where μ^{t*} denotes the map on $\mathcal{M}_1$ associated with the kernel $(\mu_x^t)_{x\in X}$).

Exercise 54. (A) Show that if we take Y to be the whole of $X^{\mathbb{T}^+}$, then Y is measurable in time if and only if either $\mathbb{T}^+ = \mathbb{N}\cup\{0\}$ or $\Sigma = \{\varnothing, X\}$. (B) Using Kolmogorov's extension theorem for standard measurable spaces, show that if (X,Σ) is standard and $Y = X^{\mathbb{T}^+}$ then the measure $\boldsymbol{\mu}_\rho^Y$ defined in Proposition 53 exists. (C) Show that if $Y \in \Sigma^{\otimes\mathbb{T}^+}$ and $\boldsymbol{\mu}_\rho^{X^{\mathbb{T}^+}}$ exists then $\boldsymbol{\mu}_\rho^Y$ exists if and only if $\boldsymbol{\mu}_\rho^{X^{\mathbb{T}^+}}(Y) = 1$.

From now until Exercise 75, assume that $\boldsymbol{\mu}_\rho^Y$ exists. The measure $\boldsymbol{\mu}_\rho^Y$ is sometimes referred to as a *Markov measure*,[9] and a Y-valued random variable whose law is equal to $\boldsymbol{\mu}_\rho^Y$ is referred to as a *(homogeneous) Markov process with initial distribution ρ and transition probabilities $(\mu_x^t)_{x\in X,\,t\in\mathbb{T}^+}$*. (For the equivalence between this definition and the more common definition via conditional probabilities, see Exercise 65 and Lemma 66.)

If ρ is a stationary probability measure of the semigroup (μ_x^t), then we write $\mathcal{I}_\rho$ to refer to the set of ρ-almost invariant sets.

[9]This is different from how the term is sometimes used in the context of random dynamical systems, to refer to a probability-measure-valued random variable whose trajectory under a given random dynamical system is a measure-valued Markov process.

Theorem 55 ("Ergodic theorem for Markov processes"). *Suppose that Y is measurable in time and ρ is stationary. Let $g : X \to \mathbb{R}$ be a function that is integrable with respect to ρ, and let $\hat{g}$ be a version of $\rho(g|\mathcal{I}_\rho)$. Then for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t)_{t\in\mathbb{T}^+} \in Y$,*

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} g(x_s)\,\lambda(ds) \;\to\; \hat{g}(x_0) \quad \textit{as } t \to \infty.$$

Exercise 56. (A) Show that, given any $\tau \in \mathbb{T}^+$, if we replace $\hat{g}(x_0)$ with $\hat{g}(x_\tau)$ in the statement of Theorem 55, the statement will remain true. (B) Assume the hypotheses of Theorem 55 and suppose moeover that $\boldsymbol{\mu}_{\delta_x}^Y$ exists for ρ-almost every $x \in X$. Show that ρ-almost every $x \in X$ has the property that for $\boldsymbol{\mu}_{\delta_x}^Y$-almost all $(x_t)_{t\in\mathbb{T}^+} \in Y$, $\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} g(x_s)\,\lambda(ds) \to \hat{g}(x_0)$ as $t \to \infty$.

Corollary 57. *Suppose that Y is measurable in time and ρ is ergodic. Let $g : X \to \bar{\mathbb{R}}$ be a measurable function with $\rho(g) \neq$ NaN. Then for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t)_{t\in\mathbb{T}^+} \in Y$,*

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} g(x_s)\,\lambda(ds) \;\to\; \rho(g) \quad \textit{as } t \to \infty.$$

Proof. Since ρ is ergodic, $\mathcal{I}_\rho$ is contained in the σ-algebra of ρ-trivial measure sets. Hence, if $\rho(g) \in \mathbb{R}$ then the constant function $x \mapsto \rho(g)$ is a version of $\rho(g|\mathcal{I}_\rho)$, and so Theorem 55 gives the desired result. Now suppose $\rho(g) = \infty$. (The case that $\rho(g) = -\infty$ is similar.) For $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t)_{t\in\mathbb{T}^+} \in Y$, we have that

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} g^-(x_s)\,\lambda(ds) \;\to\; \rho(g^-) \quad \text{as } t \to \infty.$$

It is also the case that for each $n \in \mathbb{N}$, for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t)_{t\in\mathbb{T}^+} \in Y$,

$$\liminf_{t\to\infty} \frac{1}{t}\int_{\mathbb{T}_{[0,t)}} g^+(x_s)\,\lambda(ds) \;\geq\; \lim_{t\to\infty} \frac{1}{t}\int_{\mathbb{T}_{[0,t)}} g^+(x_s) \wedge n \;\lambda(ds) \;=\; \rho(g^+ \wedge n);$$

and therefore, since $\sup_{n\in\mathbb{N}} \rho(g^+ \wedge n) = \infty$ by the monotone convergence theorem, we have that for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t)_{t\in\mathbb{T}^+} \in Y$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} g^+(x_s)\,\lambda(ds) \;\to\; \infty \quad \text{as } t \to \infty.$$

Combining these gives the desired result. □

Now the proof of Theorem 55 essentially divides into two parts: proving the pointwise ergodic theorem for dynamical systems (which we shall not do in full here), and characterising the almost-invariant sets of the "horizontal shift dynamical system". (It is perhaps worth saying now that Exercises 59, 61, 65 and 72 are included for the sake of completeness, and are not actually needed in the proof of Theorem 55.)

Recall that an autonomous dynamical system (f^t) is said to be *measurable* if the map $(t,x) \mapsto f^t(x)$ is measurable.

Proposition 58 (Ergodic theorem for dynamical systems)**.** *Let (f^t) be a measurable autonomous dynamical system on X, let μ be an invariant probability measure of (f^t), and let $\mathcal{I}_\mu$ denote the set of μ-almost invariant sets. Let $g : X \to \mathbb{R}$ be a function that is integrable with respect to μ. Then the limit*

$$\lim_{t\to\infty} \frac{1}{t} \int_{\mathbb{T}_{[0,t)}} g(f^s(x))\,\lambda(ds) \;=:\; G(x)$$

exists and is finite for μ-almost all $x \in X$; and (setting $G(x) := \mathrm{NaN}$ for all x where this limit does not exist) G is a version of $\mu(g|\mathcal{I}_\mu)$.

It will be left as an exercise to the reader to show that the function $G : X \to \mathbb{R}'$ is indeed a measurable function (although this is not such an important point—the important point, which we *will* see explicitly, is that if $\bar{g}$ is a version of $\mu(g|\mathcal{I}_\mu)$ then the limit given in Proposition 58 exists and is equal to $\bar{g}(x)$ for μ-almost all $x \in X$).

Exercise 59. Show that if $\mathbb{T}^+ = \mathbb{N} \cup \{0\}$ then G is measurable with respect to the σ-algebra $\mathcal{I} \subset \Sigma$ of sets that are strictly invariant under f^1 (and hence $\mu(g|\mathcal{I}_\mu)$ can be replaced by $\mu(g|\mathcal{I})$ in the statement of Proposition 58).

Now the proof of Proposition 58 in discrete time (i.e. when $\mathbb{T}^+ = \mathbb{N} \cup \{0\}$) can be found in virtually any textbook on ergodic theory, and so we will not give it here. However we will show how to extend the statement from discrete to continuous time.[10]

Lemma 60. *Let (f^t) be an autonomous dynamical system on X and let μ be an invariant probability measure of (f^t). Let $g : X \to \mathbb{R}$ be a function that is integrable with respect to μ. For μ-almost all $x \in X$, $\frac{1}{n} g(f^n(x)) \to 0$ as the integer n tends to ∞.*

Note that Lemma 60 is obvious in the case that g is bounded; the non-triviality specifically comes from the fact that g may be unbounded.

Proof of Lemma 60. Let $A_n(x) := \frac{1}{n}\sum_{i=0}^{n-1} g(f^i(x))$ for each $x \in X$ and $n \in \mathbb{N}$. Obviously μ is an invariant probability measure of the discrete-time dynamical system (f^n), and hence the discrete-time version of Proposition 58 gives that $\lim_{n\to\infty} A_n(x)$ exists and is finite for μ-almost all $x \in X$. For each $x \in X$ and $n \geq 2$,

$$\tfrac{1}{n} g(f^n(x)) \;=\; A_n(x) - \left(1 - \tfrac{1}{n}\right) A_{n-1}(x).$$

The desired result clearly follows. □

Proof of Proposition 58 with $\mathbb{T}^+ = [0,\infty)$. Since we can split g into positive and negative parts, it will suffice just to consider the case that g is nonnegative. Define $\tilde{g} : X \to [0,\infty]$ by

$$\tilde{g}(x) \;=\; \int_0^1 g(f^s(x))\,ds.$$

By Corollary 14, we have that $\mu(\tilde{g}) = \mu(g) < \infty$; hence, using the discrete-time version of Proposition 58, for μ-almost all $x \in X$,

$$\frac{1}{n}\sum_{i=0}^{n-1} \tilde{g}(f^i(x)) \;\to\; \bar{g}(x) \quad \text{as } n \to \infty$$

[10] We essentially follow the outline given at the start of section 1.2.2 of here.

where $\bar{g} : X \to [0, \infty)$ is a version of the conditional expectation (under μ) of $\tilde{g}$ given the σ-algebra of sets that are μ-almost invariant with respect to f^1.

Now let $\tilde{X} := \{x \in X : \tilde{g}(f^i(x)) < \infty \ \forall i\}$. (It is clear that this a μ-full set.) Letting $\lceil t \rceil$ denote the smallest integer greater than or equal to t, one can verify that for any $x \in \tilde{X}$ and $t > 0$,

$$\left| \left(\frac{1}{t} \int_0^t g(f^s(x))\, ds \right) - \left(\frac{1}{\lceil t \rceil} \sum_{i=0}^{\lceil t \rceil - 1} \tilde{g}(f^i(x)) \right) \right| \leq \left(\frac{1}{\lceil t \rceil^2} \sum_{i=0}^{\lceil t \rceil - 1} \tilde{g}(f^i(x)) \right) + \left(\frac{1}{\lceil t \rceil} \tilde{g}(f^{\lceil t \rceil - 1}(x)) \right).$$

Using the discrete-time version of Proposition 58, we see that the first of the two terms on the right-hand side tends to 0 as $t \to \infty$ for μ-almost all x; and using Lemma 60, we see that the second of the two terms tends to 0 as $t \to \infty$ for μ-almost all x. Therefore

$$\frac{1}{t} \int_0^t g(f^s(x))\, ds \ \to \ \bar{g}(x) \quad \text{as } t \to \infty$$

for μ-almost all $x \in X$. To complete the proof, we show that $\bar{g}$ is a version of $\mu(g|\mathcal{I}_\mu)$. First we show that $\bar{g}$ is $\mathcal{I}_\mu$-measurable, i.e. that it is μ-almost invariant with respect to $(f^t)_{t \geq 0}$. Fix any $\tau \in \mathbb{T}^+$. For any $x \in \tilde{X}$, letting $x^\tau := f^\tau(x)$, we have that for all $t > 0$,

$$\underbrace{\frac{1}{\tau + t} \int_0^{\tau + t} g(f^s(x))\, ds}_{①} = \underbrace{\left(1 - \frac{\tau}{t + \tau}\right)}_{\to 1 \text{ as } t \to \infty} \underbrace{\left(\frac{1}{t} \int_0^t g(f^s(x^\tau))\, ds \right)}_{②} + \underbrace{\frac{1}{\tau(t + \tau)} \int_0^\tau g(f^s(x))\, ds}_{\to 0 \text{ as } t \to \infty}.$$

For μ-almost all x, ① tends to $\bar{g}(x)$ as $t \to \infty$; since μ is f^τ-invariant, we also have that for μ-almost all x, ② tends to $\bar{g}(x^\tau)$ as $t \to \infty$. Therefore $\bar{g}$ and $\bar{g} \circ f^\tau$ agree μ-almost everywhere. So $\bar{g}$ is μ-almost invariant. Now, given any $A \in \mathcal{I}_\mu$, we have

$$\begin{aligned}
\int_A \bar{g}(x)\, \mu(dx) &= \int_A \tilde{g}(x)\, \mu(dx) \\
&= \int_0^1 \int_A g(f^s(x))\, \mu(dx)\, ds \quad \text{(by Corollary 14)} \\
&= \int_0^1 \int_{f^{-s}(A)} g(f^s(x))\, \mu(dx)\, ds \quad (\text{since } \mu(\, A \,\triangle\, f^{-s}(A)\,) = 0) \\
&= \int_0^1 \int_A g(x)\, \mu(dx)\, ds \\
&= \int_A g(x)\, \mu(dx).
\end{aligned}$$

So we are done. □

Exercise 61. Let (f^t) be a measurable autonomous dynamical system on X, let μ be an invariant probability measure of (f^t), and let $g : X \to \bar{\mathbb{R}}$ be a measurable function with $\mu(g) \in \{-\infty, \infty\}$. Show that for μ-almost all $x \in X$,

$$\lim_{t \to \infty} \frac{1}{t} \int_{\mathbb{T}_{[0,t)}} g(f^s(x))\, ds$$

exists. (Hint: if $\mu(g) = \infty$, restrict to sublevel sets of $\liminf_{t \to \infty} \frac{1}{t} \int_{\mathbb{T}_{[0,t)}} g(f^s(\cdot))\, ds$.)

We now introduce the *horizontal shift dynamical system*. For each $\tau \in \mathbb{T}^+$, define the map $\theta^\tau : Y \to Y$ by $\theta^\tau((x_t)_{t\in\mathbb{T}^+}) = (x_{\tau+t})_{t\in\mathbb{T}^+}$. It is clear that $(\theta^t)_{t\in\mathbb{T}^+}$ is an autonomous dynamical system on Y.

Lemma 62. *For all $\tau \in \mathbb{T}^+$, $\boldsymbol{\mu}^Y_{\mu^{\tau*}\rho}$ exists and is equal to $\theta^\tau_*\boldsymbol{\mu}^Y_\rho$.*

Proof. We need to show that for any $t_1 \le \ldots \le t_n$ in $\mathbb{T}^+$,

$$\pi_{0,t_1,\ldots,t_n*}(\theta^\tau_*\boldsymbol{\mu}^Y_\rho) \;=\; \boldsymbol{\mu}^{t_1,\ldots,t_n}_{\mu^{\tau*}\rho}.$$

For any $A \in \Sigma^{\otimes(n+1)}$,

$$\begin{aligned}
&\pi_{0,t_1,\ldots,t_n*}(\theta^\tau_*\boldsymbol{\mu}^Y_\rho)(A)\\
&\quad= \pi_{\tau,t_1+\tau,\ldots,t_n+\tau*}\boldsymbol{\mu}^Y_\rho(A)\\
&\quad= \boldsymbol{\mu}^{\tau,t_1+\tau,\ldots,t_n+\tau}_\rho(X\times A)\\
&\quad= \int_X\int_X\int_X\int_X\cdots\int_X \mathbb{1}_A(x_0,\ldots,x_n)\,\mu^{t_n-t_{n-1}}_{x_{n-1}}(dx_n)\,\ldots\,\mu^{t_2-t_1}_{x_1}(dx_2)\,\mu^{t_1}_{x_0}(dx_1)\,\mu^\tau_x(dx_0)\rho(dx)\\
&\quad= \int_X\int_X\int_X\cdots\int_X \mathbb{1}_A(x_0,\ldots,x_n)\,\mu^{t_n-t_{n-1}}_{x_{n-1}}(dx_n)\,\ldots\,\mu^{t_2-t_1}_{x_1}(dx_2)\,\mu^{t_1}_{x_0}(dx_1)\,\mu^{\tau*}\rho(dx_0)\\
&\quad= \boldsymbol{\mu}^{t_1,\ldots,t_n}_{\mu^{\tau*}\rho}(A).
\end{aligned}$$

So we are done. □

We immediately have the following corollary:

Corollary 63. *$\boldsymbol{\mu}^Y_\rho$ is invariant with respect to (θ^t) if and only if ρ is a stationary probability measure.*

Lemma 64. *Suppose ρ is stationary. For any $A \in \Sigma$, the set $\tilde{A} := \pi_0^{-1}(A)$ is $\boldsymbol{\mu}^Y_\rho$-almost invariant with respect to (θ^t) if and only if A is ρ-almost invariant (with respect to (μ^t_x)).*

Proof. For any $t \in \mathbb{T}^+$,

$$\boldsymbol{\mu}^t_\rho(A\times(X\setminus A)) \;=\; \boldsymbol{\mu}^Y_\rho(\pi_{0,t}^{-1}(A\times(X\setminus A)) \;=\; \boldsymbol{\mu}^Y_\rho(\tilde{A}\setminus\theta^{-t}(\tilde{A})).$$

Hence Exercise 26(B) gives the desired result. □

Now for any $S \subset [0,\infty)$, let $\mathcal{Y}_S := \sigma(\pi_s : s \in \mathbb{T}_S)$.

Exercise 65. Show that for any probability measure ν on X, if there exists a probability measure $\boldsymbol{\mu}$ on Y such that for all $\tau, s \in \mathbb{T}^+$ and $A \in \Sigma$ the map $(x_t) \mapsto \mu^s_{x^\tau}(A)$ is a version of $\boldsymbol{\mu}(\pi^{-1}_{\tau+s}(A)|\mathcal{Y}_{[0,\tau]})$, then $\boldsymbol{\mu}^Y_\nu$ exists and is equal to $\boldsymbol{\mu}$.

