THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

DEPARTMENT OF STAT. AND OR

REFRESHER COURSE, SUMMER 2020

Real Analysis

Original Author:

Oleg MAYBA

(UC Berkeley, 2006)

Last revised:

Summer, 2020

BASED ON THE NSF SPONSORED (DMS GRANT NO 0130526) VIGRE BOOT CAMP LECTURE NOTES IN THE DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, BERKELEY

August 5, 2020

Contents

Introduction

These notes are for use in the warm-up camp for incoming NC (Bio) Statistics and Operations Research graduate students. The analysis review will prepare you for the first year courses. These notes will cover some of the very basics of classical real analysis, and then some extra material which will be especially useful for statisticians and those interested in probability. The notes are modifications of previous notes used at Berkeley.

The original presentation borrowed heavily from 'Real Mathematical Analysis' by C.C. Pugh, but the material now draws from a variety of sources. Many examples will be phrased in terms of the real line, but since real-life research often involves more exotic spaces, this review will attempt to introduce some degree of generality and give examples that a working statistician or probabilist might encounter.

For this latest edition of the analysis notes, we have revised the material somewhat dramatically based on feedback from prior years. We have eschewed more material from basic undergraduate analysis and have beefed up the sections of probability spaces and measure theory as they are a recurring theme in the more theoretical courses here at UNC. Some additional material regarding function spaces has been added, along with several examples motivating some of the fundamental concepts of measure theory.

We hope you find these notes useful! Go Heels!

1 Introduction to metric spaces

1.1 Real line preliminaries

We will denote by $\mathbb{R}, \mathbb{Q}, \mathbb{Z}$, and $\mathbb{N}$ the sets of all real numbers, rational numbers, integers and positive integers, respectively. Recall the notions of finite, countably infinite, and uncountably infinite sets. These notions will be discussed later in this chapter.

The real number line is the most basic and important space to be familiar with. There are many reasons for this, not the least of which is that fundamental concepts of real numbers are used to describe much more complex spaces and structures.

> **Definition.** (Upper bound)
>
> Let $S \subset \mathbb{R}$ and $M \in \mathbb{R}$. Then M is an **upper bound** for S if $s \leq M$ for all $s \in S$. We say S is **bounded above** by M. If a set is bounded both from above and below, we say that it is **bounded**.

The lower bound is defined analogously. The notions *unbounded above* and *unbounded* are defined in the natural way, e.g. if for all $N \in \mathbb{R}$ there exists $s \in S$ s.t. $s > N$, then we say that S is *unbounded* above.

Clearly for a given set, if an upper bound exists then it may not be unique. For example the set $[0, 1]$ is upper bounded by 1, but it is also upper bounded by any number > 1. Therefore it will be convenient to define the following.

> **Definition.** (Supremum)

Let $S \subset \mathbb{R}$ be nonempty. Suppose $M^* \in \mathbb{R}$ is an upper bound for S such that $M^* \leq M$ for any upper bound M of S. Then M^* is the **supremum** (sup) of S, written $\sup(S)$.

If an upper bound for S does not exist, then we set $\sup(S) = +\infty$.

The infimum ($\inf(S)$) is defined similarly as the greatest lower bound. For example, if $S = [0, 1]$, then 1 and 4 are upper bounds, and 0, -4 are lower bounds with 0 and 1 being the infimum and supremum, respectively. Note that some authors choose to write "$\sup(S)$ does not exist" instead of $\sup(S) = +\infty$. We may use both interchangeably.

Fact. (Four properties)

1. Let $S \subset \mathbb{R}$. If S is bounded above then $\sup(S)$ exists.
2. Let $S \subset \mathbb{R}$. If $\sup(S)$ exists, then it is unique.
3. Let $S \subset \mathbb{R}$ be nonempty. Then for any $c > 0$, we can find $s \in S$ such that $s \in [\sup(S) - c, \sup(S)]$.
4. Let $A, B \subset \mathbb{R}$ such that $A \subset B$. Assume all supremums and infimums exist. Then $\sup A \leq \sup B$ and $\inf A \geq \inf B$.

And a slightly less obvious one:

Fact. Let A, B be nonempty, bounded subsets of $\mathbb{R}$, and define:

$$A + B = \{a + b : a \in A, b \in B\} \quad \text{and} \quad A - B = \{a - b : a \in A, b \in B\}$$

Then the following hold:

1. $\sup(A + B) = \sup A + \sup B, \quad \inf(A + B) = \inf(A) + \inf(B)$

2. $\sup(A-B) = \sup A - \inf B, \quad \inf(A-B) = \inf(A) - \sup(B)$

Proof. We only show the first identity—the rest are analogous and a straightforward exercise. First note that $A+B$ is upper bounded so $\sup(A+B)$ exists. It is sufficient to show that $\sup A + \sup B \le \sup(A+B)$.

Assume that $\sup A + \sup B > \sup(A+B)$. We obtain a contradiction by showing that there exists $a' \in A$ and $b' \in B$ such that $a'+b' > \sup(A+B)$. Set $c = \sup(A)+\sup(B)-\sup(A+B) > 0$. By the definition of $\sup A$ and $\sup B$, there are certainly points $a' \in A$ and $b' \in B$ such that

$$a' \in [\sup A - \frac{c}{4}, \sup A] \quad \text{and} \quad b' \in [\sup B - \frac{c}{4}, \sup B]$$

and these points satisfy $a' + b' - \sup(A+B) \ge \frac{c}{2} > 0$. $\square$

Finally, the following (sometimes called the *epsilon principle*) will often be useful:

Theorem 1.1. Let $x, y \in \mathbb{R}$. If for all $\epsilon > 0$, $x \le y + \epsilon$, then $x \le y$. Furthermore, if for all $\epsilon > 0$, $|x-y| \le \epsilon$, then in fact $x = y$.

Proof. Assume to the contrary that $x > y$. Then $x - y > 0$ and $\frac{x-y}{2} > 0$, so setting $\epsilon = \frac{x-y}{2}$ violates the assumption that $x \le y + \epsilon$ for all $\epsilon > 0$.

To show the last result, perform the same analysis assuming $x < y$ and combine. $\square$

1.2 Definition of metric space

The concept of a metric space is fundamental to probability. In the simplest case of the real line, the basic properties (e.g. the triangle inequality) are intuitively obvious. But in more complicated spaces, this may not be the case.

Definition. (Metric space)

A **metric space** (M, d) is a set M together with a function $d : M \times M \to \mathbb{R}$ (known as **metric**) that satisfies the following properties. For all $x, y, z \in M$:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0$ if and only if $x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, y) \leq d(x, z) + d(z, y)$ (*triangle inequality*)

3-dimensional Euclidean space, $\mathbb{R}^3$ is a metric space when we consider it together with the Euclidean distance. Since this is how we normally measure distances in real life, it helps to think of an arbitrary metric d as a *distance function* between elements of the set M.

When metric d is understood, we often simply refer to M as the metric space. Many metric spaces are minor variations of the familiar real line. For example, $\mathbb{R}^3$ is a metric space when we consider it together with the Euclidean distance. Similarly, $\mathbb{Q}$ with the Euclidean (absolute value) metric is also a metric space. Other metric spaces are a little less obvious. We give some examples.

Example. (Discrete metric)

Given any arbitrary set M, the **discrete metric** on that set is defined as follows:

$$d(x,y) = \begin{cases} 1, & \text{if } x \neq y \\ \\ 0, & \text{if } x = y \end{cases}$$

Our second example seems somewhat arbitrary, but it turns out that it has several important properties in probability theory:

Example. (The space $\mathbb{R}^\infty$)

Denote by $\mathbb{R}^\infty$ the space of sequences $\boldsymbol{x} = (x_1, x_2, x_3, \dots)$ where $x_i \in \mathbb{R}$ for $i = 1, 2, \dots$. Let b be the metric on $\mathbb{R}$ defined by $b(\alpha, \beta) = 1 \wedge |\alpha - \beta|$. Then a metric on $\mathbb{R}^\infty$ can be defined by:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{\infty} \frac{b(x_i, y_i)}{2^i}$$

A final example shows that we can define a metric on the space of matrices.

Example. (Metric on space of matrices) Let $M_{n,m}(\mathbb{C})$ denote the collection of all $n \times m$ matrices whose entries are complex numbers. For $A, B \in M_{n,m}(\mathbb{C})$, define the metric d by

$$d(A, B) \doteq \operatorname{rank}(A - B).$$

Finally we show (using the definition) that a rather large class of useful spaces are actually also metric spaces.

Example. (Normed vector/linear spaces)

A normed vector space is a vector space V with a norm $\|\cdot\| : V \to \mathbb{R}$ which satisfies, for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and $c \in \mathbb{R}$:

1. $\|\boldsymbol{v}\| = 0$ if and only if $\boldsymbol{v} = \boldsymbol{0}$
2. $\|c \cdot \boldsymbol{v}\| = |c| \cdot \|\boldsymbol{v}\|$
3. $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$

Then $d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|$ defines a metric on V (check for yourself).

The key about metric spaces is that, by satisfying just the four axioms, we induce a whole plethora of other results and properties which we shall discuss in more detail later on. As just a small taste of this, we prove the *reverse triangle inequality* which is often useful:

Theorem 1.2. (Reverse triangle inequality)

Let (M, d) be a metric space. Then for all $x, y, z \in M$,

$$|d(x,z) - d(z,y)| \leq d(x,y)$$

Proof. There are two cases:

1. $d(x,z) - d(y,z) \geq 0$:

   Starting with the ordinary triangle inequality, $d(x,y) + d(y,z) \geq d(x,z)$,

   $$\begin{aligned} d(x,y) &\geq d(x,z) - d(y,z) \\ &= |d(x,z) - d(z,y)| \end{aligned}$$

2. $d(x,z) - d(y,z) < 0$:

Starting with the ordinary triangle inequality $d(x, y) + d(z, x) \geq d(y, z)$,

$$\begin{aligned} d(x, y) &\geq d(y, z) - d(z, x) \\ &= |d(y, z) - d(z, x)| \\ &= |d(x, z) - d(z, y)| \qquad \square \end{aligned}$$

1.3 Finite, Countable, and Uncountable Sets

An important concept in mathematical analysis and the study of probability and statistics is finite sets and the different types of infinite sets.

Definition. Let $\mathbb{N}$ be the set of natural numbers and let $\mathbb{N}_n = \{1, 2, \ldots, n\}$. We say that that a set S is **finite** if there exists a bijective mapping $f : S \to \mathbb{N}_n$ for some n. We say S is **countably infinite** or **countable** if there exists a bijective mapping $f : S \to \mathbb{N}$. If S is not finite or countable we say it is **uncountably infinite** or **uncountable**.

We can think of a finite, countable, and uncountable sets as increasing in size. I.e. even though both countable and uncountable sets are infinite, countable sets are, in a sense, smaller. It follows that if we say a set is "at most countable" then it is either finite for countably infinite.

Definition. A function f defined on the set of natural numbers is called a **sequence**. If $f(n) = x_n$ for $n \in \mathbb{N}$ we denote the sequence as $\{x_n\}$.

From the definition of sequence it is clear that any countable set of numbers can be arranged in a sequence.

Theorem 1.3. Every infinite subset of a countable set A is countable.

Proof. Suppose $E \subset A$ is infinite. Since A is countable, it can be arranged in a sequence $\{x_n\}$. Construct a new sequence $\{n_k\}$ in the following way. Let n_1 be the smallest integer such that $x_{n_1} \in E$. Let n_2 be the second smallest integer such that $x_{n_2} \in E$. Continue in this manner. Then $g(k) = x_{n_k}$ gives a 1-1 correspondence between E and $\mathbb{N}$. Therefore E is countable. $\square$

The following is an important property of countable sets.

Theorem 1.4. Suppose A is at most countable, and, for every $\alpha \in A, B_\alpha$ is at most countable. Then $\cup_{\alpha \in A} B_\alpha$ is at most countable. Note that "at most countable" can be be replaced by countable in this theorem

Proof. See Chapter 2 of Rudin for a proof. $\square$

This property is extremely important and will come up repeatedly in the study of statistics and probability.

Example.
1. $[0, 1]$ is uncountable.
2. The set of rational numbers is countable.
3. The set of all binary sequences is uncountable.

See Chapter 2 of Rudin for proof of 2 and 3.

1.4 Sequences and convergence

Oftentimes it is useful to analyze points in a metric space considered as a sequence. We will use the notation (x_n) or $\{x_n\}$ for the sequence of points $x_1, x_2, \ldots, x_n, \ldots$ in metric space M. The members of a sequence are not assumed to be distinct, thus 1,1,1,1,... is a legitimate sequence of points in $\mathbb{Q}$. A sequence (y_k) is a **subsequence** of (x_n) if there exists sequence $1 \leq n_1 < n_2 < n_3 < \ldots$ such that $y_k = x_{n_k}$.

For example, some subsequences of the sequence 1,1,2,1,2,3,1,2,3,4,1,2,3,4,5,... are:

1. odds: 1,3,5,7,9,...
2. primes: 2,3,5,7,11,...
3. original sequence with duplicates removed: 1,2,3,4,5,...

A fundamental notion in a metric space is that of a limit of a sequence.

> **Definition.** (Convergence of a sequence)
>
> A sequence (x_n) of points in M is said to **converge to** x, if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $n \geq N$ implies $d(x_n, x) < \epsilon$. We call x the **limit** of (x_n) and write $x_n \to x$.
>
> We say a sequence (x_n) in a space M **converges in** M if there exists a point $x \in M$ such that $x_n \to x$.

A subtle but important point in the above defintion is that convergence always happens “in” some space. If the space M that a sequence (x_n) belongs to is understood, then we might simply say "(x_n) converges." Convergent sequences have some very key properties. The first is uniqueness:

Theorem 1.5. If the limit of a sequence exists, then it is unique.

Proof. Let (x_n) be a sequence in M that converges and suppose that $x_n \to x$ and $x_n \to y$. Let $\epsilon > 0$ be given. Then $\exists\, N_1$ such that $d(x_n, x) < \frac{\epsilon}{2}$ for $n \geq N_1$ and $\exists\, N_2$ such that $d(x_n, y) < \frac{\epsilon}{2}$ for $n \geq N_2$. Let $N = \max(N_1, N_2)$. Then for $n \geq N$, the triangle inequality gives

$$d(x, y) \leq d(x, x_n) + d(x_n, y) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Since our choice of $\epsilon > 0$ was arbitrary, this holds for all $\epsilon > 0$ and thus $x = y$ by the epsilon principle proved in the first section. □

Another key result which relates the convergence of the original sequence to convergence of subsequences is:

Theorem 1.6. Every subsequence of a convergent sequence converges, and converges to the same limit as the original sequence.

Proof. Exercise. □

Convergent sequences behave as one would expect, in a variety of ways. One of these ways is their behavior when added or subtracted:

Lemma 1.7. Let (x_n), (y_n) be two convergent real-valued sequences with limits x and y, respectively. Then:

1. $\lim(x_n + y_n) = x + y$

2. $\lim(x_n - y_n) = x - y$

Proof. For (1), fix $\epsilon > 0$ and note that:

$$\begin{aligned} |(x+y) - (x_n - y_n)| &= |(x - x_n) + (y - y_n)| \\ &\leq |x - x_n| + |y - y_n| \end{aligned}$$

Now we can find N_1 such that $|x - x_n| < \frac{\epsilon}{2}$ for all $n \geq N_1$, and N_2 such that $|y - y_n| < \frac{\epsilon}{2}$ for all $n \geq N_2$. Setting $N = \max\{N_1, N_2\}$ proves (1). Then (2) follows from (1) by considering the sequence $(-y_n)$ which converges to $-y$. □

This result may seem trivial, but it has an important corollary: the *squeeze theorem.*

Lemma 1.8. Suppose (x_n) and (y_n) are convergent real-valued sequences with limits x and y, respectively. Also suppose $x_n \leq y_n$ for all n. Then $x \leq y$ as well.

Proof. By the above lemma, $y - x = \lim y_n - \lim x_n = \lim(y_n - x_n)$. Since $y_n - x_n \geq 0$ for all n, then the limit is also ≥ 0.[1] □

Corollary 1.9. (Squeeze theorem)

Suppose (x_n), (y_n) and (z_n) are convergent real-valued sequences with $x_n \leq z_n \leq y_n$ for all n. Suppose that $\lim x_n = \lim y_n = c$. Then $\lim z_n = c$ as well.

The following result can be useful for determining whether a sequence converges.

[1] Actually, to be completely rigorous we need to note that the set $\{x \in \mathbb{R} : x \geq 0\}$ is **closed**. We will return to this later.

Lemma 1.10. Let $\{x_n\}$ be a sequence in a metric space M. Then x_n converges to $x \in M$ if and only if every subsequence $\{x_{n_k}\}$ has a further subsequence $\{x_{n_{k_j}}\}$ that converges to x.

Proof. Suppose that $x_n \to x$ and let $\{x_{n_k}\}$ be a subsequence of $\{x_n\}$. Then $x_{n_k} \to x$, so every subsequence of $\{x_{n_k}\}$ converges to x.

Now, in order to show that the other implication holds, suppose that every subsequence has a further subsequence which converges to x, but that x_n does not converge to x. Since x_n does not converge to x, there is some $\epsilon > 0$ such that for all $N \in \mathbb{N}$, there is some $n \geq N$ such that $d(x_n, x) \geq \epsilon$. So define the sequence $\{n_k\}$ as follows:

$$n_1 \doteq \inf\{n \in \mathbb{N} : d(x_n, x) \geq \epsilon\},$$

and for each $k \geq 2$, let

$$n_k \doteq \inf\{n > n_{k-1} : d(x_n, x) \geq \epsilon\}.$$

Then for all $k \in \mathbb{N}$, $d(x_{n_k}, x) \geq \epsilon$, so $\{x_{n_k}\}$ has no subsequences which converge to x. This is a contradiction, so it follows that $x_n \to x$. $\square$

Sometimes a sequence is not convergent in a space but it satisfies a slightly weaker property which is nevertheless very useful.

Definition. A sequence (x_n) in M is **Cauchy** if for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $n, m \geq N$ implies $d(x_n, x_m) < \epsilon$.

In other words a sequence is Cauchy if eventually all the terms are all very close to each other. These sequences have some nice properties. First of all, they are bounded:

Theorem 1.11. Let (x_n) be a Cauchy sequence in M. Then it is bounded.

Proof. Let $\epsilon > 0$ be fixed. Then since (x_n) is Cauchy we may find N such that $|x_n - x_m| < 1$ for all $n, m \geq N$. Now note that $(x_n)_{n \leq N}$ is a finite set of points, so it is bounded inside some interval $[-c, c]$. Therefore the entire sequence is bounded inside the interval $[-c-1, c+1]$. $\square$

Our next theorem shows that Cauchy-ness is a weaker condition than convergence:

Theorem 1.12. If a sequence is convergent, then it is Cauchy

Proof. Suppose $x_n \to x$ in M. Fix $\epsilon > 0$. Then there exists N such that $d(x_n, x) < \frac{\epsilon}{2}$ for all $n \geq N$. So let $n, m \geq N$. Then by the triangle inequality

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \qquad \square$$

It doesn't take long to see that the converse is not always true. Consider the sequence

$$3, 3.14, 3.141, 3.1415, \ldots$$

This sequence is clearly Cauchy. When considered as a sequence in $\mathbb{R}$, it does converge to π. However, we can also think of it as a sequence in $\mathbb{Q}$ in which case it doesn't converge (in $\mathbb{Q}$), since $\pi \notin \mathbb{Q}$. Another example is the sequence $(1/n)_{n=1,2,\ldots}$, which is clearly Cauchy. In the space $[0, 1]$, this sequence is clearly convergent. However, in the space $(0, 1)$ this sequence is not convergent since $0 \notin (0, 1)$. Spaces in which Cauchy sequences are guaranteed to converge merit a special name:

Definition. A metric space M is **complete** if all Cauchy sequences in M are convergent in M.

We will work with property more in later chapters, as it becomes key in several theorems. For now, we will content ourselves with two small theorems to build intuition:

Theorem 1.13. (Cauchy criterion)

A series $\sum_{n=1}^{\infty} a_n$ in a complete metric space M converges iff for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for any $n > k > N$,

$$\Big|\sum_{j=k}^{n} a_j\Big| < \epsilon$$

Proof. Consider the sequence of partial sums S_n as a Cauchy sequence. □

Theorem 1.14. $\mathbb{R}$ is complete.

Proof. Let (a_n) be a Cauchy sequence in $\mathbb{R}$ and consider the set:

$$S = \{s \in \mathbb{R} : \exists \text{ infinitely many } n \in \mathbb{N} \text{ for which } a_n \geq s\}.$$

Since Cauchy sequences are bouned, S is bounded above and has a supremum $b < \infty$. We will show that $a_n \to b$.

Fix $\epsilon > 0$. Since (a_n) is Cauchy, there exists N such that $|a_n - a_m| < \frac{\epsilon}{2}$ for all $m, n \geq N$. This implies that $a_n \geq a_N - \frac{\epsilon}{2}$ infinitely namy times and $a_n \geq a_N + \frac{\epsilon}{2}$ only finitely many times. So $a_N - \frac{\epsilon}{2} \in S$ and $a_N + \frac{\epsilon}{2}$ is an upper bound for S. Since b is the supremum for S, we must have $a_N - \frac{\epsilon}{2} \leq b$ and $b \leq a_N + \frac{\epsilon}{2}$.

So by the triangle inequality, for all $n \geq N$,

$$\begin{aligned}|b - a_n| &= |(b - a_N) + (a_N - a_n)| \\ &\leq |b - a_N| + |a_N - a_n| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2}.\end{aligned}$$

□

In fact, not only is $\mathbb{R}$ complete but $\mathbb{R}^n$ is complete for any $n \geq 1$. However, we will skip the proof in the interest of time.

Theorem 1.15. $\mathbb{R}^n$ is complete.

