2. Weak convergence.

Theorem 2.1. Riesz's theorem says that a nonnegative linear function on the space of continuous functions $C(\mathbf{X})$ on a compact metric space $\mathbf{X}$ can be represented as

$$\Lambda(f) = \int f(x)\mu(dx)$$

where μ is finite nonnegative countably additive measure on the Borel σ-field of $\mathbf{X}$.

Proof. It involves many steps. We can assume with out loss of generality that $\Lambda(\mathbf{1}) = 1$.

Step 1. For any closed set $A \subset \mathbf{X}$ define $\alpha(A) = \inf_{\{f\geq 0, f=1 \text{ on } A\}} \Lambda(f)$. $\alpha(\cdot)$ is finitely sub-additive on the class of closed sets and additive over disjoint unions.

Proof. If A, B are closed sets so is $A \cup B$ and $\alpha(A \cup B) \leq \alpha(A) + \alpha(B)$. There is an $f \geq \chi_A$ with $\Lambda(f) \leq \alpha(A) + \epsilon$ and a $g \geq \chi_B$ with $\Lambda(g) \leq \alpha(B) + \epsilon$. $f + g \geq \chi_{A\cup B}$ and $\alpha(A\cup B) \leq \Lambda(f+g) = \Lambda(f)+\Lambda(g) \leq \alpha(A)+\alpha(B)+2\epsilon$.. If A and B are disjoint closed sets there is a continuous function ϕ with $0 \leq \phi \leq 1$ and $\phi = 0$ on B and 1 on A (Urysohn's lemma). If $f \geq \chi_{A\cup B}$, and $\Lambda(f) \leq \alpha(A\cup B)+\epsilon$, it follows that $f\phi \geq \chi_A$ and $f(1-\phi) \geq \chi_B$. $\alpha(A) \leq \Lambda(f\phi), \alpha(B) \leq \Lambda(g)$ and $\alpha(A)+\alpha(B) \leq \Lambda(f\phi)+\lambda(f(1-\phi)) = \Lambda(f) \leq \alpha(A\cup B)+\epsilon$, proving $\alpha(A \cup B) \geq \alpha(A) + \alpha(B)$.

Step 2. For any open set G we define $\beta(G) = \sup_{A\subset G} \alpha(A)$ the supremum taken over closed sets A. $\beta(\cdot)$ is countably sub-additive on the class of open sets as well as countably additive over disjoint union of open sets.

Proof. Let $G = G_1 \cup G_2$ be a union of two open sets and $A \subset G$. Let $A_1 = A \cap \{x : d(x, G_2^c) \leq d(x, G_1^c)\}$ and $A_2 = A \cap \{x : d(x, G_1^c) \leq d(x, G_2^c)\}$. Clearly $A = A_1 \cup A_2$. Claim $A_1 \subset G_1$. If not there is $x \in A_1, x \in G_1^c$. $d(x, G_1^c) = 0$ and from the definition of A_1, $d(x, G_2^c) = 0$. This means $x \in G_1^c \cap G_2^c$. Contradicts $A \subset G_1 \cup G_2$. $A = A_1 \cup A_2$. $A_1 \subset G_1$ and $A_2 \subset G_2$. It now follows that given $\epsilon > 0$ there is A such that

$$\beta(G_1 \cup G_2) \leq \alpha(A) + \epsilon \leq \alpha(A_1) + \alpha(A_2) + \epsilon \leq \beta(G_1) + \beta(G_2) + \epsilon$$

If $G_1 \cap G_2 = \emptyset$ so is $A_1 \cap A_2$ and $\beta(G_1 \cup G_2) \geq \alpha(A_1) + \alpha(A_2)$. If $\{G_i\}$ is a countable disjoint collection then $\beta(\cup_i G_i) \geq \sum_i \beta(G_i)$. We have till now not used compactness. If $G = \cup_i G_i$ and $A \subset G$ is a compact set (closed set in a compact space) $A \subset \cup_{i=1}^n G_i$ for some n. then $\beta(G) \leq \beta(A) + \epsilon \leq \beta(\cup_{i=1}^n G_i) + \epsilon \leq \sum_{i=1}^n \beta(G_i) + \epsilon$.

Step 3. $\beta(A^c) + \alpha(A) = 1$. More generally, if $A \subset G$ then $G = A \cup (A^c \cap G)$ is a disjoint union and $\beta(G) = \alpha(A) + \beta(G \cap A^c)$.

