Analysis for Economists: ECON0118

Martin Cripps

November 2, 2020

Contents

INTRODUCTION

These notes are cobbled together from a combination of the texts: Royden (1988), Ok (2007), McDonald and Weiss (1999), the various Wikipedea and from various other sources mentioned along the way. One problem with these notes is that there are no pictures. I will try to supply lots in the lectures. I am sure there are many typos and errors — please let me know if you think you have found one so I can correct it.

1. SETS

1.1. Definition and examples

A set is a collection of elements it can written as: a capital letter, for example A; or explicitly as a collection of elements $\{1, 2, a\}$; or as a collection of objects satisfying a certain property $\{x \in \mathbb{R} : x^2 > 1\}$. The order of the elements are listed in the set A does not matter. If x is one of the elements of A we write $x \in A$. If x is not one of the elements of A we write $x \notin A$. There is one set $\emptyset$ which contains no elements. There may be very many elements in A and sets can be *too* large. One famous example of this is *the set of all sets that do not contain themselves as elements*. Such a set is ill-defined and leads to logical problems. None of the sets we consider will suffer from this problem.[1]

A subset of A is a set B, such that every element of B is in A. That is, $x \in B, \Rightarrow x \in A$. If this is the case we write $B \subset A$. For every set A it is true that $\emptyset, A \subset A$. Sometimes we want to think of the class of all subsets of the set A. This is written as 2^A, that is, 2^A is a collection of sets each one is a distinct subset of A. Two sets A and B are equal, written as $A = B$, if $B \subset A$ and $A \subset B$.[2]

Here are some important examples of sets:

[1]An example of an ill-defined set used in economics would be the set of all Blackwell experiments. Sometimes the word 'class' is used to indicate the set is too large to be clearly defined. Here we use the word 'class' to denote a collection of sets.

[2]Some authors use $B \subset A$ to exclude the case $B = A$ and write $B \subseteq A$ to allow for the subset to equal the whole set—we will not make such a distinction.

- $\mathbb{N}$ the set $\{1, 2, \dots\}$ of natural or counting numbers.
- $\mathbb{Z}_+$ the set $\{0, 1, 2, \dots\}$ of non-negative integers.
- $\mathbb{Z}$ the set $\{\dots, -2, -1, 0, 1, 2, \dots\}$ of integers.
- $\mathbb{Q}$ the set of all rational numbers (ratios of integers)
- $\mathbb{R}$ the set of all real numbers.
- $\mathbb{R}_+$ the set of non-negative reals.
- $\mathbb{R}^n$ the set n-dimensional vectors.
- $\mathcal{C}(A, B)$ the set of continuous functions from set A to set B.
- $(a, b) := \{x \in \mathbb{R} : a < x < b\}$, $(a, b] := \{x \in \mathbb{R} : a < x \leq b\}$, $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$, $[a, \infty) := \{x \in \mathbb{R} : a \leq x\}$, ...intervals of the real line.
- $\mathbb{C}$ the set of imaginary numbers.

Clearly we have $\mathbb{N} \subset \mathbb{Z}_+ \subset \mathbb{Z} \subset \mathbb{R} \subset \mathbb{C}$ and $(a, b) \subset \mathbb{R}$.

1.2. Set Operations and their Properties

We now will assume that the sets we are studying live in some ambient larger fixed set Ω. That is, $A, B \subset \Omega$. Then we define complement, intersection and union as follows:

$$
\begin{aligned}
A^c &:= \{x \in \Omega : x \notin A\}, \\
A \cap B &:= \{x \in \Omega : x \in A \text{ and } x \in B\}, \\
A \cup B &:= \{x \in \Omega : x \in A \text{ or } x \in B\}.
\end{aligned}
$$

(Note: these are definitions so we write :=. The object next to the colon is the thing being defined and the object next to the equality sign is the definition. Some authors use $\triangleq$ and others use $\equiv$ to do the same thing.) There are also two kinds of way of subtracting one set from another:

$$
\begin{aligned}
A \setminus B &:= \{x \in A : x \notin B\} = A \cap B^c, \\
A \triangle B &:= (A \cup B) \setminus (A \cap B);
\end{aligned}
$$

the second of these is called the symmetric difference to distinguish it from the more usually used first.

EXERCISE 0: *Give examples of sets A and B where $A \setminus B \neq B \setminus A$. When does $A \setminus B = \emptyset$ and $A \triangle B = \emptyset$? Draw a picture to illustrate $A \triangle B \triangle C$ in the case when all these sets have non-empty intersections.*

Two sets are said to be *disjoint* if they do not intersect. That is, A and B are disjoint if $A \cap B = \emptyset$. A collection of sets is said to be pairwise disjoint if any pair of sets from the collection is disjoint. If A and B are disjoint then: $A \setminus B = A$ and $A \triangle B = A \cup B$.

There are some simple properties of these operations. First, it does not matter what order you take intersections or unions

$$
\begin{aligned}
A \cap (B \cap C) = (A \cap B) \cap C =& A \cap B \cap C \\
A \cup (B \cup C) = (A \cup B) \cup C =& A \cup B \cup C
\end{aligned}
$$

So if A_i, where $i \in I$, is a family of sets indexed by the set I we can unambiguously write unions and intersections over the family as $\bigcup_{i \in I} A_i$ and $\bigcap_{i \in I} A_i$. For example,

$$
[0] = \bigcap_{n \in \mathbb{N}} [0, 1/n], \qquad [0, 1] = \bigcup_{n \in \mathbb{N}} [0, 1/n].
$$

Second, the distributive law applies to unions and intersections

$$
\begin{aligned}
B \cap \left(\bigcup_{i \in I} A_i \right) &= \bigcup_{i \in I} (B \cap A_i) \\
B \cup \left(\bigcap_{i \in I} A_i \right) &= \bigcap_{i \in I} (B \cup A_i)
\end{aligned}
$$

EXERCISE 1: *Prove this using the method described below.*

If $\mathcal{C}$ is a family of subsets of Ω, that is $A \subset \Omega$ for all $A \in \mathcal{C}$, then we can also write down a property that is know as *De Morgan's laws*. (These are really useful in doing probability.)

$$
\left(\bigcap_{A \in \mathcal{C}} A \right)^c = \bigcup_{A \in \mathcal{C}} A^c, \qquad \left(\bigcup_{A \in \mathcal{C}} A \right)^c = \bigcap_{A \in \mathcal{C}} A^c.
$$

To prove set equalities you must show that if x is in the set on the left then it is also in the set on the right, that is, $(\bigcap_{A \in \mathcal{C}} A)^c \subset \bigcup_{A \in \mathcal{C}} A^c$. And you must also show that if x is in the set on right then it is also in the set on the left, that is, $\bigcup_{A \in \mathcal{C}} A^c \subset (\bigcap_{A \in \mathcal{C}} A)^c$.

De Morgan: To do this proof suppose $x \in (\bigcap_{A \in \mathcal{C}} A)^c$. Then there must be at least one $A \in \mathcal{C}$ that does not contain x. This implies $x \in A^c$. So $x \in \bigcup_{A \in \mathcal{C}} A^c$.

To do the other direction suppose that $x \in \bigcup_{A \in \mathcal{C}} A^c$ then there must be at least one A such that $x \in A^c$. If $x \notin A$, then $x \notin \bigcap_{A \in \mathcal{C}} A$. So $x \in (\bigcap_{A \in \mathcal{C}} A)^c$. ■

The two final concepts in this section relate to sequences of sets. Suppose that we have a sequence of sets $(A_n)_{n=1}^{\infty}$ where $A_n \subset \Omega$ for all n. Then, $\cap_{n=k}^{\infty} A_n$ consists of elements that are in all A_n's when $n \geq k$. If x is only in finitely many of the A_n's then it is outside $\cap_{n=k}^{\infty} A_n$ for all k sufficiently large. But this says something stronger, if $x \in \cap_{n=k}^{\infty} A_n$ then all A_n have x in it not just an infinite subsequence. So the set $\cap_{n=k}^{\infty} A_n$ is a lower bound on the sets A_n for all n large. Notice that these sets get weakly bigger as k increases. If we then take the union of these

$$\liminf A_n := \bigcup_{k=0}^{\infty} \left(\bigcap_{n=k}^{\infty} A_n \right)$$

then we get the collection of all x's that are in infinitely many of the A_n *and* are in all A_n for n sufficiently large. This is one notion of the limit of a sequence of sets.

An alternative would be to first consider $\cup_{n=k}^{\infty} A_n$. These are the points that are in some A_k for $k \geq n$. Notice that if x is in a finite number of the A_n, then it will ultimately be outside $\cup_{n=k}^{\infty} A_n$ for some k sufficiently large. But that if x is in infinitely many of the A_n then it is in $\cup_{n=k}^{\infty} A_n$ for all k. Thus if we define

$$\limsup A_n := \bigcap_{k=0}^{\infty} \left(\bigcup_{n=k}^{\infty} A_n \right)$$

we find all x's that are in infinitely many of the A_n.

EXERCISE 2: *Prove that* $\liminf A_n \subset \limsup A_n$.

EXERCISE 3: *Find* $\liminf A_n$ *and* $\limsup A_n$ *when*

$$A_n := \begin{cases} [0, 1 + \frac{1}{n}] & n \text{ even}, \\ [-1 - \frac{1}{n}, 0] & n \text{ odd}, \end{cases}$$

A final notion is the <u>Cartesian product</u> of sets. This is a way of defining ordered pairs/lists/sequences from sets. If A is a set and B is a set, then

$$A \times B := \{(a, b) : a \in A, b \in B\}.$$

This can be done with many sets $\times_{k=1}^{N} A_k := A_1 \times A_2 \times \cdots \times A_n$.

EXERCISE 4: *Show that* $A \times B \neq B \times A$. *And that* $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$.

1.3. *Functions and Maps*

1.3.1. Basics

Given two sets A and B a function f from A to B is a rule that takes every element of A and assigns to it *a unique* element of B. This statement is written as

$$f : A \to B.$$

The element of B assigned to $a \in A$ is denoted $f(a) \in B$. Thus another way of writing down the function is to say that it takes $a \in A$ and maps it to $f(a)$. This is written in symbols as $a \mapsto f(a)$ — note the difference in the arrow between this and what is written above.

There are two important aspects of this definition. First, every $a \in A$ has to be assigned an object in B—it is not enough that most of them are. If a mapping does not assign an object in B to every element of A it is *not* a function. Second, f must assign only one element in B to each $a \in A$. The definition of a function given in Ok is less intuitive than this but equivalent. The image of a set $A' \subset A$ is written as $f(A') \subset B$ where

$$f(A') := \{b \in B : b = f(a) \text{ for some } a \in A'\}.$$

The set $f(A) \subset B$ is called the range of the function f. The set A is called the domain of the function f. If $f(A) = B$ the function is called surjective or onto. A function is said to be injective or one-to-one if each element of A is mapped to a unique element of B. That is $f(a) = f(a')$ if and only if $a = a'$. Finally, a function that is both injective and surjective is called a bijection.

If the function f is a bijection, then it is possible to define a new function which is its inverse. That is

$$f^{-1}(b) := \{a \in A : f(a) = b\}.$$

EXERCISE 5: *Prove that f^{-1} defined above is not a function when f is not a bijection.*

Further points:

- The graph of the function $f : A \to B$ is a subset of $A \times B$ can be written as

$$G := \{(a, b) \in A \times B : a \in A, b = f(a)\}.$$

- A function $f : A \to B$ that has been restricted to lie on a subset $C \subset A$ is written as $f|_C$.
- A function that maps $g : A \times B \to B$ so that the pair $(a, b) \mapsto b$ is called a projection map.

- If there are functions $f : A \to B$ and $g : B \to C$ then the composition function $g \circ f$ is defined as $g \circ f(a) := g(f(a))$.
- If $A' \subset A$ then the indicator function $\mathbb{1}_{A'} : A \to \{0, 1\}$ is defined as

$$\mathbb{1}_{A'}(a) = \begin{cases} 1 & a \in A' \\ 0 & a \notin A'. \end{cases}$$

1.3.2. Sequences as Functions

A sequence in the set A can be thought of as a function that takes the natural numbers $\mathbb{N} = \{1, 2, \dots\}$ and maps them into a set A. We will write such a sequence as $(a_n)_{n=1}^{\infty}$. Thus $a_n \in A$ is the nth element of the list and is the image of $n \in \mathbb{N}$ of the function. The older and still very common notation for this sequences is $\{a_n\}_{n=1}^{\infty}$ but this makes the sequence look like a set. The difference between a sequence and a set is that the order of the objects does not matter for sets but when we talk about a sequence but it definitely does matter. We will write the set of all sequences of elements from the set A as A^{∞}.

Let $(a_n)_{n=1}^{\infty}$ be a sequence in A^{∞} and let $n_1 < n_2 < \dots$ be an increasing sequence of positive integers. Then if look only at the terms $a_{n_1}, a_{n_2}, \dots$ we have built a new sequence $(a_{n_k})_{k=1}^{\infty}$ from the original sequence $(a_n)_{n=1}^{\infty}$. The new sequence $(a_{n_k})_{k=1}^{\infty}$ is called a subsequence.

1.4. The Axiom of Choice & Zorn's Lemma

Let $\mathcal{C}$ be a family of non-empty sets. The number of sets $C \in \mathcal{C}$ might be very large indeed. The Axiom of Choice says that no matter how large $\mathcal{C}$ is it is always possible to think of the operation of selecting one element from each set in C that is in $\mathcal{C}$. This property seems trivial but some very non-trivial mathematical results (Zorn's Lemma and Hausdorff Maximum Principle) are *equivalent* to it.

We now decribe Zorn's Lemma. We begin by describing a partial ordering. A preference ordering is usually assumed to be a *complete* ordering — an individual has preferences that are able to compare all possible consumption bundles. A partial ordering is a generalization of this and describes an individual who is unable to compare certain bundles. Pareto dominance is an example of a partial ordering, because it is not always possible to say that one allocation Pareto dominates another.

Definition 1 *If Ω is a set. A (binary) relation $\succeq$ on Ω is* a partial ordering *if for all $x, y, z \in \Omega$:*

1. *$x \succeq x$,*
2. *$x \succeq y$ and $y \succeq x$ implies $x = y$,*
3. *$x \succeq y$ and $y \succeq z$ implies $x \succeq z$.*

Notice that this definition does not say for any x and y we have to have either $x \succeq y$ or $y \succeq x$. So there are points in Ω that $\succeq$ simply cannot compare. But, there may be subsets of points in Ω that can always be compared. For example, if $\Omega = \{(0,0),(1,1),(2,0),(0,2),(2,2)\}$ then $(0,0) < (1,1) < (2,2)$. All the elements in this subset of Ω can be compared with each other. This is a totally ordered subset of Ω. A totally ordered subset of Ω is called a chain. If $A \subset \Omega$ and $u \succeq a$ for every $a \in A$ then u is a called an upper bound on A. Finally, $m \in \Omega$ is a maximal element of Ω if $y \succeq m$ and $y \in \Omega$ implies that $y = m$. Now we are able to state Zorn's Lemma

Lemma 1 (Zorn's Lemma) *If $(\Omega, \succeq)$ is a partially ordered set (poset) such that every chain has an upper bound in Ω. Then Ω has a maximal element.*

It seems reasonable to ask why is this useful? Fortunately a very eminent mathematician has prepared an answer for you. Just follow this link.

https://gowers.wordpress.com/2008/08/12/how-to-use-zorns-lemma/.

1.5. Countability

As I am sure you already know there are many different types/sizes of infinity. A set is countably infinite if it can be mapped by a bijection to $\mathbb{N} = \{1, 2, \dots\}$. We say that a set is countable if it is either countably infinite or finite.

If X is a countable set this means that we can always write all the elements of X as a simple list. Why? Well the above definition says that there exists a bijection $f^{-1} : \mathbb{N} \to X$. So, for all $n \in \mathbb{N}$ define $x_n := f^{-1}(n)$ as f is a bijection the list $\{x_1, x_2, \dots\}$ is equal to X which is what we wanted to prove.

Here are some basic properties that you need to know

- A subset of a countable set is countable.
- The Cartesian product of two countable sets is countable.
- The rationals $\mathbb{Q}$ are countable.
- A countable union of countable sets is countable.

- An important, and useful property is that a countable sequence of closed non-empty nested intervals has a non-empty intersection. (We need to defer the proof of this to the next section.)
- The real numbers $\mathbb{R}$ are not countable.

The proof of this last claim is quite simple. If the the real numbers $\mathbb{R}$ were countable, then its subset the interval $[0, 1]$ is also countable. Suppose that we could count $[0, 1]$, that is $[0, 1] = \{x_1, x_2, \dots\}$. Now divide $[0, 1]$ into 3 closed intervals $[0, 1/3]$, $[1/3, 2/3]$, $[2/3, 1]$ there must be one of these intervals that does not contain x_1 — call this I_1. Divide I_1 into 3 intervals, there must be one that does not contain x_2 — call this $I_2 \subset I_1$. Repeat this along the sequence (x_n) to generate a sequence of closed nested intervals (I_n) Observe that the set $\cap_{n=1}^{\infty} I_n$ is non-empty (by one of the bullet points above) and contains a point in $[0, 1]$. So we have a found a point in $[0, 1]$ that is not in $\{x_1, x_2, \dots\}$ this is a contradiction.

1.6. Sets of Sets

Here we consider a set that contains other sets as its elements. You have already encountered 2^A the set of all subsets of A. Another example is

$$\mathcal{A}^* = \{\emptyset, \{a, b\}, \{c\}, \{a, b, c\}\}$$

a set of a subsets of $\{a, b, c\}$.

To distinguish the collection from its elements we sometimes call a set of sets a *class* or a *family*. Thus we will try to use the terminology $\mathcal{A}$ is a class or a family and not refer to it as a set. (Although it is!) There are all special classes of sets that have special names — algebras, σ-algebras, and topologies—these are nevertheless sets of sets that have special properties.

1.6.1. Algebras

Recall the basic operations that can be performed on sets: intersection, union and complements. An algebra of sets is a class of sets where taking intersections, unions or complements generates a set that is still in the class. The math-y way of saying this is that the algebra is *closed* under these operations, (i.e. they don't generate objects that are "outside" the algebra).

Definition 2 *Let Ω be a set. Let $\mathcal{A}$ be collection of subsets of Ω. $\mathcal{A}$ is an* algebra *if: (a) $A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$, (b) $A, B \in \mathcal{A}$ implies that $A \cup B \in \mathcal{A}$.*

EXERCISE 6: *Show that if $\mathcal{A}$ is an algebra, then if $A, B \in \mathcal{A}$ then so too is $A \cap B$.*

Algebras and their close cousins σ-algebras are essential if you are going to do probability theory and integration. In that setting the family of sets described by the algebra are thought of as all the events that can occur and the probability of these events is an extra ingredient in the story.

EXERCISE 7: *Is the class $\mathcal{A}^*$ at the start of this section an algebra? If not how many sets must we add to make it so? What is the smallest (least number of sets) algebra of subsets of $\{a, b, c\}$? What is the largest?*

EXERCISE 8: *A finite partition of the set A is a collection of disjoint non-empty sets $\{B_i\}_{i=1}^n$ such that $\cup_i B_i = A$. Let $\mathcal{A}$ be the smallest/coarsest algebra on A that contains $\{B_i\}_{i=1}^n$. How many sets are there in $\mathcal{A}$?*

EXERCISE 9: *Suppose that $A, B \in \mathcal{A}$ where $\mathcal{A}$ is an algebra and $A \cap B \neq \emptyset$. Show that there exists $C, D \in \mathcal{A}$ satisfying $C \cap D = \emptyset$ and $C \cup D = A \cup B$.*

EXERCISE 10: *Suppose that $A_1, A_2, A_3, \ldots, A_N \in \mathcal{A}$ where $\mathcal{A}$ is an algebra. Show that there exists $\cup_{n=1}^N A_n \in \mathcal{A}$.*

There is *always* a smallest algebra (or σ-algebra) generated by a given family of sets. This is a very important concept for probability theory. Suppose that $\mathcal{C}$ is an arbitrary family of subsets of Ω. Now let $\{\mathcal{A}_i : i \in I\}$ be the set of all algebras that contain $\mathcal{C}$ (notice that this is potentially an enormous set). The set $\{\mathcal{A}_i : i \in I\}$ is non empty as 2^Ω is an algebra that contains $\mathcal{C}$. Now define the family

$$\mathcal{A} = \bigcap_{i \in I} \mathcal{A}_i$$

(Notice $\mathcal{A}_i \cap \mathcal{A}_j$ is a collection of subsets of Ω that are both in $\mathcal{A}_i$ and $\mathcal{A}_j$.) This is smaller than all the algebras $\mathcal{A}_i$ and consists of sets that are in all of the $\mathcal{A}_i$'s (like $\mathcal{C}$). Now you should check that $\mathcal{A}$ is an algebra... Thus $\mathcal{A}$ defined above is the smallest algebra generated by $\mathcal{C}$. (By replacing the word algebra above with σ-algebra we can similarly show that there is a smallest σ-algebra generated by $\mathcal{C}$.

1.6.2. σ-algebras

In the definition of an algebra we are allowed to take finite numbers of unions or intersections or complements. In a σ-algebra we are allowed to take countable numbers of unions or intersections or complements.

Definition 3 *Let Ω be a set. Let $\mathcal{A}$ be collection of subsets of Ω. $\mathcal{A}$ is a σ-*algebra *if: (a) $A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$, (b) $(A_n)_{n=1}^{\infty}$ where $A_n \in \mathcal{A}$ for all n then $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$.*

Why are algebras and σ-algebra's different? Here are two examples that should help you understand.

EXERCISE 11: *Suppose that X is an infinite set. Define $\mathcal{A} \subset 2^X$ a family of subsets of X as follows: $A \in \mathcal{A}$ if either A has a finite number of elements or if A^c has a finite number of elements. Show that $\mathcal{A}$ is an algebra but not a σ-algebra.*

EXERCISE 12: *Define $\mathcal{A} \subset 2^{\mathbb{R}}$ a family of subsets of $\mathbb{R}$ as follows. $A \in \mathcal{A}$ if it is the finite union of intervals that are of the form $[a, b)$ or $(-\infty, b)$ where $-\infty < a < b \leq +\infty$. Show $\mathcal{A}$ is an algebra but not a σ-algebra.*

1.6.3. Monotone Classes

The collection $\mathcal{D} \subset 2^{\Omega}$ of subsets of Ω is called a monotone class, if any increasing or decreasing sequence of sets in $\mathcal{D}$ converges to an element of $\mathcal{D}$. That is: (1) if $(D_n)_{n=1}^{\infty}$ satisfies $D_n \in \mathcal{D}$ and $D_n \subset D_{n+1}$ for all n, then $\cup_{n=1}^{\infty} D_n \in \mathcal{D}$. Or (2) $(D_n)_{n=1}^{\infty}$ satisfies $D_n \in \mathcal{D}$ and $D_n \supset D_{n+1}$ for all n, then $\cap_{n=1}^{\infty} D_n \in \mathcal{D}$.

