Chapter 4

Weak convergence and Compactness.

Let X be a complete separable metic space and $\mathcal{B}$ its Borel $\sigma-$field. We denote by $\mathcal{M}(X)$ the space of probability measures on $(X, \mathcal{B})$. A sequence $\mu_n \in \mathcal{M}(X)$ of probability measures converges weakly to a probability measure $\mu \in \mathcal{M}(X)$ ($\mu_n \Rightarrow \mu$) if for every bounded continuous function $f : X \to R$ we have

$$\lim_{n\to\infty} \int_X f(x) d\mu_n = \int_X f(x) d\mu \tag{4.1}$$

Theorem 4.1. *The following are equivalent.*

1. $\mu_n \Rightarrow \mu$
2. *For every bounded uniformly continuous function* f
$$\lim_{n\to\infty} \int_X f(x) d\mu_n = \int_X f(x) d\mu$$
3. *For every closed set* $C \subset X$
$$\limsup_{n\to\infty} \mu_n(C) \le \mu(C)$$
4. *For every open set* $G \subset X$
$$\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$$
5. *For every continuity set* A*, i.e.* A *suc that* $\mu(\bar{A}) = \mu(A) = \mu(A^o)$*, we have*
$$\lim_{n\to\infty} \mu_n(A) = \mu(A)$$

Proof. Clearly 1 $\Rightarrow$ 2 . To show that 2 $\Rightarrow$ 3, we consider $f_k(x) = [\frac{1}{1+d(x,C)}]^k$. $f_k(x)$ is uniformly continuous and bounded by 1. $f_k(x) \geq \mathbf{1}_C(x)$ and $f_k(x) \to \mathbf{1}_C(x)$ as $n \to \infty$. For every k

$$\limsup_{n\to\infty} \mu_n(C) \leq \lim_{n\to\infty} \int_X f_k(x) d\mu_n = \int_X f_k(x) d\mu$$

By letting $k \to \infty$, we obtain

$$\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$$

We see that 3 $\Leftrightarrow$ 4 by taking complements. For any set A, $\mu(A^c) = 1 - \mu(A)$. It is clear that 3 and 4 $\Rightarrow$ 5. Because A^o is open and $\bar{A}$ is closed,

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

If $\mu(A^o) = \mu(A) = \mu(\bar{A})$, we have equality everywhere and that implies 5. Finally to prove that 5 $\Rightarrow$ 1, we take a bounded continuous function f and approximate it uniformly by a simple function $f_k = \sum_{i=1}^k a_i \mathbf{1}_{A_i}(x)$. The natural choice for A_i is $A_i = \{x : a_i \leq f(x) \leq a_{i+1}\}$. Since the sets $\{x : f(x) = a\}$ are disjoint all but a countable number of them are of measure 0, under μ. We can pick $\{a_i\}$ avoiding this countable set. Then all our sets A_i will be continuity sets and so for every k.

$$\lim_{n\to\infty} \int_X f_k(x) d\mu_n = \int_X f_k(x) d\mu$$

Since f_k approximates f uniformly it follows that

$$\lim_{n\to\infty} \int_X f(x) d\mu_n = \int_X f(x) d\mu$$

which is 1. □

In a complete separable metric space any probability measure P, is essentially supported on a compact set. More precisely

Theorem 4.2. *Given any probability measure μ on the Borel subsets $\mathcal{B}$ of a complete separable metric space X and $\epsilon > 0$, there exists a compact set K_ϵ such that $\mu(K_\epsilon) \geq 1 - \epsilon$.*

Proof. . Consider a countable dense subset $D = \{x_j\}$ of X and spheres $\{S(x_j, \frac{1}{k})\}$ of radius $\frac{1}{k}$ around points $x_j \in D$. Clearly $\cup_j B(x_j, \frac{1}{k}) = X$ and by the countable additivity of μ, for each k, we can find n_k such that

$$\mu[\cup_{j=1}^{n_k} B(x_j, \frac{1}{k})] \geq 1 - \frac{\epsilon}{2^k}$$

The set $E = \cap_k \cup_{j=1}^{n_k} B(x_j, \frac{1}{k})$ is totally bounded and has the property $\mu(E) \geq 1 - \epsilon$. Since X is complete, the closure $\bar{E}$ of E is compact. □

