Empirical Process Reading Group Notes

Manu Navjeevan

March 12, 2021

1 Math Review

1.1 Vector Spaces and Norms

Definition 1.1 (Vector Space)**.** A vector space X is a set of elements with two operations, addition $(+)$ and scalar multiplication $(\cdot)$, and an additive identity $\mathbf{0} \in X$ satisfying:

1. $x + y = y + x$
2. $(x + y) + z = x + (y + z)$
3. $\mathbf{0} + x = x$, $\forall x \in X$
4. $\alpha(x + y) = \alpha x + \alpha y$
5. $(\alpha + \beta)x = \alpha x + \beta x$
6. $(\alpha\beta)x = \alpha(\beta)x$
7. $0x = \mathbf{0}$ and $1x = x$

Examples include $\mathbb{R}^K$ and $\mathcal{C}[a, b]$, the set of all continuous functions from $[a, b] \to \mathbb{R}$.

Definition 1.2 (Norm)**.** Let X be a vector space. A norm is a functional, $\|\cdot\| : X \to \mathbb{R}$ satisfying

1. $\|x\| \geq 0$, $\forall x \in X$ and $\|x\| = 0$ if and only if $x = \mathbf{0}$.
2. $\|x + y\| \leq \|x\| + \|y\|$ (Triangle Inequality)
3. $\|\alpha x\| = |\alpha|\|x\|$, $\forall \alpha \in \mathbb{R}, x \in X$

Examples of norms include the ℓ^p norms on $\mathbb{R}^K$ or the sup-norm on the space of all bounded, real valued, functions. On $\mathbb{R}^K$ all norms are equivalent, which is to say that for any two norms $\|\cdot\|_1, \|\cdot\|_2$ there are constants C_1, C_2 such that $C_1\|\cdot\|_2 \leq \|\cdot\|_1 \leq C_2\|\cdot\|_2$. However, this is not generally the case for functional vector spaces. For example on $\mathcal{C}[a, b]$ there is no constant c such that, for all f:

$$\sup_{x\in[a,b]} f(x) = \|f\|_\infty \leq c\|f\|_2 = \left(\int_a^b f^2(x)dx\right)^{1/2}.$$

Closely related to a norm is the concept of a metric, which is a way of defining a distance on a space.

Definition 1.3 (Metric)**.** Let X be a vector space. A metric (or distance metric) on X is a functional $d(x, y) : X \times X \to \mathbb{R}$ satisfying:

1. $d(x, y) \geq 0, \forall x, y$ and $d(x, y) = 0 \iff x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$, $\forall x, y, z$

It is straightforward to verify that, given a norm on a vector space X, we can generate a valid metric:

$$d_{\|\cdot\|}(x,y) := \|x-y\|.$$

We return to these concepts when discussing a topology.

1.2 Topology and Continuity

A topology is a general structure under which we can discuss concepts such as convergence and continuity. We can start with a general structure and then discuss spaces where the topology is generated by a metric (or norm).

Definition 1.4 (Topology)**.** A topology on a set X is a collection of subsets of X, $\boldsymbol{\tau} \subset 2^X$ satisfying:

1. $\emptyset, X \in \boldsymbol{\tau}$.
2. $\boldsymbol{\tau}$ is closed under finite intersections, if $\{A_k\}_{k=1}^K \in \boldsymbol{\tau}$ then $\bigcap_{k=1}^K A_k \in \boldsymbol{\tau}$.
3. $\boldsymbol{\tau}$ is closed under arbitrary unions, if $\{A_k\}_k \in \boldsymbol{\tau}$ then $\bigcup_k A_k \in \boldsymbol{\tau}$.

The elements of $A \in \tau$ are called open sets. A set, B, is closed if it's complement is in $\boldsymbol{\tau}$, $B^c \in \boldsymbol{\tau}$.

Some simple examples include the trivial topology, $\boldsymbol{\tau} = \left\{X, \emptyset\right\}$ and the discrete topology $\boldsymbol{\tau} = 2^{\mathcal{X}}$. Given a topology, we can define some familiar terms:

Definition 1.5 (Interior)**.** For a subset $A \subseteq X$, the interior of A, denoted A°, is the largest open set included in A (where largest is defined under the usual subset ordering). We can also express this as the union of all open sets contained by A.

$$A^\circ = \bigcup \{B : B \in \boldsymbol{\tau}, B \subseteq A\}.$$

Note that a set is open if and only if $A = A^\circ$.

Definition 1.6 (Closure)**.** For a subset $A \subseteq X$, the closure of A, denoted $\bar{A}$, is the smallest closed set the covers A. We can express this as the intersection of all closed sets containing A:

$$\bar{A} = \bigcap \{B : B^c \in \boldsymbol{\tau}, A \subseteq B\}.$$

By De-Morgan's law and closer of the topology under arbitrary union we can see that this intersection always gives a closed set. A set is closed if and only if $A = \bar{A}$.

Lemma 1.1. *Suppose $x \in \bar{A}$, then for every neighborhood of x, V_x, we have that $V_x \cap A \neq \emptyset$.*

Proof. Let $x \in \bar{A}$ and suppose for some neighborhood V_x of x we have that $V_x \cap A = \emptyset$. Then we know that $V_x^\circ \cap A = \emptyset$. Take $\tilde{A} = \bar{A} \cap (V_x^\circ)^c$. We can verify that this is a smaller closed set that also contains A. □

Definition 1.7 (Boundary)**.** The boundary of a set A, denoted δA, is $\bar{A} \setminus A^\circ$.

A useful concept when talking about convergence under a topology is that of a neighborhood of a point $x \in X$.

Definition 1.8 (Neighborhood)**.** For a point $x \in X$ a set V is a neighborhood of X if $x \in V^\circ$.

We can now use the topology to define limit points and convergence.

Definition 1.9 (Limit Point)**.** A point $x \in X$ is a limit point of a set $A \subseteq X$ if, for every neighborhood V of x,

$$A \bigcap \left(V \setminus \{x\}\right) \neq \emptyset.$$

In other words, every neighborhood of x intersects with A at a point other than x. Let A' be the set of all limit points of $A \subseteq X$.

Lemma 1.2. *If S is a subset of X, then $\bar{S} = S \cup S'$.*

Proof. First show that $\bar{S} \subseteq S \cup S'$. Let $x \in \bar{S}$. If $x \in S$ then we are done. Otherwise, suppose $x \in \bar{S} \setminus S$. This means that for all V_x we have that $S \cap V_x = S \cap \big(V_x \setminus \{x\}\big)$. By the result of Lemma 1.1, we have that $V_x \cap S \neq \emptyset$. So, $x \in S'$.