1.5 Limsup and liminf

Before introducing the critical concepts of limsup and liminf, we extend our previous notation a bit. Say that $\lim(x_n) = \infty$ if, for all $M > 0$ there exists N such that $n \geq N$ implies $x_n > M$. Define $\lim(x_n) = -\infty$ in the analogous way. To illustrate with some examples:

1. (x_n) = 3, 3.1, 3.14, 3.141, ... Then $\lim(x_n) = \pi$
2. (x_n) = 2, 3, 5, 7, 11, ... Then $\lim(x_n) = \infty$
3. (x_n) = -1, -4, -9, -16, ... Then $\lim(x_n) = -\infty$
4. (x_n) = -1, 1, -2, 2, -3, 3, ... Then $\lim(x_n)$ does not exist.

The key takeaway here is that a sequence can either have a finite limit, a limit at $\pm\infty$, or its limit may not exist at all particularly in the case of oscillatory behavior. However, there is one special case in which the limit will always exist:

Definition. Let (x_n) be a real-valued sequence. Then (x_n) is **monotone increasing** if $n > m$ implies $x_n > x_m$, and **monotone non-decreasing** if $n > m$ implies $x_n \geq x_m$. Monotone decreasing and monotone non-increasing sequences are defined similarly. Finally, a sequence is said to be **monotone** if it's either monotone non-increasing or monotone non-decreasing.

We illustrate this briefly with some examples:

1. $(x_n) =$ 3, 3.1, 3.14, 3.141, ... is monotone non-decreasing
2. $(x_n) =$ 1, 2, 3, 5, 8, 13, ... is monotone increasing
3. $(x_n) =$ -1, -4, -9, -16, ... is monotone decreasing
4. $(x_n) =$ 1, 2, 2, 3, 3, 3, 4, 4, 4, 4, ... is monotone non-decreasing
5. $(x_n) =$ -1, -1, -1, -1, ... is monotone non-decreasing and monotone non-increasing
6. $(x_n) =$ 1, 1, 2, 1, 2, 3, 1, 2, 3, 4, ... is not monotone but it has many possible monotone subsequences

The most important property of monotone sequences in $\mathbb{R}$ is:

Theorem 1.16. Let (x_n) be a real-valued non-decreasing (non-increasing) sequence which is bounded above (below). Then it is convergent in $\mathbb{R}$.

Proof. Assume WLOG that (x_n) is non-decreasing and define $b = \sup_n x_n$. We show $x_n \to b$. Fix some $\epsilon > 0$. By defintion of b, we have that there exists N such that $x_N > b - \epsilon$. But since (x_n) is non-decreasing, then $x_n > b - \epsilon$ for all $n \geq N$. Therefore

$$b - \epsilon < x_n \leq b < b + \epsilon \quad \Rightarrow \quad |x_n - b| < \epsilon \qquad \square$$

Now we arrive at the final definition of this section. We have just shown that for a bounded monotone sequence, its limit will always exist. However recall that for an *arbitrary* real-valued sequence (x_n), its limit may not exist. We now introduce two extremely useful limit concepts which *always* exist or take *one* of the values $+\infty$, $-\infty$:

Definition. (Limit superior/inferior)

The **limit superior** of a sequence (x_n) in $\mathbb{R}$ is defined as:

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} (\sup_{k\geq n} x_k)$$

Similarly, the **limit inferior** of (x_n) is defined as:

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} (\inf_{k\geq n} x_k)$$

Here are some examples:

1. $(x_n) = 1, 2, 3, 4, 5, \ldots$, then lim sup$(x_n)$ = lim inf $(x_n) = \infty$

2. $(x_n) = 1, 1, 2, 1, 2, 3, \ldots$, then lim sup$(x_n) = \infty$, lim inf$(x_n) = 1$

3. $(x_n) = 3, 3.1, 3.14, 3.141, \ldots$, then lim sup$(x_n)$ = lim inf(x_n)= π

The limsup and liminf of a sequence have several important properties, including sub- and super-additivity (see exercises). The following result provides some intution about the $\limsup$ and $\liminf$ of a sequence.

Fact. Let $\{x_n\}$ be a bounded sequence in $\mathbb{R}$. Let

$$x^* \doteq \limsup_{n\to\infty} x_n, \quad x_* \doteq \limsup_{n\to\infty} x_n.$$

Then for each $\epsilon > 0$, there are at most finitely many $n \in \mathbb{N}$ such that $x_n > x^* + \epsilon$. Similarly, for each $\epsilon > 0$, there are at most finitely many $n \in \mathbb{N}$ such that $x_n < x_* - \epsilon$.

The following is the most foundational result tying together lim sup, lim inf, and lim.

Theorem 1.17. Let (x_n) be a real-valued sequence. Then $x_n \to x$ if and only if $\limsup x_n = \liminf x_n = x$.

Proof. First define the two inf/sup sequences:

$$y_n = \sup_{k\geq n} x_k \quad \text{and} \quad z_n = \inf_{k\geq n} x_k$$

($\Rightarrow$) Suppose that $x_n \to x$. For every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $x - \epsilon < x_n < x + \epsilon$ for all $n \geq N$. Then $x - \epsilon \leq z_n \leq y_n \leq x + \epsilon$ for all $n \geq N$. So

$$x - \epsilon \leq \lim z_n \leq \lim y_n \leq x + \epsilon.$$

Therefore $\lim z_n = \lim y_n = x$, namely $\liminf x_n = \limsup x_n = x$.

($\Leftarrow$) Now suppose that $\liminf x_n = \limsup x_n = x$. In other words $z_n \to x$ and $y_n \to x$. But also note that for all n, $z_n \leq x_n \leq y_n$ so that $x_n \to x$ as well by the squeeze theorem. $\square$

To give an example of when liminf/limsup are useful, we prove the following lemma.

Lemma 1.18. Let (x_n) be a real-valued bounded sequence, so that both $a = \limsup x_n$ and $b = \liminf x_n$ are finite. Then there exists a subsequence (x_{n_k}) such that $x_{n_k} \to a$ as $k \to \infty$. Similarly, there exists (possibly another) subsequence converging to b.

Proof. We will prove the first statement by choosing a subsequence (x_{n_k}) such that $|x_{n_k} - a| \leq \frac{1}{k}$. The second statement can be proved similarly. Let $y_n = \sup_{m \geq n} x_m$. Then $y_n \downarrow a$. Let $N_1 \doteq \inf\{n \in \mathbb{N} : a \leq y_n \leq a + 1\}$, and let $n_1 \doteq \inf\{n \geq N_1 : y_{N_1} - 1 \leq x_n \leq y_{N_1}\}$. Now, recursively define the sequences $\{N_k\}$ and $\{n_k\}$ by

$$N_k \doteq \inf \left\{ n > n_{k-1} : a \leq y_n \leq a + \frac{1}{k} \right\}, \quad n_k \doteq \inf \left\{ n \geq N_k : y_{N_k} - \frac{1}{k} \leq x_{n_k} \leq y_{N_k} \right\}.$$

The sequence $\{N_k\}$ is well defined, since $y_n \downarrow a$. That the sequence $\{n_k\}$ is well defined can be argued by contradiction. For each k we have that

$$a - \frac{1}{k} \leq y_{N_k} - \frac{1}{k} \leq x_{n_k} \leq y_{N_k} \leq a + \frac{1}{k},$$

so $|x_{n_k} - a| \leq \frac{1}{k}$. It follows that $x_{n_k} \to a$ as $k \to \infty$. □

The following corollary will be used later when we talk about compactness.

Corollary 1.19. Any bounded real-valued sequence (x_n) has a convergent subsequence.

1.6 Limsup and liminf of sequences of sets

Notions of lim sup and lim inf can be defined for collections of sets as well. In particular, if M is some set, and for each $n \in \mathbb{N}$, $A_n \subseteq M$, then the lim sup and lim inf of the collection of sets $\{A_n\}$ are defined by

$$\liminf_{n\to\infty} A_n \doteq \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m, \quad \limsup_{n\to\infty} A_n \doteq \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

These defintions show up frequently in measure theory and probability. For some intution why, it helps to think carefully about what it means for an element x to belong to $\liminf\limits_{n\to\infty} A_n$

and $\limsup_{n\to\infty} A_n$. If $x \in \liminf_{n\to\infty} A_n$, then there is some $n \in \mathbb{N}$ such that for all $m \geq n$, A_m. Similarly, if $x \in \limsup_{n\to\infty} A_n$, then for each $n \in \mathbb{N}$, there is some $m \geq n$ such that $x \in A_m$. It follows that if $x \in \liminf_{n\to\infty} A_n$, then "$x$ eventually belongs to all A_n". If $x \in \limsup_{n\to\infty} A_n$, then "$x$ belongs to infinitely many A_n". This means that if each A_n represents some event of interest, then

$$\text{“}\mathbb{P}(\text{infinitely many } A_n \text{ occur}) = \mathbb{P}\left(\limsup_{n\to\infty} A_n\right).\text{”}$$

For a set $A \subseteq M$, the **indicator function** of A is the function $1_A : M \to \mathbb{R}$ given by

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

The following result helps explain the connection between the lim sup and lim inf of a sequence and the lim sup and lim inf of a collection of events.

Lemma 1.20. Let $\{A_n\}_{n=1}^{\infty}$ be a collection of subsets of a set M. Then for each $x \in M$,

$$1_{\limsup_{n\to\infty} A_n}(x) = \limsup_{n\to\infty} 1_{A_n}(x).$$

Proof. Exercise. □

1.7 Big-O and little-o notation

Sometimes we are not interested in the exact asymptotic behavior of sequences or functions, but only in some estimates. That is, we are only concerned with the *rate* of growth of a sequence or function. To focus on this, we use special notation for comparing the growth of two functions. The notation for sequences are analogous.

Definition. (Big-O, little-o)

Let f, g be two real-valued functions on $S \subset \mathbb{R}$.

1. Big-O: We say $f(x) = \mathrm{O}(g(x))$ as $x \to a$ if there exists $M < \infty$ and some $\delta > 0$ such that:

$$\left|\frac{f(x)}{g(x)}\right| < M \quad \text{for} \quad |x - a| < \delta$$

   We say $f(x) = \mathrm{O}(g(x))$ as $x \to \infty$ if there exists $M < \infty$ and some x_0 such that

$$|f(x)| \leq M|g(x)| \quad \text{for} \quad x > x_0$$

   If it is clear from the context what a is, we may simply write $f(x) = \mathrm{O}(g(x))$.

2. Little-o: We say that $f(x) = \mathrm{o}(g(x))$ as $x \to a$ if:

$$\lim_{x \to a} \left|\frac{f(x)}{g(x)}\right| = 0$$

   Again, if it is clear from the context what a is, we may simply write $f(x) = \mathrm{o}(g(x))$.

The intuitive meaning of these two terms is essentially the following. For example, if $f(x) = \mathrm{O}(g(x))$, then f eventually exhibits the same rate of growth as g. If $f(x) = \mathrm{o}(g(x))$, then f eventually grows slower than g. This leads us to one final definition that is useful for describing the asymptotic behavior of a function.

Definition. (Asymptotic equivalence) A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be admissible if there is some x_0 such that for $x > x_0$, we have $f(x) > 0$. Consider two admissible functcions $f, g : \mathbb{R}_+ \to \mathbb{R}_+$. We say that $f \sim g$ (as $x \to \infty$) if

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1.$$

An admissible sequence is defined similarly. For two admissible sequences $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, we say $a_n \sim b_n$ (as $n \to \infty$) if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

The following lemma shows that $\sim$ is an equivalence relation. Namely, for admissible functions f, g, h we have the following:

1. $f \sim f$
2. If $f \sim g$, then $g \sim f$.
3. If $f \sim g$ and $g \sim h$, then $f \sim h$.

Lemma 1.21. The relation $\sim$ is an equivalence relation on the set of admissible functions.

Proof. Let f, g, h be admissible functions such that $f \sim g$ and $g \sim h$. We begin by showing that $f \sim f$. Since f is admissible, there is some x_f such that if $x > x_f$, then $f(x) > 0$. Therefore for all $x > x_f$, we have

$$\frac{f(x)}{f(x)} = 1,$$

so

$$\lim_{x\to\infty} \frac{f(x)}{f(x)} = 1.$$

Therefore $f \sim f$. We now show that $g \sim f$. Since g is admissible, there is some $x_g > 0$ such that if $x > x_g$, then $g(x) > 0$. Then for all $x > x_f \vee x_g$, we have

$$\frac{g(x)}{f(x)} = \frac{1}{\frac{f(x)}{g(x)}},$$

which tends to 1 as $x \to \infty$. Therefore $g \sim f$. We now show that $f \sim h$. Let $x_h > 0$ be such that whenever $x > x_h$, we have $f(x) > 0$. Then for all $x > \max\{x_f, x_g, x_h\}$, we have

$$\frac{f(x)}{h(x)} = \frac{f(x)}{g(x)}\frac{g(x)}{h(x)},$$

which tends to 1 as $x \to \infty$. Therefore $f \sim h$. □

We now recall a useful fact from calculus.

Theorem 1.22. Suppose that $f : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous, positive, and decreasing, and let $a_n \doteq f(n)$. Then for each n,

$$\int_1^n f(x)dx \leq \sum_{i=1}^n a_i \leq a_1 + \int_1^n f(x)dx. \tag{1}$$

The following result quantifies the asymptotic behavior of the sequence $\{H_n\}$, where

$$H_n \doteq \sum_{i=1}^n \frac{1}{i},$$

denotes the n^{th} harmonic number. This fact is interesting in its own right and is frequently useful.

Theorem 1.23. There is some constant $\gamma \geq 0$ (called the Euler-Mascheroni constant) such that

$$\lim_{n\to\infty} (H_n - \log n) = \gamma.$$

Proof. Using (1) with $f(x) \doteq 1/x$, we can see that

$$\log n = \int_1^n \frac{1}{x}dx \leq H_n \leq 1 + \int_1^n \frac{1}{x}dx = 1 + \log n.$$

Now, let $\kappa_n \doteq H_n - \log n$, and note that the previous display ensures that $0 \leq \kappa_n \leq 1$. Additionally, $\{\kappa_n\}$ is a decreasing sequence, as

$$\begin{aligned}
\kappa_n - \kappa_{n+1} &= (H_n - \log n) - (H_{n+1} - \log(n+1)) \\
&= \log(n+1) - \log n + \left(\sum_{i=1}^{n} \frac{1}{i} - \sum_{i=1}^{n+1} \frac{1}{i} \right) \\
&= \log(n+1) - \log n - \frac{1}{n+1} \\
&= \int_n^{n+1} \frac{1}{x} dx - \frac{1}{n+1} \\
&= \int_n^{n+1} \left(\frac{1}{x} - \frac{1}{n+1} \right) dx \\
&> 0.
\end{aligned}$$

The final inequality in the previous display follows from the fact that $\frac{1}{x} > \frac{1}{n+1}$ for all $x \in [n, n+1)$. We have shown that $\{\kappa_n\}$ is bounded below (also, it is bounded above) and is decreasing, so its limit exists. This means the quantity

$$\gamma \doteq \lim_{n\to\infty} \kappa_n = \lim_{n\to\infty} (H_n - \log n),$$

is well-defined. $\square$

The first few decimals of the Euler-Mascheroni constant are given by

$$\gamma \approx 0.5772\ldots,$$

but there are many important facts about it that are not yet known. For example, it is not known whether it is irrational. The Eurler-Mascheroni constant appears in a wide range of contexts. For a partial list, see this link. The follow corollary says that $\{\log n\}$ and $\{H_n\}$ are asymptotically equivalent.

Corollary 1.24. As $n \to \infty$, $H_n \sim \log n$.

Proof. Note that

$$\frac{H_n}{\log n} = \frac{H_n - \log n}{\log n} + \frac{\log n}{\log n} = \frac{H_n - \log n}{\log n} + 1,$$

so the result follows from the previous theorem. □

2 Structure of Metric Spaces

2.1 A side note on topology

Previously, we defined metric spaces as a pair of objects: a set of points together with a function which relates pairs of points in that set. This has far-reaching consequences and implies many important properties of that set.

In this section we take a different view of metric spaces by looking at how whole subsets of our spaces interact with each other and behave under certain operations. In a metric space, the fundamental unit of analysis is:

Definition. (Open sets: metric definition)

In a metric space (X, d) the **open ball around** $c \in X$ **of radius** $r > 0$ is the set:

$$B_r(c) = \{x \in X : d(c, x) < r\}$$

Then we say a set $S \subset X$ is **open** if for all points $x \in S$, there exists some $r > 0$ such that $B_r(x) \subset S$.

This definition should be familiar to you. In general, the definition of an open set has nothing to do with a metric.

Definition. (Open sets: topological definition)

Let X be some set. Then a **topology** $\mathcal{T}$ for X is a collection of *subsets* of X which have the following properties:

1. The empty set ($\emptyset$) and the whole space X are in $\mathcal{T}$.
2. Any arbitrary union of sets in $\mathcal{T}$ is in $\mathcal{T}$.
3. Any finite intersection of sets in $\mathcal{T}$ is in $\mathcal{T}$.

Elements of $\mathcal{T}$ are called **open** sets. We refer to the pair $(X, \mathcal{T})$ as a **topological space**. When the topology is clear from context, we sometimes refer to X as a topological space.

The definition of a topology makes clear exactly what we mean by a "structure" of a space: loosely speaking, it is a way of breaking up the space into smaller units (subsets, or open sets) such that the smaller units interact with each other in a clearly-defined way (i.e. they are closed under arbitrary unions and finite intersections).

Example. (Indiscrete topology) Let X be a nonempty set. Let $\mathcal{T} \doteq \{\emptyset, X\}$.

Example. (Discrete topology) Let X be a nonempty set. Let $\mathcal{T} \doteq \{A : A \subseteq X\}$.

In this way of looking at open sets, we have not mentioned a metric at all. In other words, strictly speaking open sets are a notion related to a topology on a space, not a metric. So what does it mean to speak of open sets with respect to a metric?

If there were no connection between open sets w.r.t. a metric and open sets w.r.t. some topology, then the use of "open sets" in both definitions would represent an extremely unfortunate clash of notation. The short answer is that putting a metric on a space actually induces a special topology on that space.

Definition. (Basis for a topology)

Let X be a set. A **basis** $\mathcal{B}$ for a topology $\mathcal{T}$ on X is a collection of open sets with the property that *every* open set in $\mathcal{T}$ can be expressed as a union of sets in $\mathcal{B}$.

From here is it easy to see that:

Corollary 2.1. If $\mathcal{B}$ is a basis for a topology $\mathcal{T}$ on X, then $S \subset X$ is open if and only if for each $x \in S$, there exists $B \in \mathcal{B}$ such that $x \in B$ and $B \subset S$.

Proof. If $S \subset X$ is open, then $S = \cup_{\lambda \in \Lambda} B_\lambda$ for a collection of $B_\lambda \in \mathcal{B}$. Clearly each $B_\lambda \subset S$. So for each $x \in S$, there exists $B_\lambda \in \mathcal{B}$ such that $x \in B_\lambda$ and $B_\lambda \subset S$.

If for each $x \in S$, there exists $B_x \in \mathcal{B}$ such that $x \in B_x$ and $B_x \subset S$, then $S = \cup_{x \in S} B_x$. So S is open. $\square$

We now recall the definition of the metric topology. Recall that in a metric space (M, d), we denote the open ball of radius $\epsilon > 0$ centered at $x \in M$ by $B_\epsilon(x)$.

Definition. Let (M, d) be a metric space. Let $\mathcal{T}_d \doteq \{A \subseteq M : \text{ for every } x \in A \text{ there is some } \epsilon > 0 \text{ such that } B_\epsilon(x) \subseteq A\}$. The collection $\mathcal{T}_d$ is the **metric topol-**

ogy on (M, d).

A useful exercise is to prove directly using the definition of a topology that $\mathcal{T}_d$ is a topology on M. This finally brings us to our punch line:

Theorem 2.2. (Metric topology)

Let (X, d) be a metric space. Then the collection of all open ϵ-balls $B_\epsilon(x)$, for $x \in X$ and $\epsilon > 0$, is a basis for the **metric topology** on X.

In other words, whenever we speak of "open balls" and "open sets" w.r.t. a metric space in this way, we are referring to the same sets as those that belong to the metric topology on the space X, induced by the metric d.

This result gives us an explicit link between topological structure of a space and the metric defined on the space. It should be noted, though, that the topology generated by a metric is not necessarily unique. That is, different metrics on the same space may actually generate the same topology, in which case we say that the metric spaces are *topologically equivalent*. Often this will be fairly obvious to confirm/refute, but in many cases that will not be true.

Example. (Different metrics generating different topologies)

Consider $\mathbb{R}$ and let d_1 be the discrete metric and d_2 be the usual absolute value metric. Then $(\mathbb{R}, d_1)$ is not topologically equivalent to $(\mathbb{R}, d_2)$.

Under d_1, every singleton set $\{a\}, a \in \mathbb{R}$ is an open set. This is obviously not true under d_2.

Example. (Different metrics generating the same topology)

Consider $\mathbb{R}$ and let metrics d_3, d_4 be defined by:

$$d_3(x,y) = |x-y|, \quad d_4(x,y) = 2|x-y|$$

Then $(\mathbb{R}, d_3)$ is topologically equivalent to $(\mathbb{R}, d_4)$.

It is easy enough to see this. Every open ball in (the topology generated by) $(\mathbb{R}, d_3)$ obviously has a corresponding open ball in (the topology generated by) $(\mathbb{R}, d_4)$. And since open balls are bases for metric topologies, then the metric topologies are the same.

Finally, we close by mentioning that some properties of metric spaces we will eventually study (e.g. compactness) are purely topological in nature. That is, no matter which different metrics we are considering, as long as the two different metric spaces generate the same topology then we are guaranteed that the property holds for both spaces. However, some properties of metric spaces are not purely topological. For example, the completeness of a metric space is not a topological property. This is illustrated in the following example.