Again a math-y way of saying this is that $\mathcal{D}$ is closed under unions of increasing sequences and intersections of decreasing sequences. An important property of monotone classes is the following

Result 1 (Monotone Class Theorem) Let $\mathcal{A}$ be an algebra of subsets of Ω. Suppose that $\mathcal{D} \subset 2^{\Omega}$ is a monotone class and that $\mathcal{A} \subset \mathcal{D}$. Then, $\mathcal{D}$ also contains the σ-algebra generated by $\mathcal{A}$.

1.6.4. Topologies

Algebras are a tool for doing integration and probability in general spaces. Topologies are a way of thinking about continuity and limits in a more general way. Let Ω be a set/space. A topology on Ω is a way of saying when objects in Ω are near to each other.

Definition 4 *$\mathcal{T}$, a family of subsets of Ω, is a topology on Ω if:*

1. *$\emptyset, \Omega \in \mathcal{T}$,*
2. *If $G_1, G_2 \in \mathcal{T}$ then $G_1 \cap G_2 \in \mathcal{T}$,*
3. *If $\mathcal{S} \subset \mathcal{T}$ then $\cup_{G \in \mathcal{S}} G \in \mathcal{T}$.*

The pair $(\Omega, \mathcal{T})$ is called a topological space.

A topology is closed under finite intersections and arbitrary unions. The sets $G \in \mathcal{T}$ that make up the topology are called the open sets. The pair $(\Omega, \mathcal{T})$ is called a topological space.

The sets that are complements of the open sets are called "closed", that is, if $F = G^c$ for some $G \in \mathcal{T}$ then F is closed. Notice that this (and (1) in the definition) implies that $\emptyset$ and Ω are both open and closed! (Such sets are sometimes called clopen.)

We will later see that the most general definition of continuity will be to say that if a function maps one topological space to another, the inverse image of an open set is an open set. An example of a weird topology is the *cofinite topology* this is the collection of subsets of Ω that have a finite complement.

$$\mathcal{T}_{(0)} := \{G \subset \Omega : G = \emptyset, \text{ or } \Omega \setminus G \text{ is finite}\}$$

EXERCISE 13: *Verify that this satisfies the properties of a topology.*

There are at least three differences between a topology and a σ-algebra —the permissible number of unions, the permissible number of intersections, and the absence of complements.

One of the most important ways of generating a topology is to think of families of sets that contain a given element. This is called a neighbourhood basis.

Definition 5 *The family of subsets $\mathcal{B}_a \subset 2^\Omega$ is said to be a neighbourhood basis at $a \in \Omega$ if:*

1. *$a \in B$ for all $B \in \mathcal{B}_a$.*
2. *If $B_1, B_2 \in \mathcal{B}_a$ then there exists $B_3 \in \mathcal{B}_a$ such that $B_3 \subset B_1 \cap B_2$.*

A simple example of a neighbourhood basis would be the collection of open intervals containing a given point. That is, for $x \in \mathbb{R}$ a neighbourhood basis for x would be the family $\mathcal{B}_x := \{\ (x - \nu, x + \nu) : \forall \nu > 0\}$. The sets $(x - \nu, x + \nu)$ are also called the "open balls at x". These are sometimes written as $B(x, \nu)$ A neighbourhood basis or simply a basis is a family of sets that form a neighbourhood basis for every point in the set. That is,

Definition 6 *The family of subsets $\mathcal{B} \subset 2^\Omega$ is said to be a neighbourhood basis if for each $a \in \Omega$ the family $\{B \in \mathcal{B} : a \in B\}$ is a neighbourhood basis at a.*

There's a simple way of checking whether a $\mathcal{B}$ is a basis: $\mathcal{B}$ is a basis iff $\cup_{B \in \mathcal{B}} B = \Omega$ and if $x \in B_1 \cap B_2$ and $B_1, B_2 \in \mathcal{B}$ then there exists $B_3 \in \mathcal{B}$ such that $x \in B_3 \subset B_1 \cap B_2$.

Now we are able to use the neighbourhood basis on Ω to generate a topology on Ω. We will use the basis family $\mathcal{B} \subset 2^{\Omega}$ to test whether a larger family of sets do indeed make up the topology. We will say that the set $G \subset \Omega$ is "open with respect to $\mathcal{B}$" if for every $x \in G$ there exists a $B \in \mathcal{B}$ such that $x \in B \subset G$. The set of all G that are "open with respect to $\mathcal{B}$" form a topology on Ω. (As an exercise check that this is so.) The topology created in this way is called the topology determined by $\mathcal{B}$.

The topology you are most used to is the topology determined by the basis of open intervals $\mathcal{B} := \{(x - \nu, x + \nu) : x \in \mathbb{R}, \nu > 0\}$.

1.6.5. Borel Sets

σ-algebras and topologies do different jobs: One allows you to integrate and to do measure theory. The other allows you have talk about continuity and convergence. But suppose you don't want to do only one of these jobs but you want to do both. Then you need the Borel sets. Suppose that $(\Omega, \mathcal{T})$ is a topological space. To do probabilities on such a space you need to be able to take countable intersections—but this lead you outside the topology. Thus what you need to do is to take the smallest σ-algebra that contains all the open sets in $\mathcal{T}$. (We know this exists by the calculation at the end of Section 1.6.1 we also know how to recognize it by the monotone class theorem.) This generates a σ-algebra, which we will write as $\mathcal{B}(\Omega)$. The sets that make up this σ-algebra are called the Borel sets. The σ-algebra $\mathcal{B}(\Omega)$ is often called the Borel σ-algebra in papers.

Analysis for Economists: ECON0118

Martin Cripps

November 2, 2020

Contents

2. Real Numbers, Sequences and Functions

We have already introduced the notation $\mathbb{R}$ for the real line. Now we will look at some of the key properties of the reals. As well as sequences and series of reals, sets of the reals and functions of the reals.

2.1. Real Numbers

There are three key properties that the real numbers satisfy: they are an ordered field that satisfies a completeness axiom. It is possible to spend a lot of time trying to build the real numbers up from scratch. We will not do this we will instead point out that they have 3 essential properties

2.1.1. Field Axioms

The first is that they satisfy a list of properties that define a *field* these are as follows. For any $x, y, z \in \mathbb{R}$:

- Commutative: $x + y = y + x$ and $xy = yx$.
- Associative: $(x + y) + z = x + (y + z)$ and $(xy)z = x(yz)$.
- Distributive $x(y + z) = xy + xz$
- Identities: There exists $0 \neq 1 \in \mathbb{R}$ such that $x + 0 = x$ and $x.1 = x$.
- Inverses: There exists $-x$ such that $x + (-x) = 0$ and if $x \neq 0$ there exists x^{-1} such that $xx^{-1} = 1$.

2.1.2. Order Axioms

There exists an order $<$ on $\mathbb{R}$ such that for any $x, y, z \in \mathbb{R}$:

- Transitive: $x < y$ and $y < z$ implies $x < z$.
- $x < y$ implies $z + x < z + y$.
- $x < y$ and $0 < z$ implies $xz < yz$.
- Exactly one of $x < y$, $x = y$, and $y < x$ holds.

2.1.3. Completeness Axiom

Let $A \subset \mathbb{R}$ be a set of real numbers. u is called an upperbound on A if $x \leq u$ for all $x \in A$. (A set does not necessarily have an upper bound.) If a set has an upper bound it is called bounded above. (Lower bound and bounded below are defined equivalently.) The following axiom is sometimes called the continuum property.

Axiom 1 (Completeness) *If A is a non-empty set that is bounded above, then there exists an upper bound u^* such that $u^* \leq u$ for all upper bounds of A.*

u^* is the least upper bound of A and is called the supremum of the set. It is written as

$$\sup A = \sup\{x : x \in A\} = \sup_{x \in A} x.$$

For example, $\sup[0,2) = \sup[0,2] \cap \mathbb{Q} = 2$. The supremum generalizes the notion of the maximum to the case where it does not exist. A bounded set always has a supremum but may not have a maximum. Similarly, the greatest lower bound of a set A that is bounded below is denoted $\inf A$ and exists by the completeness axiom applied to the set $-A$. We use these property to establish 2 results:

Result 2 (Archimedean Priniciple) For any $x \in \mathbb{R}$ there is an $n \in \mathbb{N}$ such that $n > x$.

Here's the intuition for this. (The set of integers less than or equal to x has an upper bound and so by the above axiom has a least upper bound — call this u^*. It follows that $x \in (u^*, u^* + 1]$ — otherwise there would be a greater integer less than x. There must also be an integer in the interval $(u^* - 1, u^*]$. Call this n^*. Then $n^* + 2$ is an integer upper bound on this set of integers and $n^* + 2 > u^*$.)

Result 3 If $a < b$ then there is an irrational number and a rational number in the interval (a, b).

Why? We know (a, b) is uncountable but the rationals are countable. So if there were only rationals in (a, b) this would be an uncountable subset of a countable set — a contradiction.

Suppose instead there was no rational in (a, b). Then divide the line up into very short segments with an integer length so that the ends of one of these segments lies in (a, b). For example choose an integer $n > 2(b - a)$. Note that $\frac{1}{n} < \frac{1}{2(b-a)}$. Now just show there exists an integer m so that $m/n \in (a, b)$.

Sometimes we want to add infinity to our number systems. We do this by asserting that $-\infty$ is a lower bound on every subset of $\mathbb{R}$ and $+\infty$ is an upper bound every subset of $\mathbb{R}$.

2.2. Sequences

2.2.1. Sequences and Convergence

Recall (from Section 1.3.2) that a sequence is written as $(x_n)_{n=1}^{\infty}$ is an ordered subset of $\mathbb{R}$ and can be interpreted as a function from $\mathbb{N}$ into $\mathbb{R}$. One of the most important definitions in these notes is when a sequence converges to a limit x. The idea is that you can choose any margin of error $\varepsilon > 0$ and all sufficiently large elements of the sequence are within this margin of error of x.

Definition 7 *The sequence $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}$ converges to x if for any $\varepsilon > 0$ there exists N such that $|x - x_n| < \varepsilon$ for all $n > N$.*

This is written as $\lim_{n\to\infty} x_n = x$.

EXERCISE 14: *Which of the following sequences converges to 0?* $x_n := ne^{-n}$, $x_n = \frac{n^2-1}{n+1}$, $x_n = \frac{n^2-1}{(n+1)^2}$, $x_n = \frac{n^2-1}{(n+1)^3}$, $x_n = \sin(n\pi)$.

Sequences that converge to $+\infty$ and $-\infty$. Here the test of the sequence is roughly the same idea. For all n sufficiently large the sequence is "close enough" to the limit. Now, close enough means that the sequence is greater than some number.

Definition 8 *The sequence $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}$ converges to ∞ if for any $B \in \mathbb{R}$ there exists N such that $x_n > B$ for all $n > N$. The sequence $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}$ converges to $-\infty$ if for any $B \in \mathbb{R}$ there exists N such that $x_n < B$ for all $n > N$.*

This is written as $\lim_{n\to\infty} x_n = \infty$ or $\lim_{n\to\infty} x_n = -\infty$. Now we will give some important results about sequences and convergence. The sequence $(x_n)_{n=1}^{\infty}$ is <u>bounded from above</u> if there exists $K \in \mathbb{R}$ such that $x_n \leq K$ for all n. This (of course) is equivalent to $\sup_n x_n \leq K$.

EXERCISE 15: *Show that if $(x_n)_{n=1}^{\infty}$ converges to x, where x is finite, then the sequence $(x_n)_{n=1}^{\infty}$ is bounded from above.*

A sequence $(x_n)_{n=1}^{\infty}$ is monotonic if either: it is increasing $x_n \leq x_{n+1}$ for all n, or it is decreasing $x_n \geq x_{n+1}$ for all n.

Result 4 Every increasing real sequence that is bounded from above converges.

Why? Well because the set of values for the sequence $\{x_n : n \in \mathbb{N}\}$ is bounded from above. By the continuum/completeness property this set has a least upper bound. The sequence must converge to this, because if it did not you could find a strictly smaller upper bound.

Result 5 Every real sequence $(x_n)_{n=1}^{\infty}$ has a monotonic subsequence.

Why? Think of the tail of the sequence, that is, the points $S_m := \{x_m, x_{m+1}, \dots\}$. Suppose that the some set S_m has no maximal element. Then choose the first element of the subsequence to be x_m and go along the sequence of $(x_n)_{n=1}^{\infty}$ and add x_n to the subsequence every time a new largest element appears. (Generating an increasing subsequence.) Suppose that all the sets S_m have a maximal element. Make the first element of the subsequence the maximal element of S_1. (Call it x_{n_1}.) Make the second element of the subsequence the maximal element of S_{n_1} and so on. (Generating a decreasing subsequence.)

The combination of the previous two results gives

Result 6 (Bolzano Weierstrass) Every bounded real sequence $(x_n)_{n=1}^{\infty}$ has a convergent subsequence.

EXERCISE 16: *Prove this.*

One of the problems with the definition of convergence above is that you need to know what the limit of the sequence is before you can test whether it converges. There is a different definition of convergence that avoids this problem. It just checks that the sequence does not move very much in the future.

Definition 9 (Cauchy Sequence) *The sequence $(x_n)_{n=1}^{\infty}$ is called a Cauchy Sequence if for each $\varepsilon > 0$ there exists N such that $|x_n - x_m| < \varepsilon$ for all $n, m \geq N$.*

For many spaces Cauchy sequences and convergent sequences are *not* the same. However, in $\mathbb{R}$ these two notions are equivalent.

Result 7 The sequence $(x_n)_{n=1}^{\infty}$ of real numbers converges to $x \in \mathbb{R}$ if and only if it is a Cauchy sequence.

Why are they equivalent? The easy thing to see is convergent $\Rightarrow$ Cauchy. If there is a limit x for any $\varepsilon > 0$ there exists N so that $|x_n - x| < \varepsilon/2$ for all $n > N$. This implies that for any $n, m > N$

$$|x_n - x_m| = |x_n - x + x - x_m| \leq |x_n - x| + |x - x_m| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon$$

and hence the sequence is Cauchy. To show Cauchy$\Rightarrow$ convergent, suppose the sequence is Cauchy then it has to be bounded above and bounded below. If $|x_n - x_m| < \varepsilon$ for all $n, m \geq N$ then $|x_N - x_m| < \varepsilon$ for all $m > N$. So the sequence is bounded above by $\max\{x_1, \dots, x_{N-1}, x_N + \varepsilon\}$ and bounded below by $\min\{x_1, \dots, x_{N-1}, x_N - \varepsilon\}$. It therefore has a convergent subsequence $(x_{n_m})_{m=1}^{\infty}$. Let x be the limit of this subsequence. We can now show that the entire sequence converges to x. Suppose $\varepsilon > 0$ is given, choose N sufficiently large so that $|x_{n_m} - x_n| \le \varepsilon/3$ for all $n_m, n > N$. And choose M sufficiently large so that $|x_{n_m} - x| \le \varepsilon/3$ for all $n_m > M$. Then for all $n > \max\{N, M\}$

$$|x_n - x| \le |x_{n_m} - x_n| + |x_{n_m} - x| \le 2\varepsilon/3 < \varepsilon.$$

Hence the sequence converges to x.

2.2.2. Cluster Points, $\liminf$, and $\limsup$

We know sequences do not necessarily converge. What can we say about non-convergent sequences and how can we characterize their long run behaviour? Cluster points describe places a sequence re-visits infinitely many times. $\liminf$ and $\limsup$ define lower and upper bounds on the limiting behaviour of the sequence. An alternative way of thinking about them is as the largest and smallest limit of a convergent subsequence.

Definition 10 *x is a* <u>*cluster point*</u> *of the sequence $(x_n)_{n=1}^{\infty}$ of real numbers if for any $\varepsilon > 0$ and $N \in \mathbb{N}$ there exists $n > N$ such that $|x - x_n| < \varepsilon$.*

EXERCISE 17: *Find the cluster points of the sequence $x_n := \sin(n\pi/k)$ where k is a positive integer.*

You might be interested in the largest and the smallest possible cluster points of a sequence. These are called the $\liminf$ and the $\limsup$ of the sequence

$$\limsup_{n\to\infty} x_n := \inf_n \sup_{k \ge n} x_k, \qquad \liminf_{n\to\infty} x_n := \sup_n \inf_{k \ge n} x_k,$$

These alway exist (although they might be infinite). (You should compare these definitions for sequences with their definitions for sets.) Some useful properties of $\limsup$ are listed below. (There are equivalent ones for $\liminf$.)

- For any $\varepsilon > 0$, there is an N so that $x_n \le \limsup x_n + \varepsilon$ for all $n > N$.
- For any $\varepsilon > 0$, there is an M so that $x_n \ge \limsup x_n - \varepsilon$ for some $n > N$.

2.2.3. Series

A important mathematical notion is a sum of an infinite number of objects. The finite sum $\sum_{i=1}^{N} x_n$ is clearly well defined by the properties we have defined so far. This is called "finite series". However we would also like to consider infinite sums or infinite series. Thus we would like to consider $\sum_{i=1}^{\infty} x_n$ when it is defined. We define

$$\sum_{i=1}^{\infty} x_n := \lim_{N\to\infty} \sum_{i=1}^{N} x_n$$

when it exists. Sometimes this limit is finite. In this case the series is called convergent. Sometimes it is ∞, sometimes it is $-\infty$ in these cases the series is called divergent. Sometimes there is no limit at all. Then our only options are to think of lim inf and lim sup.

EXERCISE 18: *What are* $\liminf_{N\to\infty} \sum_{n=1}^{N} (-1)^n$ *and* $\limsup_{N\to\infty} \sum_{n=1}^{N} (-1)^n$*?*

One famous way of determining whether the limit exists is to look at the ratio of successive terms

$$r_n := \left| \frac{x_{n+1}}{x_n} \right| .$$

If $\lim_{n\to\infty} r_n < 1$ then the series converges. If $\lim_{n\to\infty} r_n > 1$ then the series diverges. However, if $\lim_{n\to\infty} r_n = 1$ then we can reach no definite conclusion.

This is an enormous subject. There is are many useful taxonomies of which series converge and which diverge. Wolfram Alpha, Mathematica, Wikipedia will all help with this.

2.3. *Open and Closed Sets of* $\mathbb{R}$

2.3.1. Open Sets

Here we are going to talk about the topology we impose on the real line. In Section 1.6.4 we have talked about open and closed sets and neighbourhood bases in an abstract way. Here we will see that these ideas arose from properties of the real line.

Hence we need to define a class or family of sets in $\mathbb{R}$. We will define a set $G \subset \mathbb{R}$ to be "open" if it passes the following test. For every $x \in G$ we can find an open ball $(x-\delta, x+\delta)$ surrounding this point so that $(x-\delta, x+\delta) \subset G$. (In the real line these open intervals form a basis for the topology.) This definition immediately implies that the intervals $(x-\delta, x+\delta)$ are also open. It also implies that the singleton sets $\{x\}$ or $\mathbb{N}$ are not open.

Sometimes we want to think of classes of sets G that are defined on D a subset of $\mathbb{R}$. In that case the definition of openness needs to be revised. G is said to be "open in D" if for every $x \in G$ there is an $\delta > 0$ so that $(x-\delta, x+\delta) \cap D \subset G$.

We now want to check that this definition of an open set defines a class, or family, of subsets of $\mathbb{R}$ that are a topology. Thus we need to check that this family satisfies Definition 4.

- *$\mathbb{R}$ and $\emptyset$ open:* Trivial.
- *G_1 and G_2 open implies $G_1 \cap G_2$ open:* If $x \in G_1 \cap G_2$ take the intersection of the two balls that contain it. This is a subset of $G_1 \cap G_2$.
- *G_i ($i \in I$) all open implies $G = \cup G_i$ open:* If $x \in G$ then $x \in G_i$ for some i use the ball for this set.

Here we can now see the difference between topologies and σ-algebras. If we take a countable intersection of open sets we can generate an non-open set, for example $\{0\} = \cap_n(-n^{-1}, n^{-1})$, but if we take a countable intersection in a σ-algebra we still get a set in the algebra.

Now we are able to characterize all the open sets in $\mathbb{R}$. This characterization is a special feature of $\mathbb{R}$ and its topology. It is not true in general

Result 8 If $G \subset \mathbb{R}$ is open, then G is a countable union of disjoint open intervals.

The way this proof goes is for each $x \in G$ define G_x to be the union of all the open intervals that contain x and are subsets of G. Show that G_x is itself an open interval and then show that for two different points either $G_x = G_y$ or $G_x \cap G_y = \emptyset$. Thus G can be written as a union of open intervals. To show this union is countable just pick a rational number in each interval.

2.3.2. Closed Sets

There are a number of different ways of defining a closed set. Previously we defined a set to be closed if it was the complement of an open set. Now we are going to define the closed sets in $\mathbb{R}$ in a different way and show that this is equivalent to the earlier definition.

We begin by defining the limit points of a set $A \subset \mathbb{R}$. This can be though of as the limit of a convergent sequence of points in A.

Definition 11 *x is a limit point of $A \subset \mathbb{R}$ if for any $\varepsilon > 0$ there exists $y \in A$ such that $|y - x| < \varepsilon$. The set of all limit points of A is denoted $\bar{A}$ and is called the closure of A.*

EXERCISE 19: *Show that $\bar{\mathbb{Q}} = \mathbb{R}$.*

A closed set is defined to be equal to its closure. That is, A is closed if and only if $A = \bar{A}$. We can show that this definition is equivalent to our earlier definition of closure.

Result 9 $A = \bar{A}$ if and only if A^c is open.

EXERCISE 20: *Write down this proof.*

2.4. Real Valued Functions

A function f that maps the set Ω into $\mathbb{R}$ is said to be real valued. Here we will think of $\Omega \subset \mathbb{R}$ so f maps real numbers to real numbers.