Now suppose that $x \in S \cup S'$. Clearly if $x \in S$ then $x \in \bar{S}$. Suppose then that $x \in S' \setminus S$ but $x \notin \bar{S}$. Let $\tilde{S}$ be any closed set containing S, that is $S \subseteq \tilde{S}$. For sake of contradiction, suppose that $x \notin \tilde{S}$ (x is a limit point of S that is not in $\tilde{S}$). Because $\tilde{S}$ is closed we know that $\tilde{S}^c \in \boldsymbol{\tau}$. Further, we know that $x \in \tilde{S}^c$ so that $\tilde{S}^c$ is a neighborhood of x. Since x is a limit point of S, we know that $\tilde{S}^c \cap S = \tilde{S}^c \cap S \setminus \{x\} \neq \emptyset$. However, we also know that $S \subseteq \tilde{S}$ so we have a contradiction. Therefore, it must be that $x \in \bar{S}$ which completes the proof. □

Lemma 1.3 (Characterization of Closed Sets)**.** *A set is closed if and only if it contains all of its limit points.*

Proof. This is a consequence of Lemma 1.2 and the fact that A is closed if and only if $\bar{A} = A$. □

Definition 1.10 (Convergence)**.** We say a sequence $\{x_n\}_{n=1}^{\infty}$ converges to a point $x \in X$ if for every neighborhood V_x of x, there exists a number M such that for all $m \geq M$, $x_m \in V_x$.

Note that under the trivial topology $\boldsymbol{\tau} = \big\{\emptyset, X\big\}$ all sequences converge to any point $x \in X$ whereas under the discrete topology on $\mathbb{R}$, $\boldsymbol{\tau} = 2^{\mathbb{R}}$, no sequence converges.

Definition 1.11 (Continuity)**.** Let $(\mathcal{X}, \boldsymbol{\tau}_1)$ and $(\mathcal{Y}, \boldsymbol{\tau}_2)$ be two topological spaces and $f : \mathcal{X} \to \mathcal{Y}$. We say f is continuous if $f^{-1}(A) \in \boldsymbol{\tau}_1$ for all $A \in \boldsymbol{\tau}_2$. That is, a continuous function maps open sets to open sets.

We can now get ready to combine the notions of continuity and convergence coming from a topology with the notions that we are familiar with from metric spaces. First, we need to define the topology generated by a metric.

Definition 1.12 (Generated Topology)**.** Let $\mathcal{A}$ be a collection of subsets of X. The topology generated by $\mathcal{A}$, $\langle \mathcal{A} \rangle$ is the smallest topology that contains $\mathcal{A}$:

$$\langle \mathcal{A} \rangle = \bigcap \{\boldsymbol{\tau} : \mathcal{A} \subseteq \boldsymbol{\tau}\}.$$

We will then define the topology generated by a metric as the topology generated by the collection of open balls $B(x, \epsilon)$.

Definition 1.13 (Open Ball)**.** Let $d(x, y)$ be a metric on a vector space X. For any point $x \in X$ define the open ball of size ϵ around x as:

$$B(x, \epsilon) = \big\{y : d(x, y) \leq \epsilon\big\}.$$

In a metric space, we consider the topology generated by all the open balls $\boldsymbol{\tau}_d = \langle \big\{B(x, \epsilon) : x \in X, \epsilon > 0\big\} \rangle$. In fact, the set of open balls is a basis for this topology, which means that every open set A in $\boldsymbol{\tau}_d$ and any point $x \in A$, there is an open ball B such that $x \in B \subseteq A$.[1]. Many topological properties such as continuity or convergence can be verified by simply confirming the properties for all members of a basis for the topology. This ties together the "epsilon-delta" notions of continuity and convergence with the more topological versions given above.

For the rest of this subsection we will talk about separability and compactness, but give examples using normed-metric spaces instead of talking in generality about the topology.

Definition 1.14 (Dense Subset)**.** A topological space $(X, \boldsymbol{\tau})$ has a dense subset $\mathcal{A}$ if $\bar{\mathcal{A}} = X$. Equivalent, by Lemma 1.2, every point of X is either in $\mathcal{A}$ or is a limit point of $\mathcal{A}$.

[1] In fact, the set of all open balls with rational ϵ is a basis for the topology

Informally, all points in X are either in $\mathcal{A}$ or arbitrarily "close" to $\mathcal{A}$. As an example, in the standard topology on $\mathbb{R}$ generated by the metric $d(x,y) = |x-y|$, the rationals $\mathbb{Q}$ are dense. We also have that, for the set of continuous functions under the sup norm, the set of all polynomials is dense, which means that we can approximate a function arbitrarily well with them.

Definition 1.15 (Seperable Space)**.** We say that a topological space $(X, \boldsymbol{\tau})$ is separable if it has a countable dense subset.

As we went over above, the real line with its standard topology is separable. The $L_p[a,b]$ spaces are also generally separable for $1 \leq p \leq \infty$. However L_∞ is not separable, which will cause issues (this is not the example below).

Example 1.1 (Bounded functions with the sup norm is not seperable)**.** Let $\{f_i\}_{i\in\mathbb{N}}$ be a countable set of functions on $B_\infty[a,b]$. Let $\{q_i\}_{i\in\mathbb{N}}$ be some counting of the rational numbers between a and b. Let $\tilde{f}$ be some function that is equal to 0 except on the rational numbers. For each rational number q_i define

$$\tilde{f}(q_i) = \begin{cases} 1 & \text{if } f_i(q_i) \leq 0 \\ -1 & \text{if } f_i(q_i) > 0 \end{cases}.$$

We can see that $\tilde{f}$ is bounded (and integrates to 0), but it is at least distance one from each function in $\{f_i\}_{i\in\mathbb{N}}$.

Initially I thought this example would work for $L_\infty[a,b]$, but this only forces a difference on a set of measure 0 and I believe L_∞ works with an essential supremum norm.

Another important/useful concept is that of compactness. The general notion is given below:

Definition 1.16 (Compact Set)**.** A set A is compact if for every collection of open sets $\{G_i\}$ such that $A \subset \bigcup G_i$, there is a finite subcollection that also covers A.

Example 1.2. The real-line is not compact. Consider the open cover $\big\{(n, n+1) | n \in \mathbb{Z}\big\}$

Example 1.3. The interval $(0,1]$ is not compact. Consider the open cover $\big\{(1/n, 1+1/n) | n \in \mathbb{N}\big\}$

Theorem 1.1 (Heine-Borel)**.** *For a subset S of the Euclidean Space*[2]*, $\mathbb{R}^n$, the following statements are equivalent:*

- *S is closed and bounded*
- *S is compact*

Compactness is nice because of various extreme value theorems that ensure that a supremum or infimum is attained. Heine-Borel gives a nice way of characterizing compactness for Euclidean Spaces, but in general there is no equivalent result for general metric spaces. We have to strengthen the boundedness assumption.

Definition 1.17 (Totally Bounded)**.** A set $\mathcal{A}$ is totally bounded if for each $\epsilon > 0$ there exists a finite sequence $\{a_1, \dots, a_n\}$ such that for $B_i = \big\{a \in \mathcal{A} : \|a - a_i\| \leq \epsilon\big\}$, $\bigcup_{i=1}^N B_i$ covers A.

