THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Department of Biostatistics

Department of Stat. and OR

Refresher course, Summer 2015

Real Analysis

Revised by:

Eric Friedlander

(UNC-Chapel Hill, 2015)

Original Author:

Oleg Mayba

(UC Berkeley, 2006)

Based on the NSF sponsored (DMS Grant No 0130526) VIRE Boot camp lecture notes in the Department of Statistics, University of California, Berkeley

July 28, 2015

Contents

1 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 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.

We hope you find these notes useful! Go Heels!

2 Introduction to metric spaces

2.1 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. We will take for granted the familiarity with notions of finite, countably infinite and uncountably infinite sets.

The real number line is the most basic and important space to be familiar with. There are many reasons for this, not 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 $s \in S$. Then 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 $S \subset \mathbb{R}$, 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. (Three properties)

1. Let $S \subset \mathbb{R}$. If $\sup(S)$ exists, then it is unique.

2. Let $S \subset \mathbb{R}$ be nonempty. Then for any $c > 0$, we can find $s \in S$ such that $s \in (\sup(S) - c, \sup(S))$.

3. 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

$$a' \in (\sup A - \frac{c}{2}, \sup A) \quad \text{and} \quad b' \in (\sup B - \frac{c}{2}, \sup B)$$

and these points satisfy $a' + b' > \sup(A+B)$.

$\square$

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

Theorem 2.1. 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.

$\square$

2.2 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 Euclieand (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 $\boldsymbol{x} = (x_1, x_2, x_3, \ldots)$ where $x_i \in \mathbb{R}$ for $i = 1, 2, \ldots$.

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:

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

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 $\boldsymbol{v} = \mathbf{0}$

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 three 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 2.2. (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 $d(x,y) + d(z,x) \geq d(y,z)$,

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

□

2.3 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 2.3. 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.

The following is an important property of countable sets.

Theorem 2.4. 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.

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 number is countable.
3. The set of all binary sequence in uncountable.

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

2.4 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 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 2.5. 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 2.6. Every subsequence of a convergent sequence converges, and it 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 2.7. 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:

$$|(x+y) - (x_n - y_n)| = |(x - x_n) + (y - y_n)|$$
$$\leq |x - x_n| + |y - y_n|$$

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 $n \geq N_2$. Set $N = \max\{N_1, N_2\}$. 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 2.8. 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, $y - x = \lim y_n - \lim x_n = \lim(y_n - x_n) = y - x$. Since $y_n - x_n \geq 0$ for all n, then the limit is also ≥ 0.[1]

□

Corollary 2.9. (Squeeze theorem)

Suppose (x_n), (y_n) and (z_n) are convergent 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.

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

[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.

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 2.10. 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 $[-c-1, c+1]$.

□

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

Theorem 2.11. 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

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

□

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 2.12. (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 $n > k > N$,

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

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

□

Theorem 2.13. $\mathbb{R}$ 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 $b < \infty$. Fix $\epsilon > 0$. We show that $a_n \to b$ by finding an N such that $|a_n - b| < \epsilon$ for all $n \geq N$.

Since (a_n) is Cauchy, there exists N_1 such that $|a_n - a_m| < \frac{\epsilon}{2}$ for all $m, n \geq N_1$. Since b is the supremum for S, then $b + \frac{\epsilon}{2} \notin S$, but $b - \frac{\epsilon}{2} \in S$. Therefore $a_n \geq b + \frac{\epsilon}{2}$ only finitely many times, but $a_n \geq b - \frac{\epsilon}{2}$ infinitely many times.

Thus we can find N_2 such that $a_m < b + \frac{\epsilon}{2}$ for all $m \geq N_2$. But now setting $N = \max\{N_1, N_2\}$, we also have $a_m \geq b - \frac{\epsilon}{2}$ for all $m \geq N$. So by the triangle inequality, for all $m \geq N$,

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

□

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 2.14. $\mathbb{R}^n$ is complete.

2.5 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. $(x_n) = 3, 3.1, 3.14, 3.141, \ldots$ Then $\lim(x_n) = \pi$

2. $(x_n) = 2, 3, 5, 7, 11, \ldots$ 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 2.15. Let (x_n) be a real-valued increasing (decreasing) sequence which is bounded above (below). Then it is convergent in $\mathbb{R}$.

