Probability measures on compact metric spaces

Summary

Recall the role of semi-algebras to simplify working with measures:

- A probability measure μ is uniquely determined by its restriction $\mu|_{\mathscr{S}}$ to a generating semi-algebra.
- Every normalized countably additive function $\mu_0 \colon \mathscr{S} \to [0,1]$ on a generating semi-algebra $\mathscr{S}$ extends to a probability measure.

On a compact metric space $\mathcal{X}$, continuous functions $f \colon \mathcal{X} \to \mathbb{R}$ play a similar role:

- A Borel probability measure μ on $\mathcal{X}$ is uniquely determined by the integrals $\mu(f) = \int f \,\mathrm{d}\mu$ of all $f \in C(\mathcal{X})$.
- Borel probability measures on $\mathcal{X}$ can be identified with normalized positive linear functional $J \colon C(\mathcal{X}) \to \mathbb{R}$ (Riesz representation theorem).
- The functionals $\mu \mapsto \mu(f)$ for $f \in C(\mathcal{X})$ induce a useful topology on the space $\mathscr{P}(\mathcal{X})$ of all Borel probability measures on $\mathcal{X}$.

Here, we provide a short summary, skipping most proofs. The following books are excellent references on the topic:

- R. M. Dudley, *Real Analysis and Probability*
- G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*

Probability measures and observables

Given a metric (or metrizable) space $\mathcal{X}$, we use the following notation:

- $\mathscr{P}(\mathcal{X})$: set of all Borel probability measures on $\mathcal{X}$.
- $C(\mathcal{X})$: set of all continuous functions $f \colon \mathcal{X} \to \mathbb{R}$.

The elements of $C(\mathcal{X})$ are sometimes called *continuous observables*. Together with the uniform norm

$$\|f\| \coloneqq \sup\{|f(x)| : x \in \mathcal{X}\},$$

the set $C(\mathcal{X})$ of all continuous observables forms a Banach space.

Argument. To show that $C(\mathcal{X})$ is complete, we use the fact that $\mathbb{R}$ is complete and a uniform limit of continuous functions is continuous.

The integral $\int f \,\mathrm{d}\mu$ of an observable f with respect to a probability measure μ is also denoted by $\mu(f)$. The integrals $\mu(f)$ for $f \in C(\mathcal{X})$ are enough to distinguish μ:

Proposition 1 (Sufficiency)**.** *Let $\mathcal{X}$ be a metric space and $\mu, \nu \in \mathscr{P}(\mathcal{X})$ two distinct probability measures on $\mathcal{X}$. Then, $\mu(f) \neq \nu(f)$ for some integrable $f \in C(\mathcal{X})$.*

The proof relies on the following:

Notes for Introduction to Ergodic Theory (MATH 307K), Siamak Taati
Last update: November 5, 2026

Fact from measure theory. Every Borel probability measure μ on a metric space $\mathcal{X}$ is regular. This means that, for every measurable $E \subseteq \mathcal{X}$ and every $\varepsilon > 0$, there exists an open set $U_\varepsilon \subseteq \mathcal{X}$ and a closed set $V_\varepsilon \subseteq \mathcal{X}$ such that

$$V_\varepsilon \subseteq E \subseteq U_\varepsilon \qquad \text{and} \qquad \mu(U_\varepsilon \setminus V_\varepsilon) < \varepsilon \,.$$

Idea of the proof: The set of all measurable $E \subseteq \mathcal{X}$ with the above property is a σ-algebra.

Proof of Proposition 1. Let $E \subseteq \mathcal{X}$ be a measurable set on which μ and ν disagree. Without loss of generality, let us assume that $\nu(E) < \mu(E)$. The idea is to find continuous functions that approximate the indicator of E.

Since $\mathcal{X}$ is a metric space, μ and ν are regular. Thus, for each $\varepsilon > 0$, we can find

- an open set $U_\varepsilon \supseteq E$ such that $\nu(U_\varepsilon \setminus E) < \varepsilon$,
- a closed set $V_\varepsilon \subseteq E$ such that $\mu(E \setminus V_\varepsilon) < \varepsilon$.