Remark 4.1. The space $\mathcal{M}(X)$ of probability measures on X, with weak convergence is a complete separable metric space. In other words, weak convergence can be metrized. To do this we note that weak convergence is a topological notion and is not altered if we change the metric to an equivalent one. But the notion of uniform continuity depends on the metric. We can change the metric so that the set of bounded uniformly continuous functions is separable. We imbed X into Y, the countable product of unit intervals, which is compact. Then X under the metric inherited from Y, will not be complete. But its completion, which is the closure $\bar{X}$ of X in Y is a compact metric space. The space of uniformly continuous functions on X is the same as the space of continuous functions on $\bar{X}$. Since $\bar{X}$ is a compact metric space, $C(\bar{X})$ is separable. Then we can define

$$d(\mu_1, \mu_2) = \sum_{j=1}^{\infty} \frac{1}{2^k} \frac{1}{1 + \|f_k\|} | \int f_k d\mu_1 - \int f_k d\mu_2 |$$

where $\{f_k\}$ is a dense set of uniformly continuous functions on X. By Theorem 4.1 convergence in this metric is equivalent to weak convergence.

The next theorem due to Prokhorov gives a necessary and sufficient condition for a subset $\mathcal{A} \subset \mathcal{M}(X)$ to be conditionally compact in the weak topology, i.e. every sequence from $\mathcal{A}$ has a weakly convergent subsequence.

Theorem 4.3. $\mathcal{A} \subset \mathcal{M}(X)$ *is conditionally compact in the weak topology if and only if for any* $\epsilon > 0$*, there is a compact set* K_ϵ *(independent of* μ*) such that*

$$\mu(K_\epsilon) \geq 1 - \epsilon$$

for all $\mu \in \mathcal{A}$*.*

Proof. **Necessity.** In the proof of Theorem 4.2 a crucial step was the limit

$$\lim_{n \to \infty} \mu(\cup_{j=1}^{n} S(x_j, \frac{1}{k})) = 1$$

$\cup_{j=1}^{n} S(x_j, \frac{1}{k})$ is an open set and $\mathcal{A}$ is contained in a compact set. By Theorem 4.1, $\mu(G)$ is lower semicontinuous as a function of μ in the weak topology. Dini's theorem ensures that the convergence above is uniform for all $\mu \in \mathcal{A}$ and that completes the proof of necessity. The construction of K_ϵ as in Theorem4.2 works simultaneously for all $\mu \in \mathcal{A}$.

Sufficiency. We first note that if X is compact, then $\mathcal{M}(X)$ is compact. $C(X)$ is separable and given any sequence μ_n from $\mathcal{M}(X)$, we can definitely choose, by diagonalization, a subsequence such that

$$\Lambda(f) = \lim_{n \to \infty} \int f(x) d\mu_n$$

exists for f in a a countable dense subset and therefore for all $f \in C(X)$. We appeal to the Riesz representation theorem to find a $\mu \in \mathcal{M}(X)$ such that

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

Actually if $\mathcal{M}(X)$ is taken as the set of all finite measures instead of just probability measures, only some minor changes need to be made. One has to be sure that $\mu_n(X) \to \mu(X)$ and this has to be added, if it not already implied, in any of the equivalent formulations of 4.1 (for instance in 3 and 4). In particular if X is compact and $\mu_n(X)$ is bounded there is a subsequence that converges weakly to a measure μ.