The following lemma provides the converse of Exercise 65 (by setting $n = 1$ and $B := X \times A$):

Lemma 66. *Fix any $\tau \in \mathbb{T}^+$, any $t_1 \le \ldots \le t_n$ in $\mathbb{T}^+$, and any $B \in \Sigma^{\otimes(n+1)}$. The map $(x_t) \mapsto \boldsymbol{\mu}^{t_1,\ldots,t_n}_{\delta_{x_\tau}}(B)$ is a version of $\boldsymbol{\mu}^Y_\rho(\pi^{-1}_{\tau,\tau+t_1,\ldots,\tau+t_n}(B)|\mathcal{Y}_{[0,\tau]})$.*

Remark 67. It is clear (using Lemma 8(A)) that the map $(x_t) \mapsto \boldsymbol{\mu}^{t_1,\ldots,t_n}_{\delta_{x_\tau}}(B)$ is $\mathcal{Y}_{\{\tau\}}$-measurable. It therefore follows from Lemma 66 that for any $S \subset [0,\tau]$ with $\tau \in S$, this same map is also a version of $\boldsymbol{\mu}^Y_\rho(\pi^{-1}_{\tau,\tau+t_1,\ldots,\tau+t_n}(B)|\mathcal{Y}_S)$.

Proof of Lemma 66. Let $\mathcal{C}$ be the set of all $C \in \mathcal{Y}_{[0,\tau]}$ with the property that

$$\underbrace{\boldsymbol{\mu}_\rho^Y\big(\,C \cap \pi^{-1}_{\tau,\tau+t_1,\dots,\tau+t_n}(B)\,\big)}_{Ⓐ} = \underbrace{\int_C \boldsymbol{\mu}_{\delta_{x_\tau}}^{t_1,\dots,t_n}(B)\,\boldsymbol{\mu}_\rho^Y(d(x_t))}_{Ⓑ}.$$

We first show that $\mathcal{C}$ includes every set of the form $C = \pi^{-1}_{0,s_1,\dots,s_{k-1},\tau}(E_0 \times \dots \times E_k)$ with $0 \le s_1 \le \dots \le s_{k-1} \le \tau$ and $E_0, \dots, E_k \in \Sigma$: for any such C, we have

$$\begin{aligned}
Ⓐ &= \int_{E_0}\int_{E_1}\cdots\int_{E_{k-1}}\int_{E_k}\int_X\cdots\int_X \mathbb{1}_B(z_0,\dots,z_n)\,\mu_{z_{n-1}}^{t_n-t_{n-1}}(dz_n)\dots \\
&\qquad\qquad \dots\mu_{z_0}^{t_1}(dz_1)\mu_{y_{k-1}}^{\tau-s_{k-1}}(dz_0)\mu_{y_{k-2}}^{s_{k-1}-s_{k-2}}(dy_{k-1})\dots\mu_{y_0}^{s_1}(dy_1)\rho(dy_0) \\
&= \int_{E_0}\int_{E_1}\cdots\int_{E_{k-1}}\int_{E_k} \boldsymbol{\mu}_{\delta_{z_0}}^{t_1,\dots,t_n}(B)\,\mu_{y_{k-1}}^{\tau-s_{k-1}}(dz_0)\mu_{y_{k-2}}^{s_{k-1}-s_{k-2}}(dy_{k-1})\dots\mu_{y_0}^{s_1}(dy_1)\rho(dy_0) \\
&= \int_{E_0\times\dots\times E_k} \boldsymbol{\mu}_{\delta_z}^{t_1,\dots,t_n}(B)\,\boldsymbol{\mu}_\rho^{s_1,\dots,\tau}(d(y_0,\dots,y_{k-1},z)) \quad \text{(by Lemma 50)} \\
&= Ⓑ.
\end{aligned}$$

Now using the monotone convergence theorem (and the fact that $Y \in \mathcal{C}$), we have that $\mathcal{C}$ is a λ-system. So the π-λ theorem gives that $\mathcal{C}$ is the whole of $\mathcal{Y}_{[0,\tau]}$, as required. □

Corollary 68 ("Markov property")**.** *Fix any $\tau \in \mathbb{T}^+$. For any $A \in \mathcal{Y}_{[\tau,\infty)}$, each version of $\boldsymbol{\mu}_\rho^Y(A|\mathcal{Y}_{\{\tau\}})$ is also a version of $\boldsymbol{\mu}_\rho^Y(A|\mathcal{Y}_{[0,\tau]})$.*

Proof. Let $\mathcal{A}$ be the set of all $A \in \mathcal{Y}_{[\tau,\infty)}$ with the desired property. By Remark 67, $\mathcal{A}$ includes all sets of the form $\pi^{-1}_{\tau,\tau+t_1,\dots,\tau+t_n}(B)$. Now using the monotone convergence theorem (and the fact that $Y \in \mathcal{A}$), we have that $\mathcal{A}$ is a λ-system. So the π-λ theorem gives the desired result. □

Corollary 69. *Suppose ρ is stationary, and fix any $\tau \in \mathbb{T}^+$. For any $A \in \mathcal{Y}$, if h_A is a version of $\boldsymbol{\mu}_\rho^Y(A|\mathcal{Y}_{\{0\}})$ then $h_A \circ \theta^\tau$ is a version of $\boldsymbol{\mu}_\rho^Y(\theta^{-\tau}(A)|\mathcal{Y}_{[0,\tau]})$.*

Proof. Follows immediately from Exercise 13(C) (with $(\Omega,\mathcal{F}) = (E,\mathcal{E}) = (Y,\mathcal{Y})$, $\mathbb{P} = \boldsymbol{\mu}_\rho^Y$, $f = \theta^\tau$, $\mathcal{G} = \mathcal{Y}_{\{0\}}$ and $g = \mathbb{1}_A$) and Corollary 68. □

Proposition 70. *Suppose ρ is stationary, and let $A \in \mathcal{Y}$ be a $\boldsymbol{\mu}_\rho^Y$-almost invariant set with respect to (θ^t). Then there is a ρ-almost invariant set $A^* \in \Sigma$ with $\boldsymbol{\mu}_\rho^Y(\,A \triangle \pi_0^{-1}(A^*)\,) = 0$.*

Proof. Let h_A be a version of $\boldsymbol{\mu}_\rho^Y(A|\mathcal{Y}_0)$. For each $n \in \mathbb{N}$, by Corollary 69 and the $\boldsymbol{\mu}_\rho^Y$-almost-invariance of A, $h_A \circ \theta^n$ is a version of $\boldsymbol{\mu}_\rho^Y(A|\mathcal{Y}_{[0,n]})$. Hence Lévy's upward theorem gives that $h_A \circ \theta^n(\mathbf{x}) \to \mathbb{1}_A(\mathbf{x})$ as $n \to \infty$ for $\boldsymbol{\mu}_\rho^Y$-almost all $\mathbf{x} \in Y$. But for each n, since h_A and $h_A \circ \theta^n$ agree $\boldsymbol{\mu}_\rho^Y$-almost everywhere, it follows that $h_{A*}\boldsymbol{\mu}_\rho^Y = (h_A \circ \theta^n)_*\boldsymbol{\mu}_\rho^Y$.

Therefore (e.g. by Exercise 22(C)) $h_{A*}\boldsymbol{\mu}_\rho^Y = \mathbb{1}_{A*}\boldsymbol{\mu}_\rho^Y = \boldsymbol{\mu}_\rho^Y(X \setminus A)\delta_0 + \boldsymbol{\mu}_\rho^Y(A)\delta_1$; so $h_A(\mathbf{x}) \in \{0,1\}$ for $\boldsymbol{\mu}_\rho^Y$-almost all $\mathbf{x} \in Y$. Hence $h_A(\mathbf{x}) = \mathbb{1}_A(\mathbf{x})$ for $\boldsymbol{\mu}_\rho^Y$-almost all $\mathbf{x} \in Y$.[11] Since h_A is $\mathcal{Y}_{\{0\}}$-measurable, there exists $A^* \in \Sigma$ such that $h_A^{-1}(\{1\}) = \pi_0^{-1}(A^*)$. So then, $\boldsymbol{\mu}_\rho^Y(\,A \triangle \pi_0^{-1}(A^*)\,) = 0$. Since A is $\boldsymbol{\mu}_\rho^Y$-almost invariant, it follows that $\pi_0^{-1}(A^*)$ is $\boldsymbol{\mu}_\rho^Y$-almost invariant, and therefore (by Lemma 64) A^* is ρ-almost invariant. □

[11] It is left as an exercise to the reader to verify that in general, if a version of a conditional probability $\mathbb{P}(E|\mathcal{G})$ takes value 0 or 1 $\mathbb{P}$-almost everywhere, then it is $\mathbb{P}$-a.e. equal to $\mathbb{1}_E$.

For any sub-σ-algebra $\mathcal{Z}$ of $\mathcal{Y}$, we write $\overline{\mathcal{Z}}^{(\boldsymbol{\mu}_\rho^Y)}$ to denote the smallest σ-algebra containing both $\mathcal{Z}$ and every $\mathcal{Y}$-measurable $\boldsymbol{\mu}_\rho^Y$-null set. If ρ is stationary, we write $\mathcal{I}_\rho \subset \Sigma$ to denote the set of ρ-almost invariant sets with respect to (μ_x^t), and we write $\mathcal{I}_{\boldsymbol{\mu}_\rho^Y} \subset \mathcal{Y}$ to denote the set of $\boldsymbol{\mu}_\rho^Y$-almost invariant sets with respect to (θ^t).

Corollary 71. *Suppose ρ is stationary. Then* $\mathcal{I}_{\boldsymbol{\mu}_\rho^Y} = \overline{\pi_0^{-1}\mathcal{I}_\rho}^{(\boldsymbol{\mu}_\rho^Y)}$.

Proof. By Lemma 64, $\pi_0^{-1}\mathcal{I}_\rho \subset \mathcal{I}_{\boldsymbol{\mu}_\rho^Y}$ and therefore $\overline{\pi_0^{-1}\mathcal{I}_\rho}^{(\boldsymbol{\mu}_\rho^Y)} \subset \mathcal{I}_{\boldsymbol{\mu}_\rho^Y}$. By Proposition 70, $\mathcal{I}_{\boldsymbol{\mu}_\rho^Y} \subset \overline{\pi_0^{-1}\mathcal{I}_\rho}^{(\boldsymbol{\mu}_\rho^Y)}$. □

Exercise 72. Show that as a consequence of Corollary 71, $\boldsymbol{\mu}_\rho^Y$ is ergodic with respect to (θ^t) if and only if ρ is ergodic with respect to (μ_x^t).

We need one last lemma in order to tie everything together and complete the proof of Theorem 55.

Lemma 73. *Y is measurable in time if and only if (θ^t) is measurable.*

Proof. By definition, the collection $\{\pi_s^{-1}(A) : s \in \mathbb{T}^+, A \in \Sigma\}$ generates the σ-algebra $\mathcal{Y}$, and therefore (θ^t) is measurable if and only if for all $s \in \mathbb{T}^+$ and $A \in \Sigma$ the set

$$\{(u,(x_t)) \in \mathbb{T}^+ \times Y \,:\, x_{s+u} \in A\}$$

is $(\mathcal{B}(\mathbb{T}^+) \otimes \mathcal{Y})$-measurable. This is in turn equivalent to saying that for all $s \in \mathbb{T}^+$ the map $(u,\mathbf{x}) \mapsto \pi_{s+u}(\mathbf{x})$ is measurable. Obviously the map $u \mapsto s+u$ is measurable for any s, and therefore the previous statement is equivalent to simply saying that the map $(u,\mathbf{x}) \mapsto \pi_u(\mathbf{x})$ is measurable—which is precisely the definition of what it means for Y to be measurable in time. □

Proof of Theorem 55. Obviously, since g is integrable with respect to ρ, $g\circ\pi_0$ is integrable with respect to $\boldsymbol{\mu}_\rho^Y$. Let h be a version of $\boldsymbol{\mu}_\rho^Y(g\circ\pi_0|\mathcal{I}_{\boldsymbol{\mu}_\rho^Y})$. Since (θ^t) is measurable (by Lemma 73), we may apply Proposition 58 to yield that for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_s)_{s\in\mathbb{T}^+} \in Y$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} g(x_s)\,\lambda(ds) \;\to\; h((x_s)_{s\in\mathbb{T}^+}) \quad \text{as } t \to \infty.$$

Now by Exercise 13(C), $\hat{g}\circ\pi_0$ is a version of $\boldsymbol{\mu}_\rho^Y(g\circ\pi_0|\pi_0^{-1}\mathcal{I}_\rho)$. Therefore, by Corollary 71, $\hat{g}\circ\pi_0(\mathbf{x}) = h(\mathbf{x})$ for $\boldsymbol{\mu}_\rho^Y$-almost all $\mathbf{x} \in Y$. So we are done. □

Exercise 74. We "generalise" one of the steps in the proof of Proposition 70: Let E be a separable metric space, let $f : X \to X$ be a measurable function, let $g : X \to E$ be a measurable function, and let μ be an f-invariant probability measure. Show that if $g\circ f^n$ converges μ-almost everywhere as $n \to \infty$, then in fact $g(f^n(x)) = g(x)$ for all $n \in \mathbb{N}$, for μ-almost every $x \in X$. (The Poincaré recurrence theorem may be helpful here.)

Exercise 75. We fix any $\tau \in \mathbb{T}^+$. Prove the following statement: If $\mu^{\tau*}\rho$ is absolutely continuous with respect to ρ then for any $A \in \mathcal{Y}$, letting h be any version of $\boldsymbol{\mu}_\rho^Y(A|\mathcal{Y}_{\{0\}})$, the map $(x_t) \mapsto \mu_{x_0}^\tau(h')$ is a version of $\boldsymbol{\mu}_\rho^Y(\theta^{-\tau}(A)|\mathcal{Y}_{\{0\}})$ where $h' : X \to [0,1]$ is defined by $h = h'\circ\pi_0$. Is the converse of this statement true?

Exercise 76. Let μ be a probability measure on X, and let F^+ be some set of nonnegative measurable functions $g : X \to [0,\infty]$ that includes all the *bounded* nonnegative measurable functions. Suppose we have a family $(T_n)_{n\in\mathbb{N}\cup\{\infty\}}$ of functions $T_n : F^+ \to F^+$ such that the following hold:

(i) for any $n \in \mathbb{N}\cup\{\infty\}$ and any pointwise-increasing sequence $(g_r)_{r\in\mathbb{N}}$ in F^+, if the pointwise limit $g_\infty := \lim_{r\to\infty} g_r$ is in F^+ then $T_n(g_\infty)(x) = \sup_{r\in\mathbb{N}} T_n(g_r(x))$ for μ-almost all $x \in X$;

(ii) for any bounded $g \in F^+$, $T_n(g)(x) \to T_\infty(g)(x)$ as $n \to \infty$ for μ-almost all $x \in X$.

Now suppose we have $g, h \in F^+$ (not necessarily bounded) such that $T_n(g)(x) \to h(x)$ as $n \to \infty$ for μ-almost all $x \in X$, and $\mu(T_n(g)) \to \mu(T_\infty(g))$ as $n \to \infty$. Show that $\mu(h) = \mu(T_\infty(g))$; moreover, in the case that $\mu(T_\infty(g)) < \infty$, show that $h(x) = T_\infty(g)(x)$ for μ-almost all $x \in X$.

Exercise 77. (A) Suppose we have proved Proposition 58, except with the statement "G is a version of $\mu(g|\mathcal{I}_\rho)$" weakened to the statement that $\mu(G) = \mu(g)$. Derive from there that G is a version of $\mu(g|\mathcal{I}_\rho)$. (B) Suppose we have proved that the limit given in Proposition 58 exists and is finite for μ-almost all $x \in X$. Derive from there that G is a version of $\mu(g|\mathcal{I}_\rho)$. (Hint: first take the case that g is bounded, and then extend to the general case using Exercise 76 with $T_n(g) = \frac{1}{n}\int_{\mathbb{T}_{[0,n)}} g(f^s(\cdot))\,\lambda(ds)$ for finite n.)

Viewing a semigroup of Markov kernels as a generalisation of an autonomous dynamical system, just as we had an ergodic theorem for dynamical systems (Proposition 58), so we more generally have an "ergodic theorem for semigroups of kernels".

Proposition 78 (Ergodic theorem for semigroups of kernels)**.** *Suppose (μ_x^t) is measurable and ρ is stationary. Let $g : X \to \mathbb{R}$ be a function that is integrable with respect to ρ. Then the limit*

$$\lim_{t\to\infty} \frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(g)\,\lambda(ds) \;\; =: \;\; H(x)$$

exists and is finite for ρ-almost all $x \in X$; and (setting $H(x) := \mathrm{NaN}$ for all x where this limit does not exist) H is a version of $\rho(g|\mathcal{I}_\rho)$.

We will not write out a proof. The discrete-time case is a special case of the Chacon-Ornstein theorem; and one may pass from discrete to continuous time by following the arguments in the proof of Proposition 58 for continuous time (to obtain that the limit exists and is finite almost everywhere) and Exercise 77(B) (to obtain that the limit H is a version of $\rho(g|\mathcal{I}_\rho)$).

Exercise 79. For the special case that (X,Σ) is standard and g is bounded, derive Proposition 78 from Theorem 55 in discrete time. (If stuck, see the proof of Corollary 85 for a similar argument.) Now extend this to continuous time. (Note that the boundedness of g makes the extension from discrete to continuous time very easy.)