Let $f_\varepsilon \colon \mathcal{X} \to [0,1]$ be a continuous function such that

$$\begin{aligned} f_\varepsilon(x) &= 1 \qquad \text{for } x \in V_\varepsilon, \\ f_\varepsilon(x) &= 0 \qquad \text{for } x \notin U_\varepsilon. \end{aligned}$$

For instance, we can choose

$$f_\varepsilon(x) \coloneqq \frac{d(x, \mathcal{X} \setminus U_\varepsilon)}{d(x, V_\varepsilon) + d(x, \mathcal{X} \setminus U_\varepsilon)}$$

where d denotes the metric on $\mathcal{X}$. (This is the elementary version of Urysohn's lemma.) Then,

$$\mathbb{1}_{V_\varepsilon} \leq f_\varepsilon \leq \mathbb{1}_{U_\varepsilon} \,,$$

hence

$$\begin{aligned} \mu(f_\varepsilon) &\geq \mu(V_\varepsilon) > \mu(E) - \varepsilon \,, \\ \nu(f_\varepsilon) &\leq \nu(U_\varepsilon) < \nu(E) + \varepsilon \,. \end{aligned}$$

Choosing $\varepsilon \leq \big(\mu(E) - \nu(E)\big)/2$ now ensures that $\nu(f_\varepsilon) < \mu(f_\varepsilon)$. □

Every probability measure $\mu \in \mathscr{P}(\mathcal{X})$ defines a linear functional $f \in C(\mathcal{X}) \mapsto \mu(f) \in \mathbb{R}$.

Definition. Let $J \colon C(\mathcal{X}) \to \mathbb{R}$ be a linear functional. We say that

- J is *positive* if $J(f) \geq 0$ whenever $f \geq 0$.
- J is *normalized* if $J(\mathbb{1}_\mathcal{X}) = 1$.

The functional $f \mapsto \mu(f)$ associated to a probability measure is clearly positive and normalized.

Proposition 2 (Continuity of positive linear functionals)**.** *Let $\mathcal{X}$ be a compact metric space. Every positive linear functional $J \colon C(\mathcal{X}) \to \mathbb{R}$ is continuous.*

Proof. Let $K \coloneqq J(\mathbb{1}_\mathcal{X}) \in \mathbb{R}$. Given $f \in C(\mathcal{X})$, we have

$$-\|f\| \mathbb{1}_\mathcal{X} \leq f \leq \|f\| \mathbb{1}_\mathcal{X}$$

hence

$$-K\|f\| \leq J(f) \leq K\|f\|$$

by the positivity of J. Therefore, $|J(f)| < K\|f\|$, which means J is Lipschitz continuous at 0. Since J if linear, it is continuous everywhere. □

Theorem 3 (Riesz representation theorem)**.** *Let $\mathcal{X}$ be a compact metric space. For every normalized positive linear functional $J \colon C(\mathcal{X}) \to \mathbb{R}$, there exists a (unique) Borel probability measure μ such that $\mu(f) = J(f)$ for every $f \in C(\mathcal{X})$.*

The proof of a more general version can be found in Folland's book (Theorem 7.2). The uniqueness is proven in Proposition 1 above.

A topology on the space of probability measures

Let $\mathcal{X}$ be a compact metric space. The *weak* topology* on $\mathscr{P}(\mathcal{X})$ is the smallest topology that makes all the maps $\mu \mapsto \mu(f)$ (for $f \in C(\mathcal{X})$) continuous. For us, this will be the default topology on $\mathscr{P}(\mathcal{X})$. In probability theory, this topology is also sometimes called the *weak topology* or the *vague topology*.

A base and a sub-base for the weak* topology. A sub-base for this topology is given by the sets of the form

$$\mathscr{U}(f; a; \varepsilon) \coloneqq \big\{\nu : |\nu(f) - a| < \varepsilon\big\}$$

where $f \in C(\mathcal{X})$, $a \in \mathbb{R}$, and $\varepsilon > 0$. A base for this topology is given by the sets of the form

$$\mathscr{U}(f_1, f_2, \ldots, f_n; a_1, a_2, \ldots, a_n; \varepsilon) \coloneqq \bigcap_{k=1}^{n} \mathscr{U}(f_k; a_k; \varepsilon)\ .$$

with $f_1, f_2, \ldots, f_n \in C(\mathcal{X})$, $a_1, a_2, \ldots, a_n \in \mathbb{R}$, and $\varepsilon > 0$.