Intuition: For any precision ϵ, you can find a finite set of points that describe $\mathcal{A}$ arbitrarily well. (much more demanding in infinite dimensions than just bounded).

Theorem 1.2. *In a complete metric space, the following are equivalent:*

- *$\mathcal{A}$ is a compact subset*
- *$\mathcal{A}$ is closed and totally bounded*
- *Every sequence in $\mathcal{A}$ has a convergent subsequence which converges to a point in $\mathcal{A}$.*

For a compact set T, let $C(T)$ be the set of continuous functions from T to $\mathbb{R}$ equipped with the sup norm. We may want to characterize when a subset K of $C(T)$ is compact.

[2] That is the space $\mathbb{R}^n$ equipped by the topology generated by the standard distance metric

Definition 1.18 (Equicontinuous)**.** A set of functions $K \subseteq C(T)$ is equicontinuous if for every $t_0 \in T$ and $\epsilon > 0$ there is a $\delta > 0$ such that $|f(t) - f(t_0)| < \epsilon$ whenever $\|t - t_0\| < \delta$ **for all** $f \in K$.

This is a bit like to uniformly continuity but adapted a bit to deal with a function space.

Theorem 1.3 (Arzela-Ascoli)**.** *If T is compact, then $K \subseteq C(T)$ is compact (under the sup-norm) if and only if K is bounded and equicontinuous.*

This concludes our discussion of topology and continuity. We now review measurability.

1.3 Probability Spaces and Outer Measure

Definition 1.19 (Sigma Algebra)**.** A collection of subsets $\mathcal{F}$ is a sigma-algebra (or sigma-field) if it contains the whole set and is closed under complement and under countable union.

Definition 1.20 (Borel Sigma Algebra)**.** For any collection of sets $\mathcal{A}$, we call the smallest sigma algebra containing $\mathcal{A}$, $\sigma(\mathcal{A})$, the sigma algebra generated by $\mathcal{A}$. The Borel sigma algebra on a topological space is the sigma algebra generated by all the open sets, $\mathcal{B}(X) = \sigma(\boldsymbol{\tau})$.

The Borel sigma algebra is useful as it makes all continuous functions measurable (defined below).

Definition 1.21 (Probability Space)**.** A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ consisting of a set of elements Ω, a sigma algebra on Ω, $\mathcal{F}$, and a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying:

1. $\mathbb{P}(A) \geq \mathbb{P}(\emptyset) = 0$ [Non-negativity]
2. If $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets then $\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i)$
3. $\mathbb{P}(\Omega) = 1$.

A measurable function between two spaces equipped with sigma algebra's is simply one that maps measurable sets to measurable sets, similar to the definition of a continuous function.

Definition 1.22 (Measurable Map)**.** A function $f : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ is measurable if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$

Lemma 1.4 (Lemma 1.3.1 VdV& W)**.** *The Borel σ-field on a metric space $\mathbb{D}$ is the smallest σ-field that makes all elements of $C_b(\mathbb{D})$ measurable (with respect to the Borel sets on $\mathbb{R}$).*[3]*.*

Proof. For any closed set F, F is the null set $\{x : f(x) = 0\}$ of the continuous, bounded function, $x \mapsto d(x, F) \wedge 1$. Since the singleton $\{0\}$ is a closed set in $\mathbb{R}$ (all metric spaces are Hausdorff), F must be in the sigma algebra on $\mathbb{D}$ to make $d(x, F) \wedge 1$ measurable. Since all the closed sets generate the Borel σ-field (because σ-fields are closed under complement), all Borel sets must be included in the sigma-algebra on $\mathbb{D}$. $\square$

Given this, we can abstractly think about a random variable as a measurable map from a probability space into another measurable space (typically the real-line). Measurability ensures that things like expectations and probabilities of random variables are well defined.

However, measurability becomes a problem when we are dealing with random functions. For example, if X is a map from a probability space to $L_\infty[a, b]$, the Borel-sigma algebra on $L_\infty[a, b]$ is quite large (its not separable). This means that measurable sets in $L_\infty[a, b]$ may not map back to measurable sets on the probability space $\Omega, \mathcal{F}, \mathbb{P}$.

This is a problem because L_∞ is typically a useful space to work in for empirical process theory. So we have to find a way to relax measurability. This means that we work with outer expectations and probabilities:

[3] $C_b(\mathbb{D})$ is the set of all continuous bounded functions from $\mathbb{D} \to \mathbb{R}$, where $\mathbb{R}$ is endowed with the standard topology on the real line

Definition 1.23 (Outer Measure and Inner Measure)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space $T : \Omega \to \mathbb{R}$. Define the outer expectation:

$$\mathbb{E}^\star[T] = \inf\left\{\mathbb{E}[U] : T \leq U, U \text{ is measurable}\right\}.$$

and the inner expectation:

$$\mathbb{E}_\star[T] = \sup\left\{\mathbb{E}[U] : U \leq T, U \text{ is measurable}\right\}.$$

We can use this to define inner and outer probability measures by restricting T to be the indicator function for an arbitrary set B. Inner and outer expectations are generally nicely behaved but they require modified versions of dominated and monotone convergence and Fubini's theorem breaks down.

2 Weak Convergence

We can now talk about weak convergence of random variables. Let X_n be a real-valued random variable with cdf $F_n(t)$ and let X be a random variable with cdf $F(t)$. The typical definition of weak convergence is that $X_n \overset{L}{\to} X$ if $F_n(t) \to F(t)$ pointwise at all continuity points of F. This is not super general for non-real valued random maps.

Theorem 2.1 (Portmanteau)**.** *For real random variables $X_n \overset{L}{\to} X$ is equivalent to:*

- $\mathbb{E}\left[g(X_n)\right] \to \mathbb{E}[g(X)]$ *for all bounded continuous functions.*
- *For all open sets G,* $\liminf \mathbb{P}(X_n \in G) \geq P(X \in G)$.
- *For all closed sets K,* $\limsup \mathbb{P}(X_n \in K) \leq \mathbb{P}(X \in K)$.

This motivates the theory of weak convergence for general metric spaces. Let $\mathbb{D}$ be a complete metric space with metric d. We can equip $\mathbb{D}$ with it's Borel-sigma algebra as defined in Definition 1.20 and a tight probability measure as defined in Definition 2.1. Let $C_b(\mathbb{D})$ be the set of all continuous and bounded real functions on $\mathbb{D}$. If X is a random variable, $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{D}$ then it's law is given $L = \mathbb{P} \circ X^{-1}$.

Definition 2.1 (Tight Probability Measure)**.** A probability measure is tight if for every $\epsilon > 0$ there is a compact set K_ϵ such that $P(K_\epsilon) \geq 1 - \epsilon$

This is a generalization of bounded in probability I believe.

Definition 2.2 (Borel Law)**.** For a random variable X, we say that X has a Borel Law L if

$$\mathbb{P}\left(X \in A\right) = \int_A dL.$$

for all Borel sets A.