Example. (Completeness is not a topological property)

Let $M \doteq \left\{1, \frac{1}{2}, \frac{1}{3}, \dots\right\}$. Let d_1 be the discrete metric on M. Namely, for $x, y \in M$, let

$$d_1(x,y) = \begin{cases} 1 & x \neq y \\ 0 & x = y. \end{cases}$$

Let d_2 be the metric on M given by $d_2(x,y) = |x-y|$. Let $\mathcal{T}_{d_1}$ and $\mathcal{T}_{d_2}$ be the topologies generated by d_1 and d_2, respectively. Using the definition of the metric topology, one can check that $\mathcal{T}_{d_1} = \mathcal{T}_{d_2} = \{A : A \subseteq M\}$. That is, every subset of M belongs to both $\mathcal{T}_{d_1}$ and $\mathcal{T}_{d_2}$. However, (M, d_1) is a complete metric space, while (M, d_2) is not. We begin by showing that (M, d_1) is complete. Let $\{x_n\}$ a Cauchy sequence in (M, d_1). Then there is some $N \in \mathbb{N}$ such that for all $n, m \geq N$, $d_1(x_n, x_m) < \frac{1}{2}$. This implies that $x_n = x_N$ for all $n \geq N$, so we have that $d_1(x_n, X_N) \to 0$ as $n \to \infty$. This shows that (M, d_1) is complete. We now show that (M, d_2) is not complete. Consider the sequence in (M, d_2) given by $y_n \doteq \frac{1}{n}$. Then $\{y_n\}$ is Cauchy in (M, d_2), but does not converge in (M, d_2), since $0 \notin M$. Therefore (M, d_2) is not complete.

The previous example is somewhat artificial, but the fact that completeness is not a topological property does show up frequently in probability theory. A key is example is the space $D[0, 1]$, which is the space of all real-valued functions on $[0, 1]$ that are right-continuous and have left limits. While the details are beyond the scope of this course, if you take a course on the weak convergence of probability measures, then you will see that there is in some sense a "natural" topology on $D[0, 1]$ (the Skorohod topology). However, under the most "natural" metric that generates this topology, the space is not complete. However, there is another topologically equivalent metric under which the space is complete.

2.2 Basic properties of open and closed sets

We begin our study of open and closed sets in metric spaces with the basic definitions (reiterating from above):

Definition. (Open and closed sets)

Recall that in a metric space (M, d) the open ball around $c \in M$ of radius $r > 0$ is the set $B_r(c) = \{x \in M : d(c, x) < r\}$. Let $S \subset M$.

1. $S \subset M$ is **open** if for all $x \in S$, there exists $r > 0$ such that $B_r(x) \subset S$

2. $S \subset M$ is **closed** if S^c is open

Thus a set $S \subset M$ is open if, for every point in S there exists some small neighborhood of that point contained entirely in S. It's easy to see that interval (a, b) is an open set in $\mathbb{R}$, and clearly every metric space M is an open subset of itself (recall the definition of open ball). To reinforce our intuition using this definition, we show that:

Theorem 2.3. The open ball $B_r(x_0)$ is, in fact, open (for $x_0 \in M$ and $r > 0$).

Proof. Let $x \in B_r(x_0)$, a point in the open ball. Now set $s = r - d(x, x_0) > 0$. Now consider the sub-open-ball $B_s(x)$. For $z \in B_s(x)$ we have:

$$
\begin{aligned}
d(z, x_0) &\leq d(z, x) + d(x, x_0) \\
&< r - d(x, x_0) + d(x, x_0) = r
\end{aligned}
$$

Thus $z \in B_r(x_0)$, so for any $x \in B_r(x_0)$ there exists $s > 0$ such that $B_s(x) \subset B_r(x_0)$. $\square$

It should be noted that the definition of a closed set given above (i.e. "its complement is open") is very topological. That is, in more general topologies (not just the metric topology), the same definition for a closed set is true. For our purposes, a more "metric" definition of closed set involving convergent sequences will be useful.

Definition. (Limit point)

Let M be a metric space and let $S \subset M$. We say that $x \in M$ is a **limit point** of S if there exists a sequence (x_n) of points in S (not necessarily distinct) such that $x_n \to x$.

Denote by $\lim S$ the set of all limit points of S.

Obviously, if $x \in S$ then it is a limit point of S; just take the constant sequence $x, x, x, \ldots$. But it is the points *outside* of S which are interesting. For example, let $M = \mathbb{R}$ and $S = \mathbb{Q}$. Then 2 is a limit point of S which belongs to S, and π is a an example of a limit point of S which does not belong to S: consider $(x_n) = 3, 3.1, 3.14, 3.141, 3.1415, \ldots$

The key result is:

Theorem 2.4. (Closed set: limit definition)

Let M be a metric space and let $S \subset M$. Then S is closed iff it contains all its limit points.

Proof. We prove by contradiction in both directions.

($\Rightarrow$): Let S be closed and let (x_n) be a sequence in S which converges to $x \in M$. Suppose that $x \notin S$. Then $x \in S^c$, which is open by virtue of S being closed. Thus we can find some $r > 0$ such that $B_r(x) \subset S^c$, which contradicts $x_n \to x$.

($\Leftarrow$): Assume $S = \lim S$. Suppose that S is not closed. Then S^c is not open so that there exists some $x' \in S^c$ such that, for every $r > 0$ the open ball $B_r(x')$ contains at least one point $x_r \in S$. Therefore the sequence defined by $x_n \in B_{\frac{1}{n}}(x')$ is in S and converges to $x' \in S^c$, contradicting that $S = \lim S$. $\square$

A few obvious closed sets can be identified from this definition:

1. Any singleton set $\{x\}$ is closed in M (assuming M is non-empty) since the only

possible sequence in that set is the constant sequence, which converges to x.

2. The entire space M is closed since any convergent sequence in M converges to a point in M.

3. For any $S \subset M$, $\lim S$ is closed.

Some sets are both closed and open—these are referred to as **clopen** sets. One such set is the whole space M. It follows then that $M^c = \emptyset$ is clopen as well. But there are also sets which are neither closed nor open. Consider the interval $[0, 1) \subset \mathbb{R}$. It is neither open (every r-neighborhood of 0 includes points in $[0, 1)^c \subset \mathbb{R}$) nor closed (it fails to include 1, which is the limit of the sequence (x_n) defined by $x_n = 1 - \frac{1}{n}$, in $[0, 1)$). Thus subsets of a metric space can be open, closed, both, or neither.

An important property of open and closed sets are their behavior under unions and intersections. From a topological perspective, these properties are definitional, but it is almost trivial to show these properties from the metric definitions.

Theorem 2.5. The arbitrary union of open sets is open.

Proof. Suppose $\{U_\alpha\}$ is a collection of open sets in M and let $U = \cup U_\alpha$. Then $x \in U$ implies that $x \in U_\alpha$ for some α. Since U_α is open, there exists an open ball around x of some radius which is contained within $U_\alpha \subset U$. □

Theorem 2.6. The intersection of finitely many open sets is open.

Proof. Suppose $U_1, U_2, \dots, U_n$ are open sets in M. Let $U = \cap U_k$. Assume $U \neq \emptyset$, otherwise the result is trivial. Now suppose $x \in U$ so that $x \in U_k$ for $k = 1, 2, \dots, n$. Since each U_k is open, for each $k = 1, \dots, n$, there exists $r_k > 0$ such that $B_{r_k}(x) \subset U_k$. Define r by:

$$r = \min\{r_1, r_2, \dots, r_n\} > 0$$

Then $B_r(x) \subset U_k$ for all $k = 1, \dots, n$ and so $B_r(x) \subset U$. $\square$

Notice that the arbitrary intersection of open sets is not necessarily open. For example, it's easy to see that $U_k = (-\frac{1}{k}, \frac{1}{k})$ is an open subset of $\mathbb{R}$, but $\cap U_k = \{0\}$ is clearly not open in $\mathbb{R}$. However, it is easy to see that this is true for closed sets.

Theorem 2.7. The arbitrary intersection of closed sets is closed. Also, the finite union of closed sets is closed.

Proof. Immediate from DeMorgan's laws. $\square$

Notice that the arbitrary union of closed sets is not guaranteed to be closed. For example, even though each $K_k = [0, 1 - \frac{1}{k}]$ is closed in $\mathbb{R}$, the union $\cup K_k = [0, 1)$ is not.

Finally, we close the section with an interesting result that is more or less specific to the case of the real line.

Theorem 2.8. Every nonempty open set $U \subset \mathbb{R}$ is a countable disjoint union of open intervals of the form (a, b), where a and b may take the values $-\infty$ and $+\infty$.

Proof. We give a constructive proof. To start, for each $x \in U$ we construct I_x, the "maximal" open interval $\subset U$ that contains x. So fix $x \in U$. Define the following:

$$a_x = \inf\{a : (a, x) \subset U\} \quad \text{and} \quad b_x = \sup\{b : (x, b) \subset U\}$$

Let $I_x = (a_x, b_x)$. Clearly $x \in I_x$, and $I_x \subset U$. If not, then a_x would not be the infimum of $\{a : (a, x) \subset U\}$ and likewise for b_x. Furthermore, I_x is maximal in the sense that $a_x, b_x \notin U$, so it cannot be enlarged and remain in U. (To see this, suppose otherwise that $b_x \in U$. Then there exists an open interval $J \subset U$ containing b_x, which implies (check) that there exists $b^* > b_x$ such that $b^* \in \{b : (x, b) \subset U\}$. However this contradicts b_x being the supremum of that set.) Thus we may cover U by the union

$$U = \bigcup_{x \in U} I_x$$

We show that this union is disjoint. Let $x, y \in U$ and suppose that $I_x \cap I_y \neq \emptyset$. Then $I_x \cup I_y$ is an open interval containing both x and y, but since all these intervals are maximal then $I_x = I_x \cup I_y = I_y$. Thus for all $x, y \in U$ either $I_x = I_y$ or the two intervals are disjoint. Therefore the above union is disjoint. To show that the union is countable, simply pick a rational number in each interval. The intervals are disjoint so the numbers are distinct, and their collection is therefore countable. $\square$

2.3 More structural properties

Now that we have covered the basics, we move on to some concepts of a more "sequential" nature. First, let us relate the notion of open and closed sets back to our definition of completeness.

Since closed sets contain all of their limit points, one might expect that if some sort of convergence result holds for the whole space, then it will also hold in a closed subset of the space. As it turns out, this is exactly true when speaking about completeness:

Theorem 2.9. Let M be a complete metric space and let $N \subset M$ be closed. Then N is complete as a metric space in its own right.

Proof. Let (x_n) be a Cauchy sequence in N. Since (x_n) is also a Cauchy sequence in M, and M is complete, then x_n is convergent to some $x \in M$. But N is closed, so we must have $x \in N$. $\square$

Now a few more definitions are in order. In what follows, let M be a metric space and let $S \subset M$.

Definition. (Closure, interior, boundary)

1. The **closure** of S is $\bar{S} = \cap K_\alpha$ where $\{K_\alpha\}$ is the collection of all closed sets that contain S.

2. The **interior** of S is $\mathrm{int}(S) = \cup U_\alpha$, where $\{U_\alpha\}$ is the collection of all open sets contained in S.

3. The **boundary** of S is $\partial S = \bar{S} - \mathrm{int}(S)$.

Immediate from the definition, we have the following facts:

Fact. If M is a metric space and $S \subset M$, then:

1. $\mathrm{int}(S) \subset S \subset \bar{S}$

2. $\mathrm{int}(S)$ is open, and $\bar{S}$ is closed

3. S is open iff $S = \mathrm{int}(S)$, and S is closed iff $S = \bar{S}$

Proof. Simple exercise. $\square$

For example, if $M = \mathbb{R}$ and $S = (a, b]$, then $\bar{S} = [a, b]$, $\text{int}(S) = (a, b)$, and $\partial S = \{a\} \cup \{b\}$. If $S = \mathbb{Q}$, then $\bar{S} = \mathbb{R}$, $\text{int}(S) = \emptyset$, and $\partial S = \mathbb{R}$. In addition, you can check for yourself that every subset of a discrete metric space M (for example, $\mathbb{N}$ with discrete metric) is clopen (why would it suffice to show that a singleton $\{x\}$ is open?) and that therefore $\forall S \subset M$, $\text{int}(S) = S = \bar{S}$ and $\partial S = \emptyset$.

Theorem 2.10. $\bar{S} = \lim S$.

Proof. For one inclusion, note from before that $\lim S$ is closed and also that $S \subset \lim S$ (e.g. consider the constant sequence). Therefore by definition of closure, $\bar{S} \subset \lim S$. For the other inclusion, note that $S \subset \bar{S}$ and $\bar{S}$ is closed, so therefore $\bar{S}$ must contain all the limit points of S. $\square$

Note that until now, we have sometimes said "S is open" or "S is closed" without explicitly referring to the metric space when it is understood. However, it should be kept in mind that the metric space M is essential to the openness/closedness of a subset $S \subset M$.For example, both $\mathbb{Q}$ and the half-open interval $[a, b)$ are clopen when considered as metric spaces in their own right. However, neither one is either open or closed when treated as a subset of $\mathbb{R}$.

Another example is a set $S = \mathbb{Q} \cap (-\pi, \pi)$, a set of all rational numbers in the interval $(-\pi, \pi)$. As a subset of metric space $\mathbb{Q}$, S is both closed (if (x_n) is a sequence in S, and $x_n \to x \in \mathbb{Q}$ then $x \in S$) and open (check for yourself). As a subset of $\mathbb{R}$, however, it is neither open (if $x \in S$ then every neighborhood of x contains some $y \notin \mathbb{Q}$) nor closed (there are sequences in

S converging to $\pi \in \mathbb{R} - \mathbb{Q}$).

The following few theorems establish the relationship between being open/closed in metric space M and some metric subspace N of M with the same metric from M (i.e. define the metric d_N on N by $d|_N(x, y) = d(x, y)$ for $x, y \in N$). The key takeaway is that using the same metric from M, subsets like N will inherit the topology (i.e. open and closed sets) from M in the following way:

Theorem 2.11. Let M be a metric space and suppose $S \subset N \subset M$. Then S is open in N if and only if there exists $L \subset M$ such that L is open in M and $S = L \cap N$.

The proof of this theorem follows from the following lemma:

Lemma 2.12. If $S \subset N \subset M$, then S is closed in N iff there exists $L \subset M$ such that L is closed in M and $S = L \cap N$.

Proof. For a set $S \subset N$ we will denote the closure of set S in M by $\bar{S}_M$, and the closure of S in N by $\bar{S}_N$. Note that $\bar{S}_N = \bar{S}_M \cap N$.

Suppose S is closed in N. Define $L = \bar{S}_M$. Then L is closed in M and $L \cap N = \bar{S}_N = S$, since S is closed in N. Conversely, suppose now that L as in the statement of theorem exists. Since L is closed, it contains all of its limit points and $S = L \cap N$ contains all of its limit points in N, therefore S is closed in N. $\square$

Finally, we generalize the notion of boundedness for subsets of $\mathbb{R}$ to arbitrary metric spaces.

Definition. (Bounded, boundedness)

Let M be a metric space. $S \subset M$ is **bounded** if there exists $x \in M$ and $0 < r < \infty$ such that $S \subset B_r(x)$.

In other words, S is bounded if it is contained in some ball. For example, $[-1, 1]$ is bounded in $\mathbb{R}$ since it's contained in $B_5(2)$ or $B_2(0)$. On the other hand, the graph of function $f(x) = sin(x)$ is an unbounded subset of $\mathbb{R}^2$, although the range of f is a bounded subset of $\mathbb{R}$ (range $= [-1, 1]$). In general, we say that f is a **bounded function** if its range is a bounded subset of the target space.

We close this section with an important result relating our new notion of boundedness to Cauchy sequences.

Theorem 2.13. Let (x_n) be a Cauchy sequence in M. Then

$$S = \{x \in M : x = x_n \text{ for some } n\}$$

is bounded. In other words, Cauchy sequences are bounded.

Proof. Let $\epsilon = 1$. Then there exists N such that $d(x_n, x_m) < 1$ for all $n, m \geq N$. In particular, $d(x_n, x_N) < 1$ for all $n \geq N$. Now define:

$$r = \max\{1, d(x_1, x_N), d(x_2, x_N), \ldots, d(x_{N-1}, x_N)\}$$

Then the entire sequence is contained in the closed ball centered at x_N of radius r. $\square$

Corollary 2.14. If a sequence is convergent, then it is bounded.

2.4 Continuous functions

The notion of continuity is key to so much of statistics and probability that its importance can hardly be overstated. There are two main definitions of continuity which one is likely to encounter. The first definition is the classical one from an undergraduate real analysis standpoint. When viewed from the lens of a function $f : \mathbb{R} \to \mathbb{R}$, it essentially says that the function has no breaks:

Definition. ($\epsilon - \delta$ definition)

Let M, N be two metric spaces and $f : M \to N$ a function.

1. f is **continous at** $x \in M$ if for all $\epsilon > 0$, there exists $\delta(x, \epsilon) > 0$ such that for $y \in M$,

$$d_M(x, y) < \delta(x, \epsilon) \quad \Rightarrow \quad d_N(f(x), f(y)) < \epsilon$$

   Or equivalently, $f(B_\delta(x)) \subset B_\epsilon(f(x))$.

2. f is **continuous** (on M) if it's continuous at every $x \in M$.

3. f is **uniformly continuous** (on M) if for all $\epsilon > 0$, there exists $\delta(\epsilon)$ **not** depending on x such that for $x, y \in M$,

$$d_M(x, y) < \delta(\epsilon) \quad \Rightarrow \quad d_N(f(x), f(y)) < \epsilon$$

In other words if a function is continuous, no matter how small of an ϵ-ball we consider in the target N space, we can always find a δ-ball small enough (in M space) such that the image of that δ-ball will map completely within the ϵ-ball. If this δ threshold doesn't depend on $x \in M$, then the mapping is uniformly continuous.

Here is a simple example to make things clear. For example, consider the function $f : (0,1) \to \mathbb{R}$ defined by $f(x) = 1/x$. This function is continuous on all of its domain, but is not uniformly continuous: given $\epsilon > 0$, no matter how small we choose δ to be, there are always points x, y in the interval $(0, \delta)$ such that $|f(x) - f(y)| > \epsilon$. Under the above interpretation, continuity is a condition on the smoothness of a function. A continuous function cannot suddenly jump from point to the next; it must transition smoothly in a way that is regulated by the interaction of δ and ϵ.

However, the $\epsilon - \delta$ definition hints at a deeper meaning. At its core, continuity is a topological concept which says something about how a mapping acts on **all open sets**. One can see (using the alternative formulation above) that the $\epsilon - \delta$ definition describes the action of a function on a special class of open sets: the open balls. However, once we recall that the metric topology has the collection of open balls as a basis, it is plain to see that the $\epsilon - \delta$ definition extends to a more general definition regarding all open sets.

Definition. (Topological definition)

Let X, Y be two topological spaces and let $f : X \to Y$ be a function. Then f is **continuous** (on X) if $f^{-1}(S)$ is open in X whenever S is open in Y.

Equivalently, f is continuous if $f^{-1}(S)$ is closed in X whenever S is closed in Y.

The next result is often useful for proving that a function between topological spaces is continuous. In particular, there is a basis of "simple sets" for a topology, then it is often easier to study continuity using that basis.

Lemma 2.15. Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Let $\mathcal{B}_Y$ be a basis for $\mathcal{T}_Y$. Then f is continuous if and only if $f^{-1}(U) \in \mathcal{T}_X$ for all $U \in \mathcal{B}_Y$.

Proof. Suppose that f is continuous. Since $\mathcal{B}_Y \subseteq \mathcal{T}_Y$, the result follows.

Now, suppose that $f^{-1}(U) \in \mathcal{T}_X$ for each $U \in \mathcal{B}_Y$. Let $O \in \mathcal{T}_Y$. Since $\mathcal{B}_Y$ is a basis for $\mathcal{T}_Y$, there is some index set Λ and collection $\{U_\lambda\}_{\lambda \in \Lambda} \subseteq \mathcal{B}_Y$ such that $O = \bigcup_{\lambda \in \Lambda} U_\lambda$. Observe that

$$f^{-1}(O) = f^{-1}\left(\bigcup_{\lambda \in \Lambda} U_\lambda\right) = \left\{x \in X : f(x) \in \bigcup_{\lambda \in \Lambda} U_\lambda\right\} = \bigcup_{\lambda \in \Lambda}\{x \in X : f(x) \in U_\lambda\} = \bigcup_{\lambda \in \Lambda} f^{-1}(U_\lambda).$$

The result follows on noting that $f^{-1}(U_\lambda) \in \mathcal{T}_X$ for all $\lambda \in \Lambda$. $\square$

Although we work mostly in metric spaces in statistics and probability, the topological interpretation is still worth bearing in mind: a function between two topological spaces is continuous if every element of the topology (structure) on N has a counterpart in the topology of M using the inverse mapping of f.

In this sense, continuous functions are ones that "sort of" build a bridge between the topologies of its domain and range. Then what about functions which not only map open sets in the range to open sets in the domain, but also map open sets in the domain to open sets in the range?

Definition. (Homeomorphism)

Let M, N be two topological spaces and $f : M \to N$ a *bijective* function.

If both f and $f^{-1} : N \to M$ are continuous, then f is a **homeomorphism**.

Essentially, homeomorphisms are mappings between two spaces that completely preserve the respective structures of the spaces. Continuous functions therefore a sort of go-between between homomorphisms and functions that don't preserve structure at all.

Example. (Real line is homemorphic to $(-1,1)$) Let $X = (-1,1)$ and $Y = \mathbb{R}$. Let $\mathcal{T}_X$ and $\mathcal{T}_Y$ denote the standard topologies on X and Y respectively. Then the map $f : X \to Y$ given by

$$f(x) \doteq \tan\left(\frac{\pi x}{2}\right), \quad x \in (-1,1),$$

is a homemorphism from X to Y. Its inverse is the map $g : Y \to X$ given by

$$g(y) \doteq \frac{2}{\pi} \arctan(y).$$

Both f and g are strictly increasing, and we have that

$$f^{-1}(a,b) = (g(a), g(b)), \quad a, b \in Y,$$

and

$$g^{-1}(c,d) = (f(c), f(d)), \quad c, d \in X.$$

Continuity of f and g follows on noting that open intervals in $(-1,1)$ and in $\mathbb{R}$ form bases for $\mathcal{T}_X$ and $\mathcal{T}_Y$, respectively.