Corollary 80. *Suppose (μ_x^t) is measurable and ρ is ergodic. Let $g : X \to \bar{\mathbb{R}}$ be a measurable function with $\rho(g) \neq \mathrm{NaN}$. Then for ρ-almost all $x \in X$,*

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(g)\,\lambda(ds) \;\; \to \;\; \rho(g) \quad \textit{as } t \to \infty.$$

Proof. Exactly the same as in Corollary 57. □

We can also find converses to Corollary 80:

Proposition 81. *Suppose* (μ_x^t) *is measurable. If either:*

(a) for each $A \in \Sigma$*, for* ρ*-almost all* $x \in X$*,*

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(A)\,\lambda(ds) \;\to\; \rho(A) \quad \textit{as } t \to \infty\,; \; \textit{or}$$

(b) ρ *is stationary, and there exists a* π*-system* $\mathcal{C} \subset \Sigma$ *with* $\sigma(\mathcal{C}) = \Sigma$ *such that for each* $A \in \mathcal{C}$*, for* ρ*-almost all* $x \in X$*,*

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(A)\,\lambda(ds) \;\to\; \rho(A) \quad \textit{as } t \to \infty\,;$$

then ρ *is ergodic.*

We start with the following exercise:

Exercise 82. For any $\mathcal{R} \subset \Sigma$, we define the *$\rho$-orthogonal complement* $\mathcal{R}_\rho^\perp$ *of* $\mathcal{R}$ by

$$\mathcal{R}_\rho^\perp \;:=\; \{A \in \Sigma : \rho(A \cap R) = \rho(A)\rho(R) \;\forall\, R \in \mathcal{R}\}.$$

(A) Show that $\mathcal{R}_\rho^\perp$ is a λ-system on X. (B) Show that for any $A \in \Sigma$, if the constant map $x \mapsto \rho(A)$ is a version of $\rho(A|\sigma(\mathcal{R}))$ then $A \in \mathcal{R}_\rho^\perp$ (and equivalently, $\mathcal{R} \subset \{A\}_\rho^\perp$).

Proof of Proposition 81. We will first prove that if (b) holds then ρ is ergodic; we will then prove that if (a) holds then ρ is stationary. This will complete the proof.

Suppose that (b) holds. By Proposition 78, we have that for each $A \in \mathcal{C}$, the constant function $x \mapsto \rho(A)$ is a version of $\rho(A|\mathcal{I}_\rho)$. So by Exercise 82(B) $\mathcal{C}$ is contained in the ρ-orthogonal complement of $\mathcal{I}_\rho$, and hence (by the π-λ theorem and Exercise 82(A)), the ρ-orthogonal complement of $\mathcal{I}_\rho$ is the whole of Σ. It follows in particular that for all $E \in \mathcal{I}_\rho$, $\rho(E) = \rho(E)^2$, so $\rho(E) \in \{0,1\}$.

Now suppose that (a) holds. It is clear that for any simple function $g : X \to \mathbb{R}$, for ρ-almost all $x \in X$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(g)\,\lambda(ds) \;\to\; \rho(g) \quad \text{as } t \to \infty.$$

Now given any bounded measurable $g : X \to \mathbb{R}$, since we can approximate g from above and from below by simple functions, it follows that for ρ-almost all $x \in X$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(g)\,\lambda(ds) \;\to\; \rho(g) \quad \text{as } t \to \infty.$$

Now fix any $A \in \Sigma$ and $\tau \in \mathbb{T}^+$, and let $\mu_\cdot^\tau(A)$ denote the function $y \mapsto \mu_y^\tau(A)$. For any $x \in X$, for all $t \in \mathbb{T}^+ \setminus \{0\}$,

$$\underbrace{\frac{1}{\tau + t}\int_{\mathbb{T}_{[0,\tau+t)}} \mu_x^s(A)\,\lambda(ds)}_{①}$$

$$= \underbrace{\left(1 - \frac{\tau}{t+\tau}\right)}_{\to 1 \text{ as } t\to\infty}\underbrace{\left(\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(\mu_\cdot^\tau(A))\,\lambda(ds)\right)}_{②} + \underbrace{\frac{1}{\tau(t+\tau)}\int_{\mathbb{T}_{[0,\tau)}} \mu_x^s(A)\,ds}_{\to 0 \text{ as } t\to\infty}.$$

We know that ①$\to \rho(A)$ as $t \to \infty$ for ρ-almost all x. Since the map $y \mapsto \mu_y^\tau(A)$ is a bounded measurable map, we also know that ②$\to \rho(\mu_\cdot^\tau(A)) = \mu^{\tau *}\rho(A)$ as $t \to \infty$ for ρ-almost all x. Hence $\rho(A) = \mu^{\tau *}\rho(A)$. Since A and τ were arbitrary, ρ is stationary. □

Remark 83. In case (b), the condition that ρ is stationary is not redundant. (Consider, for example, the map $f : x \mapsto x + 1$ on the space $(X, \Sigma) = (\mathbb{N}, 2^{\mathbb{N}})$, with $\mathcal{C} = 2^{\mathbb{N}\setminus\{1\}}$ and $\rho = \delta_1$.)

Using Proposition 81 we can also give a converse to Corollary 57. We first give the following exercise:

Exercise 84 (*Continuous-time conditional dominated convergence theorem*)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $h : \Omega \to [0, \infty]$ be a $\mathbb{P}$-integrable function. Let $g : \Omega \times (0, \infty] \to \mathbb{R}$ be a function such that

- for each $t \in (0, \infty]$, for $\mathbb{P}$-almost all $\omega \in \Omega$, $|g(\omega, t)| \le h(\omega)$;
- $\omega \mapsto g(\omega, t)$ is measurable for each $t \in (0, \infty]$;
- for $\mathbb{P}$-almost all $\omega \in \Omega$, $g(\omega, t) \to g(\omega, \infty)$ as $t \to \infty$.

Now suppose we have a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, and a function $\tilde{g} : \Omega \times (0, \infty] \to \mathbb{R}$ such that

- for each $t \in (0, \infty]$, $\tilde{g}(\,\cdot\,, t)$ is a version of $\mathbb{P}(g(\,\cdot\,, t)|\mathcal{G})$;
- for $\mathbb{P}$-almost every $\omega \in \Omega$ there exists $k(\omega) \in [0, \infty)$ such that the map $t \mapsto \tilde{g}(\omega, t)$ is continuous on $(k(\omega), \infty)$.

Show that for $\mathbb{P}$-almost all $\omega \in \Omega$, $\tilde{g}(\omega, t) \to \tilde{g}(\omega, \infty)$ as $t \to \infty$. (Hint: first consider convergence as t tends to infinity in $\mathbb{Q}$, and then extend to $\mathbb{R}$ using continuity. For the first part it may help to look up a proof of the conditional dominated convergence theorem.)

Corollary 85. *Suppose that (μ_x^t) is measurable, Y is measurable in time, and $\boldsymbol{\mu}_\rho^Y$ exists. If either:*

(a) for each $A \in \Sigma$, for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t) \in Y$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mathbb{1}_A(x_s)\,\lambda(ds) \;\to\; \rho(A) \quad \textit{as } t \to \infty\,; \textit{ or}$$

(b) ρ is stationary, and there exists a π-system $\mathcal{C} \subset \Sigma$ with $\sigma(\mathcal{C}) = \Sigma$ such that for each $A \in \mathcal{C}$, for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t) \in Y$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mathbb{1}_A(x_s)\,\lambda(ds) \;\to\; \rho(A) \quad \text{as } t \to \infty\,;$$

then ρ is ergodic.

Proof. Fix any $A \in \Sigma$ with the property that for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t) \in Y$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mathbb{1}_A(x_s)\,\lambda(ds) \;\to\; \rho(A) \quad \text{as } t \to \infty.$$

For each $s \in \mathbb{T}^+$, Lemma 66 (with $\tau = 0$, $n = 1$, $t_1 = s$ and $B = X \times A$) gives that

$$(x_t) \mapsto \mu_{x_0}^s(A) \quad \text{is a version of} \quad \boldsymbol{\mu}_\rho^Y\big(\,(x_t) \mapsto \mathbb{1}_A(x_s)\,\big|\,\mathcal{Y}_{\{0\}}\,\big).$$

Hence, for each $t' \in \mathbb{T}^+ \setminus \{0\}$, Exercise 15 gives that

$$(x_t) \;\mapsto\; \frac{1}{t'}\int_{\mathbb{T}_{[0,t')}} \mu_{x_0}^s(A)\,\lambda(ds) \quad \text{is a version of} \quad \boldsymbol{\mu}_\rho^Y\!\left(\,(x_t) \;\mapsto\; \frac{1}{t'}\int_{\mathbb{T}_{[0,t')}} \mathbb{1}_A(x_s)\,\lambda(ds)\,\middle|\,\mathcal{Y}_{\{0\}}\right).$$

Hence the conditional dominated convergence theorem (for discrete time) or Exercise 84 (for continuous time) gives that for $\boldsymbol{\mu}_\rho^Y$-almost all $(x_t) \in Y$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_{x_0}^s(A)\,\lambda(ds) \;\to\; \rho(A) \quad \text{as } t \to \infty.$$

Now it is clear (e.g. using Lemma 5(B)) that the set of all $x \in X$ with the property that

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(A)\,\lambda(ds) \;\to\; \rho(A) \quad \text{as } t \to \infty$$

is a measurable set. Hence we have that for ρ-almost all $x \in X$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(A)\,\lambda(ds) \;\to\; \rho(A) \quad \text{as } t \to \infty.$$

This reduces the problem to Proposition 81, which we have already proved. □

Section 4 Appendix: Processes with stationary and independent increments

Imagine we have a stochastic process taking values in a space that is naturally equipped with some group operation; for example, we could have an $\mathbb{R}^n$-valued stochastic process, where $\mathbb{R}^n$ is viewed as a group under addition, or we could have a *stochastic flow* on a smooth manifold M—that is, a Diffeo(M)-valued stochastic process, where Diffeo(M) is viewed as a group under composition. In such cases it makes sense to talk about the "increment" in the value of the stochastic process between some "start" time and some "end" time. We will now introduce the notion of "stationary and independent increments", and its relation to the "diagonal shift dynamical system".

Suppose Σ includes all the singletons in X.[12] Suppose we have a group operation $\circ$ on X such that the maps $(x,y) \mapsto y \circ x$ and $x \mapsto x^{-1}$ are measurable, and let $e \in X$ be the identity. Define $DY \subset Y$ by $DY := \pi_0^{-1}(\{e\})$; we equip DY with the σ-algebra $D\mathcal{Y}$ of $\mathcal{Y}$-measurable subsets of DY. Given two probability measures ν_1 and ν_2 on X, we define their *composition* $\nu_2 \circ \nu_1$ to be the image measure under the map $(x,y) \mapsto y \circ x$ of $\nu_1 \otimes \nu_2$, that is

$$\nu_2 \circ \nu_1(A) = \int_{X \times X} \mathbb{1}_A(y \circ x)\, \nu_1 \otimes \nu_2(d(x,y))$$

for all $A \in \Sigma$. We will say that a family $(\nu^t)_{t \in \mathbb{T}^+}$ of probability measures ν^t on X is *consistent* if $\nu^0 = \delta_e$ and $\nu^{s+t} = \nu^t \circ \nu^s$ for all $s,t \in \mathbb{T}^+$. Given a consistent family (ν^t), for any $x \in X$ and $t \in \mathbb{T}^+$ we define the probability measure $\bar{\nu}_x^t$ on X to be the image measure under the map $y \mapsto y \circ x$ of ν^t (i.e. $\bar{\nu}_x^t = \nu^t \circ \delta_x$); that is,

$$\bar{\nu}_x^t(A) = \int_X \mathbb{1}_A(y \circ x)\, \nu^t(dy).$$

Exercise 86. Let (ν^t) be a consistent family of probability measures. Show that $(\bar{\nu}_x^t)_{x \in X,\, t \in \mathbb{T}^+}$ is a semigroup of Markov kernels on X.

So, given a consistent family (ν^t), for any probability measure ρ on X we can ask whether the Markov measure $\bar{\boldsymbol{\nu}}_\rho^Y$ exists. If $\bar{\boldsymbol{\nu}}_{\delta_e}^Y$ exists then we will define the probability measure $\boldsymbol{\nu}^Y$ on DY to be the restriction of $\bar{\boldsymbol{\nu}}_{\delta_e}^Y$ to $D\mathcal{Y}$; otherwise we say that $\boldsymbol{\nu}^Y$ does not exist.

Exercise 87. Given a consistent family (ν^t), show that if $\boldsymbol{\nu}^Y$ exists then $\bar{\boldsymbol{\nu}}_\rho^Y$ exists for any probability measure ρ on X.

Exercise 88. Let $\boldsymbol{\mu}$ be a probability measure on Y, and let $(\nu^t)_{t \in \mathbb{T}^+}$ be a family of probability measures on X. We will say that $\boldsymbol{\mu}$ *has stationary and independent increments distributed according to* (ν^t) if for any $t_1, t_2 \in \mathbb{T}^+$ with $t_1 \leq t_2$ the following hold:

(a) the image measure of $\boldsymbol{\mu}$ under the map $(x_t) \mapsto x_{t_2} \circ x_{t_1}^{-1}$ is equal to $\nu^{t_2 - t_1}$;

(b) on the probability space $(Y, \mathcal{Y}, \boldsymbol{\mu})$, the random variable $(x_t) \mapsto x_{t_2} \circ x_{t_1}^{-1}$ is independent of $\mathcal{Y}_{[0,t_1]}$.

Show that the following are equivalent:

(i) $\boldsymbol{\mu}$ has stationary and independent increments distributed according to (ν^t);

(ii) (ν^t) is consistent, $\boldsymbol{\nu}^Y$ exists, and the image measure of $\boldsymbol{\mu}$ under the map $(x_t) \mapsto (x_t \circ x_0^{-1})$ from Y to DY is equal to $\boldsymbol{\nu}^Y$.

(A Y-valued random variable whose law has stationary and independent increments is itself said to be a stochastic process with stationary and independent increments.)

Now we define the *diagonal shift dynamical system* $(D\theta^t)_{t \in \mathbb{T}^+}$ to be the $\mathbb{T}^+$-indexed family of maps $D\theta^t : DY \to DY$ given by $D\theta^\tau((x_t)_{t \in \mathbb{T}^+}) = (x_{\tau + t} \circ x_\tau^{-1})_{t \in \mathbb{T}^+}$ for all $\tau \in \mathbb{T}^+$ and $(x_t)_{t \in \mathbb{T}^+} \in DY$.

[12]This assumption is not actually required to be able to formulate valid definitions of the main concepts in this appendix; however, it does allow for a more straightforward exposition.

Exercise 89. (A) Show that $(D\theta^t)$ is indeed an autonomous dynamical system on DY, and that if Y is measurable in time then $(D\theta^t)$ is measurable. (B) Let (ν^t) be a consistent family, and suppose that $\boldsymbol{\nu}^Y$ exists. Show that $\boldsymbol{\nu}^Y$ is invariant with respect to $(D\theta^t)$.

Let us mention some important examples:

(I) [Poisson process] Let $(X,\Sigma,\circ) = (\mathbb{Z},2^{\mathbb{Z}},+)$, let $\mathbb{T}^+ = [0,\infty)$, and let Y be the set of all $(x_t)_{t\geq 0} \in \mathbb{Z}^{[0,\infty)}$ such that $t \mapsto x_t$ is càdlàg and $x_t - \lim_{s\to t^-} x_s \in \{0,1\}$ for all $t \in (0,\infty)$. Fix a value $\lambda \in (0,\infty)$, and for each $t \in [0,\infty)$ define the probability measure ν^t on $\mathbb{Z}$ to be the Poisson distribution with parameter λt, that is,

$$\nu^t(\{n\}) \;=\; \begin{cases} \frac{(\lambda t)^n \exp(-\lambda t)}{n!} & n \geq 0 \\ 0 & n < 0. \end{cases}$$

One can show that $(\nu^t)_{t\geq 0}$ is consistent, and that $\boldsymbol{\nu}^Y$ exists. Specifically, $\boldsymbol{\nu}^Y$ may be obtained as follows: Let $R := \{(t_n)_{n\in\mathbb{N}} \in (0,\infty)^{\mathbb{N}} : \sum_{i=1}^{\infty} t_i = \infty\}$, with $\mathcal{R}$ being the σ-algebra of $\mathcal{B}((0,\infty))^{\otimes\mathbb{N}}$-measurable subsets of R. Define the probability measure ϵ on $(0,\infty)$ to be the exponential distribution of decay rate λ, that is, $\epsilon(A) := \lambda\int_A \exp(-\lambda t)\,dt$ for all $A \in \Sigma$. Then one can show that $\boldsymbol{\nu}^Y$ is equal to the image measure of $\epsilon^{\otimes\mathbb{N}}|_{\mathcal{R}}$ under the map

$$\begin{aligned} R &\to DY \\ (t_n)_{n\in\mathbb{N}} &\mapsto \left(\max\left\{r \in \mathbb{N}\cup\{0\} : \sum_{i=1}^{r} t_i \leq t\right\}\right)_{t\geq 0}. \end{aligned}$$

A DY-valued random variable with law $\boldsymbol{\nu}^Y$ is called a *Poisson process of intensity* λ; our construction of $\boldsymbol{\nu}^Y$ essentially states that a Poisson process can be constructed as a cumulative count of events where the time-spacings between consecutive events are i.i.d. exponentially distributed random times. (A proof of this fact can be found in many textbooks and lecture notes on stochastic processes.)