Here are some examples of real-valued functions

- $f(.)$ is an even function if $f(-z) = f(z)$ for all z in its domain. $f(.)$ is an odd function if $f(-z) = -f(z)$ for all z in its domain.
- f is a non-zero polynomial degree n if $f(z) \equiv a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$ and $a_n \neq 0$.
- f is a rational function if $f(z) \equiv p(z)/q(z)$ and p, q are non-zero polynomials.
- $f : [a, b] \to \mathbb{R}$ is called a step function if there exists $z_0 < z_1 < \cdots < z_n \in [a, b]$ such that $f(z) = f(z')$ for all $z, z' \in (z_m, z_{m+1})$.

One useful property we will state without proof is that if a polynomial is equal to zero on an open interval, then it is zero everywhere.

Probably the most important definition here is continuity. A function is continuous if the function does not change much if there is a sufficiently small change in where it is evaluated. We will give two apparently different definitions of continuity—but actually one is much more general than the other.

Definition 12 (Continuity 1) *The function $f : \Omega \to \mathbb{R}$ is continuous at $x_0 \in \Omega \subset \mathbb{R}$ if for any $\epsilon > 0$ there exists a $\delta > 0$ so that $|f(x) - f(x_0)| < \epsilon$ if $x \in \Omega$ and $|x - x_0| < \delta$. It is continuous on Ω if it is continuous on every $x_0 \in \Omega$.*

Definition 13 (Continuity 2) *The function $f : \Omega \to \mathbb{R}$ is continuous on $\Omega \subset \mathbb{R}$ if for any open set $G \subset \mathbb{R}$ the set*

$$f^{-1}(G) := \{x \in \Omega : f(x) \in G\}$$

is open in Ω.

Actually the definition that the inverse image of an open set is itself open can be applied to functions between any two topological spaces, so this is the most general notion of continuity we have.

EXERCISE 21: *If f is continuous and F is a closed subset of $\mathbb{R}$ show that $f^{-1}(F)$ is closed in Ω.*

How would one show that Definition 12 is equivalent to Definition 13? To do Definition 12 $\Rightarrow$ Definition 13 take an open set $G \subset \mathbb{R}$ by openness every $f(x_0) \in G$ can be surrounded by an open interval in G. By continuity if $f(x)$ is in this interval then x is in an open interval around x_0. Take the union of all these open intervals for all $f(x_0) \in G$. This means the inverse is the union of open sets. To do Definition 13 $\Rightarrow$ Definition 12 just observe that one kind of open set one can consider is $(f(x_0) - \epsilon, f(x_0) + \epsilon)$.

A related notion (a weakening of continuity that preserves some important features) that we will use later is that of semi-continuity.

Definition 14 *The function $f : \mathbb{R} \to \mathbb{R}$ is: (a)* upper semicontinuous *if $f^{-1}([\infty, r))$ is open for every $r \in \mathbb{R}$. (a)* lower semicontinuous *if $-f$ is upper semicontinuous.*

First observe that a continuous function is both upper and lower semicontinuous. So now consider the step function $f(x) = 0$ for all $x \leq 0$ and $f(x) = 1$ for all $x > 0$. The inverse image of the interval $[-\infty, 0.5)$ is the closed interval $[-\infty, 0]$ thus this function fails the test for upper semicontinuity. However, it passes the test for lower semicontinuity. Loosely speaking a function is upper semicontinuous if the better than set is closed. That is, if the set $\{x : f(x) \geq r\}$ is closed. Upper semicontinuous functions defined on compact sets have maxima.

Now let us describe some strengthenings of continuity. Notice first that the definition of continuity is local — the ε and the δ chosen can depend on the x_0 where the function is being tested for continuity. So to check that e^x is close to e^{x_0} a great deal depends on how large x_0 is.

The first strengthening of continuity requires that the same δ and ϵ work for all x_0. This is called uniform continuity.

Definition 15 *The function $f : \Omega \to \mathbb{R}$ is uniformly continuous if for any $\epsilon > 0$ there exists a $\delta > 0$ so that $|f(x) - f(x_0)| < \epsilon$ if $x, x_0 \in \Omega$ and $|x - x_0| < \delta$.*

The next strengthening is to link the size of ϵ to the size of $|x - x_0|$.

Definition 16 *The function $f : \Omega \to \mathbb{R}$ is α-Hölder continuous if for some $\alpha > 0$ there exists $K > 0$ such that*

$$|f(x) - f(x_0)| < K|x - x_0|^\alpha, \qquad \forall x, x_0 \in \Omega.$$

This can be strengthened further in the following ways Lipschitz Continuous ($\alpha = 1$), Nonexpansiveness ($\alpha = 1$ and $K = 1$), and contraction ($\alpha = 1$ and $K < 1$).

EXERCISE 22: *Consider the function $\sqrt{x}$ on the domain $[0, \infty)$ and on the domain $[1, \infty)$, which of these notions of continuity holds?*

We now investigate some important consequences of continuity of real valued functions that are defined on closed intervals.

Result 10 The function $f : [a, b] \to \mathbb{R}$ is continuous if and only if it is uniformly continuous.

Uniform continuity $\Rightarrow$ continuity is trivial. To show that continuity on $[a, b]$ implies uniform continuity is harder. Suppose there was a continuous function on $[a, b]$ that was not uniformly continuous then to ensure that $|f(x) - f(x')| < \epsilon$ was small we could not choose one δ that is there are sequences (x_n) and (x'_n) in $[a, b]$ such that $|x_n - x'_n| < 1/n$ but $|f(x_n) - f(x'_n)| > \epsilon$. These sequences have convergent subsequences (by Result 6). They must converge to the same limit contradicting $|f(x_n) - f(x'_n)| > \epsilon$ for all n and the continuity of f.

EXERCISE 23: *Give an example to show that this is false if f is defined on an open interval.*

The next result we use in many guises to establish existence of something.

Result 11 (Intermediate Value Theorem) If $f : [a, b] \to \mathbb{R}$ is continuous and $v \in [f(a) f(b)]$ then there exists $x^* \in [a, b]$ satisfying $f(x^*) = v$.

Why? If $f(a) = f(b)$ then we can choose $x^* = a$, so suppose $f(a) < v < f(b)$. Now consider the set $S := \{x \in [a, b] : f(x) \leq v\}$. This is non-empty $a \in S$ and bounded above by b, so $c = \sup S$ exists. There exists a sequence $(x_n) \subset S$ such $x_n \to c$. By continuity $\lim f(x_n) = f(c)$ and $f(c) \leq v$. Suppose $f(c)$ is strictly less than v. Then for all $x > c$ we have $f(x) \geq v$. Thus the function f has a discontinuity at c — a contradiction. So $f(c) = v$ and we are done.

The next result is a key property that makes maximization of continuous functions a simple matter.

Result 12 If $f : \mathbb{R} \to \mathbb{R}$ is continuous and $C \subset \mathbb{R}$ is a compact (closed and bounded), then $f(C)$ is compact.

Why? If $f(C)$ was not bounded we could find a sequence $(x_n) \in C$ so that $|f(x_n)| \to \infty$. But (x_n) is in C so it has a convergent subsequence. By continuity $f(x_n)$ also has a convergent subsequence. This contradicts $|f(x_n)| \to \infty$. Thus $f(C)$ is bounded. Now we need to show $f(C)$ is closed. Suppose we have a limit point of a sequence in $f(C)$. This is generated by a sequence with a convergent subsequence. Continuity then shows the limit point is in $f(C)$.

Finally a very important result that we will prove again more generally.

Result 13 If $f : C \to \mathbb{R}$ is continuous and $C \subset \mathbb{R}$ is compact then f attains a maximum and a minimum on C.

As $f(C)$ is compact there exists $\sup_{x \in C} f(x)$ define this to be the number M. We need to show that there exists $x^* \in C$ such that $f(x^*) = M$. There must exist a sequence $(x_n) \subset C$ so that $f(x_n) \to M$. This sequence has a convergent subsequence so we can assume that $x_n \to x^*$ and $x^* \in C$ as C is closed. Then by continuity we have that $f(x^*) = M$.

2.5. Sequences of Functions

We will now start thinking about sets and sequences of real-valued functions. Let $D \subset \mathbb{R}$ and let $\mathcal{C}(D)$ be the set of continuous functions that map D to $\mathbb{R}$. We know that this set is close under addition, multiplication by a scalar and multiplication: if $f, g \in \mathcal{C}(D)$ and $\alpha \in \mathbb{R}$ then so is $f + g$, αf and $f \times g$.

The strongest and most desirable mode of convergence for functions in $\mathcal{C}(D)$ is termed pointwise convergence. This is defined below.

Definition 17 *The sequence $(f_n)_{n=1}^{\infty}$ of functions in $\mathcal{C}(D)$ converges pointwise on D if for every $x \in D$ the real valued sequence $(f_n(x))_{n=1}^{\infty}$ converges in $\mathbb{R}$.*

If $(f_n)_{n=1}^{\infty}$ converges pointwise, then we can use the limits at each x to define a new function $f(x) := \lim_{n\to\infty} f_n(x)$. What is very important is that we do *not* know that this limiting function is also continuous.

Consider the sequence of functions $(f_n)_{n=1}^{\infty}$: $f_n : [0, 1] \to \mathbb{R}$ where $f_n(x) := x^n$. If $x < 1$ the sequence $x^1, x^2, x^3, \ldots$ converges to zero. If $x = 1$ the sequence $x^1, x^2, x^3, \ldots$ converges to one. Thus $(f_n)_{n=1}^{\infty}$ converges pointwise to a function that is zero on $[0, 1)$ and one at $x = 1$.

Now consider the same sequence of functions but on the larger domain $f_n : [-1, 1] \to \mathbb{R}$. This does not converge pointwise because $f_n(-1)$ does not converge.

One question one might want to ask is: if you take a sequence of functions in a set and then take a pointwise limit, do you generate a function that is also in that set? This property is called closure under pointwise limits. (Notice this does not require that a pointwise limit exists for every sequence of functions, only that when you have chosen a sequence with a pointwise limit then the limit is in the set.)

If you make the set of functions you consider sufficiently large, for example any function from D to $\mathcal{R}$, then it is automatic that if a pointwise limit exists then it is also in the set. The example above shows that the continuous functions are not closed under pointwise limits. You can take a sequence of continuous functions that has a pointwise limit, but the limit is not continuous. Is there a way of ensuring closure of a smaller set? The answer is yes provided you take limits in a different way…

Definition 18 *The sequence $(f_n)_{n=1}^{\infty}$ of functions on D converges uniformly to f, if for every $\varepsilon > 0$ there exists N such that $|f_n(x) - f(x)| < \varepsilon$ for all $n > N$ and all $x \in D$.*

Here uniformly means that f_n is close to f for all the x's that are possible. Uniform convergence is stronger than pointwise convergence. For example $f_n(x) = x^n$ converges pointwise on $[0, 1]$ but it does not converge uniformly. For any n we can find an $x \in [0, 1)$ so that $x^n > \varepsilon$. So although x^n converges pointwise to zero on $[0, 1)$ this convergence is not uniform.

Now we have the following result, which shows that uniform convergence preserves continuity. Or that the set of continuous functions is closed under uniform convergence.

Result 14 If a sequence $(f_n)_{n=1}^{\infty}$ of functions in $\mathcal{C}(D)$ converges uniformly to f, then $f \in \mathcal{C}(D)$ (is continuous).

Why? We need to show that f the limit is continuous. So we need to show that we can make $|f(x) - f(x')| < \epsilon$ if we make $|x - x'|$ small. We know, as f_n is continuous, that we can make $|f_n(x) - f_n(x')| < \epsilon/3$ if we make $|x - x'| < \delta$. We know we can make $|f_n(x) - f(x)| < \epsilon/3$ and $|f_n(x') - f(x')| < \epsilon/3$ if we make n large (by uniform convergence). Now we can combine these (exercise) to show $|f(x) - f(x')| < \epsilon$ if $|x - x'| < \delta$.

2.6. The Cantor Set and Function

There are two ways of describing the Cantor set. The first is an interrative process of defining a sequence of subsets of the interval $P_0 := [0, 1]$. Start by taking the open middle interval of this set $G_1 := (1/3, 2/3)$ and remove it, that is, $P_1 := P_0 \setminus G_1$. The set $P_1 = [0, 1/3] \cup [2/3, 1]$ consists of two closed intervals, take the middle intervals of these sets $G_2 := (1/9, 2/9) \cup (7/9, 8/9)$ and remove them to define the set $P_2 := P_1 \setminus G_2$. The set P_2 consists of $4 = 2^2$ distinct closed intervals of length $(1/3)^2$. If we repeat this n times the set P_n consists of 2^n intervals of length $(1/3)^n$. Thus the total length of $P_n = (2/3)^n \to 0$. Suppose this is repeated an infinite number of times and we define

$$G = \bigcup_{n=1}^{\infty} G_n, \qquad P = \bigcap_{n=1}^{\infty} P_n.$$

G is an open set — it is a countable union of open sets. G the sum of the lengths of the open sets that have been removed equals the entire interval $\sum_{i=1}^{\infty}(\frac{2^{n-1}}{3^n}) = 1$. P is a closed set—is the intersection of closed sets. It contains no interval of strictly positive length. (IF there was an interval longer than $(1/3)^n$ then it would have been removed at the $n+1$th round. P is called the Cantor Ternary Set.

We can also describe the Cantor Ternary Set using discounting. Suppose an agent has a discount factor $\delta = 1/3$ and receives an infinite stream of payoffs $u_1, u_2, \dots$, then there discounted present value would be

$$\frac{1}{3}u_1 + \left(\frac{1}{3}\right)^2 u_2 + \left(\frac{1}{3}\right)^3 u_3 + \left(\frac{1}{3}\right)^4 u_4 + \dots$$

Suppose that u_n can take one of three different values: $u_n \in \{0,1,2\} := U$. Then the largest utility this agent could get is $\sum_{i=1}^{\infty} 2(\frac{1}{3})^n = 1$ and the smallest they could get is zero. In fact by choosing sequences (u_n) of different values in $\{0,1,2\}$ you can get any real utility in the interval $[0,1]$. The sequence $(u_n) \in \{0,1,2\}^\infty$ such that $x = \sum_{i=1}^{\infty} u_n(\frac{1}{3})^n$ is called the "ternary expansion of x". If we consider the set of possible values of $x = \sum_{i=1}^{\infty} u_n \frac{1}{3^n}$ where u_1 cannot take the value 1, but must be either 0 or 2, then this consists of the two intervals $[0,1/3]$ and $[2/3,1]$. Thus the Cantor ternary set consists of points

$$x = \sum_{i=1}^{\infty} u_n \frac{1}{3^n}, \qquad (u_n) \in \{0,2\}^\infty.$$

To write this formally we could write

$$P = \bigcap_{n=1}^{\infty} P_n = \left\{ x : x = \sum_{i=1}^{\infty} \frac{u_n}{3^n}, \qquad (u_n) \in \{0,2\}^\infty \right\}.$$

Although the Cantor set appears to be very small—it has zero Lebesgue measure—it is very large in another sense because it is uncountable. To see this observe that if $x \in P$ then there is a unique (u_n) that describes it. Now define a function $h : P \to [0,1]$ that takes this (u_n) and maps it to a different sum

$$h(x) = h((u_n)) = \sum_{i=1}^{\infty} \frac{(u_n/2)}{2^n}.$$

Observe that when $u_n = 2$ for all n then $h((u_n)) = 1$ and that by choosing $(u_n) \in \{0,2\}$ appropriately the expression on the right equals the binary expansion of some number

$y \in [0,1]$. Therefore, h maps P onto the uncountable set $[0,1]$. It follows that P is uncountable.

Suppose we use $h : P \to [0,1]$ to define a function that takes $[0,1]$ to $[0,1]$ by adding in flat bits on $[0,1] \setminus P$. This is called the Cantor function —it is non-decreasing and continuous. But (we will show later) has a zero derivative almost everywhere!

2.7. *Riemann Integration*

Suppose we have an interval $[a,b]$ and subdivide it into closed intervals $[a_0,a_1]$, $[a_1,a_2]$,..., $[a_{n-1},a_n]$; where $a_0 = a$ and $a_i < a_{i+1}$ and $a_n = b$. Let $\alpha := (a_0,\dots,a_n)$ describe the particular subdivision that is chosen. By the completeness axiom if $f : [a,b] \to \mathbb{R}$ is bounded (but not necessarily continuous) we can define the numbers

$$\bar{f}_i := \sup_{x\in[a_i,a_{i+1}]} f(x); \qquad \underline{f}_i := \inf_{x\in[a_i,a_{i+1}]} f(x).$$

These numbers define boxes with base $[a_i,a_{i+1}]$ and height $\bar{f}_i$ or $\underline{f}_i$ that lie above or below the function f on the interval $[a_i,a_{i+1}]$. If we add up the areas of these boxes we get upper and lower estimates for the area under the function f. These estimates are given by the sums:

$$\bar{F}_\alpha := \sum_i \bar{f}_i(a_{i+1}-a_i), \qquad \underline{F}_\alpha := \sum_i \underline{f}_i(a_{i+1}-a_i).$$

Now we might want to think about very good ways of approximating the area under the function f. Thus we might want to think of partitions α that make this upper estimate as small as possible or partitions α that make this lower estimate as big as possible:

$$\inf_\alpha \bar{F}_\alpha, \qquad \sup_\alpha \underline{F}_\alpha.$$

EXERCISE 24: *Prove that this infimum and supremum both exist and that* $\inf_\alpha \bar{F}_\alpha, \geq \sup_\alpha \underline{F}_\alpha$.

These objects have names. $\inf_\alpha \bar{F}_\alpha$ is called the Upper Riemann Integral of f over $[a,b]$ and $\sup_\alpha \underline{F}_\alpha$ is called the Lower Riemann Integral of f over $[a,b]$.

Definition 19 (Riemann Integrable) *If f is a bounded real valued function on $[a,b]$ and $\inf_\alpha \bar{F}_\alpha = \sup_\alpha \underline{F}_\alpha$, then f is Riemann Integrable on $[a,b]$ and we denote*

$$\int_a^b f(x)\,dx := \inf_\alpha \bar{F}_\alpha, = \sup_\alpha \underline{F}_\alpha.$$

When does this definition of an integral work? We know it works for step functions with finite numbers of steps. We will also show it works for continuous functions on closed intervals and for absolutely continuous functions. Here is an example of a function where $\inf_\alpha \bar{F}_\alpha = 1$ and $0 = \sup_\alpha \underline{F}_\alpha$ so the Riemann integral is not defined. Define the function $g(x)$ so that $g(x) = 1$ when x is a rational number and $g(x) = 0$ when x is an irrational number. Then as there is both a rational and an irrational in every open interval we have that $\bar{F}_\alpha = 1$ for all α and $\underline{F}_\alpha = 0$ for all α. Clearly this function has too many discontinuities to be integrable. It is also quite famous and is sometimes called the Dirichlet Function. This function also provides a useful example of the difference between Lebesgue integrals and Riemann integrals. Although, g is not Riemann integrable it *is* Lebesgue integrable. We will see later that it has a Lebesgue integgral of zero.

EXERCISE 25: *Assume the function $f : [a, b] \to \mathbb{R}$ is uniformly continuous (for any $\epsilon > 0$ there exists δ such that if $|a_i - a_{i+1}| < \delta$ then $|f(a_i) - f(a_{i+1})| < \delta$). Use this to show that this implies $\bar{F}_\alpha - \underline{F}_\alpha < \epsilon$. Hence deduce that continuous functions on closed intervals are Riemann integrable.*

We now state a characterization result for Riemann integrable functions—the proof of this will be deferred until we develop Lebesgue measure later in the course. First we define a notion of a small set. That is a set that can be covered with a countable set of intervals that are each really small in length.

Definition 20 *$E \subset \mathbb{R}$ is said to be a set of measure zero, if for any $\varepsilon > 0$ there exists a sequence of open intervals (I_n) such that $E \subset \cup_n I_n$ and $\sum_n |I_n| < \infty$ (where $|I_n|$ is the length of I_n).*

Result 15 A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable iff the set of discontinuities is a set of measure zero.

An increasing function on $[a, b]$ has a countable number of discontinuities, thus all increasing functions on compact sets are Riemann integrable.[3]

An important practical question is when does the limit of an integral converge to the integral of the limit. Here's an example where it fails. Define the sequence of continuous functions (f_n) on $[0, 1]$ as follows: For $x \in [0, 1/n]$ let $f_n(x) = n^2 x$. For $x \in (1/n, 2/n]$ define $f_n(x) = 2n - n^2 x$. And elsewhere let $f_n(x) = 0$. Clearly

$$\int_0^1 f_n(x)\,dx = (1/2) \times \frac{2}{n} \times n = 1$$

But this sequence of functions converges pointwise to the constant function equal to zero and the integral of zero is zero. When does integration preserve limits then? You need uniform convergence not pointwise convergence.

[3] Let S_n be the set of discontinuities greater than length $1/n$. There are at most $n(b - a)$ elements in S_n if the function is monotonic. $\cup_{n=0}^{\infty} S_n$ counts all discontinuities. This is a countable union of finite sets.

Result 16 If (f_n) is a sequence of Riemann integrable functions on $[a,b]$ and f_n converges uniformly to f, then f is also Riemann integrable and

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

2.8. *Differentiation*

Suppose that f is a real valued function. If f is continuous we known that has $x' \to x$ so $f(x') \to f(x)$. Suppose that f were not continuous, then as x' approaches x there may not be a well defined limit. And how x' approaches x might also affect this limit. Hence we could define

$$\begin{aligned}
\limsup_{y\to x^+} f(y) &:= \inf_{\delta>0}\sup_{0<y-x<\delta} f(y) && \text{upper right limit}\\
\liminf_{y\to x^+} f(y) &:= \sup_{\delta>0}\inf_{0<y-x<\delta} f(y) && \text{lower right limit}\\
\limsup_{y\to x^-} f(y) &:= \inf_{\delta>0}\sup_{0<x-y<\delta} f(y) && \text{upper left limit}\\
\liminf_{y\to x^-} f(y) &:= \sup_{\delta>0}\inf_{0<x-y<\delta} f(y) && \text{lower left limit}
\end{aligned}$$

As an example of this consider the function $f(x) = \sin(1/x)$ for $y \neq 0$. This function oscillates between $+1$ and -1 infinitely many times as it approaches zero. Thus the upper left and right limits are $+1$ the lower left and right limits are -1.