Proof. Clearly

$$\beta(A^c \cap G) + \alpha(A) = \sup_{B\subset A^c\cap G} \alpha(B) + \alpha(A) = \sup_{B\subset A^c\cap G} \alpha(A \cup B) \leq \sup_{C\subset G} \alpha(C) = \beta(G)$$

On the other hand for any closed set A and any $\epsilon > 0$ there is an open set $U \supset A$ such that $\beta(U) \leq \alpha(A) + \epsilon$. To see this, note that by definition for any $\delta > 0$ there is $f \in C(\mathbf{X})$ such that $f \geq \chi_A$ and $\Lambda(f) \leq \alpha(A) + \delta$. There is a neighborhood U of A such that $f \geq 1 - \delta$ on U and with $\delta' = (1-\delta)^{-1} - 1$, $(1+\delta')f \geq \chi_U$ and for any $C \subset U$,

$$\alpha(C) \leq \Lambda((1+\delta')f) = (1+\delta')(\alpha(A) + \delta) = \alpha(A) + \epsilon$$

if we choose δ small enough. Therefore $\beta(U) \leq \alpha(A) + \epsilon$. Find V an open set such that $A \subset V \subset \overline{V} \subset U$. Then $G \subset U \cup (G \cap (\overline{V})^c)$.

$$\beta(G) \leq \beta(U) + \beta(G \cap (\overline{V})^c) \leq \alpha(A) + \beta(G \cap A^c) + \epsilon$$

Step 4. For any set E we define

$$\mu^*(E) = \inf_{G \supset E} \beta(G); \quad \mu_*(E) = \sup_{A \subset E} \alpha(A)$$

The class $\mathcal{E}$ of sets E for which $\mu^*(E) = \mu_*(E)$ is a σ-field and $\mu(E) = \mu^*(E) = \mu_*(E)$ is a countably additive measure on $\mathcal{E}$. $\mathcal{E}$ contains all open and closed sets and hence includes $\mathcal{B}$ the Borel σ-field. .

Proof. $E \in \mathcal{E}$ if and only if given any $\epsilon > 0$ there are sets A, G such that A is closed, G is open $A \subset E \subset G$ and $\beta(G) - \alpha(A) < \epsilon$. If $A \subset E \subset G$ then $G^c \subset E^c \subset A^c$ and $\beta(A^c) - \alpha(G^c) = \beta(A^c \backslash G^c) = \beta(G \backslash A) = \beta(G) - \beta(A) < \epsilon.\mathcal{E}$ is closed under finite unions and μ is additive over finite disjoint unions. Since $\mathcal{E}$ has been shown to be a field, to prove it is a σ-field, we need to consider only countable disjoin unions. We have disjoint $\{E_i\}$ and $G_i \supset E_i \supset A_i$ with $\beta(G_i \backslash A_i) \leq \epsilon 2^{-i}$. $\beta(\cup_i G_i) - \sum_i \alpha(A_i) \leq \epsilon$. $\sum \alpha(A_i)$ is convergent and therefore for some finite n, $\beta(\cup_i G_i) - \sum_{i=1}^n \alpha(A_i) \leq 2\epsilon$. That closed sets are in $\mathcal{E}$ was shown in step 3.

Step 5.

$$\Lambda(f) = \int f(x) \mu(dx)$$

Proof. Can assume $0 \leq f \leq 1$. Given $\epsilon > 0$ we can find a finite number of closed disjoint sets $A_1, A_2, \ldots, A_n$ such that $\sum_{i=1}^n \mu(A_i) \geq \mu(\mathbf{X}) - \epsilon$ and $\sup_{x \in A_i} f(x) - \inf_{x \in A_i} f(x) \leq \epsilon$ for every i. Let U_i be open sets that are again disjoint and $U_i \supset A_i$ for every i. Let g_i be continuous functions $0 \leq g_i(x) \leq 1$, $g_i(x) = 1$ on A_i and $g_i(x) = 0$ for $x \notin U_i$. We have $f \geq \sum_i [\inf_{x \in A_i} f(x)] g_i(x)$.

$$\begin{aligned}
\Lambda(f) &\geq \sum_i [\inf_{x \in A_i} f(x)] \Lambda(g_i) \\
&\geq \sum_i [\inf_{x \in A_i} f(x)] \mu(A_i) \\
&\geq \sum_i \int_{A_i} f(x) d\mu - \epsilon \mu(A_i) \\
&= \int f(x) d\mu - 2\epsilon \mu(\mathbf{X})
\end{aligned}$$

Since ϵ is arbitrary $\Lambda(f) \geq \int f d\mu$. The same is true with $1-f$. Together they imply $\Lambda(f) = \int f d\mu$.

Theorem 2.2. Let X be a complete separable metric space and μ a countably additive finite measure with $\mu(\mathbf{X}) = 1$. Then for any $\epsilon > 0$ there is a compact set K_ϵ such that $\mu(K_\epsilon) \geq 1 - \epsilon$.