(II) [Standard Brownian motion] Fix $n \in \mathbb{N}$, and let $(X,\Sigma,\circ) = (\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),+)$. Let $\mathbb{T}^+ = [0,\infty)$, and let Y_n be the set of all $(x_t)_{t\geq 0} \in (\mathbb{R}^n)^{[0,\infty)}$ such that $t \mapsto x_t$ is continuous. For each $t \in \mathbb{T}^+$, define the probability measure ${}^n\nu^t$ on $\mathbb{R}^n$ to be the Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $t\mathbf{I}^{(n)}$, that is,

$${}^n\nu^t(A) \;:=\; \frac{1}{(2\pi t)^{\frac{n}{2}}}\int_A \exp\left(\tfrac{1}{2t}|x|^2\right)\,\lambda_n(dx)$$

for $t > 0$, where λ_n denotes the Lebesgue measure on $\mathbb{R}^n$, and ${}^n\nu^0 = \delta_{\mathbf{0}}$. (Note that ${}^n\nu^t = ({}^1\nu^t)^{\otimes n}$.) Again, one can show that $({}^n\nu^t)_{t\geq 0}$ is consistent. One of the great theorems of the last century is that ${}^{\boldsymbol{n}}\boldsymbol{\nu}^{Y_n}$ exists! A proof of this fact can be found in many textbooks and lecture notes on stochastic calculus or Brownian motion. A DY_n-valued random variable with law ${}^{\boldsymbol{n}}\boldsymbol{\nu}^{Y_n}$ is called an n-dimensional *Wiener process* or *standard Brownian motion*. It is easy to show that an n-dimensional Wiener process is precisely the concatenation of n independent 1-dimensional Wiener processes. (In other words, identifying $(\mathbb{R}^n)^{[0,\infty)}$ with $(\mathbb{R}^{[0,\infty)})^n$ in the obvious manner, it is easy to show that ${}^{\boldsymbol{n}}\boldsymbol{\nu}^{Y_n} = \left({}^{\boldsymbol{1}}\boldsymbol{\nu}^{Y_1}\right)^{\otimes n}$.)

(III) [Brownian motion] Fix $n \in \mathbb{N}$, and let $(X, \Sigma, \circ)$ and Y_n be as in **(II)**. Given any $b \in \mathbb{R}^n$ and any symmetric nonnegative-definite matrix $C \in \mathbb{R}^{n\times n}$, for each $t \in \mathbb{T}^+$, define the probability measure ${}^{b,C}\nu^t$ on $\mathbb{R}^n$ to be the Gaussian distribution with mean tb and covariance matrix tC, that is, the unique probability measure on $\mathbb{R}^n$ satisfying

$$\int_{\mathbb{R}^n} e^{iu^{\mathrm{T}}x}\ {}^{b,C}\nu^t(dx) \;=\; e^{tu^{\mathrm{T}}(ib-\frac{1}{2}Cu)}$$

for all $u \in \mathbb{R}^n$. Again, one can show that $({}^{b,C}\nu^t)_{t\geq 0}$ is consistent. Using the fact that all Gaussian distributions can be expressed as affine transformations of the standard Gaussian distribution, the existence of Wiener processes introduced in **(II)** implies the existence of ${}^{b,C}\boldsymbol{\nu}^{Y_n}$. A Y_n-valued random variable whose law takes the form ${}^{b,C}\bar{\boldsymbol{\nu}}_\rho^{Y_n}$ for some probability measure ρ on $\mathbb{R}^n$ is called a *Brownian motion with initial distribution ρ, drift b and diffusion C*. Remarkably, we have the following theorem: for any consistent family $(\nu^t)_{t\geq 0}$ on Y_n, $\boldsymbol{\nu}^{Y_n}$ exists if and only if there exist $b \in \mathbb{R}^n$ and a symmetric nonnegative-definite $C \in \mathbb{R}^{n\times n}$ such that $(\nu^t)_{t\geq 0} = ({}^{b,C}\nu^t)_{t\geq 0}$. (For a proof, see Theorem 1 of here.)

Exercise 90. Let $\mathbb{T}^+ = \mathbb{N} \cup \{0\}$ and let $(X, \Sigma, \circ) = (\mathbb{Z}, 2^{\mathbb{Z}}, +)$. (A) [Simple random walk] Let (ν^n) be the consistent family such that $\nu^1 = \frac{1}{2}(\delta_{-1} + \delta_1)$. Show that if Y is the set of all sequences of integers, then $\boldsymbol{\nu}^Y$ exists, but if Y is the set of bounded sequences of integers then $\boldsymbol{\nu}^Y$ does not exist. (B) Give an example of an integer n, a set $A \subset \mathbb{Z}^n$, an increasing sequences of positive integers $(n_r)_{r\in\mathbb{N}}$ and an assignment of a set $A \subset \mathbb{Z}^{n_r}$ to each $r \in \mathbb{N}$, such that the following hold:

(i) for any $r_1 < r_2$, $(A_{r_1} \times \mathbb{Z}^{n_{r_2}-n_{r_1}}) \cap A_{r_2} = \varnothing$;

(ii) if Y is the set of bounded sequences of integers then

$$\pi^{-1}_{0,\dots,n-1}(A) \;=\; \bigcup_{r=1}^{\infty} \pi^{-1}_{0,\dots,n_r-1}(A_r)\,;$$

(iii) if Y is the set of all sequences of integers then

$$\pi^{-1}_{0,\dots,n-1}(A) \;\neq\; \bigcup_{r=1}^{\infty} \pi^{-1}_{0,\dots,n_r-1}(A_r).$$

(The purpose of Exercise 90(B) is to demonstrate that, even though Carathéodory's extension theorem guarantees the existence of $\boldsymbol{\mu}_\rho^Y$ if the map $\pi^{-1}_{0,t_1,\dots,t_n}(A) \mapsto \boldsymbol{\mu}_\rho^{t_1,\dots,t_n}(A)$ is countably additive on the class $\{\pi^{-1}_{0,t_1,\dots,t_n}(A) : n \in \mathbb{N}, A \in \Sigma^{\otimes(n+1)}, t_1 \leq \ldots \leq t_n\}$ of cylinder sets in Y, nonetheless the countable additivity of this map may depend on Y.)

5 Ergodic decomposition

In Section 2 we saw that the class of stationary probability measures is a convex set whose extreme points are precisely the ergodic probability measures. In this section we introduce a further deep fact about the "structure" of the class of stationary probability measures (under appropriate conditions).

Let us begin by saying that if Σ is countably generated—meaning that there exists a countable generator $\mathcal{A} \subset \Sigma$ of Σ—then there also exists a *countable π-system* $\tilde{\mathcal{A}}$ generating Σ. To see this, observe that for any collection of sets $\mathcal{A}$, the set $\tilde{\mathcal{A}}$ of all finite intersections of members of $\mathcal{A}$ is obviously a π-system (in fact, it is clearly the smallest π-system containing $\mathcal{A}$); and if $\mathcal{A}$ is countable then $\tilde{\mathcal{A}}$ is obviously also countable.

Exercise 91. Let $(\Omega, \mathcal{F})$ be a measurable space, and let $p_1, p_2 : \Omega \to \mathcal{M}_1$ be measurable functions. Show that if Σ is countably generated then $\{\omega \in \Omega : p_1(\omega) = p_2(\omega)\} \in \mathcal{F}$.

Recall that (X, Σ) is said to be *standard* if there exists a Polish topology on X whose Borel σ-algebra coincides with Σ. By the *Borel isomorphism theorem* (Theorem 5 of here), any two uncountable standard measurable spaces are measurably isomorphic to each other.

ASSUMPTION: Throughout the rest of this section, (X, Σ) is standard, (μ_x^t) is a measurable semigroup of kernels, and ρ is a probability measure on X.

Suppose ρ is stationary; then an *ergodic decomposition* $(I, \mathcal{I}, \nu, (\hat{\rho}_\alpha)_{\alpha \in I})$ *of* ρ consists of a probability space $(I, \mathcal{I}, \nu)$ and an I-indexed family $(\hat{\rho}_\alpha)_{\alpha \in I}$ of probability measures $\hat{\rho}_\alpha$ on X such that (i) the map $\alpha \mapsto \hat{\rho}_\alpha$ is measurable, (ii) $\hat{\rho}_\alpha$ is ergodic for ν-almost all $\alpha \in I$, and (iii) $\rho(A) = \int_I \hat{\rho}_\alpha(A)\, \nu(d\alpha)$ for all $A \in \Sigma$.

In essence, the main result of this section is that if ρ is stationary then an ergodic decomposition exists. More specifically, we will construct (up to ρ-almost everywhere equality) a particular Markov kernel $(\hat{\rho}_x)_{x \in X}$ on X, which we call the "canonical ergodic decomposition" of ρ, and we will then prove that $(X, \Sigma, \rho, (\hat{\rho}_x)_{x \in X})$ really is an ergodic decomposition of ρ.

In order to define these "canonical ergodic decompositions", we will need to introduce the notion of *conditional distributions*. (The following definition has nothing to do with the semigroup (μ_x^t); it is a general concept in probability theory.)

Definition 92. Recall that (X, Σ, ρ) is a probability space, with (X, Σ) standard. Now let $\mathcal{G}$ be a sub-σ-algebra of Σ. We will say that a Markov kernel $(\nu_x)_{x \in X}$ is *a version of* $\rho(\,\cdot\,|\mathcal{G})$ [to be read: "a version of the conditional distribution of ρ given $\mathcal{G}$"] if for every $A \in \Sigma$, the map $x \mapsto \nu_x(A)$ is a version of $\rho(A|\mathcal{G})$.

Observe, in particular, that the map $x \mapsto \nu_x$ is a $\mathcal{G}$-measurable map from X to $\mathcal{M}_1$, and that ρ is stationary with respect to the kernel (ν_x).

Proposition 93 (Existence and essential uniqueness of conditional distributions)**.** *Let $\mathcal{G}$ be a sub-σ-algebra of Σ. (A) There exists a version of $\rho(\,\cdot\,|\mathcal{G})$. (B) Let (ν_x^1) be a version of $\rho(\,\cdot\,|\mathcal{G})$, and let (ν_x^2) be another kernel on X with $x \mapsto \nu_x^2$ being a $\mathcal{G}$-measurable map. Then (ν_x^2) is a version of $\rho(\,\cdot\,|\mathcal{G})$ if and only if $\rho(x \in X : \nu_x^1 \neq \nu_x^2) = 0$.*

Proof. [Adapted from Billingsley, Theorem 33.3.] (A) First suppose X is finite or countable, and for each $x \in X$ let $h_x : X \to [0,1]$ be a version of $\rho(\{x\}|\mathcal{G})$. By the σ-additivity of conditional probabilities there exists a ρ-full set $\tilde{X} \in \mathcal{G}$ such that $\sum_{y \in X} h_y(x) = 1$ for all $x \in X'$. So for every $x \in \tilde{X}$ we may define the probability measure ν_x on X by

$$\nu_x(A) \;=\; \sum_{y \in A} h_y(x)\,;$$

and fixing some arbitrary probability measure c on X, we may set $\nu_x := c$ for all $x \in X \setminus \tilde{X}$. The σ-additivity of conditional probabilities gives that for each $A \subset X$ the map $x \mapsto \nu_x(A)$ is a version of $\rho(A|\mathcal{G})$.

Now suppose that X is uncountable. We may assume without loss of generality that $(X, \Sigma) = ([0,1], \mathcal{B}([0,1]))$. For each $a \in (0,1) \cap \mathbb{Q}$, let $H_a : X \to [0,1]$ be a version of $\rho([0,a]|\mathcal{G})$; and let $H_0(x) := 0$ and $H_1(x) := 1$ for all $x \in X$. Since $[0,1] \cap \mathbb{Q}$ is countable, there exists a ρ-full set $X' \in \mathcal{G}$ such that for all $x \in X'$ the map $a \mapsto H_a(x)$ from $[0,1] \cap \mathbb{Q}$ to $[0,1]$ is increasing. By the conditional dominated convergence theorem (together with the countability of $[0,1) \cap \mathbb{Q}$), there exists a ρ-full set $X'' \in \mathcal{G}$ such that for all $x \in X''$ and $a \in [0,1) \cap \mathbb{Q}$, $H_{a+\frac{1}{n}}(x) \to H_a(x)$ as $n \to \infty$. So, if we let $\tilde{X} := X' \cap X''$, then we have that for each $x \in \tilde{X}$ there exists a unique probability measure ν_x on X such that $\nu_x([0,a]) = H_a(x)$ for all $a \in [0,1] \cap \mathbb{Q}$. Fixing some arbitrary probability measure c on X, let us set $\nu_x := c$ for all $x \in X \setminus \tilde{X}$. Now let $\mathcal{D} \subset \Sigma$ be the set of all $A \in \Sigma$ such that the map $x \mapsto \nu_x(A)$ is a version of $\rho(A|\mathcal{G})$. It is clear that $\mathcal{D}$ includes the collection of sets of the form $[0,a]$ with $a \in [0,1] \cap \mathbb{Q}$ (which is itself a π-system generating Σ). Using the σ-additivity of conditional probabilities (together with the fact that $X \in \mathcal{D}$) we have that $\mathcal{D}$ is a λ-system. Hence the π-λ theorem gives the desired result.

(B) It is clear that if $\rho(x \in X : \nu_x^1 \neq \nu_x^2) = 0$ then (ν_x^2) is a version of $\rho(\,\cdot\,|\mathcal{G})$. Now suppose that (ν_x^2) is a version of $\rho(\,\cdot\,|\mathcal{G})$. Let $\mathcal{A}$ be a countable π-system generating Σ. It is clear that for ρ-almost every $x \in X$, for all $A \in \mathcal{A}$, $\nu_x^1(A) = \nu_x^2(A)$; so the π-λ theorem gives the desired result. $\square$

Exercise 94. Let $\mathcal{G}$ be a sub-σ-algebra of Σ, and let (ν_x) be a version of $\rho(\,\cdot\,|\mathcal{G})$. Show that for any $g : X \to \mathbb{R}'$ which is integrable with respect to ρ, the map $x \mapsto \nu_x(g)$ is a version of $\rho(g|\mathcal{G})$.

Exercise 95. Given σ-algebras $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \Sigma$, we will say that $\mathcal{G}_2$ *is a ρ-trivial extension of* $\mathcal{G}_1$ if $\mathcal{G}_2 \subset \sigma(\mathcal{G}_1 \cup \mathcal{N}_\rho)$, where $\mathcal{N}_\rho$ is the set of all Σ-measurable ρ-null sets. Show that if $\mathcal{G}_2$ is a ρ-trivial extension of $\mathcal{G}_1$ then any version of $\rho(\,\cdot\,|\mathcal{G}_1)$ is also a version of $\rho(\,\cdot\,|\mathcal{G}_2)$.

Exercise 96. Proposition 93 may be obtained as a special case of the *disintegration theorem*[13], which states the following: Continuing to assume that (X, Σ) is standard, let $(\Omega, \mathcal{F})$ be a measurable space, let μ be a probability measure on $\Omega \times X$, and let $\mathbb{P}$ be the image measure of μ under the projection $(\omega, x) \mapsto \omega$; then there exists a measurable function $\tilde{\mu} : \Omega \to \mathcal{M}_1$, unique up to $\mathbb{P}$-almost-everywhere equality, such that for every $A \in \mathcal{F} \otimes \Sigma$,

$$\mu(A) \;=\; \int_\Omega \int_X \mathbb{1}_A(\omega, x)\, \tilde{\mu}(\omega)(dx)\, \mathbb{P}(d\omega).$$

[13] The disintegration theorem may actually be proved in a very similar manner to Proposition 93: in place of *a version of $\rho(A|\mathcal{G})$*, one considers *a version of the density of the measure $E \mapsto \mu(E \times A)$ with respect to $\mathbb{P}$.*

(A) Obtain Proposition 93 as a special case of the disintegration theorem. (B) Hence, given a sub-σ-algebra $\mathcal{G}$ of Σ and a version (ν_x) of $\rho(\cdot|\mathcal{G})$, show that for any $A \in \mathcal{G} \otimes \Sigma$

$$\rho(\, x \in X \,:\, (x,x) \in A \,) \;=\; \int_X \int_X \mathbb{1}_A(x,y)\, \nu_x(dy)\, \rho(dx).$$

Now prove the same formula without reference to the disintegration theorem. (Hint: first consider $(\mathcal{G} \otimes \Sigma)$-measurable rectangles in $X \times X$.)

Lemma 97. *Let $\mathcal{G}$ be a sub-σ-algebra of Σ, and let (ν_x) be a version of $\rho(\cdot|\mathcal{G})$. Then for ρ-almost every $x \in X$, $\nu_x(y \in X : \nu_y = \nu_x) = 1$.*

Proof. Let $\mathcal{A} \subset \Sigma$ be a countable π-system generating Σ, and for each $A \in \mathcal{A}$ let

$$E_A \;:=\; \{(x,y) \in X \times X \,:\, \nu_x(A) = \nu_y(A)\}.$$

It is clear that $E_A \in \mathcal{G} \otimes \mathcal{G}$. Now if we let

$$E \;:=\; \{(x,y) \in X \times X \,:\, \nu_x = \nu_y\}$$

then the π-λ theorem gives that $E = \bigcap_{A \in \mathcal{A}} E_A$. So $E \in \mathcal{G} \otimes \mathcal{G} \subset \mathcal{G} \otimes \Sigma$. Hence Exercise 96(B) gives that

$$\int_X \int_X \mathbb{1}_E(x,y)\, \nu_x(dy)\, \rho(dx) \;=\; \rho(\, x \in X \,:\, (x,x) \in E\,) \;=\; 1$$

and therefore, for ρ-almost every $x \in X$, $\int_X \mathbb{1}_E(x,y)\, \nu_x(dy) = 1$. But $\int_X \mathbb{1}_E(x,y)\, \nu_x(dy)$ is precisely $\nu_x(y \in X : \nu_y = \nu_x)$. So we are done. □

Now recall once again that (μ_x^t) is a measurable semigroup of kernels on X. As before, if ρ is stationary then we shall write $\mathcal{I}_\rho$ to denote the set of ρ-almost invariant sets.