Convergence in the weak* topology. Let $\mu, \mu_1, \mu_2, \ldots \in \mathscr{P}(\mathcal{X})$. Then,

$$\mu_n \to \mu \qquad \text{if and only if} \qquad \mu_n(f) \to \mu(f) \text{ for every } f \in C(\mathcal{X}),$$

where $\mu_n \to \mu$ denotes convergence in the weak* topology.

Exercise. Suppose $F(\mathcal{X}) \subseteq C(\mathcal{X})$ be a family of observables whose linear span is dense in $C(\mathcal{X})$. Then,

$$\mu_n \to \mu \qquad \text{if and only if} \qquad \mu_n(f) \to \mu(f) \text{ for every } f \in F(\mathcal{X}).$$

The following theorem provides alternative characterizations of weak* convergence.

Theorem 4 (Portmanteau theorem)**.** *Let $\mathcal{X}$ be a compact metric space and $\mu, \mu_1, \mu_2, \ldots \in \mathscr{P}(\mathcal{X})$. The following conditions are equivalent:*

(i) $\mu_n \to \mu$ *in weak* topology.*

(ii) $\liminf_{n\to\infty} \mu_n(B) \geq \mu(B)$ *for every open* $B \subseteq \mathcal{X}$.

(iii) $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$ *for every closed* $C \subseteq \mathcal{X}$.

(iv) $\lim_{n\to\infty} \mu_n(E) = \mu(E)$ *for every measurable* $E \subseteq \mathcal{X}$ *such that* $\mu(\partial E) = 0$.

Remark. The characterizations in terms of open and closed sets can be remembered as follows: As $n \to \infty$, some of the probability mass of μ_n can fall off an open set B or fall in a closed set C. The proof of the theorem can be found in Dudley's book, or the book *Convergence of Probability Measures* by Billingsley. ◇

The weak* topology on a compact metric space is again compact and metrizable.

Proposition 5 (Metrizability of the weak* topology)**.** *Let $\mathcal{X}$ be a compact metric space. The weak* topology on $\mathscr{P}(\mathcal{X})$ is metrizable.*

Proof sketch. It follows from the Stone-Weierstrass theorem that $C(\mathcal{X})$ is separable (i.e., has a countable dense set). Recall that $C(\mathcal{X})$ is separable. Let $g_1, g_2, \ldots$ be dense in $C(\mathcal{X})$ and $g_k \neq 0$ for every k. For every k, let $h_k \coloneqq g_k / \|g_k\|$. Then,

$$d(\mu, \nu) \coloneqq \sum_{k=1}^{\infty} 2^{-k} \big|\mu(h_k) - \nu(h_k)\big|$$

defines a metric on $\mathscr{P}(\mathcal{X})$ that is compatible with the weak* topology. See Dudley's book for other metrics compatible with the weak* topology. □

Theorem 6 (Compactness of the weak* topology)**.** *Let $\mathcal{X}$ be a compact metric space. The weak* topology on $\mathscr{P}(\mathcal{X})$ is compact.*

Proof sketch. Since $\mathscr{P}(\mathcal{X})$ is metrizable, it is enough to show sequential compactness. Let $\mu_1, \mu_2, \ldots \in \mathscr{P}(\mathcal{X})$. We use the separability of $C(\mathcal{X})$ and a diagonal argument to find a $n_1 < n_2 < \cdots$ such that $\mu_{n_k}(f)$ converges for every $f \in C(\mathcal{X})$. We can then verify that the functional

$$\begin{aligned} L &\colon C(\mathcal{X}) \to \mathbb{R} \\ L(f) &\coloneqq \lim_{k\to\infty} \mu_{n_k}(f) \end{aligned}$$

is linear, positive, and normalized, and apply the Riesz representation theorem. □