We can also use these notions to define 4 different derivatives of a function

$$\begin{aligned}
D^+f(x) &:= \limsup_{h\to 0^+}\frac{f(x+h)-f(y)}{h}\\
D_+f(x) &:= \liminf_{h\to 0^+}\frac{f(x+h)-f(y)}{h}\\
D^-f(x) &:= \limsup_{h\to 0^-}\frac{f(x+h)-f(y)}{h}\\
D_-f(x) &:= \liminf_{h\to x^-}\frac{f(x+h)-f(y)}{h}
\end{aligned}$$

If $D^+f(x) = D_+f(x)$ and finite there is a right derivative of the function. If $D^-f(x) = D_-f(x)$ and finite there is a left derivative of the function. And if $D^+f(x) = D_+f(x) = D^-f(x) = D_-f(x)$ and all these limits are finite we say the function is differentiable at x.

There are many examples of continuous but nowhere differentiable functions - the path of a Brownian motion is one such.

There are two key results that are worth pointing out now but without a proof. First,

Result 17 (Lebesgue's Theorem) If the function $f : [a, b] \to 1$ is monotone, then it is differentiable almost everywhere.

Second the fundamental Theorem of calculus.

Result 18 For any continuous function $f : [a, b] \to \mathbb{R}$ and function $F : [a, b] \to \mathbb{R}$

$$F(x) = F(a) + \int_a^x f(t)\, dt \qquad \forall x \in [a, b],$$

if and only if F is continuously differentiable and $F' = f$.

2.9. Convex Functions

Now we will consider functions $f : D \to \mathbb{R}$, where $D \subset \mathbb{R}$ that are convex.

Definition 21 *The function $f : D \to \mathbb{R}$, on the interval $D \subset \mathbb{R}$ is convex if*

$$f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda) f(x')$$

for all $x, x' \in D$ and all $\lambda \in (0, 1)$.

The function f is concave if $-f$ is convex. An equivalent definition is:

Definition 22 *The function $f : D \to \mathbb{R}$, on the interval $D \subset \mathbb{R}$ is convex if*

$$f(x_2) \leq \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3)$$

for all $x_1 < x_2 < x_3 \in D$.

we have the following properties for convex functions.

Result 19 If the function f is convex on the interval $[a, b]$, then

- The function is continuous at every interior point of $[a, b]$.
- The left and right derivatives exist and are finite at every interior point.
- $\inf_{x \in [a,b]} f(x) > -\infty$ and $\sup_{x \in [a,b]} f(x) < \infty$.

Analysis for Economists: ECON0118

Martin Cripps

November 2, 2020

Contents

3. METRIC SPACES 1

3.1. Topology

Most of the ideas that we will develop for metric spaces are topological in character hence topologies naturally arise at this point. However, we have already introduced the fundamentals of topology at various points above: We defined the notion of a topology, a topological space, a basis, and how a basis might generate a topology in Section 1.6.4. Further we have considered how the open intervals are a basis for $\mathbb{R}$. To revise the notion of a topology complete the following exercise.

EXERCISE 26: *Suppose that $(\Omega, \mathcal{T})$ is a topological space. Show that if $\Omega' \subset \Omega$ then $(\Omega', \mathcal{T}')$ is a topological space, where $\mathcal{T}' = \{G' : G' = G \cap \Omega', G \in \mathcal{T}\}$.*

The topology $\mathcal{T}'$ is called the relative topology (on Ω'). We can also use topologies to define continuous functions. This extends Definition 13 to arbitrary topological spaces.

Definition 23 *Let $(\Omega, \mathcal{T})$ and $(\Lambda, \mathcal{S})$ be topological spaces. The function $f : \Omega \to \Lambda$ is continuous, if $f^{-1}(G) \in \mathcal{T}$ when $G \in \Lambda$.*

You should be able to use the properties of a basis to do the following exercise.

EXERCISE 27: *Suppose that we endow the real line with the topology generated basis of open intervals. Then show that Definition 12 and Definition 13 are equivalent. (Hint: Recall that any open set G can be written as a union $G = \cup_{x \in G} B_x$ where B_x is a neighbourhood of x in the basis.)*

The final issue in this section is to describe when two topological spaces are equivalent. This is the topological notion of equivalence and is why a topologist thinks a teacup and a donut are equivalent! Spaces are topologically equivalent if there is a continuous bijection between them. These continuous bijections have a special name, they are called homeomorphisms.

Definition 24 *Given two topological spaces $(\Omega, \mathcal{T})$ and $(\Lambda, \mathcal{S})$. If the function $f : \Omega \to \Lambda$ is continuous, the function $f^{-1} : \Lambda \to \Omega$ is continuous, and f is 1–1; then it is called a homeomorphism and the spaces $(\Omega, \mathcal{T})$ and $(\Lambda, \mathcal{S})$ are said to be homeomorphic.*

EXERCISE 28:

1. *Give an example of a bijection between $(0, 1)$ and $(0, 1)$ that is not a homeomorphism.*

2. *Prove that* $(0,1)$ *and* $\mathbb{R}$ *are homeomorphic.*

3. *Prove that any two open intervals in* $\mathbb{R}$ *are homeomorphic.*

4. *Suppose that* f *is a homeomorphism between the open intervals* (a,b) *and* (c,d). *Show that for any* $x \in (a,b)$ $f((a,x)) \cap f((x,b)) = \emptyset$. *Can there exist* $x < y < z$ *in* (a,b) *such that* $f(x) < f(y) > f(z)$? *Show that* f *is either strictly increasing or strictly decreasing.*

5. *Show that the intervals* $(0,1)$ *and* $(0,1]$ *are not homeomorphic. (Hint: If there was a homeomorphism, then there exists* $x \in (0,1)$ *such that* $f(x) = 1$.*)*

3.1.1. Product Topologies

Suppose you have a collection of topological spaces $(\Omega_\iota, \mathcal{T}_\iota)$ where $\iota \in I$ and I is an arbitrary set. You might want to use these topologies to create a new topology $\mathcal{T}$ on the product set $\Omega := \times_{\iota \in I}\Omega_\iota$. One important way of doing this is called the product topology. The open sets of the product topology $\mathcal{T}$ are formed from finite intersections and arbitrary unions of the product of the open sets

$$G = \times_{\iota \in I} G_\iota, \qquad G_\iota \in \mathcal{T}_\iota, \forall \iota$$

Where all of the sets in this product are open and only finitely many of the G_ι's are not equal to Ω_ι. These are often called "cylinder sets". This is the coarsest topology on Ω that makes the projections that takes the points in Ω and maps them to their ιth coordinate continuous. This is an example of a weak topology that we will define later.

3.2. Metrics and Norms

Now we will look at less abstract ways of defining nearness. A neighbourhood basis is a topological notion of nearness, that allows stretching and does not relate to our notions of distance at all. The first notion of distance we will give is a metric, which takes two points and says how close they are.

3.2.1. Metrics

Definition 25 *Let* Ω *be a set and* $\rho : \Omega \times \Omega \to \mathbb{R}$ *be a function. Then* ρ *is a metric and* (Ω, ρ) *is a metric space if*

1. $\rho(x,y) \geq 0$ *for all* $x, y \in \Omega$ *and* $\rho(x,y) = 0$ *if and only if* $x = y$.
2. $\rho(x,y) = \rho(y,x)$ *or all* $x, y \in \Omega$.
3. $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$ *or all* $x, y, z \in \Omega$.

The metric ρ can be quite abstract and not measure distances particularly well. Moreover, the relative distances between points can be quite strange. The only constraint is (3) which is called the triangle inequality for obvious reasons. Here are some potential metrics for you to investigate.

EXERCISE 29: *Which of the following functions are metrics?*

1. $\rho(x,y) = 1$ *if* $x \neq y$ $\rho(x,y) = 0$ *if* $x = y$?
2. $\rho((x_1,x_2),(y_1,y_2)) := \min\{|x_1 - y_1|, |x_2 - y_2|\}$?
3. $\rho((x_1,x_2),(y_1,y_2)) := \max\{|x_1 - y_1|, |x_2 - y_2|\}$?
4. $\rho(p,q) := p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q}$ *where* $p, q \in (0,1)$?
5. *If* $[a,b]$ *and* $[c,d]$ *are closed intervals of real numbers*
$$\rho([a,b],[c,d]) := \min_{x \in [a,b], y \in [c,d]} |x - y|.$$
6. *If* $[a,b]$ *and* $[c,d]$ *are closed intervals of real numbers*
$$\rho([a,b],[c,d]) := \max_{x \in [a,b], y \in [c,d]} |x - y|.$$
7. $$\rho(x,y) := \frac{|x-y|}{1+|x-y|}$$

3.2.2. Metrics and Topologies

Once we have a metric we can use it to define neighbourhoods of points in the usual way. Defining the open ball around the point x to be

$$B^{\rho}_{\epsilon}(x) := \{y \in \Omega : \rho(x,y) < \epsilon\}.$$

The collection of open balls is a neighbourhood basis for the arbitrary set Ω and, therefore, determines a topology on the space Ω. Thus the way we use the open intervals to define a topology on $\mathbb{R}$ extends to any set that is equipped with a metric. When we want to emphasize that the topology comes from a metric we might write $\mathcal{T}_\rho$.[4]

EXERCISE 30: *Consider the first metric in Exercise 3.2.1—what is the topology it induces? What functions are continuous with respect to this topology (if it is applied to both the image and domain).*

If two different metrics generate the same topology, then we say that the metrics are equivalent. To ensure that the metrics ρ and τ generate topologies with the same open sets it is enough to ensure that you can find a ρ-neighbourhood inside every τ neighbourhood and vice versa. That is, the metrics ρ and τ generate the same topologies on Ω if and only if for every x and ϵ there exists ϵ' and ϵ'' such that

$$B^\tau_{\epsilon'}(x) \subset B^\rho_\epsilon(x), \qquad B^\rho_{\epsilon''}(x) \subset B^\tau_\epsilon(x).$$

(To see this take an open set in $\mathcal{T}_\tau$ and then it can be written as a union of τ-open balls. We can use the above to show that it can also be written as a union of ρ-open balls. Hence that it is in $\mathcal{T}_\rho$.)

Now we observe that metrics do not need to be unbounded to be equivalent. Suppose that $\rho(x,y)$ is a metric on Ω. Then the metric

$$\tau(x,y) := \frac{\rho(x,y)}{1+\rho(x,y)}$$

is is equivalent to ρ. Thus every metric is equivalent to a bounded metric.

3.2.3. Norms

A lot of the way we use distance is to scale things up by a factor or to shrink things down. That means we need some scale for measuring things. The scale that is used to measure things is usually going to be $\mathbb{R}$ or perhaps the complex numbers $\mathbb{C}$. But in its full generality we only require the scale to satisfy the axioms of a field that we first introduced in Section 2.1.1. In the definition below there is little harm (in the economics applications) in thinking of F as being the same as $\mathbb{R}$.

We are now going to impose further structure on the spaces we are studying so that they are Vector spaces. What this does is allows us to be able to take an x in Ω say and multiply it by 2 or add it to other members of Ω. This also allows us to introduce the notion of a norm — which is more demanding than the notion of a metric.

[4]Notice that there is an inverse question—given a topology can we write down a metric that induces it? (There is an answer to this question that is a famous result due to Urysohn.)

Definition 26 (Linear/Vector Space) *This consists of: (1) a set Ω , (2) a field F, (3) a function $+$ that maps $\Omega \times \Omega$ into Ω, and (4) a function $\times$ that maps $F \times \Omega$ into Ω, such that for all $x, y, z \in \Omega$ and all $\alpha\beta \in F$:*

1. $x + y = y + x$,
2. $x + (y + z) = (y + x) + z$,
3. *There exists* $0 \in \Omega$ *such that* $x + 0 = x$,
4. *There exists* $-x \in \Omega$ *such that* $x + (-x) = 0$,
5. $\alpha(\beta x) = (\alpha\beta)x$,
6. $\alpha(x + y) = \alpha x + \alpha y$,
7. $(\alpha + \beta)x = \alpha x + \beta x$,
8. $1x = x$,

(For examples of this, let Ω be a vector space $\mathbb{R}^n$ or the space of continuous functions $f : [a, b] \to \mathbb{R}$ and let F be the real numbers.) Equipped with this we can now define a norm.

Definition 27 *Suppose that Ω is a Linear space with the scalar field $\mathbb{R}$ (or $\mathbb{C}$). A function $\| \cdot \| : \Omega \to \mathbb{R}$ is said to be a norm onΩ if for all $x, y \in \Omega$ and $\alpha \in \mathbb{R}$ (or $\alpha \in \mathbb{C}$:*

1. $\|x\| \geq 0$ *for all $x \in \Omega$ with equality iff $x = 0$.*
2. $\|\alpha x\| = |\alpha|\|x\|$ *for all $x \in \Omega$, $\alpha \in \mathbb{R}$.*
3. $\|x + y\| \leq \|x\| + \|y\|$ *for all $x, y \in \Omega$.*

The pair $(\Omega, \|.\|)$ is a normed space.

The notion of a norm is strictly stronger than a metric. To see this do the following exercise

EXERCISE 31: *If $\|.\|$ is a norm, show that $\rho(x, y) := \|x - y\|$ is a metric on Ω. Which of the metrics in Exercise 3.2.1 can be generated from norms?*

Just like metrics we can ask the question—when do two different norms generate the same topology? When are two norms equivalent? We have the following important condition. The norms $\|.\|_a$ and $\|.\|_b$ on Ω are equivalent if there are constants $A > 0$, $B > 0$ such that

$$A\|x\|_a \leq \|x\|_b \leq B\|x\|_a.$$

An extremely useful property is the following:

Result 20 Any two norms on a finite-dimensional vector space are equivalent.

To see a non-standard example of this consider $\mathcal{M}_n := \mathbb{R}^{n^2}$ the finite-dimensional vector space of all $n \times n$ matrices with real entries. Let $\{m_{ij}\} = M \in \mathcal{M}$ denote such a matrix. Here are some examples of norms on the set of $M \in \mathcal{M}$: the trace of $M^T M$; the square root of the absolute value of the largest eigenvalue of $M^T M$; the absolute value of the largest element in M. All of these objects satisfy the above relationship, so for example, there exists constants such that

$$B\text{tr}(M^T M) \leq \max_{ij} |m_{ij}| \leq C\text{tr}(M^T M)$$

Thus to find non-equivalent norms it is necessary to consider infinite dimensional spaces such as function spaces. Consider, for an example of this, $C([0,1])$, the space of continuous real-valued functions on $[0,1]$. And we define the norms on this space

$$\|f\|_1 := \int_0^1 |f(x)|\, dx, \qquad \|f\|_2 := \left(\int_0^1 |f(x)|^2\, dx\right)^{1/2},$$

(Strictly speaking we need these integrals to be Lebesgue integrals — which will be defined in the second half of this course.) And let $f_n(x) := x^n$, then

$$\|f_n\|_1 = \frac{1}{n+1}, \qquad \|f_n\|_2 = \frac{1}{\sqrt{2n+1}}.$$

These two norms are equivalent if there exists a $A > 0$ and $B > 0$ such that

$$\frac{A}{n+1} \leq \frac{1}{\sqrt{2n+1}} \leq \frac{B}{n+1}$$

for all n. This cannot be true—simply multiply through by $n+1$ and let $n \to \infty$.

3.3. Convergence, Closure and Completeness

3.3.1. Closed Sets

Recall that we defined a closed set to be the complement of an open set. Given the topological space $(\Omega, \mathcal{T})$, $F \subset \Omega$ is closed if $F^c \in \mathcal{T}$, but we also think of closed sets as being sets that contain the limits of convergent sequences. We now need to check that these two ideas are compatible with each other.

First let us define when a sequence (x_n) converges to x in the topological space $(\Omega, \mathcal{T})$. This just says if an open set contains the limit x then the sequence (x_n) is in this set always in the long run.

Definition 28 *The sequence (x_n) converges to $x \in \Omega$, if for every open set $G \in \mathcal{T}$ that contains x there exists N such that $x_n \in G$ for all $n > N$.*

We could define the closure of a set E to be the set of limits of sequences in E, but instead we use the idea that if x is a limit of a sequence of points in E then the open sets that contain x must also intersect with E.

Definition 29 *Let $E \subset \Omega$ in the topological space $(\Omega, \mathcal{T})$. The point x is a limit point of E if when $x \in G \in \mathcal{T}$, then $G \cap E \neq \emptyset$. The closure of E (denoted $\bar{E}$) consists of all limit points of E.*

EXERCISE 32: *Use the definitions to show that*

1. *If E^c is open then $E = \bar{E}$.*
2. *$\bar{E}$ is the intersection of all closed sets that contain E.*
3. *Suppose $x \in \bar{E}$ in a topological metric space. Let $B_n(x) := \{z : \rho(x, z) < 1/n\}$. Find a sequence of points in $E \cap B_n(x)$ converging to x.*
4. *Define $\mathbb{Q} \cap [0, 1]$ to be the set of rationals in $[0, 1]$. Show that $\overline{\mathbb{Q} \cap [0, 1]} = [0, 1]$.*

This last exercise is an example of a very important property the notion of a dense set. Intuitively, a set is dense in a space if it is near to, or good at approximating, every point in the set. Thus many proofs try and show that something holds for a dense subset of a space and then to argue that this extends to the whole space. A dense set does not need to be especially big in the sense of measure though.

Definition 30 *The set E is dense in the topological space $(\Omega, \mathcal{T})$ iff $\bar{E} = \Omega$.*

As you've already shown, an example of a dense set is the rationals in the real line. On the other hand, the rationals are only countable whereas there are uncountable reals, so in one sense the set of rationals is quite small when compared to the reals.

3.3.2. Completeness and Contraction Mapping

For sequences in closed sets if a sequence converges to a limit, then the limit is in the set. But you don't really worry about whether a limit exists. In complete spaces there is a class of sequences which are guaranteed to have limits and the limit is guaranteed to be in the space. We will see one very important result for economists — the contraction mapping theorem—is a simple consequence of this completeness property. We begin by reminding ourselves of what a Cauchy sequence is. . .

Definition 31 *The sequence (x_n) in the metric space (Ω, ρ) is Cauchy if for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $\rho(x_n, x_m) < \epsilon$ for all $n, m > N$.*

In $\mathbb{R}$ we know that Cauchy sequences converge, so now we will define complete metric spaces to be those metric spaces that also have this property.

Definition 32 *The metric space (Ω, ρ) is complete if every Cauchy sequence converges to an element of Ω.*

(Notice you are getting two things here — the existence of a limit and the fact that it is in the space.) This is an exceptionally useful property. If a space is complete then you know you can take a certain kind of sequence and know that its limit is still in the space. It is a simple matter to find examples of non-complete metric spaces—just delete the limit of a Cauchy sequence—for example the sequence $(1/n)$ in the space $(0, 1)$. So, you need to be careful to check completeness. Recall that if a sequence is convergent then it is also Cauchy, thus if the set E is complete, then it is also closed.

Here are some examples of complete spaces:

- $\mathbb{R}$ with the usual metric.
- Any finite product of complete spaces (Ω_n, ρ_n) where the metric on the product space is $\rho(x, y) := \sum_n \rho_n(x_n, y_n)$.
- For $p \geq 1$, the space of all p summable series:

$$\ell^p := \{(x_n) \in \mathbb{R}^\infty : \sum_{n=1}^{\infty} |x_n|^p < \infty\}.$$

  where the metric is $d_p((x_n), (y_n)) := (\sum_{n=1}^{\infty} |x_n - y_n|^p)^{1/p}$.
- The set of all bounded functions on a set T: $\mathcal{B}(T) := \{f : T \to \mathbb{R} : \sup_{x \in T} f(x) < \infty\}$ with the metric $d_\infty(f, g) := \sup_{x \in T} |f(x) - g(x)|$.
- As a consequence of the previous bullet point $\mathcal{C}[a, b]$ the set of continuous real valued functions on the interval $[a, b]$ is complete (continuous functions on closed intervals are bounded).

There are other examples of complete spaces of functions, but we need to be able to define integrals properly before we consider these.

EXERCISE 33: *Show that the space $\ell^\infty := \{(x_n) \in \mathbb{R}^\infty : \sup_n |x_n| < \infty\}$ equipped with metric $d_\infty((x_n), (y_n)) := \sup_n |x_n - y_n|$ is complete.*

It is completeness that allows us to prove the contraction mapping theorem.

Result 21 (Contraction Mapping) Let (Ω, ρ) be a complete metric space, let $\lambda < 1$, and let $f : \Omega \to \Omega$ be a function satisfying:

$$\rho(f(x), f(y)) \leq \lambda \rho(x, y), \qquad \forall x, y \in \Omega.$$

Then there exists a unique $x^* \in \Omega$ s.t. $f(x^*) = x^*$.

This pretty easy to prove now you've understood what completeness does. Just pick a point $x_0 \in \Omega$ and do f to it again and again. That is define the sequence $(x_n) \subset \Omega$ such that $x_{n+1} = f(x_n)$. The contraction property implies that this sequence is a Cauchy sequence (all the steps after x_n add up to at most $\rho(x_n, x_{n-1})/(1-\lambda)$). Completeness implies that it converges to a limit that is also in Ω. The contraction property implies f is continuous so this limit is a fixed point. Uniqueness follows by assuming two fixed points exist and then deriving a contradiction of the contraction property.

3.3.3. Interior and Baire Category

The final important concept in this section is the notion of the interior of a set. We have shown that the closure of a set can be thought of as the "smallest" closed set that contains it. The interior on the other hand is the "largest" open set that is inside it.

Definition 33 *If E is a subset of a topological space $(\Omega, \mathcal{T})$, then x is an interior point of E if there exists G an open subset of E such that $x \in G \subset E$. The set of all interior points of E is denoted E^o and is called the interior of E.*

EXERCISE 34:

1. *Show that E^o is an open set.*
2. *Show that if G is open then $G = G^o$.*

Some care is needed here because often things that look like the interior are not. For example, consider the interval $[0,1)$ as a subset of $\mathbb{R}$. Its closure is the interval $[0,1]$ and its interior is the interval $(0,1)$. However, if we consider the interval $\{0\} \times [0,1)$ as a subset of $\mathbb{R}^2$ then its closure is $\{0\} \times [0,1]$ and its interior is empty. (The set $\{0\} \times (0,1)$ can be defined as the "relative interior" of $\{0\} \times [0,1)$.) Thus because the set $\{0\} \times [0,1]$ is kind of "thin" in $\mathbb{R}^2$ it has an empty interior.

Now we introduce the notion of a nowhere dense set which formally defines the notion of thinness.

Definition 34 *A set $E \subset \Omega$ is nowhere dense if the interior of its closure is empty, that is, if $(\bar{E})^o = \emptyset$.*

Clearly any finite set of points is in $\mathbb{R}$ closed (because you can fit an interval between them) and has an empty interior. Thus this is a nowhere dense set. Now we state an important property of complete metric spaces — they cannot be made up of countable collections of nowhere dense sets.