Proof. Assume WLOG that (x_n) is increasing 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 increasing, then $x_n > b - \epsilon$ for all $n \geq N$. Therefore

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

$\square$

Now we arrive to the final definition of this section. We have just shown that for a 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)

Given a sequence (x_n) in $\mathbb{R}$, the **limit superior** of a sequence (x_n) in $\mathbb{R}$ as:

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

Similarly, the **limit inferior** of (x_n) 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$(x_n)= \pi$

The limsup and liminf of a sequence have several important properties, including sub- and super-additivity (see exercises). But first we present a useful result which will be used immediately to prove the most foundational result tying together $\limsup$, $\liminf$, and lim:

Lemma 2.16. Let (x_n) be a real-valued sequence and suppose $\liminf x_n$ and $\limsup x_n$ are finite. Then (x_n) is bounded.

Proof. Exercise.

The foundational result which has an exact analogue when we introduce liminf and limsup of *sets* is the following:

Theorem 2.17. 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 $x_n \to x$. Then $y_n \downarrow x$ and $z_n \uparrow x$, with $y_n \geq z_n$ for all n. For every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $x - \epsilon < x_n < x + \epsilon$ for all $n > N$. It follows that:

$$x - \epsilon \leq z_n \leq y_n \leq x + \epsilon \quad \text{for all } n \geq N$$

Therefore $z_n \to x$ and $y_n \to x$ also, so $\liminf x_n = \limsup x_n = x$.

($\Leftarrow$) 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 some examples of when liminf/limsup are useful, we prove the following two lemmas which play a key role in the Strong Law of Large Numbers (SLLN):

Lemma 2.18. (Cesaro's lemma)

Let $\{a_n\}$ be a sequence of strictly positive real numbers with $a_n \uparrow \infty$. Let $\{X_n\}$ be a convergent sequence of real numbers. Then:

$$\frac{1}{a_n}\sum_{k=1}^{n}(a_k - a_{k-1})X_k \to \lim X_n$$

Proof. Let $\epsilon > 0$. Since X_n is convergent, we can choose N such that

$$X_k > \lim X_n - \epsilon \quad \text{whenever } k \geq N$$

Then split the sum into the portion up to N and the portion beyond N, and apply the above inequality:

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{a_n}\sum_{i=1}^{n}(a_k - a_{k-1})X_k \\
&\qquad \geq \liminf_{n\to\infty} \left\{ \frac{1}{a_n}\sum_{k=1}^{N}(a_k - a_{k-1})X_k + \frac{a_n - a_N}{a_n}(\lim X_n - \epsilon) \right\} \\
&\qquad \geq 0 + \lim X_n - \epsilon
\end{aligned}$$

Where the last step follows from the fact that $a_n \uparrow \infty$ and a_N is finite.

This is true for all $\epsilon > 0$, so $\liminf \geq \lim X_n$. To show that $\limsup \leq \lim X_n$, follow the same argument except choose N such that $X_k < \lim X_n + \epsilon$ whenever $k \geq N$.

□

Lemma 2.19. (Kronecker's lemma)

Let $\{Y_n\}_{n\geq 1}$ be a real-valued sequence and let $\{a_n\}_{n\geq 1}$ be a sequence of strictly positive real numbers with $a_n \uparrow \infty$. If $\sum Y_n/a_n < \infty$, then $S_n/a_n \to 0$.

Proof. Define $X_n = \sum_{i=1}^n Y_n/a_n$. By assumption, X_n converges so $\lim X_n$ exists. Note $X_k - X_{k-1} = Y_k/a_k$ so that:

$$\begin{aligned} S_n &= \sum_{k=1}^n a_k(X_k - X_{k-1}) \\ &= a_n X_n - \sum_{k=1}^n (a_n - a_{n-1})X_{k-1} \end{aligned}$$

Divide by a_n, send $n \to \infty$, and apply Cesaro's lemma.

□

2.6 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 function. To focus on this, we use special notation for comparing the growth of two functions.

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.

3 Structure of Metric Spaces

3.1 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. However, it is actually a little facetious. The original (more general) 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. The union of any subcollection of $\mathcal{T}$ is in $\mathcal{T}$.
3. The intersection of a finite subcollection of $\mathcal{T}$ is in $\mathcal{T}$.

Elements of $\mathcal{T}$ are called **open** sets.

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. (African safari space)

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 $\mathcal{T}$ is equal to a union of sets in $\mathcal{B}$.