Given this setup, we can now define weak convergence:

Definition 2.3 (Weak Convergence)**.** Let $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ be a sequence of probability spaces and $X_n : \Omega_n \to \mathbb{D}$. Then we say that $X_n \overset{L}{\to} X$ if:

$$\mathbb{E}^\star\left[f(X_n)\right] \to \mathbb{E}[f(X)].$$

for every $f \in C_b(\mathbb{D})$

We can characterize this convergence using another Portmanteau theorem.

Theorem 2.2 (Portmanteau)**.** *The following are equivalent:*

1. $X_n \overset{L}{\to} X$
2. $\liminf \mathbb{P}_\star(X_n \in G) \geq \mathbb{P}(X \in G)$ *for all open sets G.*
3. $\limsup \mathbb{P}^\star(X_n \in F) \leq \mathbb{P}(X \in F)$ *for every closed set F.*

4. $\lim P(X_n \in B) = P(X \in B)$ *for every Borel set* B *with* $P(X \in \delta B) = 0$.

Question: Is X supposed to have a Borel Law? Otherwise where do open and closed sets get tied into this? Is it from the notion of convergence?

Proof. This proof is in a few steps.

(4) $\implies$ (3): Suppose that $\lim P(X_n \in B) = P(X \in B)$ for every Borel set B with $\mathbb{P}(X \in \delta B) = 0$. Let F be a closed set and let $F^\epsilon = \{x : d(x, F) < \epsilon\}$. The sets δF^ϵ are disjoint for different values of $\epsilon > 0$ (The boundary of this set is $\delta F^\epsilon = \{x : d(x, F) = \epsilon\}$), so at most countably many of them can have nonzero L-measure (otherwise the measure of the entire space would be infinite). Choose a sequence $\epsilon_m \downarrow 0$ with $L(\delta F^{\epsilon_m}) = 0$ for each m (this is possible because only countably many ϵ have $L(F^\epsilon) \neq 0$). For a fixed m, by (4) we have that:

$$\limsup P^\star\left(X_\alpha \in F\right) \leq \limsup P^\star\left(X_\alpha \in \overline{F^{\epsilon_m}}\right) = L\left(\overline{F^{\epsilon_m}}\right).$$

letting $m \to \infty$ gives (3).

(3) $\iff$ (2): Take any closed set F. Its complement F^c is open. If

$$\liminf \mathbb{P}_\star(X_n \in F^c) \geq \mathbb{P}(X \in F^c).$$

Then

$$\begin{aligned} \limsup \mathbb{P}^\star(X_n \in F) &\leq \liminf 1 - \mathbb{P}_\star(X_n \in F^c) \\ &\leq 1 - \mathbb{P}(X \in F^c) \\ &= \mathbb{P}(X \in F) \end{aligned}$$

a symmetric argument shows the backwards direction.

(2)+(3) $\implies$ (4): This is straightforward if we recall that, for any set with $L(\delta B) = 0$ we have that $L(B) = L(\bar{B})$. Then we bound the lim sup by the lim inf:

$$\limsup \mathbb{P}^\star(X \in B) \leq \limsup \mathbb{P}(X \in \bar{B}) \leq \mathbb{P}(X \in \bar{B}) = \mathbb{P}(X \in B) \leq \liminf \mathbb{P}_\star(X_n \in B).$$

which gives (4).

(1) $\implies$ (2): Take any G open and define the sequence of functions:

$$f_m(x) := \min(1, m \cdot d(x, G^c))$$

Notice that $f_m(x) \in C_b(\mathbb{D})$ and $f_m(x) \leq \mathbb{1}\{x \in G\}$. So, for every m we have that

$$\begin{aligned} \liminf \mathbb{P}_\star(X \in G) &= \liminf \mathbb{E}_\star\left[\mathbb{1}\{X \in G\}\right] \\ &\geq \liminf \mathbb{E}_\star\left[f_m(X)\right] \\ &\geq \mathbb{E}[f_m(X)] \end{aligned}$$

since $f_m(x) \uparrow \mathbb{1}\{X \in G\}$ by monotone convergence we get the result in (2).

Question: How do we know from weak convergence that this sequence converges in inner expectation?

By VdV and Wellner, weak convergence implies (is equivalent to) $\liminf \mathbb{E}_\star\left[f(X_n)\right] \geq \mathbb{E}\left[f(X)\right]$ for every bounded, Lipschitz continuous, non-negative f. I think the argument for why this is the case goes: Let $f \geq 0$ be bounded and continuous. Then by weak convergence

$$\limsup \mathbb{E}^\star[-f(X_n)] = \mathbb{E}[-f(X)].$$

Taking negatives will give:

$$\liminf \mathbb{E}_\star[f(X_n)] \geq -\limsup \mathbb{E}^\star[-f(X_n)] = \mathbb{E}[f(X)].$$

In any case, $f_m(X)$ is Lipschitz continuous which gives the result.

(2) $\implies$ (1): (SKETCH)

- Suppose $f(x) \geq 0$ is continuous and bounded
- Approximate it from above and below by indicator functions of open sets.

$\square$

Weak convergence is nice because it gives the continuous mapping theorem.

Theorem 2.3 (Continuous Mapping Theorem)**.** *Let $g : \mathbb{D} \to \mathbb{E}$ be continuous at every point $\mathbb{D}_0 \subseteq \mathbb{D}$. If $X_n \xrightarrow{L} X$ and $\mathbb{P}(X \in \mathbb{D}_0) = 0$ then $g(X_n) \xrightarrow{L} g(X)$.*

Proof. (Without Discontinuity Points): Let $Z_n = g(X_n)$ and $Z = g(X)$. We want to show that $\mathbb{E}^\star \left[f(Z_n)\right] \to \mathbb{E}\left[f(Z)\right]$ for all $f \in C_b(\mathbb{D}; \mathbb{E})$.

$$\lim_{n\to\infty} \mathbb{E}[f(Z_n)] = \lim_{n\to\infty} \mathbb{E}[f(g(X_n))] = \mathbb{E}[f(g(X))] = \mathbb{E}[f(Z)].$$

The main step here is weak convergence of X_n and the stability of $C_b(\mathbb{D}; \mathbb{E})$ under composition.

(With Discontinuity Points, from VdV&W): The set D_g of all points at which g is discontinuous can be written

$$D_g = \bigcup_{m=1}^{\infty} \bigcap_{k=1}^{\infty} \left\{x : \exists y, z \in B(x, 1/k) \text{ with } d_{\mathbb{E}}(g(y), g(x)) > 1/m\right\}.$$

Intuition: Recall that g is continuous at x if for every $m \in \mathbb{N}$ there exists a $k \in \mathbb{N}$ such that[1]

$$y, z \in B(x, 1/k) \implies d_{\mathbb{E}}(g(y), g(z)) < 1/m$$

If the function is not continuous at x you can find a counterexample for some $k, m \in \mathbb{N}$.