Proposition 98. *Suppose ρ is stationary, let $\mathcal{G}$ be any sub-σ-algebra of $\mathcal{I}_\rho$, and let (ν_x) be a version of $\rho(\cdot|\mathcal{G})$. Then ν_x is stationary for ρ-almost all $x \in X$.*

Proof. We first show that for each $t \in \mathbb{T}^+$, $(\mu^{t*}\nu_x)_{x \in X}$ is a version of $\rho(\cdot|\mathcal{G})$. Fix any $t \in \mathbb{T}^+$, $A \in \Sigma$ and $G \in \mathcal{G}$.

$$\begin{aligned}
\int_G \mu^{t*}\nu_x(A)\, \rho(dx) \;&=\; \int_G \int_X \mu_y^t(A)\, \nu_x(dy)\, \rho(dx) \\
&=\; \int_G \mu_x^t(A)\, \rho(dx) \quad \text{(by Exercise 94 with } g(x) = \mu_x^t(A)) \\
&=\; \rho(A \cap G) \quad \text{(e.g. by Exercise 26(C))}.
\end{aligned}$$

So $(\mu^{t*}\nu_x)_{x \in X}$ is a version of $\rho(\cdot|\mathcal{G})$, and therefore $\mu^{t*}\nu_x = \nu_x$ for ρ-almost all $x \in X$. This is true for any given $t \in \mathbb{T}^+$. Now by Exercise 91 (with $\Omega = \mathbb{T}^+ \times X$), the set $\{(t,x) : \mu^{t*}\nu_x = \nu_x\}$ is a measurable set; consequently, by Corollary 14, ρ-almost every $x \in X$ has the property that for λ-almost all $t \in \mathbb{T}^+$, $\mu^{t*}\nu_x = \nu_x$. By Exercise 32(A), all such x actually have the property that for every $t \in \mathbb{T}^+$, $\mu^{t*}\nu_x = \nu_x$. So we are done. □

Theorem 99. *Suppose ρ is stationary, and let $(\hat{\rho}_x)$ be a version of $\rho(\cdot|\mathcal{I}_\rho)$. Then $\hat{\rho}_x$ is ergodic for ρ-almost all $x \in X$.*

Corollary 100 (Ergodic decomposition theorem). *Suppose ρ is stationary. Letting $(\hat{\rho}_x)$ be a version of $\rho(\cdot|\mathcal{I}_\rho)$, $(X,\Sigma,\rho,(\hat{\rho}_x))$ is an ergodic decomposition of ρ. Hence there exists a probability measure Q on the set $\mathcal{M}_e$ of ergodic probability measures (equipped with its natural σ-algebra $\sigma(\mu \mapsto \mu(A) : A \in \Sigma)$) such that*

$$\rho(A) = \int_{\mathcal{M}_e} \mu(A)\, Q(d\mu)$$

for all $A \in \Sigma$.

Observe that in Corollary 100, $(\mathcal{M}_e,\mathcal{K}_e,Q,(\mu)_{\mu\in\mathcal{M}_e})$ is itself an ergodic decomposition of ρ (where $\mathcal{K}_e$ denotes the natural σ-algebra on $\mathcal{M}_e$).

One immediate important consequence of Theorem 99 / Corollary 100 is that if (μ_x^t) admits a stationary probability measure then it admits an ergodic probability measure, and if it admits more than one stationary probability measure then it admits more than one ergodic probability measure.

Proof of Corollary 100. It is clear that $(X,\Sigma,\rho,(\hat{\rho}_x))$ is an ergodic decomposition of ρ. Letting $\tilde{X} \in \Sigma$ be a ρ-full set such that $\hat{\rho}_x$ is ergodic for all $x \in \tilde{X}$, and letting $\tilde{\Sigma}$ be the set of Σ-measurable subsets of $\tilde{X}$, take Q to be the image measure of $\rho|_{\tilde{\Sigma}}$ under the map $x \mapsto \hat{\rho}_x$ from $\tilde{X}$ to $\mathcal{M}_e$. □

Remark 101. Suppose there exists a σ-locally compact metric on X whose Borel σ-algebra coincides with Σ, such that the map $x \mapsto \mu_x^t$ is continuous (with respect to the narrow topology) for all $t \in \mathbb{T}^+$. Then the second assertion in Corollary 100 is a special case of "Choquet's theorem" in convex analysis (and in fact, for it to be true, the condition that (μ_x^t) is measurable can be dropped); this can be shown by following the arguments presented here. [Semigroups of kernels with narrow-continuous dependence on the spatial parameter are said to be *Feller-continuous*; such semigroups will be discussed more in the next section.]

Remark 102. A natural question to ask is whether the probability measure Q in Corollary 100 is unique. The answer is yes (see e.g. [18] and references therein for further details).

Exercise 103. (A) Using Proposition 49, show how one can derive the continuous-time case of Theorem 99 if one already knows the discrete-time case. (B) The reference given in Remark 102 actually considers Markov *kernels* (which are equivalent to discrete-time semigroups of kernels). Using Proposition 49, prove the assertion in Remark 102 for continuous-time measurable semigroups.

Exercise 104. Let $(\Omega,\mathcal{F})$ be a measurable space, and let $(E,\mathcal{E})$ be a measurable space such that the diagonal in $E \times E$ is $(\mathcal{E}\otimes\mathcal{E})$-measurable (e.g. $\mathcal{E}$ could be the Borel σ-algebra of a second-countable Hausdorff topology). Let $g : \Omega \to E$ be a measurable function. (A) [*Measurable graph theorem*] Show that the graph of g is an element of $\mathcal{F}\otimes\mathcal{E}$. (Hint: consider the map $(\omega,x) \mapsto (g(\omega),x)$.) (B) [*Measurable image theorem*] Assume moreover that $(\Omega,\mathcal{F})$ is standard. Show that for any $A \in \mathcal{F}$, $g(A)$ is universally measurable with respect to $\mathcal{E}$. (Hint: recall the measurable projection theorem.) (C) Still taking $(\Omega,\mathcal{F})$ to be standard, let m be a measure on $(\Omega,\mathcal{F})$. Show that for any m-full $A \subset \Omega$, $g(A)$ is g_*m-full. Is it generally true that for any m-null $A \subset \Omega$, $g(A)$ is g_*m-null?

Remark 105. Suppose ρ is stationary. By Exercise 104(C), if we define $R : X \to \mathcal{M}_1$ by $R(x) = \hat{\rho}_x$ then $\mathcal{M}_e$ is an $R_*\rho$-full subset of $\mathcal{M}_1$.

Now before proving Theorem 99, we give the following exercise.

Exercise 106. Find the incorrect assertion within the following "proof" of Theorem 99, and give an actual example for which that particular assertion fails. [Hint: consider the shift map on the infinite product of a non-trivial probability space.]

"Let $\mathcal{I}^+$ be the σ-algebra generated by the forward-invariant sets under (μ_x^t), and let $(\tilde{\rho}_x)$ be a version of $\rho(\,\cdot\,|\mathcal{I}^+)$. Using Corollary 44 we have that $\mathcal{I}_\rho$ is a ρ-trivial extension of $\mathcal{I}^+$, and hence (by Exercise 95) $(\tilde{\rho}_x)$ is also a version of $\rho(\,\cdot\,|\mathcal{I}_\rho)$. So, since $\tilde{\rho}_x = \hat{\rho}_x$ for ρ-almost all $x \in X$, it will suffice to show that $\tilde{\rho}_x$ is ergodic for ρ-almost all $x \in X$. Now by Proposition 98, $\tilde{\rho}_x$ is stationary for ρ-almost all $x \in X$. So, by Theorem 39, we only need to show that for ρ-almost all $x \in X$, $\tilde{\rho}_x$ assigns trivial measure to every forward-invariant set: but since $(\tilde{\rho}_x)$ is already a version of the conditional distribution of ρ given $\mathcal{I}^+$, we have that for ρ-almost all x, for any forward-invariant set $G \in \mathcal{I}^+$,

$$\tilde{\rho}_x(G) \;=\; \begin{cases} 1 & x \in G \\ 0 & x \notin G. \end{cases}$$

QED."

Proof of Theorem 99. Let $\mathcal{C} \subset \Sigma$ be a countable π-system generating Σ. By Proposition 78, for all $A \in \mathcal{C}$, for ρ-almost all $y \in X$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_y^s(A)\,\lambda(ds) \;\to\; \hat{\rho}_y(A) \quad \text{as } t \to \infty.$$

Therefore, for all $A \in \mathcal{C}$, ρ-almost every $x \in X$ has the property that for $\hat{\rho}_x$-almost all $y \in X$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_y^s(A)\,\lambda(ds) \;\to\; \hat{\rho}_y(A) \quad \text{as } t \to \infty.$$

Since $\mathcal{C}$ is countable, it follows that ρ-almost every $x \in X$ has the property that for all $A \in \mathcal{C}$, for $\hat{\rho}_x$-almost all $y \in X$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_y^s(A)\,\lambda(ds) \;\to\; \hat{\rho}_y(A) \quad \text{as } t \to \infty.$$

By Lemma 97 it follows that ρ-almost every $x \in X$ has the property that for all $A \in \mathcal{C}$, for $\hat{\rho}_x$-almost all $y \in X$,

$$\frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_y^s(A)\,\lambda(ds) \;\to\; \hat{\rho}_x(A) \quad \text{as } t \to \infty.$$

We also know from Proposition 98 that $\hat{\rho}_x$ is stationary for ρ-almost all $x \in X$. Hence case (b) of Proposition 81 gives that $\hat{\rho}_x$ is ergodic for ρ-almost all $x \in X$. $\square$

Exercise 107 (Simultaneous ergodic decomposition)**.** Fix a Polish topology on X whose Borel σ-algebra coincides with Σ. For each $x \in X$ and $t \in \mathbb{T}^+ \setminus \{0\}$, define the probability measure $\bar{\mu}_x^t$ on X by

$$\bar{\mu}_x^t(A) \;=\; \frac{1}{t}\int_{\mathbb{T}_{[0,t)}} \mu_x^s(A)\,\lambda(ds)$$

for all $A \in \Sigma$. Now fix a probability measure c on X, and for each $x \in X$ define the probability measure $\bar{\mu}_x^\infty$ on X by

$$\bar{\mu}_x^\infty = \begin{cases} \lim\limits_{t\to\infty} \bar{\mu}_x^t & \text{if this limit exists} \\ c & \text{otherwise} \end{cases}$$

where the limit is taken in the narrow topology. Prove that for *any* stationary probability measure w, $(\bar{\mu}_x^\infty)_{x\in X}$ is a version of $w(\cdot|\mathcal{I}_w)$. (Hint: recall Lemma 19.)

Now we have said that if (μ_x^t) admits a stationary probability measure then it admits an ergodic probability measure. We will give a generalisation of this fact. Fix $n \in \mathbb{N}$. Recall that a set $S \subset \mathbb{R}^n$ is said to be *convex* if $\lambda\mathbf{a} + (1-\lambda)\mathbf{b} \in S$ for all $\mathbf{a}, \mathbf{b} \in S$ and $\lambda \in [0,1]$. We first give the following general fact about convex sets.

Proposition 108. *Let $(\Omega, \mathcal{F}, m)$ be a measure space, let $g : \Omega \to \mathbb{R}^n$ be an m-integrable function, and let $S \subset \mathbb{R}^n$ be a convex set such that $g(\omega) \in S$ for m-almost all $\omega \in \Omega$. Then $m(g) \in S$.*

This essentially states that the centre of mass of a convex object, if it exists, lies within the convex object.

With this, we will prove the following corollary of the ergodic decomposition theorem:

Corollary 109. *Suppose ρ is stationary. Let $g : X \to \mathbb{R}^n$ be a ρ-integrable function, and let $S \subset \mathbb{R}^n$ be a convex set such that $\rho(g) \notin S$. Then there exists an ergodic probability measure ρ' such that g is ρ'-integrable and $\rho'(g) \notin S$.*

To illustrate this result: Assume ρ is stationary, and suppose we have a set $A \in \Sigma$ and a ρ-integrable function $h : X \to \mathbb{R}$ such that $\rho(A) = 1$ and $\rho(h) > 0$. Then it is guaranteed that there exists an ergodic probability measure ρ' with $\rho'(A) = 1$ and $\rho'(h) > 0$; to see this, just apply Corollary 109 with $n = 2$, $g(x) = (\mathbb{1}_A(x), h(x))$ and $S = [0,1) \times (-\infty, 0]$.

We now start to prove Proposition 108. Let us use $\cdot$ for the dot-product on $\mathbb{R}^n$, let $|\cdot|$ denote the corresponding norm (the Euclidean norm), and let $d(\cdot,\cdot)$ denote the corresponding metric (the Euclidean metric).

Lemma 110. *Let $S \subset \mathbb{R}^n$ be a convex set, with S° denoting the interior of S (relative to $\mathbb{R}^n$). For any $\mathbf{a} \in \mathbb{R}^n \setminus S^\circ$ there exists $\mathbf{n} \in \mathbb{R}^n$ with $|\mathbf{n}| = 1$ such that for all $\mathbf{x} \in S$, $\mathbf{n}\cdot(\mathbf{x}-\mathbf{a}) \geq d(\mathbf{a}, S)$.*

The following exercise takes the reader through the proof:

Exercise 111 (adapted from section 7.3 of here)**.** (A) Show that for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ with $\mathbf{a}\cdot\mathbf{b} > 0$ there exists $c \in (0,1)$ such that for all $\lambda \in (0,c)$, $|\mathbf{a} - \lambda\mathbf{b}| < |\mathbf{a}|$. (B) Show that for any closed $G \subset \mathbb{R}^n$ and any $\mathbf{a} \in \mathbb{R}^n \setminus G$ there exists $\mathbf{x} \in G$ such that $d(\mathbf{a}, \mathbf{x}) = d(\mathbf{a}, G)$. (C) Prove Lemma 110 in the case that $\mathbf{a} \notin \partial S$. (D) Prove Lemma 110 in the case that $\mathbf{a} \in \partial S$ by considering a sequence $(\mathbf{a}_r)_{r\in\mathbb{N}}$ in $\mathbb{R}^n \setminus \bar{S}$ converging to $\mathbf{a}$.

Proof of Proposition 108. Let $C \subset \mathbb{R}^n$ be the support of g_*m. Since adjusting g on a null set does not affect either its integral or the support of g_*m, we may assume without loss of generality that $g(\omega) \in S \cap C$ for all $\omega \in \Omega$. Now suppose for a contradiction that $m(g) \notin S$. Let $S' \subset \mathbb{R}^n$ and $L \subset \mathbb{R}^n$ denote respectively the convex hull of $S \cap C$ and the affine subspace of $\mathbb{R}^n$ generated by $S \cap C$. (Note that $S' \subset L \cap S$.) Let r be the dimension of L, let $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ be an isometry such that $\varphi(L) = \mathbb{R}^r \times \{\mathbf{0}\}$ (where $\mathbf{0}$ is the zero-vector in $\mathbb{R}^{n-r}$), and define the isometry $\tilde{\varphi} : L \to \mathbb{R}^r$ by $\varphi(\mathbf{x}) = (\tilde{\varphi}(\mathbf{x}), \mathbf{0})$. Since $m(g) \notin S$, we must have that $r \geq 1$ (i.e. g does not map Ω into a single point). Now $\varphi(g(\omega)) \in \mathbb{R}^r \times \{\mathbf{0}\}$ for all $\omega \in \Omega$, so $m(\varphi \circ g) \in \mathbb{R}^r \times \{\mathbf{0}\}$. Since φ is an affine transformation, $m(\varphi \circ g) = \varphi(m(g))$, and so we can write $\varphi(m(g)) = (\mathbf{a}, \mathbf{0})$ with $\mathbf{a} = \tilde{\varphi}(m(g))$. Since $m(g) \notin S$, we have in particular that $m(g) \notin S'$, so $\mathbf{a} \notin \tilde{\varphi}(S')$. Hence, using Lemma 110, there exists $\mathbf{n} \in \mathbb{R}^r$ such that $\mathbf{n} \cdot (\mathbf{x} - \mathbf{a}) \geq 0$ for all $\mathbf{x} \in \tilde{\varphi}(S')$. So

$$\mathbf{n} \cdot (\tilde{\varphi}(g(\omega)) - \mathbf{a}) \;\geq\; 0$$

for all $\omega \in \Omega$; but we also have that

$$\int_\Omega \mathbf{n} \cdot (\tilde{\varphi}(g(\omega)) - \mathbf{a})\, m(d\omega) \;=\; \mathbf{n} \cdot \left(\underbrace{\int_\Omega \tilde{\varphi}(g(\omega))\, m(d\omega)}_{=\mathbf{a}} - \mathbf{a} \right) \;=\; 0.$$

It follows that $\mathbf{n} \cdot (\tilde{\varphi}(g(\omega)) - \mathbf{a}) = 0$ for m-almost all $\omega \in \Omega$. Thus the support of $(\tilde{\varphi} \circ g)_* m$ is contained in the $(r-1)$-dimensional affine space $\{\mathbf{y} : \mathbf{n} \cdot (\mathbf{y} - \mathbf{a}) = 0\}$. But the support of $(\tilde{\varphi} \circ g)_* m$ is precisely $\tilde{\varphi}(C)$, so C is contained in an $(r-1)$-dimensional affine space. However, this contradicts the fact that the smallest affine space L containing $S \cap C$ is r-dimensional. $\square$

Proof of Corollary 109. Let Q be as in Corollary 100. By Theorem 12,

$$\rho(g) \;=\; \int_{\mathcal{M}_e} \mu(g)\, Q(d\mu).$$

By Proposition 108 (with $m := Q$ and with g replaced by the map $\mu \mapsto \mu(g)$), it is *not* the case that $\mu(g) \in S$ for Q-almost all $\mu \in \mathcal{M}_e$. Nonetheless it is clear that g is μ-integrable for Q-almost all $\mu \in \mathcal{M}_e$. So there must exist $\mu \in \mathcal{M}_e$ such that g is μ-integrable and $\mu(g) \notin S$. $\square$

6 Feller-continuity

Fix a separable metrisable topology on X, with $\Sigma = \mathcal{B}(X)$. We regard $\mathcal{M}_1$ as being equipped with the narrow topology. We will say that a Markov kernel (μ_x) on X is *Feller-continuous* if the map $x \mapsto \mu_x$ is continuous. Note that if (μ_x) is Feller-continuous then for any open $U \subset X$ the set $\{x \in X : \mu_x(U) > 0\}$ is open. (Some of the important facts about Feller-continuous kernels do not actually rely on the full strength of Feller-continuity, but only on this fact.)