Result 22 (Baire Category Theorem) The complete metric space (Ω, ρ) is not the union of a countable collection of nowhere dense sets.

Why is this true? Pick an open set then there has to be a point in it and an open ball that is not in the first of these nowhere dense sets. Now pick a second point and an open ball in this open ball such that each of them is outside the second nowhere dense set. Repeat this ad infinitum. Completeness implies this sequence converges to a limit and that this limit is outside all the countable collection of nowhere dense sets.

An important consequence of this is:

Result 23 (Uniform Boundedness Principle) Suppose that $\mathcal{F}$ is a family of continuous functions from the complete metric space (Ω, ρ) to the normed space $(\Lambda, \|.\|)$. If for every $x \in \Omega$ the functions in $\mathcal{F}$ evaluated at x are bounded: $\sup_{f \in \mathcal{F}} f(x) < \infty$. Then, there exists M and an open set $G \subset \Omega$ such that $\|f(z)\| \leq M$ for all $z \in G$ and all $f \in \mathcal{F}$.

3.4. Connectedness and Separability

3.4.1. Connectedness

This topic appears to particularly abstract, so we will largely report the ideas and move on. It does have one very useful consequence — the intermediate value theorem that is left as an exercise.

Two sets A, B in a topological space $(\Omega, \mathcal{T})$ are <u>separated</u> if we can put each of the sets inside disjoint open sets. (There exists $U, V \in \mathcal{T}$ such that $U \cap V = \emptyset$ and $A \subset U$, $B \subset V$.) This leads to two special concepts of a space:

Hausdorff Space A topological space where if $x \neq y$ then $\{x\}$ and $\{y\}$ are separated.

Normal Space A topological space where disjoint closed sets are separated.

In normal spaces we know that continuous real-valued functions exist and a continuous real-valued function on a closed subset can be extended to a continuous function on the whole space.

Connectedness is the idea that one can travel in a continuous manner from one part of a set/space to any other part without leaving the set/space. Connectedness can be defined in several different ways. Here are three different and equivalent ways of defining disconnected. The space $(\Omega, \mathcal{T})$ is disconnected if:

- The space can be decomposed into two disjoint open sets: $\Omega = U \cup V$ where $U \cap V = \emptyset$ and $U, V \in \mathcal{T}$.
- There is a proper subset of the space that is open and closed.
- The space can be decomposed into two sets A and B such that $\bar{A} \cap B = A \cap \bar{B} = \emptyset$.

A set is <u>connected</u> if it is not disconnected.

Every connected set in $\mathbb{R}$ is an interval. Suppose not and W is a connected subset of $\mathbb{R}$ and $a, b \in W$ but there exists $a < c < b$ such that $c \notin W$. Then the intervals $(-\infty, c)$ and (c, ∞) contain W and they also satisfy the third bullet point condition above. So W is not connected—a contradiction.

One of the most useful properties of connectedness is that it is preserved by continuous functions, so

Result 24 Let $(\Omega, \mathcal{T})$ be a connected topological space and $f : \Omega \to \Lambda$ be a continuous function, then $f(\Omega)$ is a connected subset of Λ.

A corollary of this is the intermediate value theorem:

Result 25 (Intermediate Value Theorem) Let $f : [a, b] \to \mathbb{R}$ be a continuous function, then for any $y \in [f(a), f(b)]$ there exists an $x \in [a, b]$ such that $f(x) = y$.

EXERCISE 35: *Prove the intermediate value theorem using the previous result.*

3.4.2. Separability

This is quite a strange name for a very natural property. A very useful feature of the real line is that it can be very well approximated by the countable set of the rationals. (The rationals are dense in the reals.) This means we can check things for a countable set and then maybe if we're lucky have them hold on an uncountable set. Separability is the name we give to abstract topological spaces that have this property.

Definition 35 *The topological space $(\Omega, \mathcal{T})$ is said to be <u>separable</u> if there exists E, a countable subset of Ω such that $\bar{E} = \Omega$.*

Its quite easy to see that lots of spaces like $\mathbb{R}$ are separable, but many function spaces are not. Consider the set of all bounded functions on $[0, 1]$ with the supremum norm.

And consider the open sets of functions

$$B_t := \left\{ f : [0,1] \to \mathbb{R} : \begin{cases} \frac{3}{2} > f(x) > \frac{1}{2} & x \leq t, \\ |f(x)| < \frac{1}{2} & x > t, \end{cases} \right\}.$$

If $t \neq s$ then $B_t \cap B_s = \emptyset$ (prove this), so this gives an uncountable collection of open sets that are all disjoint. If a set is dense it must intersect with each of these open sets and hence must have an uncountable number of elements.

A second property that we can use the rationals to do is to generate a basis for the topology on the real line that is countable. Simply choose all intervals with rational endpoints and gives a countable collection of sets that generates the usual topology on $\mathbb{R}$. This is such an important property it is given a name:

Definition 36 *A topological space is second countable if it has a countable basis.*

A very useful property of metric spaces is that if they are separable, then they are automatically second countable. This follows from taking the open balls centred at the dense set with diameter $1/n$ for $n = 1, 2, \ldots$.

Result 26 If a metric space is separable then it is second countable.

3.5. Compactness and Metric Spaces

Compactness (like continuity or convergence) can be defined in lots of different ways. You are probably familiar with the closed and bounded definition of compactness for real intervals. But here we want to talk about metric spaces being compact, so a different approach is required.

First we introduce the notion of an open covering. Suppose we have a family of open sets $\mathcal{O} \subset \mathcal{T}$, such that every point in the set $E \subset \Omega$ is in at least one set in the family $\mathcal{O}$, that is,

$$E \subset \bigcup_{O \in \mathcal{O}} O.$$

Then we say that *"$\mathcal{O}$ is an open cover of E"*. A subcover is a sub collection of the sets in $\mathcal{O}$ that also covers E. It seems strange but this is how we're now going to define compactness.

Definition 37 *The set $E \subset \Omega$ is compact if every open cover of E can be reduced to a finite subcover. If Ω is compact, then (Ω, ρ) is called a compact metric space.*

EXERCISE 36: *Suppose (x_n) is a sequence in a metric space that converges to x. Show that the set $\{x_n : n = 1, 2, \ldots\} \cap \{x\}$ is compact.*

This is a definition that is almost impossible to check. So, we're going to need some other tests that are equivalent to this to enable us to recognize compactness. However, it is often very useful to know that if there is an open cover of a compact set, then we can throw away all but a finite number of these sets and still cover the original set. Finally, notice that this definition is given for metric spaces but the same definition will also work in topological spaces that do not have metrics. Thus if we talk about compact topological spaces below, this is the definition we have in mind.

Towards the end of finding a checkable condition for compactness, let us introduce the notion of a totally bounded set. Suppose you are given a radius size ε and you are asked to cover your target set with balls of radius ε, then if the set if totally bounded it is possible to do this with only a finite number of balls.

Definition 38 *$E \subset \Omega$ is totally bounded if for any ε there exists a finite set of points $x_1, \dots, x_n \in E$ such that*

$$E \subset \bigcup_{m=1}^{n} B_\varepsilon(x_m), \qquad \text{where } B_\varepsilon(x_m) := \{x : \rho(x, x_m) < \varepsilon\}.$$

Obviously the definition of compactness above implies that a set is totally bounded, because the $B_\varepsilon(x_m)$ are an open cover. But actually this is a weaker condition because there may be other kinds of open cover (not open balls) that cannot be reduced to a finite subcover. However, we now have the following very important result.

Result 27 The following are equivalent:

1. $E \subset \Omega$ is compact.
2. If (F_n) is a sequence of closed subsets of Ω and $E \cap (\cap_{i=1}^{n} F_i) \neq \emptyset$ for all n then $E \cap (\cap_{i=1}^{\infty} F_n) \neq \emptyset$.
3. Every sequence $(x_n) \subset E$ has a subsequence converging to a point in E. (Often called the Bolzano-Weierstrass result/property.)
4. E is complete and totally bounded.

Why are these equivalent?

(1) $\Rightarrow$ (2) if (F_n) is a sequence of closed sets and the limit intersection was empty then (F_n^c) is an open cover and then we cannot reduce it to a finite open cover. This contradicts (1).

(2) $\Rightarrow$ (3) Let (x_n) be a sequence in E and define the closed sets in (2) to be $F_n := \overline{\{x_{n+1}, x_{n+2}, \dots\}}$. These satisfy (2)'s conditions (check). Hence there is an $x \in E \cap (\cap_{i=1}^{\infty} F_n)$. This is the limit of a subsequence: Clearly $x \in F_1$ so there must exist an $x_k \in F_1$ which within $1/2$ of x. As $x \in F_k$ there must exist $x_{k'} \in F_k$ that is within $1/4$ of x — repeat this process forever.

(3) $\Rightarrow$ (4) First we find a finite subcover. Suppose you could not find a finite subcover then you can find an infinite sequence of balls with centres that are all ϵ apart — contradicting the fact that the centres of the balls have to have a convergent subsequence by (3). Completeness is easy as we know (from(3)) that a Cauchy sequence has to have convergent subsequence—this is automatically the limit of the whole sequence.

(4) $\Rightarrow$(3) Take a sequence of points in E and cover E with a finite open cover of balls radius $1/2$. The sequence must have a subsequence in one of these balls and all the points in this subsequence are at most 1 apart. Now find an open cover of this ball radius $1/4$ - the points in this sub-subsequence are $1/2$ apart. Iterate and find a Cauchy sequence that must converge by completion.

(4)&(3) $\Rightarrow$ (1) First we will show that (E, ρ) is separable (has a countable dense set). For each n there is a finite collection of open balls radius $1/n$ that cover E. Take the union over n of the centres of these balls. This is a countable set and is dense in E by construction. As E is a metric space we also know that E has a countable basis. Every open cover of E can be reduced to a countable subcover, by choosing for each x in the dense set one neighbourhood in on of the open sets in the cover that contains it (the Lindelöf property). If this countable cover cannot be reduced to a finite subcover, This countable subcover then we can find a sequence the violates (3).

EXERCISE 37: *Let $\ell^2(\mathbb{N})$ be the space of all sequences of real numbers (x_n) satisfying $\sum_{n=1}^{\infty} x_n^2 < \infty$ endowed with the norm $\|(x_n)\| = \sqrt{\sum_{n=1}^{\infty} x_n^2}$. If $(y_n) \in \ell^2(\mathbb{N})$ show that the set of sequences (z_n) satisfying $|z_n| \leq |y_n|$ for all n is a compact subset of $\ell^2(\mathbb{N})$.*

EXERCISE 38: *Prove a corollary to the above — a subset of $\mathbb{R}$ is compact iff it is closed and bounded. (This is called the Heine-Borel Theorem.)*

Some of the most important consequences of compactness comes when we combine it with continuity — we will delay consideration of this until the next section.

Analysis for Economists: ECON0118

Martin Cripps

November 2, 2020

Contents

4. METRIC SPACES 2

In this section we explore the metric space ideas of the previous section and see how a lot of these notions generate mathematical properties you have been using in your economics.

4.1. Continuity and Compactness

Recall that in Definition 23 we defined a function between two topological or metric spaces to be continuous if the inverse image of an open set was open. Our first set of results combine this definition with the notion of compact spaces (both metric and topological) that was described at the end of the previous section.

Result 28 Let $(\Omega, \mathcal{T})$ be a compact topological space, let $(\Lambda, \mathcal{S})$ also be a topological space, and let $f : \Omega \to \Lambda$ be continuous, then $f(\Omega) \subset \Lambda$ is compact.

Why? Well if $\mathcal{O}$ is an open cover of $f(\Omega)$ and $O \in \mathcal{O}$, then (by continuity) we know that $f^{-1}(O)$ is open and the sets $\{f^{-1}(O) : O \in \mathcal{O}\}$ are an open cover of Ω. By compactness of Ω this is reducible to a finite open cover. Then map this finite open cover of Ω to Λ using f and show that this is a finite subcover of $\mathcal{O}$.

Now we can prove (as an exercise) a result we use time and again in economics, that a continuous real-valued function on a compact set has a maximum and a minimum.

WEIERSTRASS PRINCIPLE 39: *Use this result to show that if f is a continuous real-valued function on the compact topological space Ω then there exists $x_1 \in \Omega$ and $x_2 \in \Omega$ such that*

$$f(x_2) \leq f(x) \leq f(x_1), \qquad \forall x \in \Omega.$$

Hence any continuous real-valued function on a compact set is bounded and attains it sup and inf.

A re-writing of the argument in Result 10 in this case also gives the result.

Result 29 Let (Ω, ρ) be a compact metric space, let Λ also be a metric space, and let $f : \Omega \to \Lambda$ be continuous, then f is uniformly continuous.

(Definition 15 is where you will find uniform continuity described.) There are weaker conditions for a function to attain its maximum. In particular, upper semicontinuous functions (Definition 14) also attain their maxima.

Result 30 Let (Ω, ρ) be a compact metric space, let f be a real-valued upper semicontinuous function on Ω then there exists $\bar{x} \in \Omega$ s.t. $f(\bar{x}) \geq f(x)$ for all $x \in \Omega$.

Why? If f is usc then the sets $f^{-1}((-\infty, n))$ are an open cover of Ω. There must be a finite open cover, by compactness, so there exists N such that $\Omega \subset f^{-1}((-\infty, N))$; [$\sup_{x\in\Omega} f(x) < N$ and is finite]. Now consider as sequence (x_n) such that $f(x_n) \to \sup_{x\in\Omega} f(x)$. This sequence has a convergent subsequence (by compactness again) so wlog suppose that $x_n \to \bar{x}$. The definition of usc functions allows them to jump up but not down, so for any convergent sequence (x_n) with $x_n \to \bar{x}$ we have $f(\bar{x}) \geq \limsup_{n\to\infty} f(x_n)$. Hence we must have

$$\sup_{x\in\Omega} f(x) \geq f(\bar{x}) \geq \limsup_{n\to\infty} f(x_n) = \sup_{x\in\Omega} f(x)$$

Hence f attains its supremum at $\bar{x}$.

EXERCISE 40: *Give an example of an usc function on $[a, b]$ that does not attain its infimum.*

The final part of this section is an often ignored, by economists, but fundamental result on functional equations.

Result 31 (Cauchy) Suppose we know that the function f satisfies

$$f(a + b) = f(a) + f(b) \tag{1}$$

for all $a, b \in \mathbb{R}$. Then,

1. There exists an α such that $f(q) = \alpha q$ for all $q \in \mathbb{Q}$.
2. If $f(.)$ is continuous then $f(x) = \alpha x$ for all $x \in \mathbb{R}$.
3. If $f(.)$ is bounded on an open interval then $f(x) = \alpha x$ for all $x \in \mathbb{R}$.

The first part of the proof of this is relegated to an exercise.

EXERCISE 41: *Show that if a function satisfies (1) then:*

1. *$f(0) = 0$.*
2. *Define $\alpha := f(1)$ and show that $f(p) = \alpha p$ for any integer p.*
3. *Hence show that $f(p/q) = \alpha(p/q)$ for any integers p, q.*

As $f(x) = \alpha x$ for all rationals and the rationals are dense in $\mathbb{R}$ choose a sequence of rationals (x_n) converging to some $x \in \mathbb{R}$ and observe that αx_n also converges and so must $f(x_n)$ by continuity so $\lim f(x_n) = \lim \alpha x_n = \alpha x$. (This result is also true if the function is semi-continuous, but you need to argue that if a jump in f happens just at the limit the $f(0) \neq 0$.) If the function is bounded on an open interval, (a, b) then we can find a K such that $|f(x) - \alpha x| < K$ for all $x \in (a, b)$. Consider the gap $|f(y) - \alpha y|$ for some arbitrary $y \in \mathbb{R}$. There is a rational number r' such that $y + r' \in (a, b)$. But then (as $f(r') = \alpha r'$)

$$|f(y) - \alpha y| = |f(y) + f(r') - \alpha(r' + y)| = |f(y + r') - \alpha(r' + y)| < K$$

But the functional equation implies $|f(ny) - \alpha n y| = n|f(y) - \alpha y|$ for all natural numbers n and so $n|f(y) - \alpha y| < K$ for all n. This implies $f(y) = \alpha y$.

EXERCISE 42: *Suppose the continuous function $g : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies the functional equation*

$$g(x)g(y) = g(x + y)$$

what is the form of g?

EXERCISE 43: *Suppose the continuous function $h : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies the functional equation*

$$h(x)h(y) = h(xy)$$

what is the form of h?

4.2. Function Spaces (Unbounded)

In Section 2.5 we defined pointwise and uniform convergence and the space $\mathcal{C}(D)$ of continuous real-valued functions on the set D where D is not necessarily compact. These definitions can obviously be extended to functions from one metric space Ω to another Λ. We will write this space as $\mathcal{C}(\Omega, \Lambda)$, but when we are considering the continuous functions that map Ω to the reals we will write this space as $\mathcal{C}(\Omega)$.

The result in that section (that uniform convergence preserves continuity) seems to indicate that we might be able to talk about the topological properties of spaces of functions. The problem is that uniform convergence is much too strong for the sequences of functions we might be interested in studying. Consider for example the sequence of continuous functions on the open interval $(0, 1)$

$$f_n(x) := 1 + x + x^2 + \cdots + x^n$$

This converges pointwise to the function $1/(1 - x)$ but not uniformly. Thus we introduce a weaker notion of uniform convergence.

Definition 39 *The sequence (f_n) in $\mathcal{C}(\Omega,\Lambda)$ converges uniformly on compact subsets to f if for each compact $K \subset \Omega$ and each $\varepsilon > 0$ there exists N s.t. for all $n > N$*

$$\rho_K(f_n, f) := \sup_{x\in K} \rho(f_n(x), f(x)) < \varepsilon. \tag{2}$$

Is it possible to construct a topology on $\mathcal{C}(\Omega,\Lambda)$ such that sequences converge in the topology if and only if they satisfy Definition 39? Thus we need to define open sets of functions.

For an arbitrary set $E \subset \Omega$ the function $\rho_E(f_n, f)$ defined in (2) is not quite a metric for two reasons: First because it may not be finite (if E could be open). Second because it can be zero even though $f \neq g$ (this occurs where f and g are the same on the set E but are different elsewhere in Ω). However, if we consider all compact subsets then this can be used to construct the topology we desired.

First we need to introduce the notion of a weak topology. Let X be a set and let f be a function that maps X to a topological space $(Y,\mathcal{Y})$. We could ensure the function f is continuous by taking all the inverse images of the open sets $G \in \mathcal{Y}$, that is, $f^{-1}(G) \subset X$ and defining these sets to be open in X. The weak topology on X is defined to be coarsest topology that is generated by the sets $\{f^{-1}(G) : G \in \mathcal{Y}\}$. Thus we have used the topology on the image space and the function to define a topology on X. This doesn't need to just use one function. Suppose we had a whole family of functions $f \in \mathcal{F}$ then we could define the open sets to make all of the functions in $\mathcal{F}$ continuous. This is called the weak topology determined by the family $\mathcal{F}$.

Now we go back to our space $\mathcal{C}(\Omega,\Lambda)$ (this is going to be the set X in our weak topology) and we consider functions that take $\mathcal{C}(\Omega,\Lambda)$ and maps it to the real line. This is going to be the function that takes an $f \in \mathcal{C}(\Omega,\Lambda)$ and maps it to its distance from $g \in \mathcal{C}(\Omega,\Lambda)$ on the compact set K. That is $f \mapsto \rho_K(f,g)$. As the compact sets K vary and the target function $g \in \mathcal{C}(\Omega,\Lambda)$ varies we have a very large family of functions. The Topology of Uniform Convergence on Compacts is defined to be the topology on $\mathcal{C}(\Omega,\Lambda)$ that makes the functions $f \mapsto \rho_K(f,g)$ continuous for all K and g. We will write this topology as $\mathcal{T}(\Omega,\Lambda)$. This is the topology that answers the question posed just below Definition 39.

The result below describes some really useful properties of the topology $\mathcal{T}(\Omega,\Lambda)$. First recall that in Section 3.4.1 a space was defined to be Hausdorff if any two distinct points in the set could be put inside disjoint open sets. We also define a space to be locally compact if for each x there is an open set G containing it and $\underline{G}$ is compact.

Result 32 Suppose that Ω is a locally compact, Hausdorff space and that (Λ,ρ) is a metric space, then

1. If Ω is second countable and Λ is complete, then $\mathcal{C}(\Omega,\Lambda)$ is complete with the topology $\mathcal{T}(\Omega,\Lambda)$.
2. If Ω is compact then the topology $\mathcal{T}(\Omega,\Lambda)$ is induced by the metric $\rho_\Omega(f,g)$.

Thus (by (1)) if $\Omega = \Lambda = \mathbb{R}^n$ (both complete and second countable) then any Cauchy sequence of continuous functions is guaranteed to converge to another continuous function. And (by (2)) if $\Omega = \Lambda = [a,b]^n$ then the topology on these functions has a metric that is given by the sup norm.

4.3. Function Spaces (Bounded)

In the previous section we noted that looking at compact subsets of the domain of functions is a way of ensuring that the sup norm is well defined for continuous functions on open sets. Another way to make sure the sup-norm is well defined when the domain of the functions are open sets is to look at the space of *continuous and bounded* functions. We will define $\mathcal{CB}(\Omega)$ to be the space of continuous and bounded real-valued functions on Ω this will be equipped with the norm $\sup_{x\in\Omega}|f(x)|$ for any $f \in \mathcal{CB}(\Omega)$. This space is complete by (1) in Result 32. But we will state the result again and prove it

Result 33 $\mathcal{CB}(\Omega)$ with the supremum metric is a complete metric space.

To show completeness we need to show that Cauchy sequences converge to limits in $\mathcal{CB}(\Omega)$. Let (f_n) be a Cauchy sequence in $\mathcal{CB}(\Omega)$, then these functions evaluated at a given point define a Cauchy sequence $(f_n(x))$ in a bounded interval of $\mathbb{R}$. The intervals in $\mathbb{R}$ are complete so we define a potential limit function as the pointwise limit $f^o(x) := \lim_{n\to\infty} f_n(x)$. Now we need to check that the sequence of functions does indeed converge to this limiting function: $\|f_n - f^o\| := \sup_x |f_n(x) - f^o(x)| \to 0$. The sequence (f_n) is Cauchy, so for any $\varepsilon > 0$ there exists and N s.t. for all $m > n > N$ $\|f_n - f_m\| < \varepsilon$. Hence for any x we have

$$|f^o(x) - f_n(x)| = \lim_{m\to\infty} |f_m(x) - f_n(x)| \le \lim_{m\to\infty} \|f_m - f_n\| < \varepsilon$$

As this upper bound holds for all x, we have that for all $n > N$ $\|f^o - f_n\| < \varepsilon$ and the convergence for the sequence to the limit. The boundedness of the limit is immediate.