The next example provides an interesting example of two spaces that are not homeomorphic.

Example. ($\mathbb{R}$ and $\mathbb{R}^2$ are not homemorphic) The spaces $\mathbb{R}$ and $\mathbb{R}^2$ are not homeomorphic under their standard topologies. We omit the proof, but the idea is that if $\mathbb{R}$ and $\mathbb{R}^2$ were homeomorphic under some homeomorphism $f : \mathbb{R} \to \mathbb{R}^2$, then $\mathbb{R} \setminus \{0\}$ and $\mathbb{R}^2 \setminus \{f(0)\}$ would be homeomorphic as well. However, it is clear that $\mathbb{R} \setminus \{0\}$

is not "connected" since it is made up of two nonempty disjoint open sets. However, $\mathbb{R}^2 \setminus \{f(0)\}$ is "connected" since it cannot be split into two nonempty disjoint open sets. The result then follows on proving that homeomorphims preserve this sort of "connectedness".

The following definition introduces the notion of *metrizability.* Most topological spaces studied in statistics and probability are metrizable.

Definition. (Metrizable) A topological space $(M, \mathcal{T})$ is *metrizable* if there is some metric d on M such that the open sets in (M, d) are the sets in $\mathcal{T}$.

Theorem 2.16. Let $(M, \mathcal{T}_M)$ and $(N, \mathcal{T}_N)$ be two homeomorphic topological spaces. If M is metrizable, then N is metrizable.

Proof. Let d denote the metric that induces the topology on M, and let $h : N \to M$ to be a homeomorphism. Define the function $\rho : N \times N \to \mathbb{R}_+$ by $\rho(y_1, y_2) \doteq d(h(y_1), h(y_2))$. We begin by showing that ρ is a metric on N. Reflexivity and the triangle property follow immediately from the fact that d is a metric. Additionally, it is clear that $\rho(y, y) = 0$. Now, suppose that $\rho(y_1, y_2) = 0$. Then $d(h(y_1), h(y_2)) = 0$, so $h(y_1) = h(y_2)$. Since h is bijective, it follows that $y_1 = y_2$. Therefore ρ is a metric on N.

Let $\mathcal{T}_\rho$ denote the topology generated by ρ. We begin by showing that $\mathcal{T}_N \subseteq \mathcal{T}_\rho$. Let $A \in \mathcal{T}_N$, and fix $a \in A$. We will show that there is some $\epsilon > 0$ such that $B_\epsilon^\rho(a) \subseteq A$, where

$$B_\epsilon^\rho(a) \doteq \{y \in N : \rho(a, y) < \epsilon\}.$$

Since h^{-1} is continuous, we know that $h(A) \in \mathcal{T}_M$. Since $h(a) \in h(A)$, it follows that there is some $\epsilon > 0$ such that $B^d_\epsilon(h(a)) \subseteq h(A)$, where

$$B^d_\epsilon(h(a)) \doteq \{x \in M : d(x, h(a)) < \epsilon\}.$$

We claim that $B^\rho_\epsilon(a) \subseteq A$. Fix $y \in B^\rho_\epsilon(a)$, so that $\rho(a, y) < \epsilon$. Since $\rho(a, y) = d(h(a), h(y))$, we know that $h(y) \in B^d_\epsilon(h(a)) \subseteq h(A)$. Since h is bijective,

$$y \in h^{-1}(h(A)) = A,$$

so we have shown that $B^\rho_\epsilon(a) \subseteq A$. Therefore $\mathcal{T}_N \subseteq \mathcal{T}_\rho$. The proof of the other inclusion is similar. $\square$

We now prove that the $\epsilon - \delta$ definition of continuity extends (and is equivalent to) the more general topological definition regarding all open sets.

Theorem 2.17. (Continuity in metric spaces)

Let M, N be metric spaces and $f : M \to N$ a function. The following are equivalent:

1. For all $x \in M$ and $\epsilon > 0$, there exists $\delta(x, \epsilon) > 0$ such that for $y \in M$,

$$d_M(x, y) < \delta(x, \epsilon) \quad \Rightarrow \quad d_N(f(x), f(y)) < \epsilon$$

2. For all open sets S in N, $f^{-1}(S)$ is open in M

Proof. Note that since the open balls of the $\epsilon - \delta$ definition are simply a special type of open set, one direction of the proof is trivial. Therefore we only prove the implication in the direction ($\Rightarrow$):

Let $S \subset N$ be open. Assume $f^{-1}(S) \neq \emptyset$, else the result is trivial. So let $a \in f^{-1}(S)$. To show $f^{-1}(S)$ is open, we show that there is an open ball around a contained in $f^{-1}(S)$. Now

$f(a) \in S$ and S is open, so there exists $\epsilon > 0$ such that $B_\epsilon(f(a)) \subset S$. But f is continuous at a, so there exists $\delta > 0$ such that $f(B_\delta(a)) \subset B_\epsilon(f(a))$. Therefore $B_\delta(a) \subset f^{-1}(S)$. $\square$

These two definitions alone can quickly yield some fairly useful results. For example:

Corollary 2.18. The composition of two continuous functions is continuous.

Also from these two basic definitions, we can also extract a third definition which will turn out to be handy. Essentially, continuous functions between metric spaces preserve convergent sequences. The proof will show that this ought to not be surprising, in light of all the connections we have drawn between continuity, open sets, and convergence.

Theorem 2.19. Let M, N be metric spaces and $f : M \to N$ a function. f is continuous at $x \in M$ if and only if whenever we have a sequence (x_n) in M such that $x_n \to x$ it is also true that $f(x_n) \to f(x)$ in N.

Proof. ($\Rightarrow$)

Suppose f is continuous at $x \in M$ and that $x_n \to x$ (in M). Fix $\epsilon > 0$. By continuity of f at x, there exists $\delta > 0$ such that $d_M(y, x) < \delta$ implies $d_N(f(y), f(x)) < \epsilon$ for $y \in M$. Now since $x_n \to x$, we can find N such that $d_M(x_n, x) < \delta$ for all $n \geq N$, so that $d_N(f(x_n), f(x)) < \epsilon$ for all $n \geq N$ also.

($\Leftarrow$)

Suppose that for each (x_n) in M with $x_n \to x$ we have $f(x_n) \to f(x)$ in N. Assume that f is *not* continuous at x. Then there exists $\epsilon > 0$ such that for no $\delta > 0$ is it always true that

$d_M(y, x) < \delta$ implies $d_N(f(y), f(x)) < \epsilon$. That is, for each $\delta > 0$ there exists $y \in M$ such that $d_M(y, x) < \delta$ and $d_N(f(y), f(x)) \geq \epsilon$.

In particular, letting $\delta = \frac{1}{n}$ for each $n = 1, 2, \ldots$, we can build a sequence (y_n) such that $d_M(y_n, x) < \frac{1}{n}$ for all n, however $d_N(f(y_n), f(x)) \geq \epsilon$ for all n. Thus (y_n) converges but $(f(y_n))$ does not converge. □

It is worth pointing out that while continuous functions preserve the convergent sequences, they in general do not preserve the **non-convergent** Cauchy sequences. For example, the continuous funcion $f : (0, 1] \to \mathbb{R}$ given by $f(x) = 1/x$ maps the Cauchy sequence $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ in $(0, 1]$ to the non-Cauchy sequence $1, 2, 3, 4, \ldots$ in $\mathbb{R}$. However, we are saved by the following result:

Theorem 2.20. Let M, N be metric spaces and $f : M \to N$ a uniformly continuous function. If (x_n) is a Cauchy sequence in M, then $(f(x_n))$ is a Cauchy sequence in N.

Proof. Exercise. □

You can also show fairly easily that **every** function defined on a discrete metric space is uniformly continuous (exercise).

Example: Consistent Estimates

In the theory of (statistical) estimation, one desirable property for an estimator to have is consistency. Broadly speaking, consistency means that the estimator will zero in on the true population parameter value as it is fed more and more samples. More precisely,

Definition. (Weak consistency)

Let $\theta \in \mathbb{R}$ be some parameter and let $\hat{\theta}_n$ be an estimator of θ based on sample of size n. We say $\hat{\theta}_n$ is **weakly consistent** if for all $\epsilon > 0$,

$$\mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) \to 0 \quad \text{as } n \to \infty$$

In other words, an estimator is weakly consistent if, for every $\epsilon > 0$, the probability of it being ϵ-distance away from the true value goes to zero as the sample size becomes infinitely large. Clearly this is a Good Thing. So how does one go about obtaining weakly consistent estimators?

If we are interested in the mean, then often we can look to the Law of Large Numbers (LLN) for help. The Weak LLN states that for a sequence of IID random variables $X_1, X_2, \ldots, X_n$ with expected value μ, then:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right) \to 0 \quad \text{as } n \to \infty$$

In other words, if we are interested in estimating the mean of a distribution, then the sample mean is guaranteed to be weakly consistent by the Weak LLN *no matter what the actual distributions is.* But what if we are interested in some other quantity?

Theorem 2.21. (A continuous mapping theorem)

Suppose $\hat{\theta}_n$ is a weakly consistent estimate of $\theta \in \mathbb{R}$ and let $f : \mathbb{R} \to \mathbb{R}$ be continuous at θ. Then $\hat{\rho}_n = f(\hat{\theta}_n)$ is a weakly consistent estimate of $\rho = f(\theta)$.

Proof. Suppose $f(\hat{\theta}_n)$ is not consistent. In other words, there exists $\epsilon > 0$ such that the sequence $\mathbb{P}(|f(\hat{\theta}_n) - f(\theta)| > \epsilon)$ does not converge to 0. In particular, for some $L > 0$ there is a subsequence $(\hat{\theta}_{n_k})_{k\geq 1}$ such that $\mathbb{P}(|f(\hat{\theta}_{n_k}) - f(\theta)| > \epsilon) > L$ for all k. Now since f is continuous at θ, there exists $\delta > 0$ such that the following implication holds for all n:

$$|f(\hat{\theta}_n) - f(\theta)| > \epsilon \quad \Longrightarrow \quad |\hat{\theta}_n - \theta| > \delta \tag{2}$$

Now for any two events A, B such that A implies B, it is true that $\mathbb{P}(A) \leq \mathbb{P}(B)$. Therefore by (1), we have:

$$\mathbb{P}(|\hat{\theta}_{n_k} - \theta| > \delta) \geq \mathbb{P}(|f(\hat{\theta}_{n_k}) - f(\theta)| > \epsilon) > L \quad \text{for all } k \tag{3}$$

However, our original sequence of estimators is consistent, so there exists $N \in \mathbb{N}$ such that $\mathbb{P}(|\hat{\theta}_{n_k} - \theta| > \delta) < L$ for all $n_k \geq N$, but this contradicts (2). $\square$

This result is very useful. For example, it lies at the heart of the *method of moments*. Once again, suppose we have iid random variables $X_1, X_2, \ldots, X_n$ with expected value μ, and suppose we are interested in estimation of $\mu^{4154781481226426191177580544000000}$ (assuming it exists). By the WLLN we know that $\bar{X}$ is a consistent estimate of μ, so since polynomials are continuous we immediately have that $\bar{X}^{4154781481226426191177580544000000}$ is a consistent estimate.

The above result extends to functions defined on arbitrary metric spaces, not just $\mathbb{R}$. For example, if there is a continuous function $\theta = f(\mu_1, \mu_2, \ldots, \mu_k)$, then $\hat{\theta} = f(\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_k)$ is a consistent estimate of θ.

2.5 Compactness

Now we have arrived at a core concept in analysis which will come up over and over again in applications. To provide a deep treatment of compactness, one has to work with topological spaces, not just metric spaces. But since in statistics and probability we care mostly about metric spaces, we will cover the special case of compactness in metric topologies only.

There are two distinct notions of compactness: *sequential* compactness and *covering* compactness. In arbitrary topological spaces, these properties are not the same and neither one implies the other. However, for metric spaces they are equivalent. Despite this, they are useful in different ways and so we shall study them both.

> **Definition.** (Sequentially compact)
>
> Let M be a metric space and let $S \subset M$. Then S is **sequentially compact** if every sequence (x_n) in S has a convergent subsequence (x_{n_k}) such that $x_{n_k} \to x$ for some $x \in S$.

Some examples will help to illustrate what this principle is telling us.

> **Example.** Let M be any metric space. Any finite subset $S = \{x_1, x_2, \ldots, x_n\}$ is sequentially compact.

Since any sequence in S has to repeat at least one x_k infinitely many times, then taking the subsequence of the repeated value is a constant (and therefore convergent) sequence. Notice also that the convergent subsequence or indeed the limit need not be unique, as the example of a sequence 1,2,1,2,1,2,1,2, ... in $S = \{1, 2\} \subset \mathbb{N}$ shows.

> **Example.** $(0, 1] \subset \mathbb{R}$ is not sequentially compact.

The sequence $\{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$ is a sequence in $(0, 1]$ but clearly cannot have a convergent subsequence in $(0, 1]$ since the original sequence converges to 0 and limits are unique. In the same vein, any subset $S \subset \mathbb{R}$ which contains a sequence converging to a value not in S is not sequentially compact (e.g. $\mathbb{N}$, $\mathbb{Q}$, etc).

This characterization of compactness is useful, but it does not immediately give much insight into what it means for the space, intrinsically. The notion of covering compactness helps a bit in this regard.

Definition. (Open covers, covering compact)

Let M be a metric space.

1. A collection of open subsets $\mathcal{U}$ of M is an **(open) cover** for $S \subset M$, if for all $x \in S$, there exists $U_x \in \mathcal{U}$ such that $x \in U_x$.
2. If $\mathcal{U}$ is a cover of S, then $\mathcal{V}$ is a **subcover** of $\mathcal{U}$ if $\mathcal{V}$ is also a cover of S and $\mathcal{V} \subset \mathcal{U}$.
3. We say that $S \subset M$ is **covering compact** if **every** open cover $\mathcal{U}$ has a finite subcover.

In this definition, we can start to get hints of what makes compactness such an important property. If a subset S is compact then any open cover of S *regardless of whether it is countably infinite, uncountably infinite, or otherwise*, can be reduced to a *finite* cover. This reduction of the infinite to the finite is essentially the key to compactness.

In order to see that not all subsets are covering compact, consider the cover of $(0, 1] \subset \mathbb{R}$ by open intervals $U_n = (\frac{1}{n}, 1 + \frac{1}{n})$. Clearly this open cover cannot be reduced to a finite subcover since we'd be left with $\{U_{n_1}, U_{n_2}, \ldots, U_{n_k}\}$ and letting $N = \max\{n_1, n_2, \ldots, n_k\}$, we observe that any $x \in (0, \frac{1}{N})$ is not contained in any of $U_{n_1}, U_{n_2}, \ldots, U_{n_k}$.

Now that we have two useful notions of compactness, it remains to unify them. At first glance it is not obvious that they are equivalent; one notion involves convergence of sequence, and the other involves open coverings. But as our proof will show, the connection between the two notions is actually quite simple.

To make clear the machinery behind the proof, we start with a definition and lemma:

> **Definition.** (Lebesgue number)
>
> Let $S \subset M$ and let $\mathcal{U}$ be some open cover for S. A **Lebesgue number** for $\mathcal{U}$ is a number $\delta > 0$ such that for each open ball $B_\epsilon(x) \subset S$ with $\epsilon \leq \delta$, there exists some $U \in \mathcal{U}$ with $B_\epsilon(x) \subset U$.

It's not exactly clear what it means if a cover has a larger Lebesgue number than another cover. However, not all covers have Lebesgue numbers, and it **is** important to distinguish between covers that have a Lebesgue number and those that do not.

Essentially, if an open cover does not have a Lebesgue number, the topology won't fit into the open cover. That is, no matter how many "big" open balls we throw out, we will never be left with a collection of smaller open sets that are subsets of sets in the open cover. On the other hand, if the open cover *does* have a Lebesgue number, then as long as we throw out enough big open sets (i.e. only consider $B_\epsilon(x)$ for $\epsilon \leq \delta$), we are guaranteed to be left with a collection of small open sets that "fits into" the open cover.

Example. Consider the open unit interval $(0, 1)$ with the standard Euclidean metric.

1. **Big Lebesgue number**: $\mathcal{U} = \{(0, 1)\}$

   In this case any open ball subset of $(0, 1)$ is still contained within $(0, 1)$, and the maximum such radius here is $\frac{1}{2}$. Therefore a Lebesgue number for this open cover is $\frac{1}{2}$.

2. **Smaller Lebesgue number**: $\mathcal{U} = \{B_{\frac{1}{10}}(x) : x \in (0, 1)\}$ Obviously any open ball with radius greater than $\frac{1}{10}$ will not be contained in any of the elements of the open cover. It follows that the largest possible Lebesgue number here is $\frac{1}{10}$.

3. **No Lebesgue number**: $\mathcal{U} = \{(\frac{1}{n}, 1) : n = 2, 3, \ldots\}$ In this case, it is easy to see that the open ball $B_{\frac{1}{n}}(\frac{1}{n})$ in the metric topology on $(0, 1)$, will not be contained in any of the elements of the cover.

From the last example, we can already see how the Lebesgue number concept hints at a tie to our intuition of how compactness takes the infinite back down to the finite. For example 3, we see that the open cover does indeed cover the whole interval when considering the infinite union of the elements of the cover. Yet, the open subset $(0, \frac{1}{2})$ is not contained in any *one* of the elements of the cover.

We can begin to feel that, if we could reduce the infinite cover to a finite subcover, then that might help matters. Thus the Lebesgue number lemma:

Lemma 2.22. (Lebesgue number lemma)

Let M be a metric space and let $S \subset M$. If S is sequentially compact, then every open cover of S has a Lebesgue number.

Proof. Suppose towards a contradiction that there is an open cover of S with no Lebesgue number. This means that there is an open cover $\{U_\alpha : \alpha \in \mathcal{A}\}$ for S such that for every $\lambda \in \mathbb{R}$, there exists $x \in S$ such that $B_\lambda(x) \nsubseteq U_\alpha$ for each $\alpha \in \mathcal{A}$. Now define a real sequence (λ_n) by $\lambda_n = \frac{1}{n}$, with corresponding sequence of points (x_n) defined as above.

Since S is sequentially compact, then (x_n) has a subsequence $(x_{n_k})_{k\geq 1}$ which converges to a point $x_0 \in S$. This point must lie in $x_0 \in U_\alpha$ for some α since $\{U_\alpha\}$ is a cover for S, and furthermore because the cover is open there exists $r > 0$ such that $B_r(x_0) \subset U_\alpha$.

Now, since $x_{n_k} \to x_0$, there exists N_1 such that $d(x_0, x_{n_k}) < \frac{r}{2}$ for all $k \geq N_1$. In addition, $\lambda_n \to 0$ so we can find N_2 such that $\lambda_{n_k} < \frac{r}{2}$ for all $k \geq N_2$. Set $N = \max\{N_1, N_2\}$ and pick some x_{n_M} with $M > N$. Then for $y \in B_{\lambda_{n_M}}(x_{n_M})$,

$$d(x_0, y) \leq d(x_0, x_{n_M}) + d(x_{n_M}, y) < \frac{r}{2} + \lambda_{n_M} < \frac{r}{2} + \frac{r}{2}$$

Thus $B_{\lambda_{n_M}}(x_{n_M}) \subset B_r(x_0) \subset U_\alpha$, which contradicts that for all n, $B_{\lambda_n}(x_n) \nsubseteq U_\alpha \forall \alpha$. □

Proof. Suppose not. Then there is an open cover $\{U_\alpha : \alpha \in \mathcal{A}\}$ for S such that for every $\lambda \in \mathbb{R}$, there exists $x \in S$ such that $B_\lambda(x) \nsubseteq U_\alpha$ for each $\alpha \in \mathcal{A}$. Define the sequence $\{\lambda_n\}$ by $\lambda_n \doteq \frac{1}{n}$. Then for each $n \in \mathbb{N}$, there is some $x_n \in S$ such that $B_{\lambda_n}(x_n) \nsubseteq U_\alpha$ for each $\alpha \in \mathcal{A}$.

Since S is sequentially compact, then (x_n) has a subsequence $(x_{n_k})_{k\geq 1}$ which converges to a point $x_0 \in S$. Since $\{U_\alpha : \alpha \in \mathcal{A}\}$ is a cover for S, we then have that

$$x_0 \in S \subseteq \bigcup_{\alpha \in \mathcal{A}} U_\alpha.$$

Therefore there is some $\alpha \in \mathcal{A}$ such that $x_0 \in U_\alpha$. Additionally, U_α is open, so (using the definition of the metric topology) there is some $r > 0$ such that $B_r(x_0) \subseteq U_\alpha$.

Since $x_{n_k} \to x_0$, there is some $N_1 \in \mathbb{N}$ such that $d(x_0, x_{n_k}) < \frac{r}{2}$ for all $k \geq N_1$. In addition, $\lambda_{n_k} \to 0$ so there is some $N_2 \in \mathbb{N}$ such that $\lambda_{n_k} < \frac{r}{2}$ for all $k \geq N_2$. Set $N \doteq \max\{N_1, N_2\}$ and let $k \geq N$. Then for $y \in B_{\lambda_{n_k}}(x_{n_k})$,

$$d(x_0, y) \leq d(x_0, x_{n_k}) + d(x_{n_k}, y) < \frac{r}{2} + \lambda_{n_k} < \frac{r}{2} + \frac{r}{2} = r.$$

Thus $B_{\lambda_{n_k}}(x_{n_k}) \subset B_r(x_0) \subset U_\alpha$, which contradicts the fact that for each $n \in \mathbb{N}$, $B_{\lambda_n}(x_n) \nsubseteq U_\alpha$ for all $\alpha \in \mathcal{A}$. The result follows. □

Now the main result:

Theorem 2.23. Let M be a metric space and let $S \subset M$. The set S is covering compact if and only if it is sequentially compact.

Proof. $(\Rightarrow)$

Suppose S is covering compact but not sequentially compact. Then there exists a sequence $\{x_n\}$ in S with no subsequence that converges to a point in S. Therefore for every $x \in S$ there exists $r_x > 0$ such that $B_{r_x}(x)$ contains only finitely many terms of $\{x_n\}$; otherwise there would be a subsequence converging to x.