From here is it easy to see that:

Corollary 3.1. 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 $B \subset S$.

This finally brings us to our punch line:

Theorem 3.2. (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 line 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. For finer technical points it will be useful to differentiate between properties that are purely topological and those which are not, but we shall not go any further into that in these notes.

3.2 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 3.3. 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:

$$d(z, x_0) \leq d(z, x) + d(x, x_0)$$
$$< r - d(x, x_0) + d(x, x_0) = r$$

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 $x_n \to x$.

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 3.4. (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 $\{x_{\frac{1}{n}}\}_{n \geq 1}$ formed this way is in S and converges to $x' \in S^c$, contradicting that $S = \lim S$.

□

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 a closed since any 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 3.5. 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 then there exists an open ball around x of some radius which is contained within U.

□

Theorem 3.6. The intersection of finitely many open sets is open.

Proof. Suppose $U_1, U_2, \ldots, U_n$ 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 $k = 1, 2, \ldots, n$. Since each U_k is open, for each $k = 1, \ldots n$ there exists $r_k > 0$ such that $B_{r_k}(x) \subset U_k$. Define r by:

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

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

□

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 3.7. 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 3.8. 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:

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

We claim $I_x = (a_x, b_x)$. 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 $a_x, b_x \notin S$, 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.

□

3.3 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 3.9. 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 $x \in N$.

□

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

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

Immediate from the definition, we have the following facts:

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

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

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

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

Proof. Trivial.

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 3.10. $\bar{S} = \lim S$.

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.

□

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. $d_N(x, y) = d_M(x, y)$. 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 3.11. 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 is trivial after one establishes the following lemma:

Lemma 3.12. 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 clousre 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.

□

Finally, to bring out discussion whole circle with the notions of bounds and boundedness for subsets of $\mathbb{R}$, we generalie the notion of boundedness 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 3.13. 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.

□

Corollary 3.14. If a sequence is convergent, then it is bounded.

3.4 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 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 M, N be two topological spaces and let $f : M \to N$ be a function. Then f is **continuous** (on M) if $f^{-1}(S)$ is open in M whenever S is open in N.
>
> Equivalently, f is continuous if $f^{-1}(S)$ is closed in M whenever S is closed in N.

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.

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

Theorem 3.15. (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 $B_\delta(a) \subset f^{-1}(S)$.

□

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

Corollary 3.16. 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 3.17. 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.

$\square$

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 3.18. 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 $\theta \in R$ 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 3.19. (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, there exists $\delta > 0$ such that the following implication holds for all n:

$$|f(\hat{\theta}_n) - f(\theta)| > \epsilon \quad \Longrightarrow \quad |\hat{\theta}_n - \theta| > \delta \tag{1}$$

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:

$$\mathbb{P}(|\hat{\theta}_{n_k} - \theta| > \delta) \geq \mathbb{P}(|f(\hat{\theta}_{n_k}) - f(\theta)| > \epsilon) > L \quad \text{for all } k \tag{2}$$

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.

3.5 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, $\ldots$ 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 3.20. (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 not. Then there is an open cover $\{U_\alpha\}$ for S such that for every $\lambda \in \mathbb{R}$, there exists $x \in S$ such that $B_\lambda(x) \not\subseteq U_\alpha$ for every α. 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, $B_{\lambda_n}(x_n) \not\subseteq U_\alpha \forall \alpha$.

$\square$

Now we have all we need to show the main result:

Theorem 3.21. Let M be a metric space and let $S \subset M$. The 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'd 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 $\{U_\alpha\}$ be some open cover for S. By the Lebesgue number lemma, $\{U_\alpha\}$ has a Lebesgue number $\lambda > 0$, so we can pick $x_1 \in S$ and $U_1 \in \{U_\alpha\}$ such that $B_\lambda(x_1) \subset U_1$. Since S is not covering compact, there exists an uncovered point $x_2 \in S$ (i.e. $x \in S \cap {U_1}^c$) and U_2 such that $B_\lambda(x_2) \subset U_2$. Continue in this way to obtain a sequence (x_n) in S and a sequence (U_n) of members of $\{U_\alpha\}$ with $B_\lambda(x_n) \subset U_n$ and $x_{n+1} \in S \cap (U_1 \cup \ldots \cup U_n)^c$ for all n.