Exercise 112. (A) Show that a kernel (μ_x) on X is Feller-continuous if and only if the map $\rho \mapsto \mu^*\rho$ from $\mathcal{M}_1$ to $\mathcal{M}_1$ is continuous. (B) Show that for any function $f : X \to X$,

$(\delta_{f(x)})_{x\in X}$ is a Feller-continuous Markov kernel if and only if f is continuous. (C) Show that if (μ_x) and (ν_x) are Feller-continuous kernels on X then their composition $(\nu^*\mu_x)$ is Feller-continuous.

We will say that a semigroup of Markov kernels (μ_x^t) on X is Feller-continuous[14] if the kernel $(\mu_x^t)_{x\in X}$ is Feller-continuous for all $t\in\mathbb{T}^+$. By Exercise 112(C), if $\mathbb{T}^+=\mathbb{N}\cup\{0\}$ then Feller-continuity is the same for the semigroup (μ_x^n) as for the kernel (μ_x^1).

Exercise 113. Let (μ_x^t) be a Feller-continuous semigroup. (A) Show that if $A\in\Sigma$ is a forward-invariant set then the closure $\bar{A}$ is also forward-invariant. (B) Show that if ρ is a stationary probability measure then $\operatorname{supp}\rho$ is forward-invariant.

The following theorem can be regarded as a partial converse to Exercise 113(B):

Theorem 114 (Krylov-Bogolyubov)**.** *Let (μ_x^t) be a semigroup of kernels on X that is both measurable and Feller-continuous, and let $K\subset X$ be a non-empty forward-invariant compact set. Then there exists at least one stationary (and hence at least one ergodic) probability measure ρ with $\rho(K)=1$.*

We start with the following exercise:

Exercise 115. For each $x\in X$ and $\tau\in\mathbb{T}^+\setminus\{0\}$, we define the probability measure $\bar{\mu}_x^t$ on X by

$$\bar{\mu}_x^\tau(A) = \frac{1}{\tau}\int_{\mathbb{T}_{[0,\tau)}}\mu_x^s(A)\,\lambda(ds)$$

for all $A\in\mathcal{B}(X)$. Show that for any bounded measurable $g:X\to\mathbb{R}$, any $x\in X$ and any $t,\tau\in\mathbb{T}^+$ with $\tau\neq 0$,

$$|\mu^{t*}\bar{\mu}_x^\tau(g)-\bar{\mu}_x^\tau(g)| \le \frac{2t\sup_{y\in X}|g(y)|}{\tau}.$$

Proof of Theorem 114. Pick any $x\in K$. It is clear that $\bar{\mu}_x^t(K)=1$ for all $t\in\mathbb{T}^+\setminus\{0\}$. Hence by Corollary 21, there exists an unbounded increasing sequence $(t_n)_{n\in\mathbb{N}}$ in $\mathbb{T}^+\setminus\{0\}$ such that $\bar{\mu}_x^{t_n}$ converges in the narrow topology (as $n\to\infty$) to a probability measure ρ with $\rho(K)=1$. We now show that ρ is stationary. Fix any $t\in\mathbb{T}^+$. By Exercise 112(A), since the kernel $(\mu_x^t)_{x\in X}$ is Feller-continuous, we have that $\mu^{t*}\bar{\mu}_x^{t_n}\to\mu^{t*}\rho$ as $n\to\infty$. But also, for any bounded continuous $g:X\to\mathbb{R}$ we have that $\mu^{t*}\bar{\mu}_x^{t_n}(g)-\bar{\mu}_x^{t_n}(g)\to 0$ as $n\to\infty$, by Exercise 115. Since $\bar{\mu}_x^{t_n}(g)\to\rho(g)$ as $n\to\infty$, it follows that $\mu^{t*}\bar{\mu}_x^{t_n}(g)\to\rho(g)$ as $n\to\infty$. Hence $\mu^{t*}\bar{\mu}_x^{t_n}\to\rho$ as $n\to\infty$. So then, $\mu^{t*}\rho=\rho$.

Thus we have proved that there exists at least one stationary probability measure ρ. Now it is easy to check that $\big(\mu_x^t|_{\mathcal{B}(K)}\big)_{x\in X,\,t\in\mathbb{T}^+}$ is a measurable semigroup of Markov kernels on K, with $\rho|_{\mathcal{B}(K)}$ being a stationary probability measure. Since K is compact, $(K,\mathcal{B}(K))$ is standard, and therefore the semigroup $\big(\mu_x^t|_{\mathcal{B}(K)}\big)$ must admit at least one ergodic probability measure r. It is then clear that the probability measure $A\mapsto r(A\cap K)$ on X is ergodic with respect to (μ_x^t). □

[14]We warn the reader that a "Feller semigroup" is a related but different concept from a Feller-continuous semigroup of kernels. (We will not discuss Feller semigroups here.)

Exercise 116. Given any semigroup of kernels (μ_x^t) on X, show that the set of open backward-invariant sets forms a topology on X. (Hint: note that for any second-countable topological space, the union of an arbitrary collection of open sets can be covered by a finite or countable subcollection thereof.)

Exercise 117. Let (μ_x^t) be a semigroup of kernels. We will say that an open set $U \subset X$ is *accessible* from a point $x \in X$ if there exists $t \in \mathbb{T}^+$ such that $\mu_x^t(U) > 0$. (A) Show that for any $x \in X$ there exists a *smallest* closed forward-invariant set containing x (which we shall denote G_x). (B) Show that for any $x \in X$ there is a *largest* open set that is not accessible from x, and that the complement C_x of this set is given by

$$C_x = \overline{\bigcup_{t \in \mathbb{T}^+} \operatorname{supp} \mu_x^t}.$$

(C) Show that for any $x \in X$, $C_x \subset G_x$, with equality in the case that (μ_x^t) is Feller-continuous.

Now we will say that a Markov kernel (μ_x) is *strong-Feller-continuous* if for every bounded measurable $g : X \to \mathbb{R}$ the map $x \mapsto \mu_x(g)$ is continuous. Note that this is indeed stronger than Feller-continuity.

Before continuing our discussion of strong-Feller-continuity, we introduce the following notions: Given an index set I and a family $(C_\alpha)_{\alpha \in I}$ of subsets of X, we will say that $(C_\alpha)_{\alpha \in I}$ is *disjoint* if $C_{\alpha_1} \cap C_{\alpha_2} = \varnothing$ for all distinct $\alpha_1, \alpha_2 \in I$, and we will say that $(C_\alpha)_{\alpha \in I}$ is *well-separated* if the following hold:

(i) for each $\alpha \in I$ there exists a neighbourhood U of $\overline{C_\alpha}$ such that

$$U \cap \bigcup_{\beta \in I \setminus \{\alpha\}} C_\beta = \varnothing;$$

(ii) $\bigcup_{\alpha \in I} \overline{C_\alpha}$ is closed.

Obviously, condition (i) is stronger than disjointness.

Exercise 118. Recall that we are assuming X to be a separable metric space. Show that if $(C_\alpha)_{\alpha \in I}$ is well-separated then I is at most countable, and is finite in the case that X is compact.

Proposition 119. *Let (μ_x) be a strong-Feller-continuous kernel. Then any disjoint family $(C_\alpha)_{\alpha \in I}$ of forward-invariant sets is well-separated.*

Corollary 120. *Let (μ_x) be a strong-Feller-continuous kernel. Given a mutually singular collection $\mathcal{S}$ of stationary probability measures, the family $(\operatorname{supp} \rho)_{\rho \in \mathcal{S}}$ is well-separated.*

This implies in particular that (μ_x) admits at most countably many ergodic probability measures, and that if X is compact then (μ_x) admits at most finitely many ergodic probability measures. (The same obviously also holds for any semigroup of kernels (μ_x^t) for which there exists $t \in \mathbb{T}^+$ such that the kernel $(\mu_x^t)_{x \in X}$ is strong-Feller-continuous.)

Proof of Proposition 119. Let $(C_\alpha)_{\alpha\in I}$ be a disjoint family of forward-invariant sets. To show condition (i): Fix any $\alpha \in I$. We know that $\mu_x(C_\alpha) = 1$ for all $x \in C_\alpha$, and hence strong-Feller-continuity gives that there exists a neighbourhood U of $\overline{C_\alpha}$ such that $\mu_x(C_\alpha) > 0$ for all $x \in U$. So then, for any $\beta \in I \setminus \{\alpha\}$, since $C_\beta \cap C_\alpha = \varnothing$, we have that $\mu_x(C_\beta) < 1$ for all $x \in U$; and therefore, since C_β is forward-invariant, $C_\beta \cap U = \varnothing$. This proves condition (i). Now, to show condition (ii): Suppose for a contradiction that $D := \bigcup_{\alpha\in I} \overline{C_\alpha}$ is not closed. Fix any $x \in \bar{D} \setminus D$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in D converging to x. For each $n \in \mathbb{N}$, let $A_n \in \{\overline{C_\alpha} : \alpha \in I\}$ be such that $x_n \in A_n$; let $E := \bigcup_{n=1}^{\infty} A_n$, and let $A_\infty := \bar{E} \setminus E$. Obviously E is forward-invariant, and therefore $\bar{E}$ is forward-invariant; by construction $x \in \bar{E}$, and so $\mu_x(\bar{E}) = 1$. However, for each fixed $n \in \mathbb{N} \cup \{\infty\}$, we have that $\mu_{x_r}(A_n) = 0$ for all r sufficiently large, and therefore by strong-Feller-continuity $\mu_x(A_n) = 0$; so then, since $\bar{E} = \bigcup_{n\in\mathbb{N}\cup\{\infty\}} A_n$, we have that $\mu_x(\bar{E}) = 0$. Thus we have a contradiction, and so condition (ii) holds. □

Proof of Corollary 120. By Proposition 119 and Exercise 113(B), it suffices to show that $(\operatorname{supp}\rho)_{\rho\in\mathcal{S}}$ is a disjoint family; so fix any two distinct $\rho_1, \rho_2 \in \mathcal{S}$. Let $A \in \mathcal{B}(X)$ be a ρ_1-full ρ_2-null set. Using Lemmas 41 and 43, let $C_1 \subset A$ be a forward-invariant ρ_1-full set, and let $C_2 \subset X \setminus A$ be a forward-invariant ρ_2-full set. Obviously $C_1 \cap C_2 = \varnothing$, so by Proposition 119, $\overline{C_1} \cap \overline{C_2} = \varnothing$. But also, $\operatorname{supp}\rho_1 \subset \overline{C_1}$ and $\operatorname{supp}\rho_2 \subset \overline{C_2}$; hence $\operatorname{supp}\rho_1 \cap \operatorname{supp}\rho_2 = \varnothing$. □

Exercise 121. Let (μ_x^t) be a semigroup of kernels, let ρ be a stationary probability measure of (μ_x^t), and suppose there exists $t \in \mathbb{T}^+$ such that the kernel $(\mu_x^t)_{x\in X}$ is strong-Feller-continuous. Show that ρ is ergodic if and only if it assigns trivial measure to every *closed* forward-invariant set.

We finish this section with a useful sufficient criterion for strong-Feller-continuity:[15]

Proposition 122. *Let (μ_x) be a Markov kernel on X. Suppose there exists a σ-finite measure ν on X such that μ_x is absolutely continuous with respect to ν for all x, and suppose there exists an assignment to each $x \in X$ of a version h_x of $\frac{d\mu_x}{d\nu}$ such that the map $x \mapsto h_x(y)$ is continuous for ν-almost all $y \in Y$. Then (μ_x) is strong-Feller-continuous.*

We begin with the following exercise:

Exercise 123 (Scheffé's Lemma)**.** Let $(\Omega, \mathcal{F}, m)$ be a measure space. Suppose we have an m-integrable function $g : \Omega \to [0,\infty)$ and a sequence (g_n) of m-integrable functions $g_n : \Omega \to [0,\infty)$, such that: (i) $g_n(\omega) \to g(\omega)$ as $n \to \infty$ for m-almost all $\omega \in \Omega$; and (ii) $m(g_n) \to m(g)$. Show that g_n converges in $\mathcal{L}^1(m)$ to g as $n \to \infty$.

Proof of Proposition 122. Fix any bounded measurable $g : X \to \mathbb{R}$. Let (x_n) be a sequence in X converging to a point x. For each n,

$$|\mu_{x_n}(g) - \mu_x(g)| \;\le\; \left(\sup_{y\in X} |g(y)|\right) \int_X |h_{x_n}(y) - h_x(y)| \,\nu(dy).$$

Now we know that $h_{x_n}(y) \to h_x(y)$ as $n \to \infty$ for ν-almost all $y \in X$; we also know that for any n, $\nu(h_{x_n}) = \nu(h_x) = 1$. Hence, by Exercise 123, $\int_X |h_{x_n}(y) - h_x(y)| \,\nu(dy) \to 0$ as $n \to \infty$, and therefore $\mu_{x_n}(g) \to \mu_x(g)$ as $n \to \infty$. □

[15]This is a slight generalisation of a remark made on p15 of here.

7 Random maps and random dynamical systems

The approach to "random dynamical systems" (RDS) that we shall work with is essentially based on that presented in here (only, we will work specifically with RDS that are adapted to a given one-parameter filtration on the underlying probability space). In this section, as is typical when studying RDS, we will often omit brackets around the arguments of certain functions in order to avoid an excess of brackets.

We start with the following exercise about independent σ-algebras:

Exercise 124. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}_1$ and $\mathcal{G}_2$ be sub-σ-algebras of $\mathcal{F}$ that are independent under $\mathbb{P}$ (meaning that $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2)$ for all $E_1 \in \mathcal{G}_1$ and $E_2 \in \mathcal{G}_2$). Let $g : \Omega \times \Omega \to \mathbb{R}'$ be a $(\mathcal{G}_1 \otimes \mathcal{G}_2)$-measurable function, and define the function $g_2 : \Omega \to \mathbb{R}'$ by $g_2(\omega) = g(\omega, \omega)$. (A) Show that

$$\mathbb{P}\otimes\mathbb{P}(g) \;=\; \mathbb{P}(g_2).$$

(B) Define the function $h : \Omega \to \mathbb{R}'$ by $h(\omega) = \int_\Omega g(\omega, \tilde{\omega})\,\mathbb{P}(d\tilde{\omega})$. Show that if g is integrable with respect to $\mathbb{P} \otimes \mathbb{P}$ then h is a version of $\mathbb{P}(g_2|\mathcal{G}_1)$.

A *random map* $(I, \mathcal{I}, \nu, (\varphi(\alpha))_{\alpha \in I})$ *on* X consists of a probability space $(I, \mathcal{I}, \nu)$ and an I-indexed family $(\varphi(\alpha))_{\alpha \in I}$ of functions $\varphi(\alpha) : X \to X$ such that the map $(\alpha, x) \mapsto \varphi(\alpha)x$ from $I \times X$ to X is measurable.

We may associate to any random map $(I, \mathcal{I}, \nu, (\varphi(\alpha))_{\alpha \in I})$ a corresponding Markov kernel of "transition probabilities" $(\varphi_x)_{x \in X}$, defined by

$$\varphi_x(A) \;=\; \nu(\alpha \in I : \varphi(\alpha)x \in A)$$

for all $x \in X$ and $A \in \Sigma$. In other words, for each $x \in X$ we define φ_x to be the image measure of ν under the map $\alpha \mapsto \varphi(\alpha)x$.

Remark 125. A natural question to ask is whether for every Markov kernel $(\mu_x)_{x \in X}$ there exists a random map $(I, \mathcal{I}, \nu, (\mu(\alpha))_{\alpha \in I})$ on X such that $(\mu_x)_{x \in X}$ is the Markov kernel associated with $(I, \mathcal{I}, \nu, (\mu(\alpha))_{\alpha \in I})$. It turns out that if (X, Σ) is standard then the answer is yes (Lemma 3.22 of here, or Theorem 1.1.1 of here).

Exercise 126. Show that for any probability measure ρ on X, $\varphi^*\rho$ is given by

$$\varphi^*\rho(A) \;=\; \int_I \varphi(\alpha)_*\rho(A)\,\nu(d\alpha).$$

Hence show that for any measurable $g : X \to \mathbb{R}'$, if $\varphi^*\rho(g) \neq \mathrm{NaN}$ then

$$\varphi^*\rho(g) \;=\; \int_I \rho(g \circ \varphi(\alpha))\,\nu(d\alpha).$$

We now relate composition of kernels with composition of independent random maps.