If X is not compact, but we have a collection $\mathcal{A} \subset \mathcal{M}(X)$, satisfying the uniform tightness condition, we restrict μ_n to K_k to get $\mu_{n,k}$ satisfying $1 \geq \mu_{n,k}(K_k) \geq 1 - \frac{1}{k}$. By the diagonalization procedure, we can assume that along a subsequence (that we continue to denote by μ_n), for every k,

$$\lim_{n\to\infty} \mu_{n,k} = \alpha_k$$

exists. It is not difficult to see that for $k > l$

$$\alpha_k \geq \alpha_l$$

and it follows that $\lim_{k\to\infty} \alpha_k = \alpha$ exists and $\mu_n \to \alpha$. $\square$

While one can use Garsia-Rodemich-Rumsey estimates to establish tightness there are some other methods are equally useful. There are stochastic processes that admit jumps like Poisson processes, and the space on which to put these measures is the space $D[0,1]$ of functions $x(t)$ on $[0,T]$ that have left and right limits at every point, $x(t+0) = x(t)$ and $x(T-0) = x(T)$. A sequence $x_n(t)$ of such functions converges (in Skorohod J-1 topology) to $x(t)$ in $D[0,1]$ if there are one to one continuous maps $\lambda_n(t)$ of $[0,1]$ onto itself such that

$$\sup_{0\leq t\leq T} |\lambda_n(t) - t| + \sup_{0\leq t\leq T} |x_n(\lambda_n(t)) - x(t)| \to 0$$

If $x(t)$ and $y(t)$ are close in this topology then their jumps do not have to align perfectly, but to every jump of significant size in one there be a corresponding jump of nearly the same size at a nearby location in the other. There is a version of Ascoli-Arzela theorem here. For a set $\mathcal{D}$ of functions in $D[0,1]$ to be conditionally compact it is necessary and sufficient that their jumps if any be uniformly bounded and for any $\delta > 0$, there be a uniform bound on the number of jumps of size greater than δ. More over these jumps should stay away by a uniform distance from each other. Two jumps of significant size are not allowed to come together and combine. If we denote by $\Delta_f(a,b)$ the oscillation of the function f in the open interval (a,b), i.e.

$$\Delta_f(a,b) = \sup_{a<c<d<b} |f(c) - f(d)|$$

then

$$\Delta^*_f(a,b) = \min_{a<c<b} \max\{\Delta_f(a,c), \Delta_f(c,b)\}$$

It is almost the oscillation in (a,b) except that we discount one jump at some point c in between. The modulus continuity of f in $D[0,1]$ is defines as

$$\bar{\omega}_f(h) = \max\{\sup_{\substack{0\le a<b\le 1\\ |a-b|\le h}} \Delta^*_f(a,b), \sup_{0\le a\le h}|f(0)-f(a)|, \sup_{1-h\le a\le 1}|f(1)-f(a)|\}$$

The Ascoli-Arzela theorem in this context requires $\bar{\omega}_f(h)$ to go to 0 uniformly and the functions f to be uniformly bounded. (which will follow from the behavior of the modulus of continuity if the functions are bounded at 0 and the jumps are uniformly bounded.)

Given a function f in $D[0,1]$, let us define successively

$$\begin{aligned} \tau_1 &= \{\inf t : |f(t)-f(0)| \ge \epsilon\} \\ \cdots\cdots & \\ \tau_j &= \{\inf t \ge \tau_{j-1} : |f(t)-f(\tau_{j-1})| \ge \epsilon\} \end{aligned}$$

until τ_j fails to exist, i.e $\sup_{\tau_{j-1}\le t\le 1}|x(t)-x(\tau_{j-1})| < \epsilon$. Then the modulus of continuity can be estimated. If $h = \min\{\tau_1, \tau_2-\tau_1, \ldots, \tau_{j-1}-\tau_{j-2}\}$, then it is not difficult to see that, because any interval of length less than h can contain at most one of the points $\tau_1, \ldots, \tau_{j-1}$,

$$\bar{\omega}_f(h) \le \epsilon$$

We now develop some tools to estimate the modulus of continuity.