Now, $\{B_{r_x}(x)\}_{x \in S}$ is an open cover for S and since S is covering compact, it has a finite subcover $\{B_{r_{x_1}}(x_1), B_{r_{x_2}}(x_2), \ldots, B_{r_{x_k}}(x_k)\}$. But by the construction above, this implies that S contains only finitely many terms of $\{x_n\}$, a contradiction.

$(\Leftarrow)$

Suppose S is sequentially compact but not covering compact. Let $\mathcal{U} \doteq \{U_\alpha : \alpha \in \mathcal{A}\}$ be an open cover for S with no finite subcover. By the Lebesgue number lemma, $\mathcal{U}$ has a Lebesgue number $\lambda > 0$, so if we fix $x_1 \in S$, then there is some $U_1 \in \mathcal{U}$ such that $B_\lambda(x_1) \subset U_1$. Since $\mathcal{U}$ has no finite subcover, $\{U_1\}$ is not a cover for S. Therefore there is some $x_2 \in S \setminus U_1$. Since λ is a Lebesgue number, there is some $U_2 \in \mathcal{U}$ such that $B_\lambda(x_2) \subseteq U_2$. Continue in this way to construct a sequence $\{x_n\}$ in S and a collection $\{U_n\}_{n=1}^{\infty} \subseteq \mathcal{U}$ such that $B_\lambda(x_n) \in U_n$ and $x_{n+1} \in S \setminus (U_1 \cup \cdots \cup U_n)$ for each $n \in \mathbb{N}$. Observe that if $n \neq m$, then $d(x_n, x_m) > \lambda$.

Since S is sequentially compact, $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$. This implies that $\{x_{n_k}\}$ is Cauchy, which contradicts the fact that $d(x_n, x_m) > \lambda$ for all $n \neq m$. Therefore we can conclude that every open cover of S has a finite subcover. $\square$

Now that we have fully established the equivalence of these two characterizations, we will henceforth simply refer to sets as "compact." In some circumstances, the covering characterization will be useful and in other situations the sequential characterization will be more useful. To illustrate, we show two ways of showing that $[0, 1]$ is indeed compact:

Lemma 2.24. $[a, b] \subset \mathbb{R}$ is compact.

Proof. (By sequential compactness)

Let (x_n) be a sequence in $[a, b]$. Then (x_n) is bounded. It then follows from Corollary 1.19 that (x_n) has a convergent subsequence. The limit is clearly in $[a, b]$ since the set is closed. $\square$

Proof. (By covering compactness)

Let $\{U_\alpha\}$ be some open cover for $[a, b]$ and consider set

$$C = \{x \in [a, b] : \text{ finitely many } U_\alpha \text{ would suffice to cover the interval } [a, x]\}$$

Clearly, C is nonempty and bounded so $x^* = \sup C$ exists. We will show that $x^* = b$.

Suppose instead that $x^* < b$. Since $\{U_\alpha\}$ is an open cover for $[a, b]$, there is some α_0 such that $x^* \in U_{\alpha_0}$. Since U_{α_0} is open, there exists $r > 0$ such that $B_r(x^*) \subset U_{\alpha_0}$. Now pick some $y \in (x^* - r, x^*] \cap C$. Let $U_{\alpha_1}, \ldots, U_{\alpha_n}$ be those finitely many members of $\{U_\alpha\}$ that suffice to cover $[a, y]$. Next pick some $z \in (x^*, x^* + r)$. Since $[y, z] \subset B_r(x^*) \subset U_{\alpha_0}$, it suffices to cover $[a, z]$ with $U_{\alpha_0}, \ldots, U_{\alpha_n}$, which contradicts that $x^* = \sup C$. Therefore $x^* = b$ necessarily. $\square$

To gain a deeper insight into how compactness can make our lives easier, we show a result whose proof makes especially clear what role compactness can play.

Theorem 2.25. Let $(M, d), (M', d')$ be metric spaces and $f : M \to M'$ be continuous. If M is compact, then f is uniformly continuous.

Proof. Fix $\epsilon > 0$. We want to find $\delta > 0$ such that whenever $x, y \in M$ with $d(x, y) < \delta$, we have $d'(f(x), f(y)) < \epsilon$. By continuity of f, for each $x \in M$ we can find $\delta_x > 0$ such that whenever $d(x, y) < \delta_x$, we have $d'(f(x), f(y)) < \frac{\epsilon}{2}$.

Denote by $B(x)$ the open ball centered at x of radius $\frac{\delta_x}{2}$. Then the collection $\{B(x)\}_{x \in M}$ is an open cover for M, and since M is compact there is a finite collection of points $x_1, \ldots, x_n$ such that $M = B(x_1) \cup \ldots \cup B(x_n)$. Now define:

$$\delta = \min\left\{\frac{\delta_{x_1}}{2}, \frac{\delta_{x_2}}{2}, \ldots, \frac{\delta_{x_n}}{2}\right\}$$

Now suppose $x, y \in M$ with $d(x, y) < \delta$. Then $x \in B(x_i)$ for some $i = 1, \ldots n$, and so $d(x, x_i) < \frac{\delta_{x_i}}{2} < \delta_{x_i}$. But also $d(x_i, y) \leq d(x_i, x) + d(x, y) < \frac{\delta_{x_i}}{2} + \delta \leq \delta_{x_i}$. Thus we have $d(x, x_i) < \delta_{x_i}$ and $d(x_i, y) < \delta_{x_i}$, and the result follows from the triangle inequality on (M', d'). □

The essential power of compactness: letting us reduce the infinite to the finite. Without compactness we are faced with an infinite set of δ_x's (one for each point in the space), but with compactness we reduce it to a finite number in a single stroke, whereby we can pick a minimum value to give us uniform continuity.

Our final goal in this section is to introduce the powerful Heine-Borel theorem. The definition of compactness is somewhat useless in practice for a simple reason: it doesn't *really* tell us what compact sets look like in terms that are easy to grasp. The Heine-Borel theorem solves this for us by giving a neat characterization for compact sets. We present the result in full generality first, using the notion of totally bounded-ness.

Definition. (Totally bounded)

Let M be a metric space and $S \subset M$. Then S is **totally bounded** if for any $\epsilon > 0$, there exists a finite number of points $x_1, \ldots, x_n$ of M such that their corresponding open ϵ-balls cover S:

$$\bigcup_{i=1}^{n} B_\epsilon(x_i) \supset S$$

In layman's terms, if a set is totally bounded then it can be covered by a finite union of open balls of *uniform size*. Obviously, a totally bounded set is bounded. It is a little more difficult to clearly see why bounded sets are not necessarily totally bounded. A few examples will help build intuition:

Example. (Bounded but not totally bounded)

1. $(\mathbb{R}, d_0)$, the real line with the Euclidean metric constrained at 1:

$$d_0(x, y) = 1 \wedge |x - y|$$

2. (ℓ^∞, d_∞), the space of bounded infinite sequences with the supremum norm:

$$d_\infty(x, y) = \sup_n |x_n - y_n|$$

3. Any infinite space with the discrete metric

Totally bounded-ness has an obvious similarity to compactness in the sense of finite covers, but one can see from the examples that it is missing a key ingredient which will make the two equivalent. It doesn't take long to guess that the missing piece is completeness. In fact, by now it should be trivial to show that:

Lemma 2.26. Every compact set is complete.

Proof. Exercise. □

The full Heine-Borel theorem brings everything together:

Theorem 2.27. (Heine-Borel)

Let M be a metric space and let $S \subset M$. Then S is compact if and only if it is complete and totally bounded.

Proof. The proof of the "only if" direction is trivial once we use the above lemma, so we only show the "if" direction. Assuming S is complete and totally bounded, we will show that S is sequentially compact. So let (x_n) be a sequence in S. Since S is complete, it will be sufficient to show that (x_n) has a Cauchy subsequence.

Because S is totally bounded, we can cover it by finitely many open 1-balls. One (or more) of these balls, call it B_1, must contain infinitely many points of (x_n). Define $J_1 \subset \mathbb{N}$ to be the set of indices of the sequence points $x_n \in B_1$. Similarly, we can also cover S by finitely many open $\frac{1}{2}$-balls. Since J_1 is infinite, then one of these balls must contain x_n for infinitely many $n \in J_1$. Call this ball B_2 and the corresponding subset of natural numbers J_2 ($\subset J_1$).

Continiuing this process yields a sequence of open balls $\{B_n\}$ and a sequence of natural number sequences $\{J_n\}$, which is nested (i.e. $J_1 \supset J_2 \supset \ldots$). Thus construct a Cauchy subsequence like so: Pick any $n_1 \in J_1$. Then choose $n_2 \in J_2$ such that $n_2 \geq n_1$, which is possible since J_n is infinite for all n. Continue to obtain the subsequence $(x_{n_k})_{k\geq 1}$, which is Cauchy by construction of $\{B_n\}_{n\geq 1}$. $\square$

With this theorem we have all the machinery needed to completely characterize compact sets in terms we are more familiar with. One special case is particularly famous, and we state it here but only sketch the proof since it is not especially insightful:

Corollary 2.28. (Heine-Borel for $\mathbb{R}^n$)

$S \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. (Sketch)

We show that S satisfies the conditions of the general Heine-Borel theorem. We know that $\mathbb{R}^n$ is complete, and we know that closed subsets of complete spaces are again complete. So all one needs to show is that S is totally bounded. To do this, we may simplify the task. If S is bounded then it is contained in some closed ball centered at the origin, with radius $M < \infty$. Thus if one can show this, we are done.

Fix $\epsilon > 0$ and divide $\mathbb{R}^n$ into an equally-spaced grid of points of form $(\frac{a_1}{m}, \frac{a_2}{m}, \ldots, \frac{a_n}{m})$ for some $m \in \mathbb{N}$. Select m small enough so that balls of radius ϵ around these points covers the M-ball. □

Finally we close with the important corollary, whose proof we leave as an exercise.

Theorem 2.29. (Bolzano-Weierstrass)

Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Proof. Exercise. □

2.6 Sequences of Functions

In the study of probability and statistics, we will frequently be concerned with metric spaces of functions. Understanding some of the basic properties of these spaces will be extremely helpful in your first year courses. In this section we will only discuss real-valued functions but the results can easily be extended to functions with more complex range spaces. We begin with a simple definition.

Definition. Let $\{f_n\}$ be a sequence of a real-valued functions defined on a set M.

Suppose the sequence $\{f_n(x)\}$ converges for every $x \in M$. The function f defined by,

$$f(x) = \lim_{n\to\infty} f_n(x), \qquad (x \in M)$$

is said to be the **limit function** of $\{f_n\}$ and we say the $\{f_n\}$ converges to f **pointwise** on M.

Here are a few examples.

Example. 1. Define the sequence of functions $\{f_n\}$ from $\mathbb{R} \to \mathbb{R}$ as $f_n(x) = e^{-n|x|}$. We can see that f_n converges pointwise in $\mathbb{R}$ to f defined by

$$f(x) = \lim_{n\to\infty} f_n(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases}$$

2. Define the sequence of function $\{f_n\}$ from $(0,1) \to \mathbb{R}$ as $f_n(x) = x^n$. Then $\{f_n\}$ converges pointwise in $(0,1)$ to $f(x) = 0$. It is important to note that the sequence of functions would not converges if we had chosen $\mathbb{R}$ as our domain space.

We are frequently concerned with what properties are preserved in the limit of a series of functions. As we can see from the examples above, continuity is not preserved through pointwise convergence. This leads us to a second mode of convergence.

Definition. We say that a sequence of functions $\{f_n\}$ converges **uniformly** on a set M to a function f if for every $\epsilon > 0$ there exists an integer N such that $n \geq N$ implies

$$\sup_{x\in M} |f_n(x) - f(x)| \leq \epsilon.$$

For pointwise convergence we require that for every ϵ there exists a $N(x, \epsilon)$ *depending on* x such that $|f_n(x) - f(x)| < \epsilon$ for $n > N$. For uniform convergence we can select one $N(\epsilon)$ which holds for all $x \in M$. Uniform convergence is stronger than pointwise convergence by which we mean that uniform convergence implies pointwise convergence.

Example.

1. The first example above gives a sequence of function which converges pointwise but not uniformly.
2. Consider the sequence functions $f_n(x) = xe^{-nx}$ defined on the set $[0, \infty)$. This function converges uniformly to $f(x) = 0$. To see this we can use elementary calculus to show that $f_n(x)$ achives it maximum at $x = 1/n$ and thus has a maximum of $f_n(1/n) = e^{-1}/n$. Therefore,

$$\sup_{x \in [0,\infty)} |f_n(x) - 0| = \sup_{x \in [0,\infty)} |f_n(x)| = \frac{e^{-1}}{n} \to 0.$$

The following theorem will show that continuity is preserved under uniform convergence.

Theorem 2.30. If $\{f_n\}$ is a sequence of continuous function on M, and if $f_n \to f$ uniformly on M, then f is continuous on M.

Proof. Let $\{x_n\}$ be a sequence in M such that $x_n \to x \in M$. Fix $\epsilon > 0$. Since $f_m \to f$ uniformly there exists an $m > 0$ such that $|f_m(y) - f(y)| < \epsilon/3$ for all $y \in M$ including $y = x$ and $y = x_i,\ i = 1, 2, \ldots$. We know that f_m is continuous so there must exists a $N > 0$ such that $n > N$ implies $|f_m(x_n) - f_m(x)| < \epsilon/3$. Then the triangle inequality implies,

$$|f(x_n) - f(x)| \leq |f(x_n) - f_m(x_n)| + |f_m(x_n) - f_m(x)| + |f_m(x) - f(x)| < \epsilon. \qquad \square$$

This leads to the definition of a very important metric space.

Definition. Let M be a set of real valued functions. The **sup norm** on this function set is defined as $\|f\| = \sup_x |f(x)|$. This induces a metric on M as follows. For $f, g \in M$, $d(f,g) = \|f-g\| = \sup_x |f(x) - g(x)|$.

Theorem 2.31. The set of continuous functions defined on $\mathbb{R}$ denoted $\mathcal{C}(\mathbb{R})$ with the metric induced by the sup norm defined above is a complete metric space.

Proof. Exercise. $\square$

2.7 Compactness in Function Spaces

Recall that we denote the collection of continuous functions from $[0,1]$ to $\mathbb{R}$ by $C[0,1]$. As we showed previously, when endowed with the metric

$$d(f,g) \doteq \|f-g\|_\infty = \sup_{x\in[0,1]} |f(x) - g(x)|,$$

the space $C[0,1]$ is complete. Additionally, one can readily verify $C[0,1]$ is a vector space, and therefore is a Banach Space.

Definition. A normed vector space $(V, \|\cdot\|)$ is a **Banach Space** if it is complete when endowed with the metric induced by $\|\cdot\|$.

Clearly, $C[0,1]$ is not compact. However it is less clear whether the unit ball in $C[0,1]$ is compact. Namely, is the set

$$B \doteq \{f \in C[0,1] : \|f\|_\infty \leq 1\},$$

compact? Before we answer this question, note that $C[0,1]$ is an infinite dimensional normed space[2]. In order to see this, let $\mathcal{P}^n[0,1]$ denote the space of polynomials on $[0,1]$ with real coefficients whose degree is no more than n. Then $\mathcal{P}^n[0,1]$ is a subspace of $C[0,1]$ for each n and the dimension of $\mathcal{P}^n[0,1]$ is $n+1$. The following lemma is used in our study of compactness.

Lemma 2.32. Let V be a Banach space and $W \subseteq V$ a proper, closed subspace of V. Then there is a sequence $\{v_n\} \subseteq V$ such that $\|v_n\| = 1$ for all n, and $d(v_n, W) \to 1$ as $n \to \infty$, where

$$d(v_n, W) \doteq \inf_{w \in W} d(v_n, w).$$

In particular, for each $\epsilon \in (0,1)$, there is some $v \in V$ such that $\|v\| = 1$ and $d(v, W) \geq \epsilon$.

Proof. Let $v \in V \setminus W$ and suppose towards a contradiction that $d(v, W) = 0$. Then there is a sequence $\{w_n\} \subseteq W$ such that $\|v - w_n\| \to 0$ as $n \to \infty$, which says that $w_n \to v$ as $n \to \infty$. However, W is closed, so this means that $v \in W$, which contradicts our assumption that $v \in V \setminus W$.

Now, let $\{w_n\} \subseteq W$ be a sequence such that $\|v - w_n\| \to d(v, W)$ as $n \to \infty$. Let $v_n \doteq (v - w_n)/\|v - w_n\|$, so that $\|v_n\| = 1$, and note that

$$\begin{aligned} d(v_n, W) &= \inf_{w \in W} \left\| \frac{v - w_n}{\|v - w_n\|} - w \right\| = \inf_{w \in W} \left\| \frac{v}{\|v - w_n\|} - w \right\| \\ &= \inf_{w \in W} \left\| \frac{v - w}{\|v - w_n\|} \right\| = \frac{\inf_{w \in W} \|v - w\|}{\|v - w_n\|} = \frac{d(v, W)}{\|v - w_n\|} \end{aligned}$$

The second and third equalities are due to the facts that $w_n/\|v - w_n\| \in W$ and $w - w/\|v - w_n\| \in W$, respectively. Letting $n \to \infty$, it follows that $d(v_n, W) \to 1$. The result follows. $\square$

[2]See the linear algebra notes for the defintion of the dimension of a linear space.

The following result shows that the unit ball is not compact in any infinite dimensional normed space (and therefore that the unit ball is not compact in $C[0,1]$).

Theorem 2.33. Let $(V, \|\cdot\|)$ be an infinite dimensional vector space. Then the unit ball $B \doteq \{v \in V : \|v\| \leq 1\}$ is not compact.

Proof. Recall that a set A in V is compact if and only if for every sequence $\{v_n\}$ in A, there is some $v \in V$ and some further subsequence $\{v_{n_k}\}$ such that $d(v_{n_k}, v) \to 0$ as $k \to \infty$. In order to show that B is not compact, we will construct a sequence in it that has no convergent subsequence.

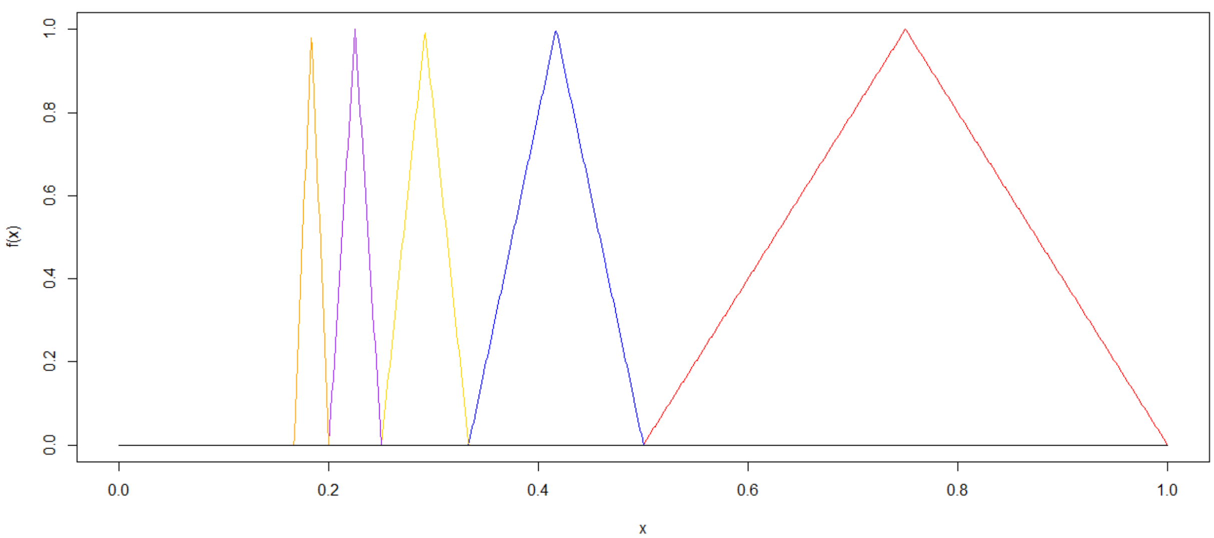
Fix $\epsilon \in (0,1)$ and let $V_1 \subseteq V_2 \subseteq V_3 \subseteq \ldots$ be subspaces of V with $\dim(V_n) = n$. Let $v_1 \in V_1$ be such that $\|v_1\| = 1$. Using the previous lemma, we can find some $v_2 \in V_2$ such that $\|v_2\| = 1$ and $d(v_2, V_1) \geq \epsilon$. By repeating this process, we can find a sequence $\{v_n\} \subseteq V$ such that $\|v_n\| = 1$ for all n, and $d(v_{n+1}, V_n) \geq \epsilon$. Now, suppose without loss of generality that $n > m$. Then $V_m \subseteq V_{n-1}$, so we have that $\|v_n - v_m\| \geq d(v_n, V_m) \geq d(v_n, V_{n-1}) \geq \epsilon$. This shows that $\{v_n\}$ has no convergent subsequences (as no subsequence is Cauchy), so the result follows. $\square$

The previous theorem shows that the unit ball in $C[0,1]$ is not compact. However, we can also explicitly show this by defining a sequence $\{f_n\}$ in $C[0,1]$ that has no convergent subsequences.

Example. For each $n \in \mathbb{N}$, let $f_n : [0,1] \to \mathbb{R}$ be given by

$$f_n(x) = \begin{cases} 0 & 0 \leq x \leq \frac{1}{n+1} \\ 2n(n+1)x - 2n & \frac{1}{n+1} \leq x \leq \frac{1}{2}\left(\frac{1}{n+1} + \frac{1}{n}\right) \\ -2n(n+1)x + 2(n+1) & \frac{1}{2}\left(\frac{1}{n+1} + \frac{1}{n}\right) \leq x \leq \frac{1}{n} \\ 0 & \frac{1}{n} \leq x \leq 1. \end{cases}$$

The functions are plotted below for $n \leq 5$.