Since S is sequentially compact, (x_n) has a subsequence $(x_{n_k})_{k\geq 1}$ with $x_{n_k} \to x \in S$ as $k \to \infty$. Therefore, there exists K such that $d(x_{n_k}, x) < \lambda$ for all $k \geq K$. In particular, $d(x_{n_K}, x) < \lambda \Rightarrow x \in B_\lambda(x_{n_K}) \subset U_{n_K}$. Since U_{n_K} is open, there exists $r > 0$ such that $B_r(x) \subset U_{n_K}$. But by construction, $x_{n_k} \notin U_{n_K}$ for all $k > K$ so $B_r(x)$ contains only finitely many terms of (x_{n_k}), which contradicts that $x_{n_k} \to x$ as $k \to \infty$.

□

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

Proof. (By sequential compactness)

Let (x_n) be a sequence in $[a, b]$. Consider the set

$$C = \{x \in [a, b] : x_n < x \text{ only finitely often}\}$$

Clearly, C is nonempty and bounded so $x^* = \sup C$ exists. We show that (x_n) has a subsequence $(x_{n_k})_{k\geq 1}$ converging to x^*. First note that if $x^* = b$, then by definition of C there exists N such that $x_n = b$ for all $n \geq N$, and we are done.

So suppose $x^* < b$. Assume that no subsequence converges to x^*. Then there exists $r > 0$ such that $r < b - x^*$ and $x_n \in B_r(x^*)$ only finitely often. But then $x^* + r \in C$ (note that $x^* + r \in [a, b]$), contradicting that $x^* = \sup C$. Thus there must exist some subsequence converging to x^*.

□

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 show that $x^* = b$.

Suppose instead that $x^* < b$. Let $U_{\alpha_1}, \ldots, U_{\alpha_n}$ be those finitely many members of $\{U_\alpha\}$ that suffice to cover $[a, x^*]$. Then $x^* \in U_{\alpha_k}$ for some k and since U_{α_k} is open, there exists $r > 0$ such that $B_r(x^*) \subset U_{\alpha_k}$. Now for any $y \in (x^*, x^* + r)$, y is covered by same $U_{\alpha_1}, \ldots, U_{\alpha_n}$ as x^*. Therefore finitely many members of $\{U_\alpha\}$ suffice to cover $[a, y]$ as well, which contradicts that $x^* = \sup C$. Therefore $x^* = b$ necessarily.

□

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 3.23. 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 find $\delta > 0$ such that for any $x, y \in M$ with $d(x, y) < \delta$, then

$d'(f(x), f(y)) < \epsilon$. 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').

$\square$

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 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 3.24. Every compact set is complete.

Proof. Exercise.

The full Heine-Borel theorem brings everything together:

Theorem 3.25. (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 be 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 $\{J_n\}$, both of which are nested (i.e. $B_1 \supset B_2 \supset \ldots$ and $J_1 \supset J_2 \supset \ldots$). 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 $\{B_n\}_{n\geq 1}$.

□

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 3.26. (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.

$\square$

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

Theorem 3.27. (Bolzano-Weierstrass)

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

Proof. Exercise.

3.6 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 chapter we will only discuss real valued function but the results can easily be extended to function 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 function 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 3.28. 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 for $m > M$ $|f_m(y) - f(y)| < \epsilon/2$ for all $y \in M$ including $y = x$ and $y = x_i,\ i = 1, 2, \ldots$. Select such an $m = m^*$. We know the 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,

$$|f(x_n) - f(x)| \leq |f(x_n) - f_{m^*}(x_n)| + |f_{m^*}(x_n) - f_{m^*}(x)| + |f_{m^*}(x) - f(x)| < \epsilon.$$

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

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

Theorem 3.29. 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.

4 Introducton to Measure Theory

4.1 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. There are two main reasons for this. Firstly, measure theory offers a clever way to rigorously represent something which seems like it shouldn't fit into any mathematical framework. 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 that the like is that they operate on the level of random variables, which take values on the real line. But in serious statistics and probability, we deal with things that are far more complex.

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:
>
> $$\mathbb{P}(X < 0 \,|\, Y = 1)$$

We know how to calculate this using undergraduate probability formula:

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

However, strictly speaking we know that $\mathbb{P}(Y = 1) = 0$ so what is going on behind the scenes? 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?

4.2 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.

Definition. (σ-algebra)

Given any set S, a **σ-algebra 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

Thus 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 together with a σ-algebra on that set, we call the pair a **measurable space** and denote it $(S, \mathcal{S})$.