Proof. By Lindelof property $\mathbf{X} = \cup_{j=1}^\infty S(x_j, \epsilon)$. By countable additivity of the measure, with $\epsilon_i = \frac{1}{i}$ there is some n_i spheres around $\{x_{i,j}\}$ of radius $\frac{1}{i}$ with their union having measure at least $1 - \epsilon 2^{-i}$. Then

$$\mu[\cap_{i=1}^\infty \cup_{j=1}^{n_i} S(x_{i,j}, \epsilon_i)] \geq 1 - \epsilon$$

and $\cap_{i=1}^\infty \cup_{j=1}^{n_i} S(x_{i,j}, \epsilon_i)$ is totally bounded and is essentially compact in a complete metric space.

Weak Convergence. We say that μ_n converges weakly to μ or $\mu_n \Rightarrow \mu$ if $\int f d\mu_n \to \int f d\mu$ for all $f \in C(\mathbf{X})$. If $\mathbf{X}$ is compact the space $\mathcal{M}$ of probability measures on $\mathbf{X}$ is compact in the weak topology. Because $C(\mathbf{X})$ is separable we can choose a subsequence such that $\int f d\mu_n$ has a convergent subsequence for a dense set of f and so for every f. The limit is a non negative linear functional $\Lambda(f)$ with $\Lambda(\mathbf{1}) = 1$ and we can use the Riesz theorem.

It is enough if most of the mass is supported on a compact set. If $\mathcal{P}$ is a collection of measures from $\mathcal{M}$ such for any $\epsilon > 0$, there is a compact set K_ϵ such that $\mu(K_\epsilon) \geq 1 - \epsilon$ for all $\mu \in \mathcal{P}$ then any sequence from $\mathcal{P}$ will have a weakly convergent subsequence. The condition is sufficient in all separable metric spaces but also necessary if the space is complete.

Problem. Can you generalize the notion of positive definiteness to functions that are not necessarily continuous? Does that characterize Fourier transforms of nonnegative functions in $L_p(R^d)$, $1 < p \leq 2$?

Problem. If probability measures $\mu_n \Rightarrow \mu$ weakly and $g \geq 0$ is a function not necessarily bounded then show that $\liminf_{n\to\infty} \int f d\mu_n \geq \int f d\mu$. If $|f| \leq Cg$ and $\lim_{n\to\infty} \int g d\mu_n = \int g d\mu$ then show that $\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$.

Problem. If probability measures $\mu_n \Rightarrow \mu$ on a compact metric space $\mathbf{X}$ and if $\mathcal{F}$ is a family of equi-continuous functions, (i.e if $x_n \to x$ then $\sup_{f\in\mathcal{F}} |f(x_n) - f(x)| \to 0$), show that

$$\sup_{f\in\mathcal{F}} \left| \int f(x) d\mu_n - \int f(x) \mu(dx) \right| \to 0$$

What if $\mathbf{X}$ is only complete and separable (not necessarily compact)?

Theorem 2.3. $\mathbf{X}$ is a separable metric space. μ_n is a sequence of probability distributions on $\mathbf{X}$.The following are equivalent.

1. $\mu_n \Rightarrow \mu$ i.e. for any $f \in C(\mathbf{X})$, $\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$.

2. For any uniformly continuous bounded function f, $\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$.

3. For any closed set C, $\mu(C) \geq \limsup_{n\to\infty} \mu_n(C)$

4. For any open set G, $\mu(G) \leq \liminf_{n\to\infty} \mu_n(G)$

5. For any continuity set A, i.e. a set for which $\mu(\bar{A}) = \mu(A^o)$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$

Proof. That **1** implies **2** is trivial. To prove **2** implies **3** let C be a closed set. Consider the function $f(x) = \frac{1}{1+d(x,C)}$. f is uniformly continuous, bounded by 1, $f = 1$ on C and $0 < f < 1$ on C^c. In particular $f_k(x) = [f(x)]^k \downarrow \mathbf{1}_C(x)$.

$$\mu(C) = \lim_{k\to\infty} \int f_k(x) d\mu = \lim_{k\to\infty} \limsup_{n\to\infty} \int f_k(x) d\mu_n \geq \limsup_{n\to\infty} \mu_n(C)$$

Taking complements **3** and **4** are equivalent.

$$\mu(\bar{A}) \geq \limsup \mu_n(\bar{A}) \geq \limsup \mu_n(A) \geq \liminf \mu_n(A) \geq \liminf \mu_n(A^o) \geq \mu(A^0)$$

If the ends are equal then there is equality everywhere. So **3** and **4** imply **5.**. Finally **5** implies **1**. Tp see this id $|f| \leq M$ the interval $[-M, M]$ can be divided into N disjoint subintervals $\{I_j\}$ such that $x : f(x) \in I_j$ are continuity sets. f can be uniformly approximated by $f_N(x) = \sum a_j \mathbf{1}_{I_j}$ and $\int f_N(x) d\mu_n \to \int f_N d\mu$.