Note that $f_n(x)$ is non-zero if and only if $\frac{1}{n+1} < x < \frac{1}{n}$. Therefore for each $m \neq n$, $d(f_n, f_m) = 1$, so $\{f_n\}$ has no convergent subsequence.

The following lemma says that any two norms on a finite dimensional vector space are equivalent.

Lemma 2.34. Let V be a finite dimensional vector space. Let $\| \cdot \|_1$ and $\| \cdot \|_2$ be two norms on V. Then there are constants $C_1, C_2 > 0$ such that

$$C_1\|v\|_1 \leq \|v\|_2 \leq C_2\|v\|_1$$

for all $v \in V$.

Proof. Exercise. $\square$

The next theorem says that the closed unit ball is compact in any finite dimensional vector space. This illustrates a fundamental difference between finite and infinite dimensional normed spaces.

Theorem 2.35. Let $(V, \|\cdot\|_V)$ be a finite dimensional vector space. Then the unit ball $\{\boldsymbol{v} \in V : \|\boldsymbol{v}\|_V \leq 1\}$ is compact.

Proof. Let $d = \dim(V)$, and let $\{v_1, \ldots, v_d\}$ be a basis for V. Let $\{r_1, \ldots, r_d\}$ be a basis for $\mathbb{R}^d$. Define the map $\varphi : V \to \mathbb{R}^d$ by

$$\varphi\left(\sum_{i=1}^{d} \alpha_i v_i\right) = \sum_{i=1}^{d} \alpha_i r_i, \quad \alpha_1, \ldots, \alpha_d \in \mathbb{R}.$$

Note that φ is a linear bijection. Now, let $\{\boldsymbol{u}_n\}_{n=1}^{\infty}$ be a sequence in $\{\boldsymbol{v} \in V : \|\boldsymbol{v}\|_V \leq 1\}$. Consider the norm on $\mathbb{R}^d$ given by

$$\|\boldsymbol{r}\| \doteq \left\|\varphi^{-1}(\boldsymbol{r})\right\|_V, \quad \boldsymbol{r} \in \mathbb{R}^d.$$

Then $\{\varphi(\boldsymbol{u}_n)\}_{n=1}^{\infty}$ belongs to the unit ball in $(\mathbb{R}^d, \|\cdot\|)$. Using our previous lemma, we can find some $C_1, C_2 > 0$ such that $C_1\|\boldsymbol{r}\| \leq \|\boldsymbol{r}\|_2 \leq C_2\|\boldsymbol{r}\|$ for all $\boldsymbol{r} \in \mathbb{R}^d$, where $\|\cdot\|_2$ denotes the standard Euclidean norm on $\mathbb{R}^d$. This implies that $\{\varphi(\boldsymbol{u}_n)\}_{n=1}^{\infty}$ in contained in $\{\boldsymbol{r} \in \mathbb{R}^d : \|\boldsymbol{r}\|_2 \leq C_2\}$, and therefore has a convergent subsequence in $(\mathbb{R}^d, \|\cdot\|_2)$. Denote this subsequence by $\{\varphi(\boldsymbol{u}_{n_k})\}_{k=1}^{\infty}$, and its limit (under $\|\cdot\|_2$) by $\boldsymbol{\eta}$. Again, since $\|\cdot\|_2$ and $\|\cdot\|$ are equivalent, we have that $\{\varphi(\boldsymbol{u}_{n_k})\}_{k=1}^{\infty}$ converges to $\boldsymbol{\eta}$ in $(\mathbb{R}^d, \|\cdot\|)$. It just remains to be seen that $\{\boldsymbol{u}_{n_k}\}_{k=1}^{\infty}$ converges to $\varphi^{-1}(\boldsymbol{\eta})$ in $(V, \|\cdot\|_V)$. Note that for each $\boldsymbol{v} \in V$, we have

$$\|\boldsymbol{v}\|_V = \|\varphi^{-1}(\varphi(\boldsymbol{v}))\|_V = \|\varphi(\boldsymbol{v})\|,$$

so

$$\|\boldsymbol{u}_{n_k} - \varphi^{-1}(\boldsymbol{\eta})\|_V = \|\varphi(\boldsymbol{u}_{n_k} - \varphi^{-1}(\boldsymbol{\eta}))\| = \|\varphi(\boldsymbol{u}_{n_k}) - \boldsymbol{\eta}\|,$$

which tends to 0 as $k \to \infty$. Therefore $\{\boldsymbol{u}_{n_k}\}_{k=1}^{\infty}$ converges to $\varphi^{-1}(\boldsymbol{\eta})$ as $k \to \infty$, so the result follows. □

Since the unit ball is not compact in $C[0,1]$, it is natural to wonder which sets are compact. Clearly, any finite set is compact in $C[0,1]$, but what about more general sets? For a compact metric space (M, d), let $C(M)$ denote the set of continuous functions from M to $\mathbb{R}$. Recall the following definitions.

> **Definition.** Let M be a compact metric space and let $\mathcal{F} \subseteq C(M)$. Then $\mathcal{F}$ is **uniformly bounded** if there is some $C < \infty$ such that
>
> $$\sup_{f \in \mathcal{F}} \|f\|_\infty \doteq \sup_{f \in \mathcal{F}} \sup_{x \in M} |f(x)| \leq C.$$
>
> Additionally, $\mathcal{F}$ is **equicontinuous** if for each $\epsilon > 0$ there is some $\delta > 0$ such that for all $x, y \in M$ such that $d(x,y) < \delta$, we have $\sup_{f \in \mathcal{F}} |f(x) - f(y)| < \epsilon$. We say that $\mathcal{F}$ is **relatively compact** if its closure $\bar{\mathcal{F}}$ is compact.

Note that $\mathcal{F}$ is relatively compact if and only if for each sequence $\{f_n\} \subseteq \mathcal{F}$, there is some f and some subsequence $\{f_{n_k}\}$ such that $\|f_{n_k} - f\|_\infty \to 0$ as $k \to \infty$. The function f need not belong to $\mathcal{F}$, namely if $\mathcal{F}$ is not closed, then we may have $f \in \bar{\mathcal{F}} \setminus \mathcal{F}$.

We now recall the notion of a separable metric space.

Definition. Let (M, d) be a metric space. A subset $S \subseteq M$ is **dense** in M if for each $x \in M$ and $\epsilon > 0$, $B_\epsilon(x) \cap S \neq \emptyset$. Equivalently, S is dense in M if $\bar{S} = M$. We say that M is **separable** if it has a countable dense subset.

Example. ($\mathbb{R}$ is separable)

Recall that for each $x \in \mathbb{R}$, there is some sequence $\{q_n\}_{n=1}^{\infty}$ in $\mathbb{Q}$ such that $q_n \to x$ as $n \to \infty$. This says that $\mathbb{Q}$ is dense in $\mathbb{R}$. Additionally, $\mathbb{Q}$ is countable, so $\mathbb{R}$ is separable.

The following lemma will be used in the proof of the Arzela-Ascoli Theorem.

Lemma 2.36. Let (M, d) be a compact metric space. Then M is separable.

Proof. Note for each $n \in \mathbb{N}$, the collection $\{B_{\frac{1}{n}}(x)\}_{x \in M}$ is an open cover for M. Since M is compact, for each $n \in \mathbb{N}$, there is a finite subcover $\mathcal{U}_n \doteq \{B_{\frac{1}{n}}(x_1^n), B_{\frac{1}{n}}(x_2^n), \ldots, B_{\frac{1}{n}}(x_{m(n)}^n)\}$ of $\{B_{\frac{1}{n}}(x)\}_{x \in M}$. Let $S \doteq \bigcup\limits_{n=1}^{\infty} \{x_1^n, \ldots, x_{m(n)}^n\}$, and observe that S is countable. Additionally, if we fix $\epsilon > 0$, $x \in M$, and $n \in \mathbb{N}$ such that $\frac{1}{n} < \epsilon$, then using the fact that $\mathcal{U}_n$ covers M, we can find some $k \in \{1, \ldots, m(n)\}$ such that $x \in B_{\frac{1}{n}}(x_k^n)$. Owing to our choice of n, we can conclude that $x_k^n \in B_\epsilon(x)$, and therefore that S is dense in M. $\square$

The Arzela-Ascoli Theorem provides necessary and sufficint conditions to determine whether a set in $C(M)$ is relatively compact.

Theorem 2.37. (Arzela-Ascoli)

Let (M, d) be a compact metric space. Then $\mathcal{F} \subseteq C(M)$ is relatively compact if and only if it is uniformly bounded and equicontinuous.

Proof. Suppose that $\mathcal{F} \subseteq C(M)$ is relatively compact and fix $\epsilon > 0$. It follows from the Heine-Borel theorem that $\bar{\mathcal{F}}$ is complete and totally bounded. Therefore there is a finite subset $\{f_n\}_{n=1}^N$ of $C(M)$ such that

$$\mathcal{F} \subseteq \bar{\mathcal{F}} \subseteq \bigcup_{n=1}^{N} B_{\frac{\epsilon}{3}}(f_n).$$

In order to see that $\mathcal{F}$ is uniformly bounded, let $C \doteq \sup\limits_{1 \leq n \leq N} \|f_n\|_\infty + \frac{\epsilon}{3}$. Then for each $f \in \mathcal{F}$, there is some $n \in \{1, \ldots, N\}$ such that $\|f - f_n\|_\infty \leq \frac{\epsilon}{3}$, which ensures that

$$\|f\|_\infty \leq \|f_n - f\|_\infty + \|f_n\|_\infty \leq C.$$

We now show that $\mathcal{F}$ is equicontinuous. Since M is compact and each f_n is continuous, we know that each f_n is in fact uniformly continuous. Therefore for each $n \in \{1, \ldots, N\}$, there is some $\delta_n > 0$ such that if $d(x, y) < \delta$ then $|f_n(x) - f_n(y)| < \frac{\epsilon}{3}$. Let $\delta \doteq \min\limits_{1 \leq n \leq N} \delta_n$ and note that if $d(x, y) < \delta$, then for all $n \in \{1, \ldots, N\}$, $|f_n(x) - f_n(y)| < \frac{\epsilon}{3}$. Additionally, for each $f \in \mathcal{F}$, there is some $n \in \{1, \ldots, N\}$ such that

$$\|f - f_n\|_\infty < \frac{\epsilon}{3},$$

so if we have $d(x, y) < \delta$, then

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &\leq \|f - f_n\|_\infty + \frac{\epsilon}{3} + \|f - f_n\|_\infty \\ &< \epsilon. \end{aligned}$$

Therefore $\mathcal{F}$ is equicontinuous.

Now suppose that $\mathcal{F}$ is uniformly bounded and equicontinuous. Let $\{f_n\}$ be a sequence in $\mathcal{F}$. Our previous lemma ensures that M is separable, so it has a countable dense subset $S \doteq \{x_n\}_{n=1}^{\infty}$. Since the sequence $\{f_n(x_1)\}_{n=1}^{\infty}$ in $\mathbb{R}$ is bounded, it has a convergent subsequence $\{f_{1,n}(x_1)\}_{n=1}^{\infty}$. Note that each $f_{1,n}$ denotes some function from the original sequence. Consequently, using the fact that the sequence $\{f_{1,n}(x_2)\}_{n=1}^{\infty}$ in $\mathbb{R}$ is bounded, we can find a further subsequence $\{f_{2,n}(x_2)\}_{n=1}^{\infty}$ which converges. Note that the sequence $\{f_{2,n}\}_{n=1}^{\infty}$ is convergent at x_1 and x_2. The convergence at x_1 is due to the fact that it is a subsequence of our original subsequence $\{f_{1,n}(x_1)\}_{n=1}^{\infty}$. Continuing this process, for each $k \in \mathbb{N}$ we can find a subsequence $\{f_{k,n}\}_{n=1}^{\infty}$ of $\{f_{k-1,n}\}_{n=1}^{\infty}$ that converges at x_i for all $i \in \{1, \ldots, k\}$. We now consider the sequence $\{f_{n,n}\}$. Note that the sequence $\{f_{n,n}\}$ converges at x_i for each $i \in \mathbb{N}$, and recall that $S \doteq \{x_n\}_{n=1}^{\infty}$ is dense in M. This means that we have constructed a sequence $\{f_{n,n}\}_{n=1}^{\infty}$ that converges along a countable dense subset of M. It just remains to be seen that $\{f_{n,n}\}_{n=1}^{\infty}$ in uniformly convergent on M. Note that if we view $\{f_{n,k}\}_{n,k=1}^{\infty}$ as an infinite array, then the sequence $\{f_{n,n}\}_{n=1}^{\infty}$ consists of the diagonal elements of the array.[3]

For notational simplicity, let $g_n \doteq f_{n,n}$. Fix $\epsilon > 0$ and using the fact that $\{f_n\}_{n=1}^{\infty}$ is equicontinuous, choose $\delta > 0$ such that if $d(x, y) < \delta$, then

$$|g_n(x) - g_n(y)| < \frac{\epsilon}{3},$$

for all $n \in \mathbb{N}$. As in the proof of the previous lemma, let S_δ denote a finite subset of S such

[3] Arguments of this sort are often referred to as diagonalization arguments.

that

$$M = \bigcup_{s \in S_\delta} B_\delta(s).$$

Since $\{g_n\}$ converges at each point in S_δ (and therefore is Cauchy at each point in S_δ) and S_δ is finite, there is some $N \in \mathbb{N}$ such that for all $n, m \geq N$,

$$|g_n(s) - g_m(s)| < \frac{\epsilon}{3}, \tag{4}$$

for all $s \in S_\delta$. Now, fix $x \in M$. Then there is some $s \in S_\delta$ such that $d(x, s) < \delta$, so if $n, m > N$, then

$$|g_n(x) - g_m(x)| \leq |g_n(x) - g_n(s)| + |g_n(s) - g_m(s)| + |g_m(s) - g_m(x)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}.$$

The first and third inequalities come from the fact that $d(x, s) < \delta$ (and our choice of δ via equicontinuity). The second inequality is due to our choice of $m, n > N$ and (4). We have shown that if $n, m > N$, then

$$|g_n(y) - g_m(y)| < \epsilon,$$

for all $y \in M$. In order to see that $\{g_n\}_{n=1}^\infty$ is uniformly convergent on M, note that there is some N_1 such that if $m, n \geq N_1$, then for all $y \in M$

$$|g_n(y) - g_m(y)| < \frac{\epsilon}{2}.$$

This means that $\{g_n\}_{n=1}^\infty$ is uniformly Cauchy on M. For each $y \in M$, let $g(y) \doteq \lim_{n\to\infty} g_n(y)$, and note that this limit exists since $\mathbb{R}$ is complete. Fix $n \geq N_1$ and let $\ell(k) \doteq n + k$ for each $k \in \mathbb{N}$. Then for each $y \in M$,

$$|g_n(y) - g(y)| = \lim_{k\to\infty} |g_n(y) - g_{\ell(k)}(y)| \leq \frac{\epsilon}{2} < \epsilon.$$

Consequently, whenever $n \geq N_1$,

$$\|g_n - g\|_\infty = \sup_{y \in M} |g_n(y) - g(y)| < \epsilon,$$

so we can conclude that $\|g_n - g\|_\infty \to 0$ as $n \to \infty$. The result follows. □

Example. For $\alpha, \beta > 0$ let

$$\mathcal{F}_{\alpha,\beta} \doteq \{f \in C[0,1] : \|f\|_\infty \leq \alpha,\ f \text{ is differentiable, and } \|f'\|_\infty \leq \beta\}.$$

Note that $\mathcal{F}_{\alpha,\beta}$ is the set of differentiable functions which are uniformly bounded by α and whose derivatives are uniformly bounded by β. Then $\mathcal{F}_{\alpha,\beta}$ is relatively compact.

Proof. Exercise. □

Example. For $\alpha > 0$, let

$$\mathcal{F}_\alpha \doteq \{f \in C[0,1] : \|f\|_\infty \leq \alpha, \text{ and } f \text{ is differentiable}\}.$$

Is the unit ball compact in $\mathcal{F}_\alpha$?

2.8 Problems with Riemann integrals

Recall from undergraduate probability that for a continuous random variable X with probability density function (pdf) f, we have that for $a < b$,

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \in [a,b]) = \int_a^b f(x)dx,$$

where the integral is a Riemann integral. This suggests that if we would like to calculate the probability that X belongs to some $A \subseteq \mathbb{R}$, then we "should" have

$$\mathbb{P}(X \in A) = \int_A f(x)dx, \tag{5}$$

where, once again, the integral is a Riemann integral. While this formula will work for many sets in $\mathbb{R}$ (for example, sets of the form $(a, b), [c, d)$, and so on), in general it is not capable of handling many of the situations that one might encounter in statistics and probability. For example, let $X \sim \text{Unif}(0, 1)$ be a random variable which is uniformly distributed over $(0, 1)$. The pdf of X is given by

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that we would like to calculate the probability that X is a rational number. Recall from undergraduate probability that if $\{A_n\}_{n=1}^{\infty}$ is a sequence of disjoint events, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

If we enumerate the rational numbers in $[0, 1]$ by $\mathbb{Q} \cap [0, 1] = \{q_n\}_{n=1}^{\infty}$ and let $A_n \doteq \{X = q_n\}$, then the $\{A_n\}_{n=1}^{\infty}$ are disjoint, and

$$\mathbb{P}(X \in \mathbb{Q} \cap [0, 1]) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \mathbb{P}(X = q_n) = 0,$$

where the final equality is due to the fact that X is a continuous random variable. Let us try to show this using the pdf of X. Recall the function $1_{\mathbb{Q}} : [0, 1] \to \mathbb{R}$ given by

$$1_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q}, \end{cases}$$

and observe that if (5) held with $A = \mathbb{Q} \cap [0, 1]$, then we would have

$$\mathbb{P}(X \in \mathbb{Q} \cap [0, 1]) = \int_{\mathbb{Q} \cap [0,1]} 1 dx = \int_0^1 1_{\mathbb{Q}}(x) dx.$$

We now show that the integral above is not well-defined as a Riemann integral. Recall that a function $\phi : [0, 1] \to \mathbb{R}$ is Riemann integrable if and only if for any $\epsilon > 0$ there is a partition

$P \doteq \{x_0 = 0 < x_1 < \cdots < x_{n-1} < x_n = 1\}$ of $[0,1]$ such that

$$U(\phi, P) - L(\phi, P) < \epsilon,$$

where

$$L(\phi, P) \doteq \sum_{i=0}^{n-1} \inf_{t \in [x_i, x_{i+1}]} \phi(t)(x_{i+1} - x_i),$$

and

$$U(\phi, P) \doteq \sum_{i=0}^{n-1} \sup_{t \in [x_i, x_{i+1}]} \phi(t)(x_{i+1} - x_i).$$

However, for any partition $P \doteq \{x_0 = 0 < x_1 < \cdots < x_{n-1} < x_n = 1\}$ of $[0,1]$, we have

$$\inf_{t \in [x_i, x_{i+1}]} 1_{\mathbb{Q}}(t) = 0$$

for all $0 \leq i \leq n-1$, as there is an irrational number in each interval $[x_i, x_{i+1}]$. Therefore

$$L(1_{\mathbb{Q}}, P) \doteq \sum_{i=0}^{n-1} \inf_{t \in [x_i, x_{i+1}]} 1_{\mathbb{Q}}(t)(x_{i+1} - x_i) = \sum_{i=0}^{n-1} 0(x_{i+1} - x_i) = 0.$$

Similarly, there is a rational number in each interval $[x_i, x_{i+1}]$, so we have

$$\sup_{t \in [x_i, x_{i+1}]} 1_{\mathbb{Q}}(t) = 1,$$

and it follows that

$$U(1_{\mathbb{Q}}, P) \doteq \sum_{i=0}^{n-1} \sup_{t \in [x_i, x_{i+1}]} 1_{\mathbb{Q}}(t)(x_{i+1} - x_i) = \sum_{i=0}^{n-1} (x_{i+1} - x_i) = x_n - x_0 = 1.$$

Thus for each partition P of $[0,1]$ we have

$$U(1_{\mathbb{Q}}, P) - L(1_{\mathbb{Q}}, P) = 1,$$

so $1_{\mathbb{Q}}$ is not Riemann-integrable. In order to remedy this "problem", so that we can express probabilities such as $\mathbb{P}(X \in \mathbb{Q} \cap [0,1])$ as integrals, one needs to develop a different notion of integration. The Lebesgue integral, which is covered in a measure theory course, reconciles this, as we have

$$\int_0^1 1_{\mathbb{Q}} dx = 0,$$

when the integral is interpreted in the sense of Lebesgue. This example also illustrates some of the difficulties that arise when one wants to study the limiting behavior of the Riemann integrals of a sequence of functions.

Example. Enumerate the rational numbers in $\mathbb{Q} \cap [0,1]$ by $\{q_n\}_{n=1}^{\infty}$. For each $n \in \mathbb{N}$, let $f_n : [0,1] \to \mathbb{R}$ be given by

$$f_n(x) \doteq 1_{\{q_1,\ldots,q_n\}}(x).$$

Then f_n converges pointwise to the function $f : [0,1] \to \mathbb{R}$ given by

$$f(x) \doteq 1_{\mathbb{Q}}(x).$$

Furthermore, each f_n has only a finite number of discontinuities and therefore is Riemann integrable. Additionally, each f_n is bounded, and f is bounded as well, so the integrals do not diverge as $n \to \infty$. However, we have

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx = 0,$$

but even in this simple case, we are unable to exchange the limit with the integral, since

$$\int_0^1 \lim_{n\to\infty} f_n(x)dx = \int_0^1 f(x)dx,$$

is not defined. Some key theorems from measure theory ensure that this problem does not arise when one works with Lebesgue integrals.