Definition. (Measure)

A **measure** on a measurable space $(S, \mathcal{S})$ is a function $\mu : \mathcal{S} \to \mathbb{R} \cup \{\infty\}$ with the following properties:

1. Non-negativity: $\mu(E) \geq 0$ for all $E \in \mathcal{S}$
2. Empty set has measure 0: $\mu(\emptyset) = 0$

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

$$\mu\Big(\cup_{i=1}^{\infty} E_i \Big) = \sum_{i=1}^{\infty} \mu(E_i)$$

If $\mu(S) = 1$, 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 $\mathbb{P}(5) = 1$. Suppose we have the following probability space.

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

4.3 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:

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

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 4.1. 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.

Theorem 4.2. Let $(E_n)_{n\geq 1}$ be an increasing (decreasing) sequence of sets. Then (E_n) is convergent and $\lim E_n = \cup_1^\infty E_n$ $(\cap_1^\infty E_n)$.

Proof. 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.

To tie all this back to σ-rings and measure theory, we have the following important result:

Theorem 4.3. (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.

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 4.4. (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 4.5. (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. $\mathcal{L} \subset \mathcal{A}$ 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.

$\square$

4.4 Random variables

In the past section 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 $X = 0$.

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 $A \in B(\mathbb{R})$.

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 $A \in B(\mathbb{R})$.

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.

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

Theorem 4.6. 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 $\mathcal{B} \supset \mathcal{S}$.

□

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. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$

The power of this measurability is that, once we are given a random variable X and a probability measure on $(\mathbb{R}, B(\mathbb{R}))$, then we do not actually need to know any details about the underlying probability space to describe probabilities on those events. Measurability guarantees that any measure we describe on $(\mathbb{R}, B(\mathbb{R}))$ will map directly to $\mathcal{F}$. But also, a sort of opposite implication is also at work here:

Theorem 4.7. Given any distribution function μ on $(\mathbb{R}, B(\mathbb{R}))$, there exists a random variable X which has distribution μ.

Proof. Hint: just define an appropriate probability space.

This relationship between probability spaces and distributions, via measurability, is exactly what lets us work with random variables abstractly. Given distribution 1 and distribution 2, the measure theoretic construct tells us that we can posit the existence of random variables X and Y having those distributions, and furthermore solve for probabilities or prove theorems about them all without ever specifying what the underyling events refer to.

Example. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space of tossing two coins. Define $X : \Omega \to \mathbb{R}$ such that $X(\omega)$ is the number of heads in the outcome ω. For example, $X(HT) = 1$.

Now let $(\Omega', \mathcal{F}', \mathbb{P}')$ be another probability space where $\Omega' = \{0, 1, 2\}$, $\mathcal{F}' = 2^{\Omega'}$ and $\mathbb{P}'$ is such that

$$\mathbb{P}'(\{0\}) = \mathbb{P}'(\{2\}) = 1/4, \text{ and } \mathbb{P}'(\{1\}) = 1/2$$

Define the random variable $X' : \Omega' \to \mathbb{R}$ such that $X'(\omega') = \omega'$ when $\omega' = 0, 1, 2$.

Then X and X' have the same CDF and distribution on $\mathbb{R}$, even though they are defined on different probability spaces.

5 Exercises

1. Prove Theorem 2.4: 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 and suppose $\liminf x_n < \infty$ and $\limsup x_n < \infty$. Prove that (x_n) is bounded.

6. Prove the following statements about lim sup and lim inf:

$$\limsup_{n\to\infty} x_n = \inf_{n\geq 0} \sup_{m\geq n} x_m \quad \text{and} \quad \liminf_{n\to\infty} x_n = \sup_{n\geq 0} \inf_{m\geq n} x_m$$

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 $\mathrm{int}(S) = S = \bar{S}$ 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 continuous

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.

6 References

1. Pugh, C.C., 'Real Mathematical Analysis', Springer, 2002

2. Rice, John A., 'Mathematical Statistics and Data Analysis', 2nd ed., Duxbury Press, 1995

3. Ross, Kenneth A., 'Elementary Analysis: The Theory of Calculus', Springer, 2000

4. Royden, H. L., 'Real Analysis' 3rd ed., Macmillan Publishing Company, 1988

5. Rudin, Walter, 'Principles of Mathematical Analysis', 3rd ed., McGraw-Hill, 1976