We also need to show that the limit is continuous. We know we can find N such that $\|f^o - f_n\| < \varepsilon/3$ for all $n > N$. We know f_n is continuous, so we can find δ such that $|f_n(x) - f_n(x')| < \varepsilon/3$ if $\rho(x,x') < \delta$. But this implies

$$|f^o(x) - f^o(x')| \le |f^o(x) - f^n(x)| + |f^n(x) - f^n(x')| + |f^n(x') - f^o(x')| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

when $\rho(x, x') < \delta$. Thus f^o is continuous at an arbitrary x' and we are done.

EXERCISE 44: *Repeat this proof for functions that map into $\mathbb{R}^2$ with the norm $\|(x, y)\| = x^2 + y^2$.*

EXERCISE 45:

Consider the space of continuous functions $\mathcal{CB}([0, 1])$ and equip this space with the metric

$$\rho(f, g) = \int_0^1 |f(x) - g(x)|\, dx.$$

1. *Verify that this is a metric space. (check that if f and g are continuous then $\rho(f, g) = 0$ implies that $f = g$.)*
2. *Is (f_n) (where $f_n(x) := x^n$) a Cauchy sequence in this space? Does this sequence have a convergent subsequence?*
3. *Is this space complete?*
4. *Which step in the above proof fails?*

If we require convergence to be uniform (because of the sup norm), then we do get completeness. But, the above exercise shows that integral metrics will not give us completeness. Recall how we defined uniform convergence in Definition 18.

Uniform convergence is often difficult to check, so it would be nice to have an easier condition that guarantees uniform convergence. Clearly pointwise convergence is not enough, but if there is some sort of monotonicity this result helps.

Result 34 (Monotone Convergence) Let Ω be a compact metric space. Suppose that $f \in \mathcal{C}(\Omega)$ and (f_n) is a sequence in $\mathcal{C}(\Omega)$ that converges pointwise to f^o and such that $f_n(x) \geq f_{n+1}(x)$ for all n and all $x \in \Omega$. Then f_n converges to f^o uniformly.

We do this proof for the case $f^o = 0$, all other cases are the same just require a lot of subtraction. This proof is going to demonstrate the power of the open cover properties of compact spaces. Let $\varepsilon > 0$ be given. We first show that for any given $x \in \Omega$ there is a $\delta_x > 0$ and N_x such that

$$f_n(x') < \varepsilon \qquad \text{for all } x' \text{ satisfying } \rho(x, x') < \delta_x, \text{ and all } n > N_x.$$

As $f_n(x)$ decreases to zero there exists an N_x such that $f_n(x) < \varepsilon/2$ for all $n \geq N_x$. As $f_{N_x}(.)$ is continuous at x there exists δ_x such that $|f_{N_x}(x') - f_{N_x}(x)| < \varepsilon/2$ for all $\rho(x, x') < \delta_x$. This, combined with $f_n(x) < \varepsilon/2$, implies $f_{N_x}(x') < \varepsilon$ for all $\rho(x, x') < \delta_x$. As $f_n(x')$ decreases in n the displayed claim follows.

Now we get to use the open cover thing. The collection of open balls $\rho(x, x') < \delta_x$ (one for each initial point $x \in \Omega$ and each with a different δ_x) is an open cover of Ω. Compactness implies it is reducible to a finite open cover. So there are only a finite number of (x, δ_x, N_x) -tuples required for the balls to $\rho(x, x') < \delta_x$ to cover the whole of Ω. On these balls only we can ensure that $f_n(x') < \varepsilon$ for all $\varepsilon \in \Omega$, by choosing N to be the maximum of the finite number of the N_x's.

4.3.1. Compactness of $\mathcal{C}(\Omega)$ and Arzelà-Ascoli

Now we will move on to thinking about the compactness of function spaces. Completeness is enough for convergent sequences to have limits but sometimes we would like more — the existence of a convergent subsequence for an arbitrary sequence of functions. That is, to know some part of a sequence is Cauchy and converges, without having to prove it. From above, we do know that monotone sequences converge and this is sometimes useful but another tool would also help. In this section we are going to provide necessary and sufficient conditions for a set of functions to be compact.

As this is going to enable us to be able to find convergent subsequences, we are going to need to rule out function spaces that contain sequences like the one in the previous Exercise. Thus we are going to need to restrict the ability of sequences of continuous functions to become increasingly variable in small intervals.

This leads us to the following definition. In this definition Ω is taken to only be a topological space. The equicontinuity can be defined for metric spaces using the topology induced by the metric. Thus the treatment here is a little more general than in OK/

Definition 40 *The set $\mathcal{F} \subset \mathcal{C}(\Omega, \Lambda)$ from the topological set Ω to the metric space Λ is said to be equicontinuous (on Ω) if for any $x \in \Omega$ and $\varepsilon > 0$ there is an open set $G \subset \Omega$ containing x such that*

$$\rho(f(x), f(x')) < \varepsilon, \qquad \forall x' \in G, \forall f \in \mathcal{F}.$$

This says that at every x every function in the set $\mathcal{F}$ does not change that much if we keep within the open neighbourhood G of x.

EXERCISE 46: *Are the Hödler continuous functions (for a given value of α) on $(0, 1)$ (Definition 16) an equicontinuous set?*

We now state the main result of this section — it is a very important result and has many versions.

Result 35 (Arzelà-Ascoli) Suppose that Ω is a separable space and Λ is a metric space. Let $\mathcal{F} \subset \mathcal{C}(\Omega, \Lambda)$ be equicontinuous. Let (f_n) be a sequence in $\mathcal{F}$ such that for each

$x \in \Omega$ the closure of the set $\{f_1(x), f_2(x), \dots\}$ is compact. Then there is a subsequence of (f_n) that converges to a continuous function $f^* \in \mathcal{C}(\Omega, \Lambda)$ and the convergence is uniform on each compact subset of Ω.

We divide this proof up into 3 steps

The first step is to get convergence on a countable set:

Step 1: If (m_n) is a sequence of maps from a *countable* set D into Y (a topological space) and the closure of the sets $\{m_1(x), m_2(x), \dots\}$ is (sequentially) compact. Then there is subsequence of these maps (m_{n_k}) that converges for all $x \in D$.

Proof: The proof of this step is an example of a famous kind of argument attributed to Cantor. As D is countable we can write it as a list $D = \{x_\ell\}$. Now consider the sequence of maps (m_n) evaluated in the first point of D: $\{m_1(x_1), m_2(x_1), \dots\}$. This is compact so it has a convergent subsequence $(m_{n_1}(x_1)) \subset (m_n(x_1))$.

Now consider the subsequence of maps (m_{n_1}) but evaluated at the second point in the list x_2. This is again a sequence in a compact set to it has a convergent subsequence. Call this sub-subsequence $(m_{n_2}(x_2)) \subset (m_1(x_2))$. Iterating this process we can generate (m_{n_k}) $k = 1, 2, \dots$ an infinite sequence of subsequences of (m_n). Now pick the nth map from the sequence (m_{n_k}) and consider the sequence of maps (m_{n_n}) this has the property that every point in D is eventually being forced to converge along a subsequence. ∎

Convergence of equicontinuous functions on a dense set implies they converge everywhere and the limit is continuous.

Step 2: If (f_n) is an equicontinuous sequence from Ω (topological) to Λ (complete metric). And the sequences $(f_n(x)) \subset \Lambda$ converge for all x in a dense $D \subset \Omega$, then (f_n) converges at every $x \in \Omega$ and the limit is continuous.

Proof: Let $x \in \Omega$ and $\varepsilon > 0$ be given. By equicontinuity there exists $G \subset \Omega$ containing x such that $\rho(f_n(x'), f_n(x)) < \varepsilon/3$ for all $x' \in G$. But D is dense and G is open, so there exists $y \in G \cap D$. As $y \in D$ we know the sequence $(f_n(y))$ converges (and thus is Cauchy). By the convergence, there exists N so that $\rho(f_n(y), f_m(y)) < \varepsilon/3$ for all $n, m > N$. Hence

$$\rho(f_m(x), f_n(x)) \leq \overbrace{\rho(f_m(x), f_m(y))}^{\text{equicontinuity}} + \underbrace{\rho(f_m(y), f_n(y))}_{\text{Cauchy}} + \overbrace{\rho(f_n(x), f_n(y))}^{\text{equicontinuity}} \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}$$

Hence $(f_n(x))$ is Cauchy and converges as Λ is complete. Let $f*(.)$ be the pointwise limit this defines. To check the continuity of $f*$ at some x we just need to check that there is a $G \ni x$ such that $\rho(f^*(x), f^*(x')) < \varepsilon$ if $x' \in G$. A similar argument to the above works here. ∎

Pointwise convergence of an equicontinuous sequence on a compact set implies uniform convergence.

Step 3: Let (f_n) be an equicontinuous sequence of functions from Ω (Compact and Topological Space) to Λ (Metric Space). Suppose that (f_n) converges pointwise to the functionf^*. Then (f_n) converges uniformly to f^* on Ω.

Proof: This again uses the power of compactness to turn an infinity of conditions into a finite set of conditions. Let $\varepsilon > 0$ be given. As (f_n) is equicontinuous for each $x \in \Omega$ there exists $G_x \ni x$ such that $\rho(f_n(x'), f_n(x)) < \varepsilon$ for all $x' \in G_x$. (This must also hold for f^* for all $x' \in G_x$.)

The sets $\{G_x : x \in \Omega\}$ are an open cover of Ω. By compactness we can choose a finite open cover $\{G_{x_1}, \dots, G_{x_k}\}$. Then we can choose N so large that f_n is close to its limit on each of these finite points: for all $n > N$ $\rho(f_n(x_i), f^*(x_i)) < \varepsilon/3$ for $i = 1, \dots, k$. Now for all $n > N$ and any $x' \in \Omega$ we have

$$\begin{aligned}\rho(f_n(x'), f^*(x')) &\le \overbrace{\rho(f_n(x'), f_n(x_i))}^{\text{equicontinuity}} + \overbrace{\rho(f_n(x_i), f^*(x_i))}^{\text{uniform on finite set}} + \overbrace{\rho(f^*(x_i), f^*(x'))}^{\text{equicontinuity}} \\ &\le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}\end{aligned}$$

Thus we have uniform convergence. ∎

EXERCISE 47: *Is the set of functions $\mathcal{C}([0,1])$ equicontinuous?*

EXERCISE 48: *Suppose that $\mathcal{F} \subset \mathcal{C}([0,1])$ is equicontinuous. Show that $\overline{\mathcal{F}}$ is also equicontinuous.*

Finally it is worth noting there are many versions of the Arzelà-Ascoli Theorem (see the Wikipedia page for some more on this.) Another version that is also useful but less general is the following.

Result 36 (Arzelà-Ascoli Again) Suppose that Ω is a compact metric space and $\mathcal{F} \subset \mathcal{C}(\Omega)$. Then $\overline{\mathcal{F}} \subset \mathcal{C}(\Omega)$ if and only if $\mathcal{F}$ is bounded and equicontinuous.

4.4. Compactness of Product Spaces

We have already defined a topology on product spaces using the topologies of the coordinate spaces in Section 3.1.1. Suppose that $\{\Omega_\iota, \mathcal{T}_\iota\}$ are a collection of topological spaces and $\Omega := \prod_{\iota \in I}$ is a product space. Consider a set in Ω that consists of one basis set $B_{\iota'}$ for $\Omega_{\iota'}$ and the product of the remaining sets $\prod_{\iota \in I \setminus \{\iota'\}}$. This is a basis set for the product topology. As one can only take finite intersections, this means that the open sets in the product topology can only be different from Ω_ι for a finite number of coordinates.

A *very* important and useful result which we will not prove is

Result 37 (Tychonoff) The product of any family of compact sets is compact in the product topology.

Analysis for Economists: ECON0118

Martin Cripps

November 2, 2020

Contents

5. Normed Spaces: Banach Spaces and Hilbert Spaces

When a space has a norm it also has a linear, or vector space, structure (see Section 3.2.3 for the formal details of this). Thus when we move away from metric spaces to normed spaces, one can start to think about linearity and how it interacts with concepts such as continuity and compactness. The notion of a complete normed space is so important that it is given a special name—a Banach Space.

The most used norms on $\mathcal{C}([0,1])$ have special notation. They are defined here and we will use these definitions below:

$$\|f\|_\infty := \sup_{x\in[0,1]} |f(x)|, \qquad \|f\|_p := \left\{ \int_0^1 |f(x)|^p \, dx \right\}^{1/p}.$$

We know that $\mathcal{C}([0,1])$ with the $\|.\|_\infty$ norm is a Banach space. For more general function spaces let (Ω, Σ, μ) be a measure space and let $L^p(\Omega)$ be the space of real-valued measurable functions on Ω and we use the norm $\|f\|_p := \left\{ \int_0^1 |f(x)|^p \, d\mu \right\}^{1/p}$. (In a space of measurable functions setting $\sup f$ may not be very important, so $L^\infty(\Omega)$ is defined using the essential supremum later in this course.)

This notation is extended to the spaces of sequences of real numbers $\mathbb{R}^\mathbb{N}$. This space can be turned into a complete metric space if we use the metric

$$\rho((x_n),(y_n)) := \sum_{n=1}^\infty 2^{-n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}.$$

The space $\mathbb{R}^\mathbb{N}$ is not a normed space however.

Exercise 49: *What properties of a norm does this metric violate?*

Let $\ell^\infty \subset \mathbb{R}^\mathbb{N}$ denote the space of all bounded real-valued sequences, that is, $\ell^\infty := \{(x_n) \in \mathbb{R}^\mathbb{N} : \sup_n |x_n| < \infty\}$. And let ℓ^p denote the space of all p-summable real-valued sequences, that is, $\ell^p := \{(x_n) \in \mathbb{R}^\mathbb{N} : \sum_n |x_n|^p < \infty\}$. Then for $p \geq 1$ we can define norms on these spaces in the same way

$$\|(x_n)\|_\infty := \sup_n |x_n|, \qquad \|(x_n)\|_p := \left\{ \sum_{n=1}^\infty |x_n|^p \right\}^{1/p} \qquad (p \geq 1).$$

Exercise 50: *Show with an example that $\{\sum_{n=1}^\infty |x_n|^p\}^{1/p}$ does not satisfy the triangle inequality if $p < 1$ and hence is not a norm in this case.*

Other important subspaces of $\mathbb{R}^\mathbb{N}$ are c the space of convergent sequences and c^0 the space of sequences that converge to zero. These subspaces are ordered in the following way

$$\ell^1 \subset \ell^p \subset \ell^q \subset c^0 \subset c \subset \ell^\infty \qquad (p < q)$$

5.0.1. Finite-Dimensional Normed Spaces

A vector space Ω is finite dimensional if every element in the space can be written as a linear combination of a given finite list of vectors $e_1, \ldots, e_n \in \Omega$. This finite list of vectors is called a basis. There are other conditions we can require of a basis, such as that the vectors that make it up are orthogonal and that they have a unit length (orthonormal). You should be familiar with these spaces from your study of linear algebra. Here we will record some key properties of these spaces.

- Any two norms on a finite dimensional vector space are equivalent.
- A finite dimensional vector space with a norm is a complete metric space (a Banach Space).
- If Ω' is a finite dimensional subspace of Ω (a normed vector space) the Ω' is closed.

5.1. Banach Spaces

Now we consider all (finite and infinite dimensional) complete normed spaces, these are called Banach spaces.

Definition 41 *A Banach Space is a complete normed space.*

EXERCISE 51: *Hunt the Banach Space: Go back through these notes and find all the Banach Spaces that you can.*

(Hint: ℓ^p is a Banach space for $1 \leq p \leq \infty$. $L^p(\Omega)$ is a Banach space for $1 \leq p \leq \infty$ but you should think about why.) Many of the properties of finite-dimensional vector spaces do not carry over to infinite dimensional ones. You have already seen an example of a normed infinite-dimensional space with two non-equivalent norms. A second problem is that subspaces are not necessarily closed. This is dealt with in the next paragraph.

What is a subspace? It is exactly what you would expect it to be: a piece of a vector space that is itself a little vector space in its own right. Formally, if Ω is a vector space then a set $\Omega' \subset \Omega$ is a subspace if it is also a vector space. If $\Omega' \neq \Omega$ it is called a "proper" subspace. For example, suppose that $c_0 := \{(x_n) : \lim_n x_n = 0\} \subset \ell^\infty$ is the vector space of sequences of real numbers that converge to zero and it is equipped with the norm $\|.\|_\infty$. A linear subspace of c_0 is the set of all sequences that are zero after some finite time

$$U := \{(x_n) \subset c_0 : \exists N \text{ s.t. } x_n = 0, \forall n > N\}.$$

Notice, however, that unlike the case where spaces have finite dimension, the linear subspace U is open not closed: $\overline{U} = c_0$.

Closed linear subspaces are going to be useful when we want to define hyperplanes, so if Ω' is an arbitrary linear subspace of Ω we would like to take its closure $\overline{\Omega'}$ and hope that this too is a subspace. Indeed this is the case.

Result 38 If Ω' is an arbitrary linear subspace of the normed vector space Ω then so too is $\overline{\Omega'}$.

EXERCISE 52: *Consider $x, y \in \overline{\Omega'}$ and sequences $(x_n), (y_n) \subset \Omega'$ such that $x_n \to x$ and $y_n \to y$. Show that $\alpha x + \beta y \in \overline{\Omega'}$.*

A final problem with infinite dimensional Banach spaces is that neither the unit ball $\|x\| \leq 1$ nor the unit sphere $\|x\| = 1$ are compact. You can find sequences in both that do not have convergent subsequences. A simple example of this would be in ℓ^∞ and the sequence $(1, 0, 0, \dots)$, $(0, 1, 0, 0, \dots)$, $(0, 0, 1, 0, 0, \dots)$,.... In this space each of these points lie on the unit sphere and each of these points is distance 1 from each other. A more complicated argument using the following Lemma establishes this for all infinite dimensional Banach spaces.

Lemma 2 (Riesz) *If: Ω is a normed vector space, $Y \neq \Omega$ is a closed linear subspace, and $\alpha \in (0, 1)$, Then, there exists $x_\alpha \in \Omega$ such that $\|x_\alpha\| = 1$ and $\|x_\alpha - y\| > \alpha$ for all $y \in Y$.*

Observe that if $Y \neq \Omega$ then there exists $y \in \Omega \setminus Y$ and as Y is closed there is an $\varepsilon > 0$ such that $\|y - x\| < \varepsilon$ Try to prove this and draw some pictures!

5.2. Linear Operators

Let us suppose that the spaces Ω and Λ are both normed vector spaces with the same field of scalars. Then we define

Definition 42 *A function $L : \Omega \to \Lambda$ is said to be a Linear Operator if for all $x, y \in \Omega$ and all scalars α two conditions hold*

1. $L(x + y) = L(x) + L(y)$,
2. $L(\alpha x) = \alpha L(x)$.

In the special case $\Lambda = \mathbb{R}$ the linear operator above is given a special name; it is called a Linear Functional. An example of a linear functional on the space $\mathcal{L}^\infty[0, 1]$, of Lebesgue measurable (this will be defined later) functions on $[0, 1]$, is the integral[5]

$$L(f) := \int_0^1 f(x)\,dx. \tag{3}$$

It takes functions in $\mathcal{L}^\infty[0, 1]$ and maps them to numbers.

A linear operator is said to be bounded if there exists a constant M such that $\|L(x)\| \leq M\|x\|$ for all $x \in \Omega$. If a linear operator is bounded then we can define its norm to be the smallest value M that can be chosen,

$$\|L\| := \sup_{\substack{x\in\Omega \\ x\neq 0}} \frac{\|L(x)\|}{\|x\|}.$$

From this follows the useful principle: $\|L(x)\| \leq \|x\|\|L\|$ for all $x \in \Omega$. Here are some examples of norms of linear operators: Consider the $n \times n$ matrix M that is defined below.

$$M = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$$

The map $x \mapsto Mx$ is a linear operator from $\mathbb{R}^2$ to itself. (Check this.) Suppose we use the Euclidean norm, $\|(x_1, x_2)\|_2 := \sqrt{(x_1)^2 + (x_2)^2}$, then $\|M\|_2$ is

$$\|M\|_2 := \sup_{x\neq 0} \frac{\sqrt{4x_1^2 + 9x_2^2}}{\sqrt{x_1^2 + x_2^2}} = 3$$

(Explain how we get the last equality.) Suppose we use the sup norm, $\|(x_1, x_2)\|_\infty := \max\{|x_1|, |x_2|\}$, then $\|M\|_\infty$ is

$$\|M\|_\infty := \sup_{x\neq 0} \frac{\max\{2|x_1|, 3|x_2|\}}{\max\{|x_1|, |x_2|\}} = 3$$

(Explain how we get the last equality.)

EXERCISE 53: *What is the norm of the linear functional described in (3)?*

EXERCISE 54: *Let $\mathcal{C}^1([0, 1])$ be the set of bounded continuously differentiable real valued functions on $[0, 1]$. Let $\mathcal{C}([0, 1])$ be the space of continuous functions on $[0, 1]$. Consider the operator L that maps the functions in $\mathcal{C}^1([0, 1])$ to their derivatives: $L(f(.)) := f'(.)$. Check that this is a linear operator. If we equip $\mathcal{C}^1([0, 1])$ and $\mathcal{C}([0, 1])$ with the norm $\|.\|_\infty$, is L a bounded operator?*

EXERCISE 55: *Is $\mathcal{C}^1([0, 1])$ with the norm $\|.\|_\infty$ a Banach space?*

[5]The superscript $\mathcal{L}^\infty$ also indicates that this function space is equipped with the norm $\|.\|_\infty$.

EXERCISE 56: *Now equip $\mathcal{C}^1([0,1])$ with the norm $\|f\| := \|f\|_\infty + \|f'\|_\infty$. What is the norm the linear operator described by the derivative? (This is a crude example of what is called a Sobolev Space—where the regularity of a function is controlled as well as its magnitude.)*

EXERCISE 57: *Let $g \in \mathcal{C}([0,1])$. Define the linear operator $L : \mathcal{C}([0,1]) \to \mathcal{C}([0,1])$ by $L(f) = g(.) \times f(.)$. Find $\|L\|$ and show L is continuous.*

Boundedness is essentially equivalent to the continuity of linear operators as the following result shows. What is interesting here is that if a linear operator is discontinuous the discontinuities are everywhere — if a linear operator is not bounded then it is discontinuous at *every* point.