We now discuss a more mathematical motivation for studying the Lebesgue integral. For $p \geq 1$, consider the vector space

$$R^p[0,1] \doteq \left\{ f : [0,1] \to \mathbb{R} : f \text{ is Riemann integrable and } \int_0^1 |f(x)|^p dx < \infty \right\}.$$

Consider the "norm" on $R^p[0,1]$ given by

$$\|f\|_p \doteq \left(\int_0^1 |f(x)|^p dx \right)^{1/p}.$$

We say "norm" because there are functions $f \in R^p[0,1]$ such that $f(x) \neq 0$ for some $x \in [0,1]$, but $\|f\|_p \neq 0$. However, for the purposes of this example, we can ignore this technical detail, and henceforth will refer to $R^p[0,1]$ as a normed vector space. It is not immediately clear that $\|\cdot\|_p$ defines a norm. In particular, the triangle inequality is not obvious. However, as you will see in other courses, the triangle inequality does hold here. We can endow $R^p[0,1]$ with the metric d_p given by

$$d_p(f,g) \doteq \|f - g\|_p, \quad f, g \in R^p[0,1].$$

As we have mentioned, complete spaces are convenient to work with, so it is natural to wonder whether $R^p[0,1]$ is complete. The following example shows that $R^2[0,1]$ is not complete, which suggests that another notion of integration is needed in order to sure that the space of square "integrable" functions is complete.

Example. For each $n \in \mathbb{N}$, define $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = x^{-1/4} 1_{[1/n,1]}(x),$$

where $1_{[1/n,1]}$ denotes the indicator function of $[1/n, 1]$. Note that $f_n \in R^2[0,1]$ for each

$n \in \mathbb{N}$. For each $m, n \in \mathbb{N}$ such that $n < m$, we have

$$f_m(x) - f_n(x) = x^{-1/4} 1_{[1/m,1/n)}.$$

Observe that

$$\|f_m - f_n\|_2 = \left(\int_0^1 \left| x^{-1/4} 1_{[1/m,1/n)} \right|^2 dx \right)^{1/2} = \left(\int_{1/m}^{1/n} x^{-1/2} dx \right)^{1/2} = \left(\frac{2}{\sqrt{n}} - \frac{2}{\sqrt{m}} \right)^{1/2},$$

so $\{f_n\}$ is Cauchy in $R^2[0,1]$. However, $\{f_n\}$ does not converge in $R^2[0,1]$. In order to see that, observe that the only possible limit (why?) is $f(x) \doteq x^{-1/4}$. However, f is unbounded and therefore is not Riemann integrable, so $\{f_n\}$ does not converge in $R^2[0,1]$.

Lemma 2.38. The space $R^p[0,1]$ is not complete for any $p \geq 1$.

Proof. Exercise. $\square$

As you may see in a measure theory course, the analogous spaces defined using another method of integration (the Lebesgue integral) are in fact complete. This is another deficiency of the Riemann integral that is remedied by the Lebesgue integral.

3 Introducton to Measure Theory

3.1 The Point of Measure Theory

If you are a graduate student in statisics or probability, you will eventually have to face up to the fact that measure theory is key to a deep understanding of your work. It offers a clever way to rigorously represent something which seems like it shouldn't fit into any mathematical framework. Also, the Lebesgue integration built on measure theory is much more convenient to use than Riemann integration.

Manipulating math statements is a deterministic exercise. How, then, can we use it to deal with random events? At the undergraduate level, the bread and butter of probability theory is the mass function or the density function. Using these, one can easily calculate probabilities, expectations, moments, etc. But it is easy to find situations where these are not enough:

> **Example.** (Random functions)
>
> Consider the space of continuous real-valued functions defined on the interval $[0,1]$, denote $C[0,1]$. Let X be a *random function* which takes values in this space.
>
> What is the "density" of X? For that matter what is the expectation of X?

One key limitation of densities and mass functions is that they operate on the level of *real-valued* random variables. But in statistics and probability, we frequently deal with more complex situations.

The $C[0,1]$ space is not a toy example. A famous random function, *the Brownian bridge*, lives in this space. But there are more concrete examples. In social network analysis we often study *random graphs* that express relationships between nodes. Often the set of nodes is in the millions, or the set of nodes changes across time. How then do we model these?

> **Example.** (Conditional probabilities?)
>
> Consider two random variables X, Y distributed iid $N(0,1)$. How do we calculate:
>
> $$\mathbb{P}(X \leq 0 \,|\, Y \leq 1) \text{ and } \mathbb{P}(X \leq 0 \,|\, Y = 1)?$$

We know how to calculate them using undergraduate probability formula:

$$\mathbb{P}(X \leq 0 \,|\, Y \leq 1) = \frac{\mathbb{P}(X \leq 0, Y \leq 1)}{\mathbb{P}(Y \leq 1)} = \frac{\int_{-\infty}^{0} \int_{-\infty}^{1} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x}{\int_{-\infty}^{1} f_Y(y)\,\mathrm{d}y}$$

and

$$\mathbb{P}(X \leq 0 \,|\, Y = 1) = \int_{-\infty}^{0} \frac{f_{X,Y}(x,1)}{f_Y(1)}\,\mathrm{d}x.$$

Here we cannot write $\mathbb{P}(X \leq 0 \,|\, Y = 1) = \frac{\mathbb{P}(X \leq 0, Y=1)}{\mathbb{P}(Y=1)}$ since $\mathbb{P}(Y = 1) = 0$. So what is going on behind the second calculation? What does a density really mean in terms of probabilities of events, and why does it work when we plug them into convenient formulas like this?

Another difficulty arises with the notion of conditional expectation.

> **Example.** (Conditional expectation) Let $\{X_n\}_{n=1}^{\infty}$ be iid mean 0 random variables. Let $X \doteq X_1 + X_2$. In undergraduate probability we show that
>
> $$\mathbb{E}(X|X_1) = \mathbb{E}(X_1 + X_2|X_1) = \mathbb{E}(X_1|X_1) + \mathbb{E}(X_2|X_1) = X_1 + \mathbb{E}(X_2) = X_1.$$
>
> In this case we have an intutitive sense of what the expected value of X given Z_1 "should" be. But how can we rigorously define the conditional expectation of one random variable given another? Fix $n \in \mathbb{N}$ and consider the random variable $S_n \doteq X_1 + \cdots + X_n$. Then can we evaluate $\mathbb{E}[X_1|S_n]$?

3.2 σ-algebras and measures

The basic concept in measure theory is *measure*. Roughly speaking, given any set S the measure is a special function that tells us how "big" certain subsets of that set are. Ultimately we will identify S with the set of all possible outcomes of a random variable (process, function, etc...), and the probability of any subset of outcomes $A \subset S$ will be the measure of A.

The key here is that a measure is a *set* function on S. We will see that measures are defined to satisfy certain properties. However, to avoid certain unpleasant pathologies, measures cannot be defined on *any* arbitrary collection of subsets of S.

3.2.1 The problem of "measure"

Let $2^{\mathbb{R}} \doteq \{A : A \subseteq \mathbb{R}\}$ denote the **power set** of $\mathbb{R}$. Suppose that we are interested in defining a function $\lambda : 2^{\mathbb{R}} \to [0, \infty]$ that assigns a "size" to each $A \subseteq \mathbb{R}$. Our intuition about how "size" should be defined leads us to the following requirements for λ:

(a) If $a < b$, then $\lambda((a, b)) = b - a$. Namely, the size of the interval (a, b) is simply its length.

(b) If $\{A_n\}_{n=1}^{\infty}$ is a collection of disjoint subsets of $\mathbb{R}$, then

$$\sum_{n=1}^{\infty} \lambda(A_n) = \lambda\left(\bigcup_{n=1}^{\infty} A_n\right).$$

(c) For $x \in \mathbb{R}$ and $A \subseteq \mathbb{R}$, let $x + A \doteq \{x + a : a \in A\}$. Then $\lambda(x + A) = \lambda(A)$. That is, λ is translation invariant.

The following theorem tells us that it is impossible to define such a function on every set $A \subseteq \mathbb{R}$.

Theorem 3.1. There is no function $\lambda : 2^{\mathbb{R}} \to [0, \infty]$ satisfying conditions (a), (b), and (c) above.

Proof. Suppose towards a contradiction that such a function λ does exist. Consider the relation on $(0,1)$ given by

$$x \sim y \iff x - y \in \mathbb{Q}.$$

It is clear that $\sim$ is an equivalence relation. Namely, the following hold for $x, y, z \in (0,1)$:

1. $x \sim x$.

2. If $x \sim y$, then $y \sim x$.

3. If $x \sim y$ and $y \sim z$, then $x \sim z$.

Let $(0,1)/\sim$ denote the collection of equivalence classes in $(0,1)$ under $\sim$. Note that any two distinct equivalence classes in $(0,1)/\sim$ are in fact disjoint. Using the axiom of choice, we can write

$$(0,1) = \bigcup_{\alpha \in I} E_\alpha,$$

where each E_α is a distinct equivalence class and I is some index set. For each $\alpha \in I$, let $e_\alpha \in E_\alpha$, and let

$$S \doteq \bigcup_{\alpha \in I} \{e_\alpha\}.$$

Then S contains exactly one element from each equivalence class. Our goal is to show that $\lambda(S)$ is not well defined.

Recall that $\mathbb{Q} \cap (-1,1)$ is countable, so we can enumerate it by $\{r_n\}_{n=1}^{\infty}$. For each $n \in \mathbb{N}$, let

$$S_n \doteq r_n + S \doteq \{r_n + s : s \in S\}.$$

We begin by showing that $(0,1) \subseteq \bigcup\limits_{n=1}^{\infty} S_n$. Let $x \in (0,1)$. Since

$$(0,1) = \bigcup_{\alpha \in I} E_\alpha,$$

there is some $\alpha \in I$ such that $x \in E_\alpha$. It follows that there is some $s_\alpha \in S$ such that $x \sim s_\alpha$. In particular, $x - s_\alpha \in \mathbb{Q}$. Additionally, $|x - s_\alpha| < 1$, so $x - s_\alpha \in (-1,1) \cap \mathbb{Q}$. Thus there is some $n \in \mathbb{N}$ such that $x - s_\alpha = r_n$. Therefore

$$x = r_n + s_\alpha \in S_n \subseteq \bigcup_{n=1}^{\infty} S_n,$$

so we have shown that $(0,1) \subseteq \bigcup\limits_{n=1}^{\infty} S_n$.

We now show that the S_n are disjoint. Suppose that $x \in S_m \cap S_n$. Then there are some $s_\alpha, s_\beta \in S$ such that

$$x = r_m + s_\alpha = r_n + s_\beta.$$

It follows that

$$s_\alpha - s_\beta = r_n - r_m \in \mathbb{Q},$$

so $s_\alpha \sim s_\beta$. However, S contains only one element from each equivalence class, so this shows that $s_\alpha = s_\beta$. Therefore $r_m = r_n$, so $S_m = S_n$. This shows that the S_n are disjoint.

By construction, $S_n \subseteq (-1,2)$ for each $n \in \mathbb{N}$, which, along with the fact that $(0,1) \subseteq \bigcup\limits_{n=1}^{\infty} S_n$, ensures that

$$(0,1) \subseteq \bigcup_{n=1}^{\infty} S_n \subseteq (-1,2).$$

Note that if $A \subseteq B \subseteq \mathbb{R}$, then $B \setminus A$ and A are disjoint, so (b) ensures that

$$\lambda(A) \leq \lambda(B \setminus A) + \lambda(A) = \lambda((B \setminus A) \cup A) = \lambda(B).$$

Therefore we can use (a) to see that

$$1 = \lambda((0,1)) \leq \lambda\left(\bigcup_{n=1}^{\infty} S_n\right) \leq \lambda((-1,2)) = 3. \tag{6}$$

Since the S_n are disjoint, we have that

$$\lambda\left(\bigcup_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} \lambda(S_n). \tag{7}$$

Additionally, (c) ensures that since $S_n = r_n + S$, for each $n \in \mathbb{N}$ we have

$$\lambda(S) = \lambda(S_n). \tag{8}$$

Combine (6), (7), and (8) to see that

$$1 \leq \sum_{n=1}^{\infty} \lambda(S) \leq 3.$$

If $\lambda(S) = 0$, then this shows that $1 \leq 0 \leq 3$, which is a contradiction. Similarly, if $\lambda(S) > 0$, then this shows that $1 \leq \infty \leq 3$, which is a contradiction. We conclude that no such λ exists. □

We have shown that there is no function that can assign "size" to every subset of $\mathbb{R}$. The function λ satisfying conditions (a), (b), and (c) is known as the Lebesgue measure - the previous theorem suggests that insread of assigning a Lebesgue measure to every subset of $\mathbb{R}$, we should instead instead assign it only to some subsets of $\mathbb{R}$. This idea is developed in the next section, where we introduce the abstract notions of measures and σ-algebras.

3.2.2 Addressing the problem of measure

Definition. (σ-algebra)

Given any set S, a **σ-algebra (also called a σ-field) on** S is a nonempty collection $\mathcal{S}$ of subsets of S such that:

1. $S \in \mathcal{S}$

2. It is closed under complements: if $E \in \mathcal{S}$, then $E^c \in \mathcal{S}$

3. It is closed under countable union: if $E_1, E_2, \ldots \in \mathcal{S}$, then $\cup_{i=1}^{\infty} E_i \in \mathcal{S}$

Note that S is *any* set. We have not specified whether this set is finite, countable, or uncountable. What's important is that given this set, we can always form a σ-algebra on that set.

The observant reader will notice that σ-algebras bear a resemblance to topologies on a set. In fact, the only difference between a topology on S and a σ-algebra on S is that a topology is closed under arbitrary unions, while a σ-algebra is closed only under countable unions.

Example. (Two stupid examples of σ-algebras)

Let S be any set. Two σ-algebras on S include:

1. The trivial σ-algebra: $\{\emptyset, S\}$

2. The power set of S: 2^S

The main point of σ-algebras is to allow us to define measures on subsets of S without running into technical issues. Therefore when we have a set S together with a σ-algebra $\mathcal{S}$ on that set, we call the pair a **measurable space** and denote it $(S, \mathcal{S})$.

Definition. (Measure)

A **measure** on a σ-algebra $\mathcal{E}$ (which contains the empty set $\emptyset$) of subsets of S is a function $\mu : \mathcal{E} \to \mathbb{R} \cup \{\infty\}$ with the following properties:

1. Non-negativity: $\mu(E) \geq 0$ for all $E \in \mathcal{E}$

2. Empty set has measure 0: $\mu(\emptyset) = 0$

3. Countable additivity: if $(E_n) \in \mathcal{E}$ is any sequence of *disjoint* sets and $\cup E_n \in \mathcal{E}$, then

$$\mu\Big(\cup_{n=1}^{\infty} E_n \Big) = \sum_{n=1}^{\infty} \mu(E_n)$$

If $S \in \mathcal{E}$ and $\mu(S) = 1$, then μ is called a **probability measure**.

A measurable space $(S, \mathcal{S})$ together with a measure μ on that space is called a **measure space** and is denoted $(S, \mathcal{S}, \mu)$. When spaces are small, we can define measures easily:

Example. (Simple space)

Let $(S, \mathcal{S})$ be defined as $S = \{1, 2, 3, 4\}$ and $\mathcal{S} = 2^S$. Two measures on $(S, \mathcal{S})$ include μ and ν defined by:

$$\mu(E) = \begin{cases} 3/4, & E = \{1\} \\ 0, & E = \{2\} \\ 1/4, & E = \{3\} \\ 0, & E = \{4\} \end{cases} \qquad \nu(E) = \begin{cases} 1/2, & E = \{1\} \\ 1/4, & E = \{2\} \\ 0, & E = \{3\} \\ 0, & E = \{4\} \end{cases}$$

But there are measures that can be defined on more general spaces. The simplest of these (and the most important, in some ways) is the counting measure:

Example. (Counting measure on a space)

Let $(S, \mathcal{S})$ be an arbitrary measurable space and define the the set function $\# : \mathcal{S} \to \mathbb{R}$ by $\#(E) = |E|, \quad E \in \mathcal{S}$. Then $\#$ is a measure on $\mathcal{S}$ and $(S, \mathcal{S}, \#)$ is a measure space.

We call # the **counting measure** on $(S, \mathcal{S})$.

We will discuss more advanced properties of these spaces in the next section. For now, we tie the concepts we just studied back to probability.

Definition. A **probability space** is a special measure space denoted $(\Omega, \mathcal{F}, \mathbb{P})$:

- Ω is an arbitrary set of **outcomes**.
- $\mathcal{F}$ is a σ-algebra on Ω, whose elements are **events**.
- $\mathbb{P}$ is a **probability measure** on $\mathcal{F}$

Example. (The probability space of tossing two coins)

Let T represents tails and H heads. Define $(\Omega, \mathcal{F}, \mathbb{P})$ by:

- $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$
- $\mathcal{F} = 2^{\Omega}$
- $\mathbb{P}(A) = \frac{1}{4} \cdot |A|$

Here it is key to emphasize that the probability measure $\mathbb{P}$ is a measure *on events, or subsets of outcomes*. Unless those events are themselves numbers (i.e. the space of outcomes is numbers), then it makes no sense to make statements like $\mathbb{P}(5) = 1$.

Later, we will introduce random variables and show how having a probability space will allow us to induce a new measure on the real line that will give us the ability to use undergraduate notation like $\mathbb{P}(X = 5) = 1/2$ rigorously.

3.3 Properties of measures and σ-algebras

Now that we have established that probabilities are really functions defined on sets, it is clear that basic set theory will be key to understanding some basic properties of σ-fields and measures. *It is assumed that you are familiar with basic set operations and concepts like DeMorgan's laws.*

Our first task is to define some set-theoretic analogues of real sequences and to prove some results about their convergence. In all that follows we assume that a sequence of sets are all subsets of the same unmentioned space, S.

Definition. (inf/sup of sets)

Let $(E_n)_{n\geq 1}$ be a sequence of sets. Then

$$\inf_{k\geq 1} E_k = \bigcap_{k\geq 1} E_k \quad \text{and} \quad \sup_{k\geq 1} E_k = \bigcup_{k\geq 1} E_k$$

To get some intuition for this definition, note that set inclusion $\subset$ induces a partial ordering of sets. Thus a rough set analogue of selecting a when $a < b$ for $a, b \in \mathbb{R}$ is to select $A \cap B$ when $A \subset B$.

Definition. (liminf/limsup/limit of sets)

Let $(E_n)_{n\geq 1}$ be a sequence of sets. Then the **liminf/limsup** of E_n are defined as:

$$\liminf_n E_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} E_k \quad \text{and} \quad \limsup_n E_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k$$

If $\liminf E_n = \limsup E_n$, then we say E_n **converges to** $\lim E_n$.

These definitions share many useful traits with their real analogues. In particular, as a sequence in n, $\inf_{k \geq n} E_k$ and $\sup_{k \geq n} E_k$ are both monotone sequences of sets.

When we interpret a sequence of sets as a sequence of events (as in probability), there is another intuitive way to understand the concept of limsup/liminf of sets:

Definition. (infinitely often, eventually)

Let $(E_n)_{n \geq 1}$ be a sequence of events (i.e. subsets of an outcome space, Ω).

1. The event consisting of outcomes happening infinitely often in the sequence (E_n) is:
$$\{E_n \text{ i.o.}\} = \limsup E_n$$
2. The event consisting of outcomes happening for all but finitely many times in (E_n) is:
$$\{E_n \text{ eventually}\} = \liminf E_n$$

To see why this connection makes sense, note that:

$$\begin{aligned}
\{A_n, \text{i.o.}\} &= \text{the event that } \forall N \geq 1,\, \exists n \geq N \text{ such that } A_n \text{ happens} \\
&= \text{the event that } \forall N \geq 1,\, (\cup_{n \geq N} A_n) \text{ happens} \\
&= \cap_{N \geq 1} (\cup_{n \geq N} A_n) \\
&= \limsup A_n
\end{aligned}$$

The exact same line of reasoning can be used to show the identity for $\{E_n \text{ eventually}\}$.

Example. Let $\Omega = \{\omega = (\omega_n)_{n \geq 1} : \omega_n = H \text{ or } T, n \geq 1\}$, and $A_n = \{w : w_n = H\} =$

"the event that the n'th trial is a head ". Then $\{A_n \text{ i.o.}\}$ is the event that there are infinitely many heads in the coin-tossing sequence and $\mathbb{P}\{A_n \text{ i.o.}\} = 1$.

Two simple results which will be useful are:

Theorem 3.2. Let (E_n) be a sequence of sets. Then $\liminf E_n \subset \limsup E_n$.

Proof. Suppose $\omega \in \liminf E_n$. Then there exists an $N \geq 1$ such that $\omega \in E_n$ for all $n \geq N$. Therefore $\omega \in E_n$ for infinitely many n and so $\omega \in \limsup E_n$. There result follows. $\square$

Theorem 3.3. Let $(E_n)_{n\geq 1}$ be an increasing (decreasing) sequence of sets.

(1) (E_n) is convergent and $\lim E_n = \cup_1^\infty E_n$ $(\cap_1^\infty E_n)$.

(2) Furthermore, if $E_n \in \mathcal{F}$ for some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathbb{P}(\lim E_n) = \lim_{n\to\infty} \mathbb{P}(E_n)$.

Proof. It suffices to prove above results for increasing sequences. For any decreasing sequence (F_n), the results follow from DeMorgan's law and the observation that (F_n^c) is increasing.

For (1), note that if $E_n \uparrow$, then $\limsup E_n = \cap_{n=1}^\infty \cup_{m=n}^\infty E_m = \cap_{n=1}^\infty \cup_{m=1}^\infty E_m = \cup_{m=1}^\infty E_m$. For the lower limit, note that if $E_n \uparrow$, then $\cap_{m=n}^\infty E_m = E_n$ and the result follows.

For (2), let $E_0 = \emptyset$ and $F_n = E_n \cap E_{n-1}^c$. Since $E_n \in \mathcal{F}$ is increasing, we have $F_n \in \mathcal{F}$, $\cup_1^\infty E_n = \cup_1^\infty F_n$, (F_n) are disjoint, $\mathbb{P}(E_n) = \mathbb{P}(F_n) + \mathbb{P}(E_{n-1})$ and $\lim_{n\to\infty} \mathbb{P}(E_n)$ exists. So

$$\mathbb{P}(\lim E_n) = \mathbb{P}(\cup_1^\infty E_n) = \mathbb{P}(\cup_1^\infty F_n) = \sum_{n=1}^\infty \mathbb{P}(F_n) = \sum_{n=1}^\infty (\mathbb{P}(E_n) - \mathbb{P}(E_{n-1})) = \lim_{n\to\infty} \mathbb{P}(E_n). \quad \square$$

The above result says that a probability measure is continuous from above and below, and the key in the proof is the property that a probability measure is countably additive.