Let $G_k^m = \left\{x : \exists y, z \in B(x, 1/k) \text{ with } d_{\mathbb{E}}(g(y), g(x)) > 1/m\right\}$. Every G_k^m is open (if x is in G_m^k the points just around x will be as well so that we can write G_m^k as a union of open balls) so that D_g is a Borel set. For every closed F we then have that:

$$\overline{g^{-1}(F)} \subseteq g^{-1}(F) \cup D_g.$$

By Portmanteau:

$$\begin{aligned}
\limsup \mathbb{P}^\star\left(g(X_n) \in F\right) \leq \limsup P^\star\left(X_n \in \overline{g^{-1}(F)}\right) &\leq \mathbb{P}\left(X \in \overline{g^{-1}(X)}\right) \\
&= \mathbb{P}\left(X \in g^{-1}(F)\right) \\
&= \mathbb{P}\left(g(X) \in F\right)
\end{aligned}$$

Applying Portmanteau again gives weak convergence. $\square$

Example 2.1. Take $\mathbb{G}_n \in L^\infty(\mathbb{R})$:

$$\mathbb{G}_n(t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathbb{1}\{X_i \leq t\} - \mathbb{E}\left[\mathbb{1}\{X \leq t\}\right]\right)$$

and suppose that $\mathbb{G}_n \xrightarrow{L} \mathbb{G}$ where $\mathbb{G}$ is some other element of $L^\infty(\mathbb{R})$. Let $Z : L^\infty(\mathbb{R}) \to \mathbb{R}$ be defined as:

$$Z(f) := \sup_t |f(t)|.$$

this function is continuous. Applying the continuous mapping theorem to Z allows us to build uniform confidence intervals.

[1] Topologically, this is saying that the inverse map of every open neighborhood of $f(x)$ is an open neighborhood of x

Let $\gamma_{1-\alpha}$ be the $1-\alpha$ quantile of $Z := \sup_t |\mathbb{G}(t)|$ and construct a confidence interval (at each t):

$$\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \le t\} - \gamma_{1-\alpha}/\sqrt{n}\,,\ \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \le t\} + \gamma_{1-\alpha}/\sqrt{n}\right].$$

Then:

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \le t\} - \gamma_{1-\alpha}/\sqrt{n} \le \mathbb{E}\left[\mathbb{1}\{X \le t\}\right] \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{X_i \le t\} + \gamma_{1-\alpha}/\sqrt{n}\ :\ \text{ for all } t\right)\\
&= \mathbb{P}\left(\left|\mathbb{G}_n(t)\right| \le \gamma_{1-\alpha}\ \forall t\right)\\
&= \mathbb{P}\left(\sup_t |\mathbb{G}_n(t)| \le \gamma_{1-\alpha}\right)
\end{aligned}$$

But by continuous mapping theorem and Portmanteau, if $\mathbb{P}(\sup_t |\mathbb{G}| = \gamma_{1-\alpha}) = 0$:

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_t \left|\mathbb{G}_n(t)\right| \le \gamma_{1-\alpha}\right) = \mathbb{P}\left(\sup_t \left|\mathbb{G}(t)\right| \le \gamma_{1-\alpha}\right) = 1-\alpha.$$

This sort of argument can be applied more generally to functions $\mathbb{G}_n(t) = \hat{m}(t) - m(t)$ to construct uniform confidence intervals.

This shows the usefulness of Portmanteau and Continuous Mapping Theorem. For finite dimension vectors we can use the central limit theorem to establish weak convergence to a normal distribution. However, when X_n is a random element in L^∞ it may be harder to show that $X_n \rightsquigarrow X$ for some other $X \in L^\infty$.

- Don't want to check $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all $f \in C_b(L^\infty)$ [There are at least 20 functions in this class]

Instead we will try to use the structure of L^∞ to show the result.

Definition 2.4 (Asymptotic Tightness)**.** A sequence X_n of random maps is asymptotically tight if for every $\epsilon, \delta > 0$ there is a compact K_ϵ such that

$$\liminf P_\star\left(X_n \in K_\epsilon^\delta\right) \ge 0.$$

where $K_\epsilon^\delta = \{y \in \mathbb{D} : d(y, K_\epsilon) < \delta\}$ is the "δ-enlargement" around K_ϵ.

Definition 2.5 (Asymptotic Measurability)**.** A sequence X_n of random maps is asymptotically measurable if for all $f \in C_b(\mathbb{D})$:

$$\mathbb{E}^\star f(X_n) - \mathbb{E}_\star f(X_n) \to 0.$$

We would like for a sequence X_n that weakly converges to an element X to inherit some properties from X:

Lemma 2.1 (Lemma 1.3.8 VdV& W)**.** *The following are true:*

1. *If $X_n \xrightarrow{L} X$ then X_n is asymptotically measurable*
2. *If $X_n \xrightarrow{L} X$ then X_n is asymptotically tight if and only if X is tight.*

Proof. (1): Take any function $f \in C_b(\mathbb{D})$. By definition of weak convergence we know that

$$\lim \mathbb{E}^\star\left[f(X_n)\right] = \mathbb{E}[f(X)] \quad\text{and}\quad \lim \mathbb{E}^\star\left[-f(X_n)\right] = \mathbb{E}[-f(X_n)].$$

I think we should have that $-\mathbb{E}_\star\left[f(X_n)\right] \ge \mathbb{E}^\star\left[-f(X_n)\right]$ for any f which give the result (I think this holds with equality but I leave it as an inequality since this is all we need for the result).

(2): Fix $\epsilon > 0$. If X is tight then there is a compact K with $\mathbb{P}(X \in K) > 1 - \epsilon$. By Portmanteau:

$$\liminf \mathbb{P}_\star(X_n \in K^\delta) \geq \mathbb{P}(X \in X^\delta).$$

which is larger than $1 - \epsilon$ for every $\delta > 0$.

Conversely, suppose that X_n is asymptotically tight. Then there exists a compact K with $\liminf P_\star(X_n \in K^\delta) \geq 1 - \epsilon$. By Portmanteau,

$$1 - \epsilon \leq \liminf \mathbb{P}_\star(X_n \in K^\delta) \leq \limsup \mathbb{P}^\star(X_n \in \overline{K^\delta}) \leq \mathbb{P}\left(X \in \overline{K^\delta}\right).$$

Let $\delta \to 0$ by monotone convergence to complete the result. [2] ∎

The converse is not generally true. Let $X_n = -1$ if n is odd and $X_n = 1$ if n is even. This sequence is asymptotically measurable and asymptotically tight but clearly does not converge. However, it does converge among a subsequence. This is the idea behind the partial converse to this theorem provided by Pohorov's Theorem.

Theorem 2.4 (Pohorev's Theorem, Theorem 1.3.9 VdV& W)**.** *Let X_n be an asymptotically tight and asymptotically measurable sequence. Then there is a subsequence X_{n_j} that converges weakly to a tight Borel law.*

Now a review problem

Example 2.2 (Problem 7; Ch 1.3 VdV& W)**.** Let X_n be a sequence of random elements in $\mathbb{D}$ and $g : \mathbb{D} \to \mathbb{E}$ a continuous function. Want to show that:

1. If X_n is asymptotically tight then $g(X_n)$ is asymptotically tight.
2. If X_n is asymptotically measurable then $g(X_n)$ is asymptotically measurable.

