10. Probability measures on compact metric spaces

§10.1 The space $M(X)$

In all of the examples that we shall consider, X will be a compact metric space and $\mathcal{B}$ will be the Borel σ-algebra.

We will also be interested in the space of continuous $\mathbb{R}$-valued functions

$$C(X,\mathbb{R}) = \{f : X \to \mathbb{R} \mid f \text{ is continuous}\}.$$

This space is also a metric space. We can define a metric on $C(X,\mathbb{R})$ by first defining

$$\|f\|_\infty = \sup_{x \in X} |f(x)|$$

and then defining

$$\rho(f,g) = \|f - g\|_\infty.$$

This metric turns $C(X,\mathbb{R})$ into a complete[1] metric spaces. Note also that $C(X,\mathbb{R})$ is a vector space.

An important property of $C(X,\mathbb{R})$ that will prove to be useful later on is that it is *separable*, that is, it contains countable dense subsets.

Rather than fixing one measure on $(X,\mathcal{B})$, it is interesting to consider the totality of possible (probability) measures. To formalise this, let $M(X)$ denote the set of all probability measures on $(X,\mathcal{B})$. The following simple fact will be useful later on.

Proposition 10.1
The space $M(X)$ is convex: if $\mu_1, \mu_2 \in M(X)$ and $0 \leq \alpha \leq 1$ then $\alpha\mu_1 + (1-\alpha)\mu_2 \in M(X)$.

Exercise 10.1
Prove the above proposition.

§10.2 The weak* topology on $M(X)$

It will be very important to have a sensible notion of convergence in $M(X)$; this is called *weak* convergence*. We say that a sequence of probability measures μ_n *weak* converges* to μ, as $n \to \infty$ if, for every $f \in C(X,\mathbb{R})$,

$$\int f \, d\mu_n \to \int f \, d\mu, \quad \text{as } n \to \infty.$$

[1] Recall that a metric space is said to be *complete* if every Cauchy sequence is convergent.

If μ_n weak* converges to μ then we write $\mu_n \rightharpoonup \mu$. (Note that with this definition it is not necessarily true that $\mu_n(B) \to \mu(B)$, as $n \to \infty$, for $B \in \mathcal{B}$.) We can make $M(X)$ into a metric space compatible with this definition of convergence by choosing a countable dense subset $\{f_n\}_{n=1}^{\infty} \subset C(X)$ and, for $\mu, m \in M(X)$, and setting

$$d(\mu, m) = \sum_{n=1}^{\infty} \frac{1}{2^n \|f_n\|_\infty} \left| \int f_n \, d\mu - \int f_n \, dm \right|.$$

However, we will not need to work with a particular metric: what is important is the definition of convergence.

Notice that there is a continuous embedding of X in $M(X)$ given by the map $X \to M(X) : x \mapsto \delta_x$, where δ_x is the *Dirac measure* at x:

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A, \end{cases}$$

(so that $\int f \, d\delta_x = f(x)$).

Exercise 10.2
Show that the map $\delta : X \to M(X)$ is continuous. (Hint: This is really just unravelling the underlying definitions.)

Exercise 10.3
Let X be a compact metric space. For $\mu \in M(X)$ define

$$\|\mu\| = \sup_{f \in C(X), \|f\|_\infty \le 1} \left| \int f \, d\mu \right|.$$

We say that μ_n converges strongly to μ if $\|\mu_n - \mu\| \to 0$ as $n \to \infty$. The topology this determines is called the strong topology (or the operator topology).

(i) Show that if $\mu_n \to \mu$ strongly then $\mu_n \rightharpoonup \mu$ in the weak topology.*

(ii) Show that $X \hookrightarrow M(X) : x \mapsto \delta_x$ is not *continuous in the strong topology.*

(iii) Prove that $\|\delta_x - \delta_y\| = 2$ if $x \neq y$. (You may use Urysohn's Lemma: Let A and B be disjoint closed subsets of a metric space X. Then there is a continuous function $f \in C(X, \mathbb{R})$ such that $0 \le f \le 1$ on X while $f \equiv 0$ on A and $f \equiv 1$ on B.)