Suppose for now that we are given a measure μ on some collection of intervals of $\mathbb{R}$. For example,

$$\mu((a,b)) = \mu([a,b]) = \mu((a,b]) = \mu([a,b)) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x$$

for $a \le b$. Is it possible to view μ as a (probability) measure on $(\mathbb{R}, \mathcal{F})$ for some σ-field $\mathcal{F}$ (which contains above intervals) on $\mathbb{R}$? The answer is yes and actually we have the following two important theorems:

Theorem 3.4. (Existence of generated σ-field)

Let S be a set and let $\mathcal{E}$ be any collection of subsets of S. Then there exists a unique σ-field denoted $\sigma(\mathcal{E})$ such that:

1. $\sigma(\mathcal{E}) \supset \mathcal{E}$
2. If $\mathcal{F}$ is any other σ-field containing $\mathcal{E}$, then $\mathcal{F} \supset \sigma(\mathcal{E})$

$\sigma(\mathcal{E})$ is called the **σ-field generated by $\mathcal{E}$**.

Proof. First note the following: if $\mathcal{F}_\gamma$ is a σ-field for each $\gamma \in \Gamma$, then $\mathcal{F} = \cap_\gamma \mathcal{F}_\gamma$ is a σ-field also. This follows directly from the definition of "$\cap$".

Now denote the set of all σ-fields containing $\mathcal{E}$ by $\{\mathcal{G}_\gamma\}_{\gamma \in \Gamma}$. This is nonempty since 2^S is a σ-field. Now set $\mathcal{G} = \cap_\gamma \mathcal{G}_\gamma$. We leave to the reader to confirm that $\mathcal{G} = \sigma(\mathcal{E})$. $\square$

Thus given any collection of subsets of a space, there always exists a unique *smallest* σ-algebra containing that set. One special type of σ-field is defined using the concept of generated

σ-fields, and is particularly important for probability theory, as it naturally "comes with" metric spaces:

Example. (Borel σ-field)

Given a metric space M, the **Borel σ-field on** M denoted $B(M)$ is the σ-field generated by the collection of open sets in M.

Since the collection of open sets in M is precisely the metric topology on M, the Borel σ-field on M is the smallest σ-field on M which contains the metric topology.

Fact. Let $\mathcal{E} = \{(a, b) : a < b\}$. Then $B(\mathbb{R}) = \sigma(\mathcal{E})$. The result also holds if one replaces in the definition of $\mathcal{E}$ the open intervals (a, b) by $[a, b]$, $(a, b]$ or $[a, b)$.

Now recall the question that whether one can view μ as a (probability) measure on $(\mathbb{R}, \mathcal{F})$ for some σ-field $\mathcal{F}$ on $\mathbb{R}$, where μ is a measure on intervals such that

$$\mu((a, b)) = \mu([a, b]) = \mu((a, b]) = \mu([a, b)) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x$$

for $a \leq b$. It's natural to guess that $\mathcal{F}$ should be at least $B(\mathbb{R})$ (and also $B(\mathbb{R})$ is rich enough for daily use), as the next theorem says.

Theorem 3.5. (Extension of measures)

Let M be some space and $\mathcal{E}$ be a collection of subsets of M. Suppose $\forall A, B \in \mathcal{E}$, $A \cap B \in \mathcal{E}$ and $A \setminus B = \cup_{i=1}^n E_i$ for some n and disjoint $E_i \in \mathcal{E}$. If μ is a finite measure on $\mathcal{E}$, then there exists a unique measure ν on $\sigma(\mathcal{E})$ such that $\mu = \nu$ on $\mathcal{E}$.

In the above scenario, $M = \mathbb{R}$, $\mathcal{E} = \{(a, b] : a \leq b\}$. Then μ can be extended to a unique (probability) measure on $\mathbb{R}$.

3.4 Random variables

In the past sections we saw how to model random phenomena at the core level of outcomes and events. But oftentimes, it helps to have an extra angle with which to study probabilistic events. We do this by defining a special function mapping outcomes in Ω to numbers in $\mathbb{R}$.

Definition. (Random variable)

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a **random variable** (r.v.) on that space is a function $X : \Omega \to \mathbb{R}$ such that $X^{-1}(B) \in \mathcal{F}$ for all $B \in B(\mathbb{R})$.

One way to make sense of this definition is to think of random variables as *numerical measurements* of outcomes. The condition that $X^{-1}(B) \in \mathcal{F}$ for all $B \in B(\mathbb{R})$ is called **measurability**, and allows us to not only look at X as a map between Ω and $\mathbb{R}$, but also as a set map between $\mathcal{F}$ and $B(\mathbb{R})$, the Borel σ-algebra on $\mathbb{R}$. In other words, measurability makes X a map that *respects the structures* of Ω and $B(\mathbb{R})$.

Again, it is easy to think up of easy toy examples for finite spaces:

Example. (Coin toss random variable)

Consider the probability space of tossing two coins, $(\Omega, \mathcal{F}, \mathbb{P})$:

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

Suppose we are interested in the number of heads. Then we might construct the random variable:

$$X(\omega) = \begin{cases} 2, & \text{if } \omega = (H, H) \\ 1, & \text{if } \omega = (H, T), (T, H) \\ 0, & \text{if } \omega = (T, T) \end{cases}$$

Other examples are equally simple, but are deceptively useful. For example:

Example. (Indicator random variable)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an arbitrary probability space and let $A \in \mathcal{F}$. Then a valid random variable is:

$$X(\omega) = \mathbb{1}_A(\omega)$$

That is, if the event A occurs, then $X = 1$. If not, then $X = 0$.

Note that in the definition of random variables, measurability is required. This guarantees that we can know well about the random variable based on the probability space on which it is defined. Otherwise, we may meet with troubles like the following example.

Example (Coin toss)**.** Consider the following smaller probability space.

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

$$\mathcal{S} = \{\Omega, \emptyset, A, A^c\}, \text{ where } A = \{(H, H)\}$$

$$\mathbb{P}(A) = \frac{1}{4}, \mathbb{P}(A^c) = \frac{3}{4}.$$

Let $X : \Omega \to \mathbb{R}$ be the coin toss map as before. Then X is not measurable and we don't know the value of $\mathbb{P}(X = 0)$.

Thus we see that measurability is central to the notion of probability. Because of this, given a σ-algebra and a measure on that algebra, we are often interested in checking that the function is in fact measurable. But how can we check this if our σ-algebras can be infinitely large? Our tools from the last section come to the rescue:

Theorem 3.6. If $\{\omega : X(\omega) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{A}$ and $\sigma(\mathcal{A}) = \mathcal{S}$, then X is measurable.

Proof. Note that

$$\{\omega : X(\omega) \in \cup_{i=1}^{\infty} B_i\} = \cup_{i=1}^{\infty} \{\omega : X(\omega) \in B_i\}$$

$$\{\omega : X(\omega) \in B^c\} = \{\omega : X(\omega) \in B\}^c$$

Therefore the class of sets $\mathcal{B} = \{B : \{\omega : X(\omega) \in B\} \in \mathcal{F}\}$ is a σ-field. Then since $\mathcal{B} \supset \mathcal{A}$ and $\mathcal{A}$ generates $\mathcal{S}$, then $\mathcal{B} \supset \mathcal{S}$. □

As we can see, since random variables are functions from a probability space to the real line, technically we cannot speak of random variables without explicitly identifying some underlying probability space. But in statistics we often make statements such a "Let X be a Uniform random variable" without ever referring to measures. What allows us to do this?

As it turns out, one thing measurability of X buys us is that it allows us to use the measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ to associate a probability measure on the target measure space $(\mathbb{R}, B(\mathbb{R}))$. Once this association is made, we can speak about $\mathbb{P}$ using the space $(\mathbb{R}, B(\mathbb{R}))$ only.

Definition. (Distribution of a random variable)

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X, the **distribution** of X is the measure μ on $(\mathbb{R}, B(\mathbb{R}))$ induced by $\mathbb{P}$ and X:

$$\mu(B) = \mathbb{P}(X^{-1}(B)) \quad \text{for all } B \in B(\mathbb{R})$$

In other words, we use the measure-preserving map X to "push foward" the measure $\mathbb{P}$ onto the target space $(\mathbb{R}, B(\mathbb{R}))$. Thus, the distribution is also called the **pushforward measure** of $\mathbb{P}$ using X.

Example. (Indicator example)

Let X be the indicator random variable defined above as $X = \mathbb{1}_A(\omega)$. Then the distribution μ of X is defined by:

$$\mu(\{1\}) = \mathbb{P}(A), \quad \mu(\{0\}) = 1 - \mathbb{P}(A)$$

Where $\mu(A) = 0$ for all other $A \in B(\mathbb{R})$ that does not contain 0 or 1.

Example. (Coin toss random variable)

Let X be the random variable defined on the coin toss probability space above. The distribution of X, μ is thus defined by:

$$\mu(\{0\}) = \frac{1}{4}, \quad \mu(\{1\}) = \frac{1}{2}, \quad \mu(\{2\}) = \frac{1}{4}$$

Where $\mu(A) = 0$ for all other $A \in B(\mathbb{R})$ that does not contain 0, 1 or 2.

The notation of probability requires an aside here. It has become convention to write $\mathbb{P}(X \in B)$ as shorthand for the more technically correct $\mathbb{P}(X^{-1}(B))$. It bears repeating, though, that $\mathbb{P}$ is a measure on the measurable space of outcomes, whereas X takes values on the real line.

Returning to distribution functions, one convenient form for the distribution of a random variable is the *cumulative distribution function*:

Definition. (Cumulative distribution function) The *cumulative distribution function* (CDF) of a r.v. X is the function

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x]) = \mathbb{P}(X^{-1}(-\infty, x])$$

Some basic properties of CDFs which can be proved directly from the properties of measures include:

1. (non-decreasing) If $x \leq y$, then $F(x) \leq F(y)$
2. (right-continuous) If $x_n \downarrow x$, then $F(x_n) \downarrow F(x)$
3. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$

Although above nice properties hold for CDF, sometimes a CDF can be very user-unfriendly.

Example. Let X be a random variable whose CDF is the Cantor function F, which is defined as follows.

- For $x \leq 0$, $F(x) = 0$.
- For $x \geq 1$, $F(x) = 1$.
- For $x \in [0, 1]$, express x in base 3. If x contains a 1, replace every digit after the first 1 by 0. Replace all 2s with 1s. Interpret the result as a binary number. The

result is $F(x)$.

It can be shown that

- $F(x)$ is a continuous CDF so that X does not have a mass function.
- X does not have a density either.
- $\mathbb{P}(X \in C) = 1$, where C is the Cantor set with "length" 0.

3.5 The π-λ theorem

To sum up, the idea of a generated σ-algebra is essentially the idea that a small class of subsets can *generate* a larger class of subsets. If we take this idea further, we arrive at a very important theorem for proving a certain class of results.

To start, we define two special types of "small classes" which, we shall see, also "generates" a σ-algebra in a special sort of way.

Definition. (π-class, λ-class)

Let S be some space and let $\mathcal{L}, \mathcal{A}$ be collections of subsets of S.

$\mathcal{A}$ is a π**-class** if it is closed under intersections.

$\mathcal{L}$ is a λ**-class** if

1. $S \in \mathcal{L}$
2. If $A, B \in \mathcal{L}$ with $A \supset B$, then $A \setminus B \in \mathcal{L}$
3. If $\{A_n\}$ is increasing with $A_i \in \mathcal{L}$, then $\lim_{n\to\infty} A_n \in \mathcal{L}$.

As with many things in measure theory, these definitions seem pretty arbitrary unless we see how they are used. Essentially the only way that π- and λ-classes are employed is through the following famous theorem. The proof is not difficult but it is notationally annoying, so we omit it here.

Theorem 3.7. (Dynkin's $\pi - \lambda$ theorem)

Let $\mathcal{L}$ be a λ-class and $\mathcal{A}$ a π-class. If $\mathcal{L} \supset \mathcal{A}$, then also $\mathcal{L} \supset \sigma(\mathcal{A})$.

But what does this mean? Suppose we want to show that some property holds for an entire σ-field $\mathcal{F}$. This can be difficult because $\mathcal{F}$ can not only be large, but also contain weird sets. The π-λ theorem tells us that:

1. If the class of sets for which the property holds is a λ-class, and
2. If the above class contains a smaller π-class which generates $\mathcal{F}$,

Then the class of sets for which the property holds in fact contains not only just the π-class, but also the entire σ-algebra generated by the π-class, $\mathcal{F}$. The concept is not difficult to grasp. All that is required is the right interpretation.

A classic example of the π-λ theorem's utility is the situation of showing that two measures on a measurable space are actually the same:

Lemma 3.8. (Identification lemma)

Suppose p_1, p_2 are probability measures on $(S, \mathcal{S})$ and $p_1(A) = p_2(A)$ for all $A \in \mathcal{A}$. If $\mathcal{A}$ is a π-class and $\sigma(\mathcal{A}) = \mathcal{S}$, then $p_1(A) = p_2(A)$ for all $A \in \mathcal{S}$.

Proof. In this case, the key is to view the problem in a way which will let us apply the π-λ theorem: we want to show that all sets in $\mathcal{S}$ have the property that p_1 and p_2 are equal on that set.

Therefore, define $\mathcal{L} = \{A \in \mathcal{S} \,|\, p_1(A) = p_2(A)\}$. Note that:

1. $\mathcal{A} \subset \mathcal{L}$ is a π-class by assumption.

2. If we show that $\mathcal{L}$ is a λ-class, then the result follows.

To show the three properties of λ-classes:

1. $S \in \mathcal{L}$ because p_1, p_2 are both probability measures.

2. To show that $A \supset B \in \mathcal{L} \Rightarrow A \setminus B \in \mathcal{L}$, use countable additivity.

3. To show that $\lim A_n \in \mathcal{L}$ for increasing $\{A_n\} \in \mathcal{L}$, use the continuity property of measures.

So by the π-λ theorem, $\mathcal{L} \supset \sigma(\mathcal{A}) = \mathcal{S}$. □

Another classic example of the π-λ theorem is the following important theorem about independence between random variables.

> **Theorem 3.9.** Let X, Y be two real-valued random variables. We say X is independent of Y is for every $A, B \subset B(\mathbb{R})$, $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$. Then X is independent of Y iff for every $a, b \in \mathbb{R}$, $\mathbb{P}(X \leq a, Y \leq b) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b)$.

Proof. " $\Rightarrow$ " is clear. Sketch of proof for " $\Leftarrow$ ":

1. Let $\mathcal{E} = \{(-\infty, a] \mid a \in \mathbb{R}\}$. Note that $\mathcal{E}$ is a π-class and $\sigma(\mathcal{E}) = B(\mathbb{R})$.

2. Fix $b \in \mathbb{R}$. Define $\mathcal{L} = \{A \in B(\mathbb{R}) \mid \mathbb{P}(X \in A, Y \leq b) = \mathbb{P}(X \in A)\mathbb{P}(Y \leq b)\}$.

   (a) Note that $\mathcal{E} \subset \mathcal{L}$ is a π-class.

   (b) Check that $\mathcal{L}$ is a λ-class.

   So by the π-λ theorem, $\mathcal{L} \supset \sigma(\mathcal{E}) = B(\mathbb{R})$. This means $\mathbb{P}(X \in A, Y \leq b) = \mathbb{P}(X \in A)\mathbb{P}(Y \leq b)$ for every $A \in B(\mathbb{R})$ and $b \in \mathbb{R}$.

3. Fix $A \in B(\mathbb{R})$. Define $\mathcal{G} = \{B \in B(\mathbb{R}) \mid \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\}$.

   (a) Note that $\mathcal{E} \subset \mathcal{G}$ is a π-class, by part 2 above.

   (b) Check that $\mathcal{G}$ is a λ-class.

   So by the π-λ theorem, $\mathcal{G} \supset \sigma(\mathcal{E}) = B(\mathbb{R})$.

So we have proved that $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ for every $A, B \in B(\mathbb{R})$. $\square$

4 Exercises

1. Prove Theorem 1.6: Every subsequence of a convergent sequence converges, and it converges to the same limit.

2. Prove that convergence of (s_n) implies convergence of $(|s_n|)$. Is the converse true?

3. Calculate the limit of $(\sqrt{n^2 + n} - n)$ as $n \to \infty$.

4. Show that any space with the discrete metric is complete.

5. Let (x_n) be a real-valued sequence. Suppose both $\liminf x_n$ and $\limsup x_n$ are

finite. Prove that (x_n) is bounded.

6. Prove the following statements about lim sup and lim inf:

   (a) $\limsup_{n\to\infty} x_n = \inf_{n\geq 0} \sup_{m\geq n} x_m$ and $\liminf_{n\to\infty} x_n = \sup_{n\geq 0} \inf_{m\geq n} x_m$

   (b) For any two real sequences $\{a_n\}$, $\{b_n\}$, prove that

$$\limsup_{n\to\infty}(a_n+b_n) \leq \limsup_{n\to\infty}(a_n) + \limsup_{n\to\infty}(b_n)$$

$$\liminf_{n\to\infty}(a_n+b_n) \geq \liminf_{n\to\infty}(a_n) + \liminf_{n\to\infty}(b_n)$$

   provided the sum on the right is not of the form $\infty - \infty$.

   (c) $\limsup_{n\to\infty} x_n = -\liminf_{n\to\infty}(-x_n)$

7. Prove the following basic results about sequences in $\mathbb{R}$:

   (a) Let $x_n \to x \in \mathbb{R}$, and let $k \in \mathbb{R}$. Then $y_n = kx_n \to kx$

   (b) Let $x_n \to x$ and $y_n \to y$ in $\mathbb{R}$. Then $z_n = x_n y_n \to xy$

   (c) Let $x_n \to x \neq 0$ in $\mathbb{R}$, s.t. $x_n \neq 0\ \forall n$. Then $y_n = 1/x_n \to 1/x$

   (d) Let $x_n \to x \neq 0$ in $\mathbb{R}$, s.t. $x_n \neq 0\ \forall n$, and let $y_n \to y$ in $\mathbb{R}$. Then $z_n = y_n/x_n \to y/x$

8. Let f, g be real-valued sequences defined on $S \subset \mathbb{R}$. Prove that $f(x) = \mathrm{O}(g(x))$ if and only if:

$$\limsup_{x\to a} \left|\frac{f(x)}{g(x)}\right| < \infty$$

9. Show that if $x \in \lim S$, then for all $r > 0$ it is true that $B_r(x) \cap S \neq \emptyset$

10. Show that every subset of a discrete metric space M is clopen (why would it suffice to show that a singleton $\{x\}$ is open?) and that therefore for all $S \subset M$, it is true that $\int(S) = S = \bar{S}$ and $\partial S = \emptyset$

11. Prove that the infimum and supremum of a non-empty bounded subset $S \subset \mathbb{R}$ belong to the closure of S.

12. Prove that if $S \subset N \subset M$, then S is open in N if and only if there exists $L \subset M$ such that L is open in M and $S = L \cap N$.

    Hint: Notice that the complement of $L \cap N$ in N is $L^c \cap N$, where L^c is the complement of L in M

13. Is every point of every open set $E \subset \mathbb{R}^2$ a limit point of E? Answer the same question for closed sets in $\mathbb{R}^2$

14. A metric space is called separable if it contains a countable dense subset. A subset E of X is dense in X if every point of X is a limit point of E or a point of E (or both). Show that $\mathbb{R}^k$ is separable.

    Hint: Consider the set of points which have only rational coordinates.

15. Prove that the composition of continuous functions is continuous

16. Prove the following:

    Let M, N be metric spaces, $f : M \to N$ a uniformly continuous function, and (x_n) a Cauchy sequence in M. Then $(f(x_n))$ is a Cauchy sequence in N.

17. Show that **every** function defined on a discrete metric space is uniformly contin-

uous

18. Let f, g be real-valued funcitons $M \to \mathbb{R}$, continuous at some $x \in M$. Then $f+g$, fg, and f/g (asusming $g(x) \neq 0$) are all continuous at x

19. Suppose f is a real function defined on $\mathbb{R}$ which satisfies

$$\lim_{h \to 0} [f(x+h) - f(x-h)] = 0$$

for every $x \in \mathbb{R}$. Then is f necessarily continuous?

20. Let f and g be continuous mappings of a metric space X into a metric space Y, and let E be a dense subset of X. Prove that $f(E)$ is dense in $f(X)$. If $g(p) = f(p)$ for all $p \in E$, prove that $g(p) = f(p)$ for all $p \in X$. (In other words, a continuous mapping is determined by its values on a dense subset of its domain.)

21. Prove that every compact set is complete

22. Prove that the continuous image of a compact set is compact.

23. Prove that any continuous real-valued function defined on a compact set assumes its maximum and minimum.

24. Let $K \subset \mathbb{R}$ consist of 0 and the numbers $1/n$, for $n = 1, 2, 3, \ldots$. Show that K is compact without using the Heine-Borel theorem.

25. Let X be a totally bounded metric space, and $f : X \longrightarrow Y$ a uniformely continuous map onto Y. Show that Y is totally bounded. Is this result still true if f is only required to be continuous?

26. Show that the Heine-Borel theorem easily implies the Bolzano-Weierstrass theorem:

Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

5 References

1. C.C. Pugh, 'Real Mathematical Analysis', Springer, 2002
2. John A. Rice, 'Mathematical Statistics and Data Analysis', 2nd ed., Duxbury Press, 1995
3. Kenneth A. Ross, 'Elementary Analysis: The Theory of Calculus', Springer, 2000
4. H. L. Royden, 'Real Analysis' 3rd ed., Macmillan Publishing Company, 1988
5. Walter Rudin, 'Principles of Mathematical Analysis', 3rd ed., McGraw-Hill, 1976
6. Ross Leadbetter, Stamatis Cambanis, Vladas Pipiras, 'A Basic Course in Measure and Probability: Theory for Applications', Cambridge University Press, 2014