Proof. 1) Suppose that X_n is asymptotically tight. Fix $\epsilon > 0$. We know that there exists a compact set K such that, $\forall \delta_1 > 0$

$$\liminf \mathbb{P}_\star\left(X_n \in K^{\delta_1}\right) \geq 1 - \epsilon.$$

The event $\left\{X_n \in K^{\delta_1}\right\}$ is a subset of the event that $\left\{g(X_n) \in g(K^{\delta_1})\right\}$ so

$$\liminf \mathbb{P}_\star\left(g(X_n) \in g(K^{\delta_1})\right) \geq \liminf \mathbb{P}_\star\left(X_n \in K_1^\delta\right) \geq 1 - \epsilon.$$

To finish recall that $g(K)$ is a compact set and choose δ_1 such that $g(K^{\delta_1}) \subseteq g(K)^\delta$ (always possible to do so by continuity of g).

2) Suppose that X_n is asymptotically measurable. This means that, for any $f \in C_b(\mathbb{D})$:

$$\mathbb{E}^\star\left[f(X_n)\right] - \mathbb{E}_\star\left[f(X_n)\right] \to 0.$$

Let $\tilde{f} \in C_B(\mathbb{E})$. For any continuous $g : \mathbb{D} \to \mathbb{E}$, $f \circ g$ is a continuous and bounded function from $\mathbb{D} \to \mathbb{R}$. This completes the proof. ∎

[2] This proof relies on compact sets being closed in metric spaces. The proof of this is as follows: Let A be compact in a metric space. We wish to show that A is closed. Take a point $x \in X \setminus A$. To show that A is closed, we want to show that there is an open neighborhood of x that is not in A (this will show that A contains all of its limit points). For every $a \in A$, let $U_a = B\left(a, \frac{d(a,x)}{2}\right)$ and $V_a = B\left(x, \frac{d(a,x)}{2}\right)$. By triangle inequality, U_a and V_a are disjoint. The union of all the sets U_a for all points $a \in A$ is an open cover of A. By compactness of A, we can get a finite subcover $U_{a_1}, \ldots, U_{a_n}$. But then $V_{a_1} \cap \cdots \cap V_{a_n}$ is an open neighborhood of x that is disjoint from A. So A is closed. Actually this argument holds in general Hausdorff spaces.

2.1 Weak Convergence in Space of Bounded Functions

So far, we have defined weak convergence. But, how do we show that $X_n \xrightarrow{L} X$? In $\mathbb{R}^K$ we have the central limit theorem, but no direct analog for random maps into L^∞.

First, some definitions.

Definition 2.6 (Marginal Random Variable)**.** Let X_n be a random map into $L^\infty(T)$ (the space of all bounded functions from $T \to \mathbb{R}$). Then, $X_n(t)$ is the marginal distribution of X_n at t. We can view $X_n(t)$ as the composition of X_n with π_t or directly as a real-valued random variable.

A general strategy will be to deal with the marginals directly. By the central limit theorem, we have conditions for the weak convergence of $X_n(t)$. Want to know what these results imply for the random map X_n.

Lemma 2.2 (Lemma 1.5.1, VdV&W)**.** *Let $X_n : \Omega \to L^\infty(T)$ be asymptotically tight. Then it is asymptotically measurable if and only if $X_n(t)$ is asymptotically measurable for every $t \in T$.*

Lemma 2.3 (Lemma 1.5.3, VdV&W)**.** *Let X and Y be tight Borel measurable maps into $L^\infty(T)$. Then $X \overset{L}{=} Y$ if and only if $X(t) \overset{L}{=} Y(t)$ for all $t \in T$.*

Theorem 2.5 (Theorem 1.5.4, VdV&W)**.** *Let $X_n : \Omega_n \to L^\infty(T)$ be arbitrary. Then X_n weakly converges to a tight limit if and only if X_n is asymptotically right and the marginals $\big(X_n(t_1), \dots, X_n(t_k)\big)$ converge weakly to a limit for every finite subset $t_1, \dots, t_k$.*

Proof. Forward direction is simple, backwards direction requires more work:

($\Longrightarrow$) Suppose that $X_n \xrightarrow{L} X$ and X is tight. By Lemma 2.1, this means that X_n is asymptotically tight. Let $T_k : L^\infty(T) \to \mathbb{R}^k$ be the projection onto the coordinates $t_1, \dots, t_K$. This is a continuous function so by continuous mapping theorem we have convergence of the marginals for any finite collection.

($\Longleftarrow$) Suppose that X_n is asymptotically tight and the marginals converge. Then, by Lemma 2.2, X_n is asymptotically measurable. By Pohorov's theorem, there is a subsequence $X_{n_k} \xrightarrow{L} X$ for some X. Suppose $X_n \overset{L}{\not\to} X$. Then, there is a subsequence $X_{n'_k}$ that stays away from X (in law). However, the marginals converge. This means that the marginals of Y are the same as the marginals of X. By Lemma 2.3, $X \overset{L}{=} Y$. □

Intuition: Why is Tightness + Convergence of Marginals Enough?

- Tightness: $P(X \in K) \geq 1 - \epsilon$ for some <u>*compact*</u> set K.
  - In a metric space, compact means for any $\epsilon > 0$ there are a finite set of points that approximate the whole set well.
    - But! For a finite set of points we have convergence of marginals

Showing convergence of marginal distributions is straightforward by CLT. Next, we cover how to show tightness. Then Theorem 2.5 gives convergence of the entire process.To verify tightness we want a better description than the definition of asymptotic tightness. Two approaches

1. Finite Approximation $\to$ simpler
2. Arzela-Ascoli Theorem $\to$ larger interest (asymptotic equicontinuity)

2.1.1 Finite Approximation

The general idea here is that, for any $\epsilon > 0$, we can partition the index set T (as in $\ell^\infty(T)$) into a finite number of sets T_i so that the variation in each set is $< \epsilon$. Formally, for any $\eta > 0$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\max_i \sup_{s,t\in T_i} \left|X_n(s) - X_n(t)\right| > \epsilon\right) < \eta.$$

Intuition: Why should we expect this to work?