Result 39 Suppose that L is a linear operator. If L is bounded then L is uniformly continuous. If L is continuous at one point, then L is bounded.

Proof: If L is bounded then $\|L(x_1) - L(x_2)\| \le \|L\|\|x_1 - x_2\|$. So if $\|x_1 - x_2\| \le \varepsilon/\|L\|$ then $\|L(x_1) - L(x_2)\| \le \varepsilon$, so L is uniformly continuous.

If L is continuous at x_0 then there is a δ such that $\|L(x_0) - L(x)\| < 1$ if $\|x_0 - x\| < \delta$. Let $w := \frac{1}{2}\delta z/\|z\|$ for any z. Notice that $\|(x_0 + w) - x_0\| = \|w\| = \frac{1}{2}\delta$, so

$$1 > \|L(x_0 + w) - L(x_0)\| = \|L(w)\| = \|\frac{\delta}{2\|z\|}L(z)\| = \frac{\delta}{2\|z\|}\|L(z)\|$$

Taking the extremes of the above inequality gives $\frac{2}{\delta}\|z\| > \|L(z)\|$ and L is bounded. ∎

We can now think about the space of all linear operators from Ω to Λ. Let $\mathcal{B}(\Omega, \Lambda)$ be the space of all bounded linear operators from the space Ω to Λ we have the following result:

Result 40 If Ω is a normed vector space and Λ is a Banach space, then $\mathcal{B}(\Omega, \Lambda)$ is a Banach space.

An important example of the set $\mathcal{B}(\Omega, \Lambda)$ occurs when $\Lambda = \mathbb{R}$ and $\mathbb{R}$ is the scalar field for the space Ω. In this case the set of linear functionals $\mathcal{B}(\Omega, \mathbb{R})$ is called the dual space of Ω and is sometimes written as Ω^*. Here is an example of a space and its dual.

Take $\mathbb{R}^n$ with the norm $\|.\|_\infty$. A vector $v \in \mathbb{R}^n$ defines a linear functional $L : \mathbb{R}^n \to \mathbb{R}$ by taking an inner product $L(x) := v^T x$. Actually every linear functional $L : \mathbb{R}^n \to \mathbb{R}$ can be described by an $v \in \mathbb{R}^n$, thus the dual space of $\mathbb{R}^n$ is also $\mathbb{R}^n$. (The fact that every linear functional can be described in this way has not yet been proved.)

We will end this section by reporting without proof or much comment some important properties of linear operators. First, we will define what it is to be an open mapping. It essentially maps open sets to open sets.

Definition 43 *A function $f : \Omega \to \Lambda$ is an open mapping if for every open set G in Ω the set $f(G)$ is open in Λ.*

EXERCISE 58: *Suppose that $f : \Omega \to \Lambda$ is continuous and a bijection. Is f^{-1} an open mapping?*

Now we will state a result that says a certain class of linear operators is an open mapping.

Result 41 (Open Mapping Theorem) If L is a continuous linear operator from a Banach space Ω onto a Banach space Λ, then L is an open mapping.

EXERCISE 59: *Suppose that $f : \Omega \to \Lambda$ is continuous and a bijection. Is f^{-1} an open mapping?*

EXERCISE 60: *Give an example of a linear mapping that is not onto that fails to be an open mapping.*

A strengthening of this is

Result 42 (Inverse Mapping Theorem) If L is a 1:1 mapping from a Banach space Ω onto a Banach space Λ, then L is a continuous linear operator iff L^{-1} is a continuous linear operator.

5.3. Hilbert Space

This discussion leads us very naturally to a new kind of space. Thus far we have considered vector spaces, that is, spaces where elements can be added together or scaled up and scaled down by a factor. We will now add to this list of properties the ability to multiply two elements together to measure the "angle" between them.

Definition 44 (Inner Product Space) *An inner product on the linear space Ω with the scalar field $\mathbb{R}$ (or $\mathbb{C}$) is a function $\langle .,. \rangle : \Omega \times \Omega \to \mathbb{R}$ satisfying:*

1. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$
2. $\langle x, z \rangle = \langle z, x \rangle$ *(when the scalar field is $\mathbb{C}$ this becomes $\langle x, z \rangle = \overline{\langle z, x \rangle}$)*
3. $\langle x, x \rangle \geq 0$

4. $\langle x, x \rangle = 0$ *iff* $x = 0$

the pair $(\Omega, \langle , \rangle)$ *is called an inner product space.*

EXERCISE 61: *Some elementary properties of inner products for you to establish*

1. $\langle x - ty, x - ty \rangle = \langle x, x \rangle - 2t\langle x, y \rangle + t^2 \langle y, y \rangle \geq 0$ *for all* $t \in \mathbb{R}$ $x, y \in \Omega$.
2. *Use the fact that the quadratic equation in* t *above is not negative (and has at most one solution) to deduce that its coefficients satisfy the condition*

$$\langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle$$

the Cauchy Schwartz inequality.

3. *Use (4) in the above definition to deduce that if* $\langle x, y \rangle^2 = \langle x, x \rangle \langle y, y \rangle$, *then* $x = ty$ *and the vectors* x *and* y *are colinear.*

Once we have an inner product, we can use it to define a norm $\|x\| := \sqrt{\langle x, x \rangle}$. We will show that $\|.\|_\infty$ cannot be defined in this way, so $\mathcal{C}([0,1])$ with the norm $\|.\|_\infty$ is a Banach space but not a Hilbert space. (Fortunately, the norm $\|.\|_\infty$ can be approximated by the norm $\|.\|_p$ as $p \to \infty$. Moreover, it is still possible to describe some of the geometric notions like hyperplanes in spaces equipped with the norm $\|.\|_\infty$.) Notice that because of the Cauchy Schwartz inequality we have that

$$-1 \leq \frac{\langle x, y \rangle}{\|x\| \|y\|} \leq 1,$$

so we can use this ratio to define an angle between these two vectors $\theta := \cos^{-1} \frac{\langle x,y \rangle}{\|x\| \|y\|}$. We will say that x and y are orthogonal if $\langle x, y \rangle = 0$.

EXERCISE 62: *Verify that* $\sqrt{\langle x, x \rangle}$ *satisfies all the conditions required of a norm.*

EXERCISE 63: *What is the norm on* $\mathcal{C}([0,1])$ *that is induced by the inner product* $\langle f, g \rangle := \int_0^1 f(x) g(x)\, dx$?

One property of inner products is the parallelogram rule

$$\langle x + y, x + y \rangle + \langle x - y, x - y \rangle = 2\langle x, x \rangle + 2\langle y, y \rangle.$$

This implies that the norm derived from the inner product satisfies

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + \|y\|^2.$$

So if a norm does not satisfy this condition it cannot have been derived from an inner product.

EXERCISE 64: *Consider $f(x) = x$ and $g(x) = 1$ in $\mathcal{C}([0,1])$ with the norm $\|.\|_\infty$. Verify that this norm is not induced by an inner product.*

Now we are able to formally define a Hilbert space as a Banach space (complete and normed) where the norm is derived from an inner product.

Definition 45 *An inner product space that is complete with respect to the norm induced by the inner product is called a Hilbert space.*

EXERCISE 65: *Suppose that (x_n) and (y_n) are two convergent sequences in the Hilbert space Ω, such that $x_n \to x$ and $y_n \to y$. Show that $\langle x_n, y_n \rangle \to \langle x, y \rangle$.*

Once we have an inner product we have a way of thinking about geometry. For $u, v \in \mathbb{R}^n$ we define a hyperplane by the equation $0 = v^T(x - u)$. This takes the $n - 1$ dimensional plane/subspace of all vectors that orthogonal to v and translated the origin to the point u. The definition of a hyperplane in a Hilbert space is therefore

$$\langle v, (x - u) \rangle = 0.$$

From this definition is seems that you need an inner product (and a Hilbert space) to define hyperplanes. This could be a problem because some extremely important spaces are Banach but not Hilbert. However, it is possible to define the notion of a hyperplane without an inner product you just take a subspace and translate it.

5.4. Bases

Recall that the set of vectors $\{x_1, \ldots, x_n\} \subset \mathbb{R}^k$ is said to be linearly independent if

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n = 0$$

implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$. This set is said to a basis for $\mathbb{R}^k$ if every $x \in \mathbb{R}^k$ can be written as a unique linear combination of the x's. This obviously implies $k = n$.

We will now extend these ideas to infinite dimensions. The sequence $(x_n) \subset \Omega$ is said to be a (Schauder) basis for Ω if for every $x \in \Omega$ there is a unique sequence of scalars (a_n) such that $x = \sum_{n=1}^\infty a_n x_n$, where the series $\sum_{n=1}^\infty a_n x_n$ converges in the norm to x. It may be possible to find an infinite sequence of linearly independent vectors that do not span the whole space. Such a sequence is called a basic sequence.

There are two famous basis's one for $\mathcal{C}[0,1]$ that is based on tent maps (triangles with base length 2^{-n} centered at the points $k2^{-n}$ for different values of k. The other (called the Haar system) is basis for the Lebesgue measurable functions on $[0,1]$. These are step functions and their reflection in the x axis: for example the function $+1$ on a short interval followed by a -1 on an equal short interval.

If $(x_n) \subset \Omega$ is a basis for Ω, then the subspaces that are spanned by the finite collections of the basis vectors are dense in Ω. This implies that the space Ω is separable. Thus a non-separable space cannot have a basis, but if it is infinite dimensional it does contain a basic sequence.

5.5. Hahn Banach Theorem and the Theorem of the Separating Hyperplane

Here we will end this section of the course by simply stating this theorem — and then describe an important application of it to Economics.

Theorem 1 (Hahn-Banach) *Let Ω be a vector space with real scalars. Let $p : \Omega \to \mathbb{R}$ satisfy $p(\alpha u) = \alpha p(u)$ and*

$$p(u+v) \leq p(u) + p(v) \qquad \forall u, v \in \Omega.$$

Suppose that Ω_0 is a linear subspace of Ω and $L_0 : \Omega_0 \to \mathbb{R}$ is a linear functional on Ω_0 that satisfies $p(u) \geq L_0(u)$ for all $u \in \Omega_0$. Then there exists a linear functional $L : \Omega \to \mathbb{R}$ on Ω_0 that satisfies $L_0(u) = L(u)$ on Ω_0 and $L(u) \leq p(u)$ everywhere.

This shows that if there is a relatively well-behaved linear functional on a subspace it can be extended to a well behaved functional on the whole space.

When there is no Hilbert space structure we cannot talk about hyperplanes. However, linear functionals will do the same job for us. We can think of all the u that satisfy $L(u) = 0$ as a subspace and all the u that satisfy $L(u) = \beta$ as a translated version of this subspace. Thus the separation theorem below uses this version of the separating hyperplane idea.

Theorem 2 (Separating Hyperplane) *Let Ω be a vector space with real scalars and suppose that A and B are disjoint convex sets in Ω and A has an interior point. Then there is L, non-zero linear functional on Ω such that*

$$\inf_{x \in A} L(x) \geq \sup_{x \in B} L(x)$$

References

McDonald, J. N., and N. A. Weiss (1999): *A course in real analysis*. Taylor & Francis US.

Ok, E. (2007): *A (2007), Real Analysis with Economic Applications*. Princeton University Press, Princeton and London.

Royden, H. L. (1988): *Real Analysis*. Prentice Hall, Englewood Cliffs, NJ, 3rd edn.

Probability for Economists: ECON0118

Martin Cripps

March 2, 2022

Contents

INTRODUCTION

These notes are closely related to the text **?**, but I have also used Wikipedea and various other sources along the way. (Notably: **?**, **?**.) I strongly recommend buying **?**— it is excellent. One problem with these notes is that there are no pictures. I will try to supply lots in the lectures. I am sure there are many typos and errors. Please let me know if you think you have found one so I can correct it.

6. PROBABILITY AND MEASURE

Why do we need measure theory to do probability? Measure theory places restrictions on the events/sets to which we can give probabilities. It places these restrictions in such a way as to ensure we can do most of the things we want to do when we talk about probabilities. Some restrictions on the events we are allowed to consider are absolutely necessary, otherwise it is possible to come up with very perverse sets that simultaneously have zero and positive probability. In Section 6.6.1 we give an example of what can go wrong if you try to do probability without placing restrictions of the events you can consider.

6.1. π-Systems, Algebras and σ-Algebras

Most of this section revises the definitions of the earlier part of the course, but it is included to acknowledge the different notation we are now using. First we introduce the notion of a π-system which is weaker than an algebra (as it doesn't allow unions). It is a collection of subsets that is closed under finite intersections.

Definition 1 *Let S be a set. Let $\mathcal{I}$ be collection of subsets of S. $\mathcal{I}$ is a $\underline{\pi\text{-system}}$ if: $A, B \in \mathcal{I}$ implies that $A \cap B \in \mathcal{I}$.*

An important example of a π-system is the set of open intervals on $\mathbb{R}$. An algebra of sets is a class of sets which is closed under finite intersections, unions or complements.

Definition 2 *Let S be a set. Let Σ_0 be collection of subsets of S. Σ_0 is an* algebra *if: (a) $A \in \Sigma_0$ implies that $A^c \in \Sigma_0$, (b) $A, B \in \Sigma_0$ implies that $A \cup B \in \Sigma_0$.*

In the definition of an algebra we are allowed to take finite numbers of unions or intersections or complements. In a σ-algebra we are allowed to take countable numbers of unions or intersections or complements.

Definition 3 *Let S be a set. Let Σ be collection of subsets of S. Σ is a σ-*algebra *if: (a) $A \in \Sigma$ implies that $A^c \in \Sigma$, (b) $(A_n)_{n=1}^{\infty}$ where $A_n \in \Sigma$ for all n then $\cup_{n=1}^{\infty} A_n \in \Sigma$.*

Recall that there is always a smallest algebra (or σ-algebra) generated by a given family of sets. If $\mathcal{C}$ is a family/class of subsets of S we will use $\sigma(C)$ to denote the smallest σ-algebra generated by $\mathcal{C}$. For example, we could think of the class $\mathcal{C}$ as being the π-system of open intervals on $\mathbb{R}$. The Borel σ-Algebra is the smallest σ-algebra on $\mathbb{R}$ that is generated by the open intervals. This will be written now as $\mathcal{B}$ (notation that was previously used for a basis). An alternative description/definition of $\mathcal{B}$ is that it is the smallest σ-algebra generated by the collection of intervals $\{(-\infty, x] : x \in \mathbb{R}\}$.

EXERCISE 0: *Write out the proof that the set (a, b) is in the smalles σ-algebra generated by the sets $\{(-\infty, x] : x \in \mathbb{R}\}$.*

Definition 4 *The pair (S, Σ), where S is a set and Σ is a σ-algebra on S, is called a measurable space.*

Here is an example of a way of mapping sets to numbers that does not satisfy the conditions of an algebra.

EXERCISE 1: *Suppose that $\mathcal{S} = \mathbb{N}$. For any set $V \subset \mathbb{N}$ define*

$$m(V) := \lim_{n\to\infty} \frac{\sharp(V \cap \{1, 2, \ldots, n\})}{n}.$$

if this limit exists. Let us define the class $\mathcal{V}$ to be the set of subsets of $\mathbb{N}$ such that this limit exists. Give an example of a set $V_0 \subset \mathbb{N}$ such that $V_0 \notin \mathcal{V}$. Hence, give examples of two sets $V_1, V_2 \in \mathcal{V}$ such that $V_1 \cap V_2 \notin \mathcal{V}$ and show that $\mathcal{V}$ is not an algebra.

6.2. Set Functions and Measures

Consider a set S equipped with an algebra Σ_0. One can define a function that takes the sets in the algebra and maps them to non-negative numbers $\mu_0 : \Sigma_0 \to [0, \infty]$. (Here we have used $[0, \infty]$ and not $[0, \infty)$, because we explicitly want to allow the possibility that this function gives the value ∞ to a set. This is called the extended positive interval.) This set function is called additive if: $\mu_0(\emptyset) = 0$ and for all $F, G \in \Sigma_0$

$$F \cap G = \emptyset \qquad \Rightarrow \qquad \mu_0(F) + \mu_0(G) = \mu_0(F \cup G).$$

The map is called countably-additive if it is additive and for any sequence $(F_n)_{n=1}^{\infty}$ of disjoint sets in Σ_0 with $\cup_n F_n \in \Sigma_0$ it is true that

$$\sum_{n=1}^{\infty} \mu_0(F_n) = \mu_0(\cup_n F_n).$$

Definition 5 *The triple (S, Σ, μ) is called a measure space if (S, Σ) is a measurable space and μ is a countably additive measure.*

- *It is called finite if $\mu(S) < \infty$.*
- *It is called σ-finite if there exists a sequence (F_n) such that $S \subset \cup F_n$ and $\mu(F_n) < \infty$ for all n.*
- *It is called a probability measure/space if $\mu(S) = 1$.*

In probability spaces, there are frequently sets that are not empty but are nevertheless given zero probability: $F \subset S$ s.t. $\mu(F) = 0$. These sets/events are called null. (An example of such a set would be a countable collection of points in $[0,1]$ where μ is given by the uniform distribution/Lebesgue measure.) The complement of a null set is an event that is given probability one, although it is not equal to S. A property that holds on an event that occurs with probability one is said to hold almost everywhere. This is frequently abbreviated to *a.e.* . Something that is a.e. true is not necessarily always true but is true for the purposes of probability statements.

If (S, Σ, μ) is a measure space and $F_i \in \Sigma$, then the following properties hold:

$$\mu(F_1 \cup F_2 \cup \cdots \cup F_N) \leq \mu(F_1) + \mu(F_2) + \cdots + \mu(F_N),$$

$$\mu(\cup_{i=1}^N F_i) = \sum_{i=1}^n \mu(F_i) - \sum_{j<i} \mu(F_j \cap F_i) + \sum_{k<j<i} \mu(F_k \cap F_j \cap F_i) + \cdots + (-1)^{N-1} \mu(\cap_{i=1}^N F_i)$$

The final formula here is called the inclusion-exclusion formula. It can be proved using induction on the number of sets N.

EXERCISE 2: *Prove the inclusion exclusion formula for $N = 3$.*

Here are some further results on the convergence of sequences of measures that follow from the fundamental properties of measure spaces.

Result 1 (Monotone Convergence) Suppose that (S, Σ, μ) is a measure space and there are sequences of events: $(F_n) \subset \Sigma$, $(G_n) \subset \Sigma$, and $(H_n) \subset \Sigma$.

- If $F_n \subset F_{n+1}$ for all n and $F := \cup F_n$ then $\mu(F_n) \to \mu(F)$ as $n \to \infty$.
- If $G_n \supset G_{n+1}$ for all n, $\mu(G_n) < \infty$ for some n, and $G := \cap G_n$ then $\mu(G_n) \to \mu(G)$ as $n \to \infty$.
- If $\mu(H_n) = 0$ for all n and $H := \cup H_n$ then $\mu(H) = 0$.

The basic trick to establish all these results is to break unions down into disjoint unions of different sets (that is $A \cup B = A \cup (B \setminus A)$). Thus we can write F as a countable union of disjoint events. The σ-additivity of the measure then implies that $\mu(F)$ equals this countable sum.

EXERCISE 3: *Do the proof of the second claim in this result.*

6.3. Fundamental Results on Probability Spaces

Checking that two probability measures are the same appears to be a difficult job—there are awful lot of sets/events to check. The following result makes this a lot easier. It says that if two probability measures agree on a π-system that generates a σ-algebra then they must also agree on the σ-algebra.

Result 2 If $\mathcal{I}$ is a π-system on S and there are two probability measures μ_0 and μ_1 on $\Sigma := \sigma(\mathcal{I})$ such that $\mu_0(I) = \mu_1(I)$ for all $I \in \mathcal{I}$, then

$$\mu_0(I) = \mu_1(I), \qquad \forall I \in \Sigma.$$

This says that if two probability measures agree on the π-system of open intervals, for example, then they must agree on all the Borel sets — a much larger class.

Our next fundamental result explains how we can "extend" a measure that works for finite operations (an algebra) to a measure for countable properties (a σ-algebra).

Result 3 (Carathéodory Extension Theorem) Σ_0 is an algebra on S and $\Sigma = \sigma(\Sigma_0)$ is the σ-algebra it induces. If $\mu_0 : \Sigma_0 \to [0, \infty]$ is a countably additive map, then there exists a measure μ on (S, Σ) such that $\mu(F) = \mu_0(F)$ for all $F \in \Sigma_0$. If $\mu_0(S) < \infty$, then μ is unique (by the previous result).

EXERCISE 4: *Let (S, Σ) be a measurable space and (μ_n) be a sequence of measures on Σ and (α_n) a sequence in $\mathbb{R}_+$ Define the set functions:*

$$(\mu_1 + \mu_2)(F) := \mu_1(F) + \mu_2(F), \qquad (\alpha_1\mu_1)(F) := \alpha_1\mu_1(F),$$
$$\left(\sum_n \alpha_n\mu_n\right)(F) := \sum_n \alpha_n\mu_n(F);$$

$\forall F \in \Sigma$. *Show that the functions $(\mu_1 + \mu_2)$, $(\alpha\mu_1)$, and $\sum_n \alpha_n\mu_n$ are also measures on Σ.*

6.4. Lebesgue Measure

This is the usual measure we apply to $\mathbb{R}^n$ when we speak of uniform distributions and distances. Here we will only describe how it can be defined in the case of $\mathbb{R}$. This avoids a lot of the technicalities associated with "outer" and "inner" Lebesgue measure that arise when we cover sets with open rectangles or fill sets with compact rectangles. We will assume the extension of Lebesgue measure from $\mathbb{R}$ to $\mathbb{R}^n$ where ever it is necessary.

To derive Lebesgue measure we will let $S := (0, 1]$ and then consider the algebra generated by the intervals $(a, b]$ where $0 \leq a < b \leq 1$. If we take finite unions or intersections of sets of this kind we get sets of the form $F := (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$, where n is finite. Thus these sets make up an algebra Σ_0 on $(0, 1]$. On this algebra, Σ_0, we define a measure μ_0 as

$$\mu_0(F) := \sum_{i=1}^{n}(b_i - a_i).$$

The problem is now to show that we can extend this additive measure on an algebra to countably additive measure on a σ-algebra. Thus Carathéodory Theorem will do the job — provided we can show that μ_0 is countably additive.

Result 4 μ_0 is σ-additive.