Lemma 4.4. *Let $\{S_j : 1 \le j \le N\}$ be a collection of random variables such that $P[S_0 = 0] = 1$, S_j is adapted to the filtration $\mathcal{F}_j$ and*

$$\sup_{0\le j\le k\le N} \operatorname{esssup}_\omega P[|S_k - S_j| \ge \epsilon | \mathcal{F}_j] \le p(\epsilon) < 1$$

Then

$$P[\sup_{0\le j\le N}|S_j| \ge 2\epsilon] \le \frac{p(\epsilon)}{1-p(\epsilon)}$$

Proof.

$$\begin{aligned} P[\sup_{0\le j\le N}|S_j| \ge 2\epsilon, |S_N|\le\epsilon] &= \sum_{k=1}^{N-1} P[\sup_{0\le j\le k-1}|S_j| < 2\epsilon, |S_k| \ge 2\epsilon, |S_N| \le \epsilon] \\ &\le \sum_{k=1}^{N-1} P[\sup_{0\le j\le k-1}|S_j| < 2\epsilon, |S_k| \ge 2\epsilon, |S_N - S_k| \ge \epsilon] \\ &\le p(\epsilon)\sum_{k=1}^{N-1} P[\sup_{0\le j\le k-1}|S_j| < 2\epsilon, |S_k| \ge 2\epsilon] \\ &= p(\epsilon) P[\sup_{0\le j\le N}|S_j| \ge 2\epsilon] \end{aligned}$$

$$P[\sup_{0\le j\le N} |S_j| \ge 2\epsilon, |S_N| > \epsilon] \le p(\epsilon)$$

Adding the two bounds

$$P[\sup_{0\le j\le N} |S_j| \ge 2\epsilon] \le p(\epsilon) + p(\epsilon)P[\sup_{0\le j\le N} |S_j| \ge 2\epsilon]$$

which proves the lemma. □

This lemma will allow us to estimate $P[\tau_j - \tau_{j-1} \le h|\mathcal{F}_{\tau_{j-1}}]$ and can be used as input for the next lemma.

Lemma 4.5. *Let* $\operatorname{esssup}_\omega P[\tau_j - \tau_{j-1} \le h|\mathcal{F}_{\tau_{j-1}}] \le \phi_\epsilon(h)$. *Then*

$$P[\bar{\omega}_f(h) \le \epsilon] \le \inf_{j,s}[j\,\phi_\epsilon(h) + e\big[\phi_\epsilon(s) + e^{-s}[1-\phi_\epsilon(s)]\big]^j]$$

Proof. Clearly for any $s > 0$,

$$\begin{aligned}
P[\bar{\omega}_f(h) \ge \epsilon] &\le P[\min\{\tau_1, \tau_2 - \tau_1, \ldots, \tau_{j-1} - \tau_{j-2}\} \le h] + P[\tau_j \le 1] \\
&\le (j-1)\phi(h) + eE[e^{-\tau_j}] \\
&\le (j-1)\phi_\epsilon(h) + e\big[\phi_\epsilon(s) + e^{-s}[1-\phi_\epsilon(s)]\big]^j
\end{aligned}$$

□

Remark 4.2. If a sequence in $D[0,1]$ converges to a limit that is a continuous function, then the limit is uniform. If the size of the largest jump goes to zero, then the convergence is to a limit in $C[0,1]$.

Example 4.1. Consider i.i.d. random variable $\{X_i\}$ with mean zero and variance 1. Let $S_n = X_1 + X_2 + \cdots + X_n$ and $X_n(t) = \frac{1}{\sqrt{n}} S_[nt]$. By central limit theorem the finite dimensional distributions of $X_n(t)$ converge to the finite dimensional distributions of $X(t)$ the Brownian motion. The process $X_n(t)$ can be realized in $D[0,1]$ and we want to show weak convergence of the processes P_n, in $D[0,1]$. First we note that for $k \ge j$,

$$P_n[|x(\frac{k}{n}) - x(\frac{j}{n})| \ge \epsilon|\mathcal{B}_{\frac{j}{n}}] \le \frac{k-j}{n\epsilon^2}$$

and as $n \to \infty$,

$$P[\max_{1\le j\le n} |X_j| \ge \epsilon\sqrt{n}] \le nP[|X| \ge \epsilon\sqrt{n}] \to 0$$