- Tightness means that you concentrate on a compact set
    - Compact set is well described by a finite # of functions

Theorem 2.6 (Theorem 1.5.6 VdV&W)**.** *A sequence of random maps $X_n \in \ell^\infty(T)$ is asymptotically tight if and only if $X_n(t)$ is asymptotically tight in $\mathbb{R}$ for every t and, for all $\epsilon, \eta > 0$ there is a partition $T = \cup_{i=1}^n T_i$ such that*

$$\limsup_{n\to\infty} \mathbb{P}^\star\left(\max_i \sup_{s,t\in T_i} \left|X_n(s) - X_n(t)\right| > \epsilon\right) < \eta \tag{FA-1}$$

Proof. Cover sufficiency. Necessity follows from Theorem 1.5.7 in Van DerVaart and Wellner. Suppose that (FA-1) holds. Fix $\epsilon > 0$ and let the partition $T = \bigcup_{i=1}^k T_i$ satisfy (FA-1) for some $\eta > 0$. We want to show that $\sup_t \left|X_n(t)\right|$ is asymptotically tight. Then:

$$\begin{aligned}
\limsup \mathbb{P}^\star\left(\sup_{t\in T}\left|X_n(t)\right| > M\right) &\leq \limsup \mathbb{P}^\star\left(\sup_{t\in T} > M, \text{ and (FA-1) holds }\right) \\
&\quad + \limsup \mathbb{P}^\star\big(\text{(FA-1) doesn't hold}\big) \\
&\leq \limsup \mathbb{P}^\star\left(\max_{1\leq i\leq k}\left|x_n(t_i)\right| + \epsilon > M\right) + \eta
\end{aligned}$$

Where in the last line we use the bounded variation within each set T_i and pick some arbitrary elements $t_i \in T_i$. Now note that each $X_n(t_i)$ is asymptotically tight by assumption so that $\max_{1\leq i\leq k_i}\left|X_n(t_i)\right|$ is asymptotically tight.[3]. This means that we can pick M so that

$$\limsup \mathbb{P}^\star\left(\sup_t \left|X_n(t)\right| > M\right) < \eta.$$

or, to put this another way, for every $\eta > 0$ we can show that there is an M such that:

$$\limsup \mathbb{P}^\star\left(\sup_t \left|X_n(t)\right| > M\right) < \eta.$$

So we have shown that $\sup_t \left|X_n(t)\right|$ is bounded in probability. Since $\sup_t \left|X_n(t)\right|$ is a map onto the real line, bounded in probability coincides with asymptotic tightness (Heine-Borel).

Now we want to construct a candidate compact set K for the process X_n. Fix $\zeta > 0$ and a sequence $\epsilon_n \downarrow 0$. First, pick an M such that

$$\limsup_{n\to\infty} \mathbb{P}^\star\left(\sup_t \left|X_n(t)\right| > M\right) < \zeta.$$

[3] Couple of quick arguments to get this one:

1. If each $X_{i,n}$ in $\{X_{i,n}\}_{i=1}^K$ is asymptotically tight then the vector $\begin{bmatrix}X_1 & \ldots & X_K\end{bmatrix}$ is asymptotically tight. This is because the Cartesian product of a finite number of compact sets is compact (with respect to the product topology).
2. If X_n is asymptotically tight and g is a continuous function then $g(X_n)$ is asymptotically tight. This is shown in Example 2.2 and basically follows from the fact that a continuous function applied to a compact set yields a compact set. The maximum operator is continuous.

we know such an M exists by the above argument. For each ϵ_m partition $T = \bigcup_{i=1}^{K(m)} T_i$ such that

$$\limsup_{n\to\infty} \mathbb{P}^\star \left(\sup_{1\le i\le K(m)} \sup_{s,t\in T_i} \left|X_n(s) - X_n(t)\right| > \epsilon_m \right) < \frac{\zeta}{2^m}.$$

For each ϵ_m let $\{z_1, \ldots, z_{p(m)}\}$ be the set of functions in $\ell^\infty(T)$ that are constant on T_i and only take values $0, \pm\epsilon_m, \pm 2\epsilon_m, \ldots, M$. It is only important for now that, for any m, $p(m)$ is finite (though large). Let

$$K_m = \bigcup_{i=1}^{p(m)} \overline{B}(z_i, \epsilon_m).$$

where $\overline{B}(z_i, \epsilon_m)$ is the closed ball of radius ϵ_m around z_i. Note that if $\sup_t \left|X_n(t)\right| \le M$ and

$$\sup_{1\le i\le k(m)} \sup_{s,t\in T_i} \left|X_n(s) - X_n(t)\right| \le \epsilon_m$$

then $X_n \in K_m$. Let $K = \bigcap_{m=1}^\infty K_m$. Then K is closed and totally bounded. Closure follows because each K_m is closed (finite union of closed sets) and an arbitrary intersection of closed sets is closed (because the arbitrary union of open sets is open). To see totally bounded fix $\delta > 0$. Then for each $\epsilon_m < \delta$ we have that $K_m = \bigcup_{i=1}^{p(m)} \bar{B}(z_i, \epsilon_m)$. Since $K_m \supset K$ these balls cover K.

We now have a candidate K. We now want to show that, for every $\delta > 0$, $K^\delta \supset \bigcap_{i=1}^m K_i$ for some m. Suppose not. Then there is a sequence $\{z_m\}$ with $z_m \notin K^\delta$ and $z_m \in \bigcap_{i=1}^m K_i$ for every m.[4] This sequence has a subsequence contained in one of the balls making up K_1, this subsequence in one of the balls in K_1 has a further subsequence contained in one of the balls making up K_2, that subsequence contains a subsequence eventually contained in K_3, and so on. [5] Consider the "diagonal" sequence formed by taking the first element of the first subsequence, the second element of the second sequence, and so on. Eventually, this would be contained in a ball of radius ϵ_m for any m.[6]Because $\epsilon_m \downarrow 0$ this means the sequence is Cauchy. Since $\ell^\infty(T)$ is a complete (Banach) space this sequence converges and must converge to an element in K. This contradicts the fact that $d(z_m, K) \ge \delta$ for every m.

Finally, combining our previous results, we want to show that $\liminf \mathbb{P}_\star\left(X_n \in K^\delta\right) \ge 1 - 2\zeta$. for every $\delta > 0$. This is equivalent to saying that $\limsup \mathbb{P}^\star\left(X_n \notin K^\delta\right) < 2\delta$. Recall that

$$\sup_t \left|X_n(t)\right| \le M \quad \text{and} \quad \sup_i \sup_{s,t\in T_i} \left|X_n(s) - X_n(t)\right| \le \epsilon_m \implies X_n \in K_m.$$

Then, to show asymptotic tightness:

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}^\star\left(X_n \notin \bigcup_{i=1}^n K_i\right) &\le \limsup \mathbb{P}^\star\left(X_n \notin \bigcup_{i=1}^m K_i; \sup_t \left|X_n(t)\right| \le M\right) + \underbrace{\limsup \mathbb{P}^\star\left(\sup_t \left|X_n(t)\right| > M\right)}_{<\zeta} \\
&\le \limsup \mathbb{P}^\star\left(\sup_i \sup_{s,t\in T_i} \left|X_n(s) - X_n(t)\right| > \epsilon_m \text{ for some } m\right) + \zeta \\
&\le \sum_{j=1}^m \limsup \mathbb{P}^\star\left(\sup_i \sup_{s,t\in T_i} \left|X_n(s) - X_n(t)\right| > \epsilon_j\right) + \zeta \\
&\le \sum_{j=1}^m \frac{\zeta}{2^j} + \zeta \\
&< 2\zeta
\end{aligned}$$

□

[4]Pick $z_m \in \bigcap_{i=1}^m K_i \setminus K^\delta$

[5]Why? Each $\{z_m\}$ is in $\bigcap_{i=1}^m K_i$. Fix some n, then eventually the sequence is contained in $\bigcap_{i=1}^n K_n$ and so is contained in K_n since $K_n \supset \bigcap_{i=1}^n K_n$. This means the sequence $\{z_m\}$ has infinite members in K_n. K_n is the union of a finite number of sets, so one of these sets must contain infinite members

[6]Key here is the boundedness of the functions we are considering.