Hence prove that $M(X)$ is not compact in the strong topology when X is infinite.

Exercise 10.4
Give an example of a sequence of measures μ_n and a set B such that $\mu_n \rightharpoonup \mu$ but $\mu_n(B) \not\to \mu(B)$.

§10.3 $M(X)$ is weak* compact

We can use the Riesz Representation Theorem to establish another important property of $M(X)$: that it is compact.

Theorem 10.2
Let X be a compact metric space. Then $M(X)$ is weak* compact.

Proof. In fact, we shall show that $M(X)$ is sequentially compact, i.e., that any sequence $\mu_n \in M(X)$ has a convergent subsequence. For convenience, we shall write $\mu(f) = \int f \, d\mu$.

Since $C(X, \mathbb{R})$ is separable, we can choose a countable dense subset of functions $\{f_i\}_{i=1}^{\infty} \subset C(X)$. Given a sequence $\mu_n \in M(X)$, we shall first consider the sequence of real numbers $\mu_n(f_1) \in \mathbb{R}$. We have that $|\mu_n(f_1)| \leq \|f_1\|_\infty$ for all n, so $\mu_n(f_1)$ is a bounded sequence of real numbers. As such, it has a convergent subsequence, $\mu_n^{(1)}(f_1)$ say.

Next we apply the sequence of measures $\mu_n^{(1)}$ to f_2 and consider the sequence $\mu_n^{(1)}(f_2) \in \mathbb{R}$. Again, this is a bounded sequence of real numbers and so it has a convergent subsequence $\mu_n^{(2)}(f_2)$.

In this way we obtain, for each $i \geq 1$, nested subsequences $\{\mu_n^{(i)}\} \subset \{\mu_n^{(i-1)}\}$ such that $\mu_n^{(i)}(f_j)$ converges for $1 \leq j \leq i$. Now consider the diagonal sequence $\mu_n^{(n)}$. Since, for $n \geq i$, $\mu_n^{(n)}$ is a subsequence of $\mu_n^{(i)}$, $\mu_n^{(n)}(f_i)$ converges for every $i \geq 1$.

We can now use the fact that $\{f_i\}$ is dense to show that $\mu_n^{(n)}(f)$ converges for all $f \in C(X, \mathbb{R})$, as follows. For any $\varepsilon > 0$, we can choose f_i such that $\|f - f_i\|_\infty \leq \varepsilon$. Since $\mu_n^{(n)}(f_i)$ converges, there exists N such that if $n, m \geq N$ then

$$|\mu_n^{(n)}(f_i) - \mu_m^{(m)}(f_i)| \leq \varepsilon.$$

Thus if $n, m \geq N$ we have

$$\begin{aligned} |\mu_n^{(n)}(f) - \mu_m^{(m)}(f)| &\leq |\mu_n^{(n)}(f) - \mu_n^{(n)}(f_i)| + |\mu_n^{(n)}(f_i) - \mu_m^{(m)}(f_i)| \\ &\quad + |\mu_m^{(m)}(f_i) - \mu_m^{(m)}(f)| \\ &\leq 3\varepsilon, \end{aligned}$$

so $\mu_n^{(n)}(f)$ converges, as required.

To complete the proof, write $w(f) = \lim_{n\to\infty} \mu_n^{(n)}(f)$. We claim that w satisfies the hypotheses of the Riesz Representation Theorem and so corresponds to integration with respect to a probability measure.

(i) By construction, w is a linear mapping: $w(\lambda f + \mu g) = \lambda w(f) + \mu w(g)$.

(ii) As $|w(f)| \leq \|f\|_\infty$, we see that w is bounded.

(iii) If $f \geq 0$ then it is easy to check that $w(f) \geq 0$. Hence w is positive.

(iv) It is easy to check that w is normalised: $w(1) = 1$.

Therefore, by the Riesz Representation Theorem, there exists $\mu \in M(X)$ such that $w(f) = \int f d\mu$. We then have that $\int f d\mu_n^{(n)} \to \int f d\mu$, as $n \to \infty$, for all $f \in C(X, \mathbb{R})$, i.e., that $\mu_n^{(n)}$ converges weak* to μ, as $n \to \infty$. □