We saw that in a complete separable metric space any probability measure is essentially supported on a compact set, in the sense that for any positive $\epsilon > 0$ there is a compact set K_ϵ such that $\mu(K_\epsilon) \geq 1 - \epsilon$. We are interested in characterizing compact subsets of the space of probability distributions under weak convergence on a complete separable metric space.

Theorem 2.4. Let $\mathcal{P}$ be a subset of the space of probability distributions $\mathcal{M}(\mathbf{X})$ on a separable metric space $\mathbf{X}$, such that given any $\epsilon > 0$, there is a compact set $K_\epsilon \subset \mathbf{X}$ such that $\mu(K_\epsilon) \geq 1 - \epsilon$ for all $\mu \in \mathcal{P}$. Then given any sequence μ_n from $\mathcal{P}$, there is a subsequence that converges weakly to a limit $\mu \in \mathcal{M}(\mathbf{X})$. The condition is also necessary if the space $\mathbf{X}$ is complete.

Proof. First let us observe that if $\mathbf{X}$ is compact then $\mathcal{M}(\mathbf{X})$ is compact under he weak topology. Given μ_n we consider the linear functionals $\Lambda_n(f) = \int f d\mu_n$. If $\mathbf{X}$ is compact $C(\mathbf{X})$ is separable we can chose a subsequence of Λ_n (which we will continue to denote by Λ_n) that converges for a dense set of continuous functions, which will then converge for all continuous functions. This limit is a nonnegative linear functional with $\Lambda(\mathbf{1}) = 1$ and by Riesz theorem is represented by a measure. The subsequence of probability distributions clearly converges to μ in the weak topology. For each ϵ we can define μ_n^ϵ as the restriction of μ_n to K_ϵ, normalized to be a probability distribution. $\mu_n^\epsilon(E) = \frac{1}{\mu_n(K_\epsilon)} \mu_n(K_\epsilon \cap E)$. For each ϵ, $\mu_n^\epsilon(E)$ are supported on the compact set K_ϵ and will have a convergent subsequence. By diagonalization we can assume that choosing a sequence $\epsilon_j \to 0$, $\lim_{n\to\infty} \mu_n^j = \mu^j$ and because $\|\mu_n^i - \mu_n^j\| \leq \epsilon_i + \epsilon_j$, it follows that $\|\mu^i - \mu^j\| \leq \epsilon_i + \epsilon_j$ and $\mu = \lim_{j\to\infty} \mu^j$ exists and $\mu_n \Rightarrow \mu$.

To prove the converse we will use Dini's theorem which says that if a sequence of upper semi continuous functions $f_n(x)$ on $\mathbf{X}$ decreases monotonically to 0, the convergence is uniform over compact subsets of $\mathbf{X}$. Given $\epsilon > 0$ for each $x \in \mathbf{X}$ there is $n_\epsilon(x)$ such that $f_{n_\epsilon(x)}(x) < \epsilon$. By upper semi continuity $f_{n_\epsilon(x)}(y) < 2\epsilon$ for y in a neighborhood $N_{\epsilon,x}$ around x. Given a compact set $K \subset \mathbf{X}$ there is a finite sub cover N_{ϵ,x_j} of K and by the monotonicity of the sequence $f_n(y) \leq 2\epsilon$ for $n \geq \sup_j N_{\epsilon,x_j}$ on K.

Proceeding with the proof of the converse, we use the Lindelof property to write for any k, $\mathbf{X} = \cup_j S(x_j, \frac{1}{k})$. Then with $G_{n,k} = \cup_{j=1}^n S(x_j, \frac{1}{k})$, for each k, $\mu(G_{n,k}) \uparrow 1$ as $n \to \infty$. Since $\mu(G)$ is a lower semicontinuous function of μ for every open set G, for every k,

$$\lim_{n\to\infty} \inf_{\mu \in \mathcal{P}} \mu(G_{n.k}) = 1$$

Given ϵ and k, there is a $N(k,\epsilon)$ such that

$$\inf_{\mu \in \mathcal{P}} \mu(G_{N(k,\epsilon),k}) \geq 1 - \epsilon 2^{-k}$$

For every $\epsilon > 0$, the set $\cap_k G_{N(k,\epsilon)}$ is totally bounded and

$$\inf_{\mu \in \mathcal{P}} \mu[\cap_k G_{N(k,\epsilon)}] \geq 1 - \epsilon$$