Proof is involved but useful as it shows the equivalence between asymptotic tightness and a finite approximation notion. The proof also builds some intuition for why tightness is important, at each step we are essentially showing that the whole behavior of the set is well describes (up to a tolerance of size ϵ) by a finite set of marginals. Weak convergence of the marginals is much easier to show.

This being said, the condition in Theorem 2.6 is hard to check. In particular, there is no guidance given on how to select the partition $\{T_i\}_{i=1}^m$. The next way to characterize tightness builds on asymptotic equicontiuity. The idea is the correct way to pick the partition is linked to some form of continuity: pick small T_i so that X_n does not move much on T_i.

Definition 2.7 (Asymptotic ρ-equicontinuity in probability)**.** Suppose ρ is a semimetric on T. Then a sequence of maps $X_n : \Omega_n \to \ell^\infty(T)$ is asymptotically ρ-equicontinuous if for every $\epsilon, \eta > 0$ there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}^\star \left(\sup_{d(s,t)<\delta} \left| X_n(s) - X_n(t) > \epsilon \right| \right) < \eta.$$

Remark. This is basically setting $T_i = \{(s,t) : p(s,t) < \delta\}$

Example. Let $X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\mathbb{1}\{X_i \le t\} - \mathbb{P}(X \le t) \right]$. Then $\left| X_n(t) - X_n(t') \right| \approx 0$ for all $|t - t'| < \delta$. Note that here, for every n, $X_n(t)$ is still a discontinuous function of t, it's just that the jumps get closer together or smaller.

Example. Suppose that $\gamma = g(X, \beta_0) + \epsilon$ with $\mathbb{E}[\epsilon|X] = 0$. By the vector LLN, we can say that $\hat\beta - \beta_0 \to_p 0$.

In contrast, *asymptotic equicontinuity* will allow to say that:

$$\hat\beta \to_p \beta_0 \implies \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(g(x_i, \hat\beta) - \mathbb{E}[g(x, \hat\beta)] \right) - \left(g(x_i, \beta_0) - \mathbb{E}\left[g(x, \beta_0) \right] \right) \right\} \right| = o_p(1).$$

which is a more powerful result.

Theorem 2.7 (Theorem 1.5.7 Vdv&W)**.** *A sequence of random maps, $X_n : \Omega_n \to \ell^\infty(T)$ is asymptotically tight if and only if $X_n(t)$ is asymptotically tight in $\mathbb{R}$ for each t and there exists a semimetric ρ on T such that (T, ρ) is totally bounded and X_n is asymptotically uniformly ρ-equicontinuous.*

Proof. First prove sufficiency then neccesity:

($\Longleftarrow$) Fix $\epsilon, \eta > 0$. Then, there is a $\delta > 0$ such that

$$\limsup \mathbb{P}^\star \left(\sup_{\rho(s,t)<\delta} \left| X_n(s) - X_n(t) \right| > \epsilon \right) < \eta.$$

Since T is totally bounded, then there are finitely many balls of radius δ that cover T, $B_1, \ldots, B_{K(\delta)}$. Make these balls disjoint by taking succesive "set-minuses" and then we have a partition of T. Then

$$\limsup \mathbb{P}^\star \left(\max_i \sup_{s,t\in T_i} \left| X_n(s) - X_n(t) \right| > \epsilon \right) \le \limsup \mathbb{P}^\star \left(\sup_{\rho(s,t)<\delta} \left| X_n(s) - X_n(t) \right| > \epsilon \right) < \eta$$

and we can apply the results of Theorem 2.6.

($\Longrightarrow$) If X_n is asymptotically tight, then $g(X_n)$ is asymptotically tight for each continuous function g. Let $K_1 \subset K_2 \subset \ldots$ be compact sets with:

$$\liminf \mathbb{P}_\star \left(X_n \in K_m^\epsilon \right) \ge 1 - 1/m.^7$$

[7] We can choose nested compact sets with this property because the union of a finite number of compact sets is compact and the probability functional is increasing with respect to the subset ordering.

For each m define a semimetric ρ_m on T by:

$$\rho_m(s,t) = \sup_{z \in K_m} \left| z(s) - z(t) \right| .$$

Then (T, ρ_m) is totally bounded. How? Cover K_m by finitely many balls of arbitrarily small radius η centered at $z_1, \dots, z_k$.[8] Partition $\mathbb{R}^k$ into cubes of edge η and for every cube pick at most one $t \in T$ such that $\big(z_1(1), \dots, z_k(t)\big)$ is in the cube. Since $z_1, \dots, z_k$ are uniformly bounded,[9] this gives finitely many points $t_1, \dots, t_p$. Now, the balls $\{t : p_m(t, t_i) < 3\eta\}$ cover T: t is in the ball around t_i for which $\big(z_1(t), \dots, z_k(t)\big)$ and $\big(z_1(t_i), \dots, z_k(t_i)\big)$ fall in the same cube. This in turn follows from the fact that $\rho_m(t, t_i)$ can be bounded by $2 \sup_{z \in K_m} \inf_i \|z - z_i\|_T + \sup_j \left| z_j(t_i) - z_j(t) \right|$ [10]. (COMPLETE THIS) □

Remark. Important not to forget the totally bounded part of the theorem. For example, in the example of the emprical CDF case, we need to show that $\mathbb{R}$ is totally bounded. The good news is we have choice of semi-metric.

Remark (Connection to Arzela-Ascoli)**.** Arzela-Ascoli: Let T be a set with metric ρ that is compact. Tet $C(T)$ be the set of all real valued continous functions on T. Then $A \subset C(T)$ is compact under $|\cdot|_\infty$ if and only if it is equicontinuous and bounded.

We can think of Theorem 2.7 as a stochastic version of this. That is for

$$\liminf \mathbb{P}_\star \left(\sup_{p(s,t) < \delta} \left| X_n(s) - X_n(t) \right| \le \epsilon \right) \ge 1 - \eta.$$

The set of functions satisfying this condition is equicontinuous. So then, if X_n falls here it is in a compact set by Arzela-Ascoli (Theorem 1.3). Showing this is a focus later.

[8] This is possible by compactness. Cover K_m by balls of radius η and then take a finite subcover.

[9] Recall that each z_i is in $\ell^\infty(T)$ which is the space of all bounded functions from $T \to \mathbb{R}$. A finite connection of bounded functions is uniformly bounded

[10]