Proposition 127. *Let $(I,\mathcal{I},\nu)$ be a probability space, and let $\mathcal{J}_1$ and $\mathcal{J}_2$ be sub-σ-algebras of $\mathcal{I}$ that are independent under ν. Let $(\varphi(\alpha))_{\alpha\in I}$ be a family of functions from X to X such that the map $(\alpha,x)\mapsto\varphi(\alpha)x$ is $(\mathcal{J}_1\otimes\Sigma)$-measurable, and let $(\psi(\alpha))_{\alpha\in I}$ be a family of functions from X to X such that the map $(\alpha,x)\mapsto\psi(\alpha)x$ is $(\mathcal{J}_2\otimes\Sigma)$-measurable. Let (φ_x) and (ψ_x) be the Markov kernels associated with the random maps $(I,\mathcal{I},\nu,(\varphi(\alpha))_{\alpha\in I})$ and $(I,\mathcal{I},\nu,(\psi(\alpha))_{\alpha\in I})$ respectively. Then $(\psi^*\varphi_x)_{x\in X}$ is the Markov kernel associated to the random map $(I,\mathcal{I},\nu,(\psi(\alpha)\circ\varphi(\alpha))_{\alpha\in I})$.*

Proof. Fix any $x\in X$ and $A\in\Sigma$.

$$\begin{aligned}
\nu(\alpha\in I : \psi(\alpha)\varphi(\alpha)x\in A) &= \int_I \mathbb{1}_A(\psi(\alpha)\varphi(\alpha)x)\,\nu(d\alpha)\\
&= \int_{I\times I}\mathbb{1}_A(\psi(\tilde\alpha)\varphi(\alpha)x)\,\nu\otimes\nu(d(\alpha,\tilde\alpha))\\
&\qquad\text{(by Exercise 124(A))}\\
&= \int_I \nu(\tilde\alpha\in I : \psi(\tilde\alpha)\varphi(\alpha)x\in A)\,\nu(d\alpha)\\
&\qquad\text{(by Corollary 14)}\\
&= \int_I \psi_{\varphi(\alpha)x}(A)\,\nu(d\alpha)\\
&= \int_X \psi_y(A)\,\varphi_x(dy)\\
&= \psi^*\varphi_x(A).
\end{aligned}$$

So we are done. □

Exercise 128. Fix a separable metrisable topology on X, with $\Sigma=\mathcal{B}(X)$. Let $(I,\mathcal{I},\nu,(\varphi(\alpha))_{\alpha\in I})$ be a random map on X such that $\varphi(\alpha)$ is continuous for every $\alpha\in I$. (A) Show that (φ_x) is Feller-continuous. (B) Show that there exists a ν-full set $J\in\mathcal{I}$ such that for every open $U\subset X$,

$$\{x\in X:\varphi_x(U)>0\} = \bigcup_{\alpha\in J}\varphi(\alpha)^{-1}(U).$$

(Hint: let $\{U_n\}_{n\in\mathbb{N}}$ be a countable base; for each n, let $N_n=\{x:\varphi_x(U_n)=0\}$, let S_n be a countable dense subset of N_n, and let $J_n=\{\alpha:\forall\,x\in S_n,\ \varphi(\alpha)x\notin U_n\}$; then take $J=\bigcap_{n=1}^\infty J_n$.)

Now just as a Markov kernel can naturally appear as the transition probabilities of a random map, so likewise a semigroup of Markov kernels can naturally appear as the transition probabilities of a "random dynamical system with memoryless noise".

Recall that in essence, a "dynamical system" is a rule specifying the future evolution of the state of a system given its initial state. A "*random* dynamical system" is a rule specifying the future evolution of the state of a system given *both* its initial state *and* the (moment-by-moment) behaviour of some random process which influences the (simultaneous moment-by-moment) evolution of the system. Now just as the term "dynamical system" is often used to refer specifically to an *autonomous* dynamical system (i.e. a dynamical system that is *homogeneous* in time), so likewise the term "random dynamical system" is generally used to refer to a random dynamical system that is "homogeneous in time". Formally this "homogeneity" may be defined by the following properties: (i) the behaviour of the random process (which we often call "noise") is statistically stationary in time; and (ii) the rule itself which determines how the evolution of the system over a time interval $[t_0, t_0 + \Delta t]$ is affected by the behaviour of the random process over the time interval $[t_0, t_0 + \Delta t]$ does not depend on t_0.

Let us first discuss the "noise" (i.e. the random process affecting the system). The "classical" way of describing a random process is essentially as follows: we have a measurable space $(\Omega, \mathcal{F})$ representing all possible information about how the process might behave (and possibly more), together with a probability distribution $\mathbb{P}$ for this; we have a "state space" M; and we have, associated to each instant in time t, a measurable function $X_t : \Omega \to M$ which represents "projecting" from the space of "all information about the process" to a specific piece of information about the "state" of the process *at* the instant in time t.

However, not all mathematical models of random processes lend themselves to this description. More specifically, we may have a model of noise which describes the "behaviour of the noise over a time interval $(t, t+h)$" (for any $h > 0$), but does not describe the "*state*" of the noise *at* any instant in time t. Accordingly, we will take the following more general approach to describing noise:

Definition 129. A *noise space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}^+}, \mathbb{P}, (\theta^t)_{t\in\mathbb{T}^+})$ consists of a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}^+}, \mathbb{P})$ and an autonomous dynamical system $(\theta^t)_{t\in\mathbb{T}^+}$ on $(\Omega, \mathcal{F})$ such that $\theta^{-t}\mathcal{F}_s \subset \mathcal{F}_{s+t}$ for all $s, t \in \mathbb{T}^+$. We say that the noise space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}^+}, \mathbb{P}, (\theta^t)_{t\in\mathbb{T}^+})$ is *stationary* if $\mathbb{P}$ is invariant with respect to (θ^t); we say that it is *memoryless* if for any $s, t \in \mathbb{T}^+$, $\mathcal{F}_s$ and $\theta^{-s}\mathcal{F}_t$ are independent under $\mathbb{P}$.

$(\Omega, \mathcal{F})$ represents the set all possibilities for how the noise will behave over the whole timeline. (This "timeline" will generally be either $\mathbb{T}^+$ or $\mathbb{T}$, where $\mathbb{T} = \mathbb{T}^+ \cup \{-t : t \in \mathbb{T}^+\}$; the latter case is achieved when θ^t is measurably invertible for all $t \in \mathbb{T}^+$). $\mathbb{P}$ is the probability distribution for how the noise will behave over the whole timeline. $\mathcal{F}_t$ represents all information regarding how the noise will behave over the time interval from 0 to t. (θ^t) is the *time-shift dynamical system*: for any $\tau \in \mathbb{T}^+$, θ^τ represents shifting the "reference time $t = 0$" backward by τ. (So $\theta^{-t}\mathcal{F}_s$ represents all information regarding how the noise will behave over the time interval from t to $s + t$.)

Exercise 130. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}^+})$ be a filtered measurable space, and let $(\theta^t)_{t\in\mathbb{T}^+}$ be an autonomous dynamical system on $(\Omega, \mathcal{F})$ such that $\theta^{-t}\mathcal{F}_s \subset \mathcal{F}_{s+t}$ for all $s, t \in \mathbb{T}^+$. Let

$$\mathcal{F}_\infty := \sigma\left(\bigcup_{t\in\mathbb{T}^+} \mathcal{F}_t\right),$$

and for all $s, t \in \mathbb{T}^+$ with $s \le t$, let $\mathcal{F}_s^t := \theta^{-s}\mathcal{F}_{t-s}$. (A) Show that for any $s \in \mathbb{T}^+$,

$$\theta^{-s}\mathcal{F}_\infty = \sigma\left(\bigcup_{t\in\mathbb{T}_{[s,\infty)}} \mathcal{F}_s^t\right).$$

Accordingly, we will write $\mathcal{F}_s^\infty := \theta^{-s}\mathcal{F}_\infty$. We will also write $\mathcal{F}_\infty^\infty := \bigcap_{s\in\mathbb{T}^+} \mathcal{F}_s^\infty$. (B) Show that for any $r, s, t, u \in \mathbb{T}^+ \cup \{\infty\}$ with $r \le s \le t \le u$, $\mathcal{F}_s^t \subset \mathcal{F}_r^u$.

Exercise 131. Let $\boldsymbol{\theta} = (\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}, (\theta^t))$ be a noise space. (A) Show that $\boldsymbol{\theta}$ is memoryless if and only if for every $s \in \mathbb{T}^+$, $\mathcal{F}_s$ and $\theta^{-s}\mathcal{F}_\infty$ are independent under $\mathbb{P}$. (B) Suppose that $\boldsymbol{\theta}$ is stationary, and that for all $t \in \mathbb{T}^+$, θ^t is bijective and its inverse $\theta^{-t} : \Omega \to \Omega$ is measurable. Show that $\boldsymbol{\theta}$ is memoryless if and only if $\mathcal{F}_\infty$ (the "future") and $\sigma\left(\bigcup_{t \in \mathbb{T}^+} \theta^t \mathcal{F}_t\right)$ (the "past") are independent under $\mathbb{P}$.

The following is a version of *Kolmogorov's 0-1 law*:

Proposition 132. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}, (\theta^t))$ be a memoryless noise space, and let $\mathcal{F}_\infty^\infty$ be as in Exercise 130. Then $\mathbb{P}(A) \in \{0,1\}$ for all $A \in \mathcal{F}_\infty^\infty$.*

Proof. It is clear that $\mathcal{F}_\infty^\infty$ is independent of $\mathcal{F}_s$ under $\mathbb{P}$ for all $s \in \mathbb{T}^+$. Obviously $\bigcup_{s \in \mathbb{T}^+} \mathcal{F}_s$ is a π-system, and so by Exercise 82(A) and the π-λ theorem, $\mathcal{F}_\infty^\infty$ is independent of the whole of $\mathcal{F}_\infty$ under $\mathbb{P}$. But $\mathcal{F}_\infty^\infty$ is itself contained in $\mathcal{F}_\infty$, and so $\mathcal{F}_\infty^\infty$ is independent of itself under $\mathbb{P}$. $\square$

Note that (θ^t) may be regarded not only as an autonomous dynamical system on $(\Omega, \mathcal{F})$, but also as an autonomous dynamical system on the "restricted" space $(\Omega, \mathcal{F}_\infty)$. (In particular, for each $t \in \mathbb{T}^+$, θ^t may be viewed as a measurable self-map of the "restricted" space $(\Omega, \mathcal{F}_\infty)$.) With this in mind, we state the following corollary of Proposition 132.

Corollary 133. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}, (\theta^t))$ be a memoryless stationary noise space, and let $\mathcal{F}_\infty$ be as in Exercise 130. Then $\mathbb{P}|_{\mathcal{F}_\infty}$ is ergodic with respect to θ^t for every $t \in \mathbb{T}^+ \setminus \{0\}$ (and is obviously therefore ergodic with respect to $(\theta^t)_{t \in \mathbb{T}^+}$.)*

Proof. Fix $t \in \mathbb{T}^+ \setminus \{0\}$, and let $A \in \mathcal{F}_\infty$ be a set that is strictly invariant under θ^t. Then $A \in \mathcal{F}_{nt}^\infty$ for all $n \in \mathbb{N}$, and therefore $A \in \mathcal{F}_\infty^\infty$. Hence $\mathbb{P}(A) \in \{0,1\}$ by Proposition 132, and so Remark 40 yields that $\mathbb{P}|_{\mathcal{F}_\infty}$ is ergodic with respect to θ^t. $\square$

Exercise 134. On the basis of Corollary 133, derive the strong law of large numbers as a special case of the pointwise ergodic theorem for discrete-time autonomous dynamical systems.

Exercise 135. (A) Let $(I, \mathcal{I}, \nu)$ be a probability space. Let $\Omega := I^{\mathbb{N}}$. For each $n \in \mathbb{N} \cup \{0\}$, let $\mathcal{F}_n$ be the σ-algebra on Ω given by

$$\mathcal{F}_n \ := \ \sigma(\, (\alpha_r)_{r \in \mathbb{N}} \mapsto \alpha_m \, : \, m \in \{1, \ldots, n\} \,).$$

(So $\mathcal{F}_0$ is the σ-algebra generated by the empty set, namely the trivial σ-algebra $\{\Omega, \varnothing\}$.) Let $\mathcal{F} := \mathcal{I}^{\otimes \mathbb{N}} = \sigma\left(\bigcup_{n=0}^\infty \mathcal{F}_n\right)$. Let $\mathbb{P} := \nu^{\otimes \mathbb{N}}$. Define the function $\theta : \Omega \to \Omega$ by $\theta((\alpha_n)_{n \in \mathbb{N}}) = (\alpha_{n+1})_{n \in \mathbb{N}}$. Taking $\mathbb{T}^+ = \mathbb{N} \cup \{0\}$, show that $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P}, (\theta^n))$ is a memoryless stationary noise space. (B) Returning to the appendix of Section 4: For any $S \subset [0, \infty)$, let $D\mathcal{Y}_S$ be the set of $\mathcal{Y}_S$-measurable subsets of DY. Let (ν^t) be a consistent family of probability measures on X such that $\boldsymbol{\nu}^Y$ exists. Show that $(DY, D\mathcal{Y}, (D\mathcal{Y}_{[0,t]}), \boldsymbol{\nu}^Y, (D\theta^t))$ is a memoryless stationary noise space. (In the case that $\boldsymbol{\nu}^Y$ is the law of a Wiener process, this noise space represents a Gaussian white noise process.)

Remark 136. Exercise 135(A) does also work in continuous time: if we let $\Omega = I^{(0,\infty)}$, $\mathcal{F} = \mathcal{I}^{\otimes(0,\infty)}$, $\mathbb{P} = \nu^{\otimes(0,\infty)}$, $\mathcal{F}_t = \sigma((\alpha_r)_{r>0} \mapsto \alpha_s : 0 < s \leq t)$, and let (θ^t) be the horizontal shift dynamical system on $\mathcal{I}^{\otimes(0,\infty)}$, then $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}, (\theta^t))$ is a memoryless stationary noise space. However, such a model of continuous-time noise will rarely if ever be useful in practice.

We now go on to define a (time-homogeneous) random dynamical system:

Definition 137. Let $\boldsymbol{\theta} = (\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}, (\theta^t))$ be a stationary noise space. A *random dynamical system (RDS) φ on X over $\boldsymbol{\theta}$* is a $(\mathbb{T}^+ \times \Omega)$-indexed family $\varphi = (\varphi(t,\omega))_{t \in \mathbb{T}^+, \omega \in \Omega}$ of functions $\varphi(t,\omega) : X \to X$ such that:

(i) $\varphi(0,\omega) = \mathrm{id}_X$ for all $\omega \in \Omega$;

(ii) $\varphi(s+t,\omega) = \varphi(t,\theta^s\omega) \circ \varphi(s,\omega)$ for all $s,t \in \mathbb{T}^+$ and $\omega \in \Omega$;

(iii) for each $t \in \mathbb{T}^+$, the map $(\omega, x) \mapsto \varphi(t,\omega)x$ from $\Omega \times X$ to X is $(\mathcal{F}_t \otimes \Sigma)$-measurable.

X is the state space of our system. $\varphi(t,\omega)x$ represents what the state of the system will be at time t, if it is x at time 0 and the realised behaviour of the noise is ω. (In particular, this obviously justifies property (i).) Property (iii) tells us that $\varphi(t,\omega)x$ will only be affected by how the noise behaves on the time-interval from 0 to t. Property (ii) represents the fact that the manner by which the noise affects the evolution of the system is homogeneous in time.

Exercise 138. Show that for any $s,t \in \mathbb{T}^+$ the map $(\omega, x) \mapsto \varphi(t,\theta^s\omega)x$ is $(\mathcal{F}_s^{s+t} \otimes \Sigma)$-measurable (where $\mathcal{F}_s^{s+t}$ is as in Exercise 130).

Exercise 139. Can you see how a random map on X naturally generates a discrete-time random dynamical system?

Let φ be a RDS over a stationary noise space $\boldsymbol{\theta}$. If $\boldsymbol{\theta}$ is memoryless, then for each $t \in \mathbb{T}^+$ we will write $(\varphi_x^t)_{x \in X}$ to denote the Markov kernel associated to the random map $(\Omega, \mathcal{F}, \mathbb{P}, (\varphi(t,\omega))_{\omega \in \Omega})$. Note that $(\varphi_x^t)_{x \in X}$ is also the Markov kernel associated to the random map $(\Omega, \mathcal{F}, \mathbb{P}, (\varphi(t,\theta^s\omega))_{\omega \in \Omega})$ for any $s \in \mathbb{T}^+$.

Proposition 140. *Suppose $\boldsymbol{\theta}$ is memoryless. Then $(\varphi_x^t)_{x \in X, t \in \mathbb{T}^+}$ is a semigroup of Markov kernels.*

Proof. $(\varphi_x^0)_{x \in X}$ is the Markov kernel associated to the identity map, which is clearly the identity kernel. Now fix any $s,t \in \mathbb{T}^+$. We know that $(\varphi_x^s)_{x \in X}$ is the Markov kernel associated to $(\varphi(s,\omega))_{\omega \in \Omega}$ and that $(\varphi_x^t)_{x \in X}$ is the Markov kernel associated to $(\varphi(t,\theta^s\omega))_{\omega \in \Omega}$. Memorylessness implies that $\mathcal{F}_0^s$ and $\mathcal{F}_s^{s+t}$ are independent under $\mathbb{P}$. Hence by Proposition 127 (together with Exercise 138), we have that $(\varphi^{t*}\varphi_x^s)_{x \in X}$ is the Markov kernel associated to $(\varphi(s+t,\omega))_{\omega \in \Omega}$. So we are done. $\square$

Exercise 141 ("Memoryless noise implies Markovian trajectories")**.** Suppose $\boldsymbol{\theta}$ is memoryless, and let $Y = X^{\mathbb{T}^+}$. Show that for any probability measure ρ on X, $\boldsymbol{\varphi}_\rho^Y$ exists and is equal to the image measure of $\mathbb{P} \otimes \rho$ under the map $(\omega, x) \mapsto (\varphi(t,\omega)x)_{t \in \mathbb{T}^+}$. (Here, following the notation scheme in Section 4, $\boldsymbol{\varphi}_\rho^Y$ denotes the Markov measure associated to the semigroup (φ_x^t) and initial distribution ρ.)

Now, for each $t \in \mathbb{T}^+$, define the map $\Theta^t : \Omega \times X \to \Omega \times X$ by

$$\Theta^t(\omega, x) = (\theta^t\omega, \varphi(t,\omega)x).$$

Exercise 142. Show that $(\Theta^t)_{t\in\mathbb{T}^+}$ is an autonomous dynamical system on $(\Omega\times X, \mathcal{F}\otimes\Sigma)$, and is also an autonomous dynamical system on the "restricted space" $(\Omega \times X, \mathcal{F}_\infty \otimes \Sigma)$.

Theorem 143 (based on Kifer, Lemma 1.2.3 and Theorem 1.2.1)**.** *Suppose $\boldsymbol{\theta}$ is memoryless, and let ρ be a probability measure on X. Then (i) ρ is stationary with respect to (φ_x^t) if and only if $\mathbb{P}|_{\mathcal{F}_\infty} \otimes \rho$ is invariant with respect to (Θ^t), and (ii) ρ is ergodic with respect to (φ_x^t) if and only if $\mathbb{P}|_{\mathcal{F}_\infty} \otimes \rho$ is ergodic with respect to (Θ^t).*

Proof. (i) We will show that for any $t \in \mathbb{T}^+$, $\mathbb{P}|_{\mathcal{F}_\infty} \otimes \varphi^{t*}\rho = \Theta_*^t(\mathbb{P}|_{\mathcal{F}_\infty} \otimes \rho)$. (The result is then immediate). By the π-λ theorem, it suffices to consider sets of the form $E \times A$ for $E \in \mathcal{F}_\infty$ and $A \in \Sigma$.

$$\begin{aligned}
\mathbb{P} \otimes \varphi^{t*}\rho(E \times A) &= \mathbb{P}(E)\,\varphi^{t*}\rho(A) \\
&= \mathbb{P}(\theta^{-t}(E)) \int_X \varphi_x^t(A)\,\rho(dx) \\
&= \left(\int_\Omega \mathbb{1}_{\theta^{-t}(E)}(\tilde\omega)\,\mathbb{P}(d\tilde\omega)\right)\left(\int_X \int_\Omega \mathbb{1}_A(\varphi(t,\omega)x)\,\mathbb{P}(d\omega)\,\rho(dx)\right) \\
&= \int_X \left(\int_\Omega \int_\Omega \mathbb{1}_{\theta^{-t}(E)}(\tilde\omega)\mathbb{1}_A(\varphi(t,\omega)x)\,\mathbb{P}(d\tilde\omega)\,\mathbb{P}(d\omega)\right)\rho(dx) \\
&= \int_X \left(\int_\Omega \mathbb{1}_{\theta^{-t}(E)}(\omega)\mathbb{1}_A(\varphi(t,\omega)x)\,\mathbb{P}(d\omega)\right)\rho(dx) \quad \text{(Exercise 124(A))} \\
&= \int_X \int_\Omega \mathbb{1}_{E\times A}(\Theta^t(\omega,x))\,\mathbb{P}(d\omega)\,\rho(dx) \\
&= \mathbb{P} \otimes \rho(\Theta^{-t}(E \times A)).
\end{aligned}$$

(ii) Obviously, the map $p \mapsto (\mathbb{P}|_{\mathcal{F}_\infty}) \otimes p$ from $\mathcal{M}_1$ to the set of probability measures on $(\Omega\times X, \mathcal{F}_\infty\otimes\Sigma)$ is linear (i.e. respects convex combinations); hence, by characterisation (iv) of ergodicity in Theorem 34, if $(\mathbb{P}|_{\mathcal{F}_\infty})\otimes\rho$ is ergodic then ρ is ergodic. Conversely, suppose that ρ is ergodic, and take any $A \in \mathcal{F}_\infty \otimes \Sigma$ that is $(\mathbb{P} \otimes \rho)$-almost invariant under (Θ^t). For each $x \in X$, let $A_x := \{\omega \in \Omega : (\omega, x) \in A\} \in \mathcal{F}_\infty$, and let $g(x) := \mathbb{P}(A_x) = \mathbb{P}(\theta^{-t}(A_x))$ (for any t). Fix any $t \in \mathbb{T}^+$; for any $x \in X$ we have that,

$$\begin{aligned}
\varphi_x^t(g) &= \int_\Omega g(\varphi(t,\omega)x)\,\mathbb{P}(d\omega) \\
&= \int_\Omega \mathbb{P}(\theta^{-t}(A_{\varphi(t,\omega)x}))\,\mathbb{P}(d\omega) \\
&= \int_\Omega \int_\Omega \mathbb{1}_A(\theta^t\tilde\omega, \varphi(t,\omega)x)\,\mathbb{P}(d\tilde\omega)\,\mathbb{P}(d\omega) \\
&= \int_\Omega \mathbb{1}_A(\theta^t\omega, \varphi(t,\omega)x)\;\mathbb{P}(d\omega) \quad \text{(by Exercise 124(A))}.
\end{aligned}$$

But since A is $(\mathbb{P} \otimes \rho)$-almost invariant, we know that ρ-almost every $x \in X$ has the property that for $\mathbb{P}$-almost all $\omega \in \Omega$, $\mathbb{1}_A(\theta^t\omega, \varphi(t,\omega)x) = \mathbb{1}_A(\omega, x)$; and hence, for ρ-almost every $x \in X$,

$$\varphi_x^t(g) = \int_\Omega \mathbb{1}_A(\omega, x)\,\mathbb{P}(d\omega) = g(x).$$

This was true for any $t \in \mathbb{T}^+$. So (by Theorem 27(B)), g is ρ-almost invariant. Hence, since ρ is ergodic, it follows that there exist $c \in \mathbb{R}$ and a ρ-full set $\tilde{X} \in \Sigma$ such that $g(x) = c$ for all $x \in \tilde{X}$. Now $\mathbb{P} \otimes \rho(A) = \int_X g(x)\, \rho(dx) = c$. So it remains to show that c is equal to either 0 or 1.

Let $X' = \{\, x \in X \,:\, \varphi_x^n(\tilde{X}) = 1 \text{ for all } n \in \mathbb{N} \cup \{0\} \,\} \subset \tilde{X}$. It is easy to show (e.g. as a special case of the discrete-time version of Lemma 43) that $\rho(X') = 1$; so, in particular, since X' is not ρ-null, there must exist $x \in X'$ such that for each $n \in \mathbb{N}$, for $\mathbb{P}$-almost all $\omega \in \Omega$, $\mathbb{1}_A(\theta^n\omega, \varphi(n,\omega)x) = \mathbb{1}_A(\omega, x)$. Fix such an x. For each $n \in \mathbb{N}$, define the function $h_n : \Omega \to [0,1]$ by

$$h_n(\omega) \;=\; \int_\Omega \mathbb{1}_A(\theta^n\tilde{\omega}, \varphi(n,\omega)x)\, \mathbb{P}(d\tilde{\omega}).$$

On the one hand, by Exercise 124(B), h_n is a version of $\mathbb{P}(\omega \mapsto \mathbb{1}_A(\theta^n\omega, \varphi(n,\omega)x)|\mathcal{F}_n)$, which is the same as being a version of $\mathbb{P}(\omega \mapsto \mathbb{1}_A(\omega, x)|\mathcal{F}_n)$, which is the same as being a version of $\mathbb{P}(A_x|\mathcal{F}_n)$. But on the other hand, for $\mathbb{P}$-almost all ω, $\varphi(n,\omega)x \in \tilde{X}$ and so

$$\begin{aligned} h_n(\omega) \;&=\; \int_\Omega \mathbb{1}_A(\tilde{\omega}, \varphi(n,\omega)x)\, \mathbb{P}(d\tilde{\omega}) \\ &\qquad \text{(since } \mathbb{P} \text{ is } \theta^n\text{-invariant)} \\ &=\; g(\varphi(n,\omega)x) \\ &=\; c. \end{aligned}$$

Therefore, for every $n \in \mathbb{N}$ the constant map $\omega \mapsto c = \mathbb{P}(A_x)$ is a version of $\mathbb{P}(A_x|\mathcal{F}_n)$. Since $A_x \in \mathcal{F}_\infty$ it follows that $\mathbb{P}(A_x)$ must be either 0 or 1.[16] So we are done. $\square$

Appendix: Markov operators

Let ρ be a probability measure on (X, Σ). We will say that two measurable functions $f, g : X \to \mathbb{R}'$ are *ρ-equivalent* if $f(x) = g(x)$ for ρ-almost all $x \in X$. For any measurable $f : X \to \mathbb{R}'$, we will write $[f]$ to denote the ρ-equivalence class of f.

Let $L^1(\rho)$ be the set of ρ-equivalences classes of ρ-integrable functions from X to $\mathbb{R}'$. For $f, g \in L^1(\rho)$ and $c \in \mathbb{R}$, we may define $f + g$, cf, f^+, f^-, $|f|$ and $\int_A f\, d\rho$ (where $A \in \Sigma$) in the natural way (by considering representatives of f and g). We also equip $L^1(\rho)$ with its natural partial ordering $\leq$ (where we say that $f \leq g$ if for any representative $\tilde{f}$ of f and any representative $\tilde{g}$ of g, $\tilde{f}(x) \leq \tilde{g}(x)$ for ρ-almost all $x \in X$). Finally, for any $c \in \mathbb{R}$, we will also write c to denote the ρ-equivalence class represented by the constant function $x \mapsto c$.

Note that $L^1(\rho)$ is a vector space; we can define a norm on $L^1(\rho)$ by $\|f\| = \int_X |f|\, d\rho$. It is well-known that this is indeed a norm, and under this norm $L^1(\rho)$ is a Banach space.

There are different definitions for a "Markov operator", but the one that we shall use is

[16] There are a few ways of justifying this; perhaps the most elementary is as follows: By Exercise 82(B), $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n$ is contained in the $\mathbb{P}$-orthogonal complement of $\{A_x\}$; and therefore (by the π-λ theorem and Exercise 82(A)) $\mathcal{F}_\infty$ is contained in the $\mathbb{P}$-orthogonal complement of $\{A_x\}$. Hence, in particular, A_x is in the $\mathbb{P}$-orthogonal complement of $\{A_x\}$.

as follows: A *Markov operator over* (X, Σ, ρ) is a function $P : L^1(\rho) \to L^1(\rho)$ with the following properties:

(i) P is linear;

(ii) $Pf \geq 0$ for all $f \in L^1(\rho)$ with $f \geq 0$;

(iii) $\int_X Pf\, d\rho = \int_X f\, d\rho$ for all $f \in L^1(\rho)$;

(iv) $P1 = 1$ (and hence $Pc = c$ for all $c \in \mathbb{R}$).

Exercise 144. Show that properties (i), (ii) and (iii) imply that P is continuous.

Exercise 145. Show that the set of Markov operators over (X, Σ, ρ) forms a monoid under composition. (So we can define a *Markov semigroup* as a $\mathbb{T}^+$-indexed family of Markov operators $(P^t)_{t \in \mathbb{T}^+}$ such that the map $t \mapsto P^t$ is a monoid homomorphism.)

As an important special case: Let $\theta : X \to X$ be a ρ-preserving measurable function. We may associate to θ the linear operator $P_\theta : L^1(\rho) \to L^1(\rho)$ given by $P_\theta[f] = [f \circ \theta]$ for any ρ-integrable $f : X \to \mathbb{R}'$. It is easy to show that P_θ is indeed a Markov operator. Note that this association $\theta \mapsto P_\theta$ of a Markov operator to a measure-preserving map *reverses* order of composition: $P_{\theta_2 \circ \theta_1} = P_{\theta_1} \circ P_{\theta_2}$.

Now just as a self-map on X can naturally give rise to a Markov operator, so (more generally) a Markov kernel on X can give rise to a Markov operator: Let (μ_x) be a Markov kernel on X with respect to which ρ is stationary. Recall from Exercise 24 that for any two ρ-equivalent functions $f, g : X \to \mathbb{R}'$ the maps $x \mapsto \mu_x(f)$ and $x \mapsto \mu_x(g)$ are ρ-equivalent. Hence we may define the function $P_\mu : L^1(\rho) \to L^1(\rho)$ by $P_\mu([f]) = [\mu.(f)]$ for any ρ-integrable $f : X \to \mathbb{R}'$. Again, it is easy to show that P_μ is a Markov operator. Also note once again that this association $(\mu_x) \mapsto P_\mu$ reverses order of composition: the Markov operator associated to $(\nu^* \mu_x)_{x \in X}$ is equal to $P_\mu \circ P_\nu$.

Definition 146. We will say that two Markov kernels (μ_x) and (ν_x) are ρ-*equivalent* if $\mu_x = \nu_x$ for ρ-almost all $x \in X$. Given a kernel (μ_x), we write $[\mu]$ to denote the ρ-equivalence class of (μ_x).

It is clear that for any two ρ-equivalent kernels (μ_x) and (ν_x), ρ is stationary with respect to (μ_x) if and only if ρ is stationary with respect to (ν_x). So:

Definition 147. For any kernel (μ_x) on X, we will say that $[\mu]$ is *measure-preserving* if ρ is stationary with respect to (μ_x).

It is also clear that for any two ρ-equivalent kernels (μ_x) and (ν_x) with respect to which ρ is stationary, $P_\mu = P_\nu$. Thus, if we let K denote the set of all ρ-equivalences classes of Markov kernels and let O denote the set of Markov operators over (X, Σ, ρ), then we can define a map $F : K \to M$ by $F([\mu]) = P_\mu$.

Exercise 148. Show that if (X, Σ) is countably generated then $F : K \to M$ is injective. (Hint: Show that, in general, if $P_\mu = P_\nu$ then for *each* $A \in \Sigma$, for ρ-almost all $x \in X$, $\mu_x(A) = \nu_x(A)$. Then, as in the proof of Proposition 93(B), use countable generation and the π-λ to give the desired result.)

Now recall once again that a measurable space (X, Σ) is said to be *standard* if there exists a Polish topology on X whose Borel σ-algebra coincides with Σ, and that every uncountable standard measurable space is isomorphic to $([0,1], \mathcal{B}([0,1]))$.

Proposition 149. *If* (X, Σ) *is standard then* $F : K \to M$ *is bijective.*

The proof of surjectivity is essentially identical to the proof of Proposition 93(A), replacing "version of $\rho(A|\mathcal{G})$" with "representative of $P[\mathbb{1}_A]$", and replacing countable additivity of conditional probabilities and the conditional dominated convergence theorem with continuity of Markov operators (Exercise 144). (The full proof is left as an exercise.)

References

[1] Kifer, Y., *Ergodic Theory of Random Transformations*, Birkhäuser, Boston (1986).

[2] Rogers, L. C. G., Williams, D., *Diffusions, Markov processes and Martingales, vol. 1: Foundations*, Wiley, Chichester (1994).

[3] Crauel, H., *Random Probability Measures on Polish Spaces*, Stochastics Monographs 11, Taylor & Francis, London (2002).

[4] Arnold, L., *Random Dynamical Systems*, Springer, Berlin Heidelberg New York (1998).

[5] Kunita, H., *Stochastic Flows and Stochastic Differential Equations*, Cambridge Studies in Advanced Mathematics 24, Cambridge University Press, Cambridge (1990).

[6] Krengel, U., *Ergodic Theorems* (with a supplement by A. Brunel), de Gruyter Studies in Mathematics 6, de Gruyter, Berlin New York (1985).

[7] Fremlin, D. H., *Measure Theory, vol. 4: Topological Measure Spaces* (Chapter 42, section 424, version of 21st March 2008) https://www.essex.ac.uk/maths/people/fremlin/chap42.pdf (2013).

[8] Billingsley, P., *Probability and Measure*, Wiley, New York (1995).

[9] Kallenberg, O., *Foundations of Modern Probability*, Springer, Berlin Heidelberg New York (2002).

[10] Aliprantis, C. D., Border, K. C., *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, Berlin Heidelberg (2006).

[11] Kandiller, L., *Principles of Mathematics in Operations Research*, Springer, Ankara (2007).

[12] Dellacherie, C., Meyer, P. A., *Probabilities and Potential*, North Holland Mathematical Studies 29, North Holland, Amsterdam (1978).

[13] Bhattacharya, R., Waymire, E. C., *A Basic Course in Probability Theory*, Springer, New York (2007).

[14] Sun, R., Lecture notes on Probability Theory (National University of Singapore, 2012), http://www.math.nus.edu.sg/~matsr/MA5260.html

[15] Albarrán, J., A Guide to the Ergodic Decomposition Theorem: Ergodic Measures as Extreme Points (2004), http://www.cimat.mx/~albarran/documentos/ergodicdesc.pdf

[16] Van Gaans, O. W., Probability measures on metric spaces (seminar notes, Delft University of Technology, 2002/03), https://www.math.leidenuniv.nl/~vangaans/jancol1.pdf

[17] Tao, T., The strong law of large numbers (2008) http://terrytao.wordpress.com/2008/06/18/the-strong-law-of-large-numbers/

[18] Lowther, G., Continuous Processes with Independent Increments (2010) https://almostsure.wordpress.com/2010/06/16/continuous-processes-with-independent-increments/

[19] Zimmermann, D., Uniqueness in ergodic decomposition of invariant probabilities, *Illinois Journal of Mathematics* **36**(2), pp325–344 (1992).

[20] Rao, B. V., Srivastava, S. M., An Elementary Proof of the Borel Isomorphism Theorem, *Real Analysis Exchange* **20**(1), pp. 347–349 (1994/95).

[21] Lamb, C. W., A comparison of methods for constructing probability measures on infinite product spaces, *Canadian Mathematical Bulletin* **30**(3), pp282–285 (1987).

[22] Seidenfeld, T., Schervish, M. J., Kadane, J. B., Improper Regular Conditional Distributions, *Annals of Probability* **29**(4), pp1612–1624 (2001). [Corrections to this paper can be found here.]