Why: Let (F_n) be a sequence of disjoint sets and $F := \cup_n F_n$. We want to show that $\mu_0(F) = \sum_n \mu_0(F_n)$. We begin with two of re-definings of the the question.

First, define the increasing sequence of sets (G_n) where $G_n = \cup_{m=1}^{n} F_m$. We (equivalently) want to show that $\mu_0(G_n) \to \mu_0(F)$. Second, define the decreasing sequence of sets (H_n), where $H_n := F \setminus G_n$. We now want to show that $\mu_0(H_n) \to 0$ if $H_n \to \emptyset$.

Suppose that it didn't and $\mu_0(H_n) > 2\epsilon$ for all n, then we want to show that $\cap_{n=1}^{\infty}(H_n) \neq \emptyset$. Notice the H_n are not closed sets but we are going to take slightly smaller sets than H_n that are closed. First take J_n such that $\overline{J}_n \subset H_n$ and such that $\mu_0(H_n \setminus J_n) < \epsilon/2^n$. You can do this as $\mu_0(H_n) > 2\epsilon$ and H_n is just a finite collection of intervals so you can shrink the length of each of the intervals in H_n by the factor $1 - \epsilon/2^n$.

If the intervals $\overline{J}_n$ were nested and had strictly positive measure, then we would know they that would have a non-empty intersection because the properties of closed intervals that we described in the previous part of this course. Thus we would have a contradiction. What we need to check is whether $\cap_{k=1}^{n} J_k \neq \emptyset$. To see this notice that

$$H_n \setminus (\cap_{k=1}^{n} J_k) \subset \cup_{k=1}^{n}(H_n \setminus J_k) \subset \cup_{k=1}^{n}(H_k \setminus J_k)$$

Hence

$$\mu_0(H_n \setminus \cap_{k=1}^n J_k) \leq \sum_{k=1}^n \mu_0(H_k \setminus J_k) < \sum_{k=1}^n \epsilon/2^k < \epsilon$$

If $\mu_0(H_n) > 2\epsilon$ and $\mu_0(H_n \setminus \cap_{k=1}^n J_k) < \epsilon$ then $\mu_0(\cap_{k=1}^n J_k) > \epsilon$. Which implies that $\cap_{k=1}^n J_k \neq \emptyset$ for all n. And hence that $\cap_{k=1}^n \bar{J}_k \neq \emptyset$ for all n. This closed sequence of nested sets must, therefore have a non-empty limit. Hence we have shown that H_n do not converge to an empty set.

6.5. *Borel Probability Spaces and Regular Measures*

Previously, we introduced the notation $\mathcal{B}$ for the Borel σ-Algebra on $\mathbb{R}$. That is, the smallest σ-algebra on $\mathbb{R}$ that is generated by the open intervals. This notion can be extended to arbitrary measurable spaces. Suppose that S is a metric space with the associated topology of open neighbourhoods. Then $\mathcal{B}(S)$ denotes the Borel σ-algebra on S, that is the σ-algebra on S that is generated by the open neighbourhoods. If μ is a probability measure on $\mathcal{B}(S)$ it is usually called a Borel probability measure on S.

Borel probability measures μ have a useful approximation property for any event $X \in \mathcal{B}(S)$ there is a sequence of open sets $G_n \supset X$ and a sequence of closed sets $F_n \subset X$ such that $\mu(G_n) \to \mu(X)$ and $\mu(F_n) \to \mu(X)$.

Definition 6 *The measure μ on the metric space S is said to be* regular *if for any $\varepsilon > 0$ and any $X \in \mathcal{B}(S)$ there exists a compact set $K \subset X$ and an open set $G \supset X$ such that $\mu(G \setminus K) < \varepsilon$.*

We can now state the result we want very succinctly.

Result 5 A Borel probability measure is regular.

6.6. *Events and Probability Spaces*

Thus far we have been able to define a space S, a σ-algebra of sets and a countably additive measure μ. We will use these notions to define a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Here the space $\Omega \ni \omega$ is interpreted as a set of outcomes/states of the world. The σ-algebra $\mathcal{F}$ is the set of events that can be observed or considered by the model. Finally, $\mathbb{P}$ is the probability measure describing the likelihood of each outcome or set of outcomes.

Example 1 (Coin Tossing) *You are likely to have seen many examples of spaces like these. One important example would be the case where the set of outcomes is all infinite sequences of* *heads and tails: $\Omega = \{H, T\}^{\mathbb{N}}$. In this case we think a coin be tossed infinitely many times. The set of events we are allowed to consider is not any subset of Ω. We will use the σ-algebra that is generated by the product topology on Ω. (Check that you recall what the product topology is.) Let $H_n \subset \Omega$ be the collection of all $\omega \in \Omega$ that have H in the nth entry. (Note that the complement of H_n is the event that ω has T in the nth entry.) Then we will define $\mathcal{F}$ to be the smallest σ-algebra generated by the sets $\{H_n : n \in \mathbb{N}\}$. Finally we need a way of attaching probabilities to events in Ω. This we leave till later.*

6.6.1. A Non-Measurable Set

In this setting it is possible to give a nice example of a non-measurable set and why we need to restrict the set of events we attach probabilities/measures to. Suppose we are able to give σ-additive probabilities to any arbitrary subset in 2^{Ω}. Now we will show that this leads us to a contradiction. For $\omega \in \Omega$ write the explicit infinite sequence of H's and T's as $\omega := (\omega_1, \omega_2, \dots)$.

Now define the set $[\omega] \subset \Omega$ to be the set of all $\omega' \in \Omega$ that only differ from ω in a finite number of entries. First let $\mathcal{K}$ denote the (countable) family of all finite subsets of $\mathbb{N}$.

$$[\omega] := \{\omega' \in \Omega : \exists K \in \mathcal{K} \text{ s.t. } \omega_n = \omega'_n \Leftrightarrow n \notin K\}.$$

As the set of all finite subsets of $\mathbb{N}$ is countable, the set $[\omega]$ is countable. We can generate any $\omega' \in [\omega]$ by flipping the entries of ω in the finite set $K \subset \mathbb{N}$. Let $f_K : \Omega \to \Omega$ denote the function that flips the entries in the K locations, so that $f_K(\omega) \in [\omega]$ for all K. We also have that

$$\cup_{K \in \mathcal{K}} f_K(\omega) = [\omega].$$

Now we will generate a contradiction. First notice that as ω varies the sets $[\omega]$ form a partition the space Ω, that is either $[\omega] = [\omega']$ or $[\omega] \cap [\omega'] = \emptyset$. Let us now pick one ω from each element of this partition (Axiom of Choice!). We will call this collection of histories the set V. This set is uncountable, because the sets $[\omega]$ are countable and their union makes the uncountable set Ω.

If we do the function f_K to the set V we generate a new selection of histories —one from each element of the partition. But we know that if we do all the functions f_K (as $K \subset \mathbb{N}$ varies) to a point in $[\omega]$ we generate the entire set so,

$$\cup_{K \in \mathcal{K}} f_K(V) = \Omega.$$

We also know that if we are able to attach probabilities μ we must have $\mu(f_K(V)) = \mu(V)$, because the events V and $f_K(V)$ differ only on a finite set and the coin tosses are independent. We know that the sets $K \in \mathcal{K}$ are countable and that the events V and

$f_K(V)$ are disjoint, so for any σ-additive probability we have

$$\begin{aligned} \mu(\cup_{K\in\mathcal{K}} f_K(V)) &= \mu(\Omega) \\ \sum_{K\in\mathcal{K}} \mu(f_K(V)) &= 1 && \sigma\text{-additivity} \\ \sum_{K\in\mathcal{K}} \mu(V) &= 1, && \mu(f_K(V)) = \mu(V) \end{aligned}$$

This is a contradiction. If $\mu(V) = 0$ then the above gives $0 = 1$. If $\mu(V) > 0$ then the above gives $\infty = 1$. The problem is we have introduced and event V to which we should not be able to give a measure. From the other results in the course, we can deduce that V is not an event in the σ-algebra generated by the product topology.

6.7. Sequences of Events

6.7.1. Almost Surely

In probability spaces we will not be able to say that something is true in all possible cases. Unfortunately, this opens up the possibility for logical contradictions and paradoxes. Instead we will content ourselves with statements that are true "almost always" or "almost surely". These are statements that hold with probability one but not necessarily in every state of the world. Suppose that $S(\omega)$ is a statement about the outcome $\omega \in \Omega$. We will say that $S(\omega)$ is true *almost surely* or a.s. if first the event that S is true is in the σ-algebra. And second that the probability that the event S is true equals one.

An immediate consequence of the Result 1 and De Morgans Laws is that the intersection of any sequence of almost sure events is itself an almost sure event.

Result 6 If there is a sequence of events $(F_n) \subset \mathcal{F}$ such that $\mathbb{P}(F_n) = 1$ for all n, then $\mathbb{P}(\cap_n F_n) = 1$.

EXERCISE 5: *Do this proof.*

6.8. Sequences of Events

Now let us recall the notions of lim inf and lim sup as applied to both sequences of numbers $(x_n) \subset \mathbb{R}$ and sequences of sets (F_n).

$$\limsup_{n\to\infty} x_n := \inf_n \sup_{k\geq n} x_k, \qquad \liminf_{n\to\infty} x_n := \sup_n \inf_{k\geq n} x_k,$$

Intuitively the lim sup is an upper bound on all very large values of x_n and the lim inf is a lower bound. An alternative interpretation is they are the largest and smallest cluster points of the sequence, so if $z > \limsup$ then eventually all points in the sequence are below z.

$$\liminf F_n := \bigcup_{k=0}^{\infty}\left(\bigcap_{n=k}^{\infty} F_n\right), \qquad \limsup F_n := \bigcap_{k=0}^{\infty}\left(\bigcup_{n=k}^{\infty} F_n\right)$$

Here the lim inf is the set of points that are in infinitely many of the F_n's and in all F_n for n large enough. The lim sup is just the set of points that is in infinitely many of the F_n's. The lim sup and lim inf are very important when applied to events. For example, it might be very important to know that a tossed coin generates an infinite sequence of heads. If H_n is the event that the coin comes up heads in period n then the event that an infinite sequence of heads occurs is the event $\limsup H_n$.

We now list and prove some important results on infinite sequences of events. The first gives a lower bound on the probability of an infinite number of events being true

Result 7 If $(F_n) \subset \mathcal{F}$ then

$$\mathbb{P}(\limsup_n F_n) \geq \limsup_n \mathbb{P}(F_n)$$

Proof: Define the decreasing sequence of events $G_n := \cup_{m\geq n} F_m$. These converge monotonically to $\limsup_n F_n$, thus by Result 1, $\mathbb{P}(G_n)$ converges monotonically to $\mathbb{P}(\limsup_n F_n)$. As $F_m \subset G_n$ for all $m \geq n$ we know that

$$\sup_{m\geq n} \mathbb{P}(F_m) \leq \mathbb{P}(G_n), \qquad \forall n.$$

Taking the limits of both sides of these in n

$$\lim_{n\to\infty} \sup_{m\geq n} \mathbb{P}(F_m) \leq \lim_{n\to\infty} \mathbb{P}(G_n).$$

OR (using the fact that $\mathbb{P}(G_n)$ converges monotonically to $\mathbb{P}(\limsup_n F_n)$)

$$\limsup_{n\to\infty} \mathbb{P}(F_n) \leq \mathbb{P}(\limsup_{n\to\infty} F_n).$$

■

How might you use this result? Suppose you have an infinite sequence of events, like the toss of an infinite sequence of coins, and you know that there is a subsequence of these events that have a probability of occurring that converges to one. (They could converge to one arbitrarily slowly but nevertheless converge to one.) Then we know $\limsup_{n\to\infty} \mathbb{P}(F_n) = 1$, so we also know that with probability one an infinite number of these events is certain to occur $\mathbb{P}(\limsup_{n\to\infty} F_n) = 1$. Notice we are not requiring that these events are independent or anything else, we are just focussing on their individual probabilities.

The next result is so important it has a name that you ought to know. It is a form of converse to the previous result. It says that if the sum of the probabilities of the individual events is finite, then only a finite number of them can ever occur. (Or, that there is zero probability of an infinity of the events happening.)

Result 8 (First Borel-Cantelli Property) If $(F_n) \subset \mathcal{F}$ and $\sum_n \mathbb{P}(F_n) < \infty$ then

$$\mathbb{P}(\limsup_n F_n) = 0.$$

Proof: Clearly if $\sum_n \mathbb{P}(F_n) < \infty$ then $\mathbb{P}(F_n) \to 0$ as $n \to \infty$. More is true, the whole tail of the sequence must shrink to zero $\sum_{m>n} \mathbb{P}(F_m) \to 0$ as $n \to \infty$. This implies

$$\sum_{m>n} \mathbb{P}(F_m) \geq \mathbb{P}(\cup_{m>n} F_m) \to 0,$$

as $n \to \infty$. Now we notice that $\cup_{m>n} F_m \supset \limsup_n F_n$. So we have that

$$\sum_{m>n} \mathbb{P}(F_m) \geq \mathbb{P}(\cup_{m>n} F_m) \geq \mathbb{P}(\limsup_n F_n) \geq 0,$$

As the term on the left converges to zero so must all the intermediate terms. ∎

This is a fantastically useful result. Again it does not rely on independence, so suppose we have a sequence of events (F_n) each with probability $1/n^2$ of occurring, then we know for sure that only a finite number of them can ever occur.

Here's an example of the use of this result that has considerable importance. Suppose you are studying a Markov chain that is currently in state i and you have calculated the probability the chain will return to that state in n periods — call this p_n. If $\sum_n p_n < \infty$ you know that the state is transient and will only be visited a finite number of times. (Note: The events in question here are highly dependent — visiting state i in period m has a big influence on the probability it will be visited in $m+1$.)

The final result applies to the lim inf of our sequence of events. Thus we are estimating the probability that all sufficiently large events in our sequence have a property.

Result 9 If $(F_n) \subset \mathcal{F}$, then

$$\mathbb{P}(\liminf_n F_n) \leq \liminf_n \mathbb{P}(F_n)$$

Proof: Define the increasing sequence of events $G_n := \cap_{m \geq n} F_m$. These converge monotonically to $\liminf_n F_n$, thus by Result 1, $\mathbb{P}(G_n)$ converges monotonically to $\mathbb{P}(\liminf_n F_n)$. As $F_m \supset G_n$ for all $m \geq n$ we know that

$$\inf_{m \geq n} \mathbb{P}(F_m) \geq \mathbb{P}(G_n), \qquad \forall n.$$

Taking the limits of both sides of these in n

$$\lim_{n\to\infty} \inf_{m\geq n} \mathbb{P}(F_m) \geq \lim_{n\to\infty} \mathbb{P}(G_n).$$

OR (using the fact that $\mathbb{P}(G_n)$ converges monotonically to $\mathbb{P}(\liminf_n F_n)$)

$$\liminf_{n\to\infty} \mathbb{P}(F_n) \geq \mathbb{P}(\liminf_{n\to\infty} F_n).$$

$\blacksquare$

7. Random Variables

We begin by defining what it is for a function to be "measurable"—this does not require a measure to be present it is really a property of the function and its relationship to a σ-algebra.

7.1. Measurable Functions

Suppose we have a space and an associated σ-algebra (S, Σ). Now consider a real-valued function that is defined on S, that is, $f : S \to \mathbb{R}$ (we are not restricting f in any way — for example it can be highly discontinuous). The function f is called measurable if observing some feature of its value generates no more information than observing some event in Σ.

The formal definition of this property is much less intuitive than this statement. First we need to be precise about what features of the outcome of the function f one is actually able to observe. To that end, recall what the Borel sets in $\mathbb{R}$ are — this is $\mathcal{B}$ the smallest σ-algebra that is generated by the open intervals. Whether $f(s) \in B$ for some Borel sets $B \subset \mathbb{R}$ is what one is able to observe. Now we can write down the definition.

Definition 7 *$f : S \to \mathbb{R}$ is said to be a Σ-measurable function if $f^{-1}(B) \in \Sigma$ for every Borel set $B \subset \mathbb{R}$.*

Here is an example. Suppose S that S is the space of all tosses of two coins

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

but the σ-algebra describes the information of a person who only observes the outcome of the first coin toss.

$$\Sigma := \{S, \emptyset, \{(H,H), (H,T)\}, \{(T,H), (T,T)\}\}$$

Then the function $f^* : S \to \mathbb{R}$ defined below

$$f^*(H,H) = 1, \qquad f^*(H,T) = 2, \qquad f^*(T,H) = 3, \qquad f^*(T,T) = 4;$$

is not Σ measurable because observing $f^*(s) \in (0, 1.5)$ tells the observer that the state (H,H) occurred. This is more than they knew from observing Σ. This is captured by the property that $(f^*)^{-1}((0, 1.5)) = (H,H) \notin \Sigma$.

There are a lot of quite boring properties of measurable functions that underpin a lot of the ways we use them. You can find proofs of these facts in the texts we use—but we will just record them here. Results on measurable functions.

- f is Σ-measurable iff for all $a \in \mathbb{R}$ $f^{-1}((-\infty, a)) \in \Sigma$.
- f is Σ-measurable iff for all $a \in \mathbb{R}$ $f^{-1}((a, \infty)) \in \Sigma$.
- f is Σ-measurable iff for all $a \in \mathbb{R}$ $f^{-1}((-\infty, a]) \in \Sigma$.
- f is Σ-measurable iff for all $a \in \mathbb{R}$ $f^{-1}([a, \infty)) \in \Sigma$.
- The set of Σ measurable functions form an algebra: For any $\alpha, \beta \in \mathbb{R}$ and f, g that are Σ measurable, then so are $f.g$ and $\alpha f + \beta g$.
- If f, g that are Σ measurable, then so is $f(g(.))$.

There are also some quite interesting and useful properties of measurable functions that repay some thought. Most of the properties below can be used to establish the only if conditions above.

1. Suppose that f is a Σ-measurable function. If A and B are Borel sets, then $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.
2. Suppose that f is a Σ-measurable function. If A is a Borel set, then $f^{-1}(A^c) = (f^{-1}(A))^c$.
3. If $\mathcal{C}$ is a sub-class of Borel sets of $\mathbb{R}$ and the smallest σ-algebra generated by $\mathcal{C}$ are the Borel sets themselves, then if $f^{-1}(C) \in \Sigma$ for every $C \in \mathcal{C}$, then f is Σ-measurable.

EXERCISE 6: *Prove the first 2 statements in this list.*

The final results on measurable functions we will involve the taking of limits. Suppose that (f_n) is a sequence of Σ-measurable functions. (To deal with limits we need to increase the range of the functions from $\mathbb{R}$ to $[-\infty, \infty]$ hence we add two new points that are denoted $\{-\infty\}$ and $\{\infty\}$ to allow for the fact that the limits of the sequences may not be unboundedly large or small.)

Result 10 Then: (i) $\inf f_n$ is Σ-measurable, (ii) $\liminf f_n$ is Σ-measurable, (iii) $\limsup f_n$ is Σ-measurable, and (iv) $\{s \in S : \lim_n f_n(s) \text{ exists } \}$ is in Σ.

Proof: For the proofs of these statements recall that is is sufficient to show that the sets of the form $\{s : h(s) \geq a\}$ are in Σ for all a. To prove (i) we take $h = \inf f_n$ and observe that

$$\{s : \inf f_n \geq a\} = \cap_n \{s : f_n \geq a\}.$$

The sets on the right are in Σ by assertion.

To prove (ii) observer that by (i) the functions $g_n := \inf_{r \geq n} f_r$ are measurable for all n. But as the functions g_n are increasing

$$\liminf_n f_n = \lim_{n \to \infty} g_n = \sup_n g_n$$

the fact that the supremum of a bunch of measurable functions follows from an argument similar to (i). (iii) requires the same argument as (ii).

In (iv) for a limit to exist we must have one of three things be true: that $\limsup f_n = \liminf f_n$, or that $\liminf f_n = \infty$, or that $\limsup f_n = -\infty$. The first of these is the condition that the difference of two measurable functions equals zero, $\limsup f_n - \liminf f_n = 0$, which is a measurable set. The second and third are also conditions on measurable functions taking values in the extended real line—again measurable sets. ■

7.2. *Random Variables, Laws, and Distribution Functions*

A random variable is just a measurable function that is defined on a probability space. So suppose we have the triple $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable is just a $\mathcal{F}$-measurable function. (Notice that we could change the probability measure $\mathbb{P}$ and leave the definition of the random variable exactly the same.)

Now let us revisit Example 1 and define a family of random variables. We will define: $f_n(\omega) = 1$ if ω has a H in the nth entry and $f_n(\omega) = 0$ if ω has a T in the nth entry. Thus f_n counts whether there is a head at the nth toss of the coin.

EXERCISE 7: *Verify that f_n is a random variable, that is, it is $\mathcal{F}$-measurable.*

The results in the previous section clearly establish that $S_n := f_1 + f_2 + \cdots + f_n$ is also a random variable. This is not very surprising. They also ensure that $\liminf_n(S_n/n)$ and $\limsup_n(S_n/n)$ are random variables. Thus we also know that the event that $p = \liminf_n(S_n/n) = \limsup_n(S_n/n)$ is a measurable event. This does seem a surprising and new property — our measure theory is powerful enough to allow us to think of the event that the empirical probability of heads is p in the limit.

7.2.1. The Law of a Random Variable and the σ-Algebra of a Random Variable

A random variable $X : \Omega \to \mathbb{R}$ that is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ implicitly defines a probability measure on $\mathbb{R}$ for the Borel sets $\mathcal{B}$. Just take any Borel set $B \subset \mathbb{R}$, then first take its inverse in Ω, that is, $X^{-1}(B) \in \mathcal{F}$. As we are able to give probabilities to sets in $\mathcal{F}$ we can give a probability to the set B that is equal to $\mathbb{P}(X^{-1}(B))$. This new probability measure is sometimes all we care about —not necessarily the underlying space Ω. Hence, this is sometimes called the "Law" of the random variable X and is written $\mathbb{P} \circ X^{-1}$. (We grew up calling this the "distribution" of X.)

Sometimes we have a given random variable X (or a set of random variables $X_1, ..., X_n$) and we want to find a σ-algebra that makes these measurable. (This is similar to the notion of a weak topology that is chosen to make some functions continuous.) How you do this is fairly obvious. You consider the sets $X^{-1}(B) \subset \Omega$, where B is a Borel set. Then you pick the smallest σ-algebra of Ω that is generated by these sets. This is denoted $\sigma(X)$.

7.2.2. Distribution Functions

Once we have the distribution/law of the random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ we can think about its distribution function. That is, the probability $X(\omega) \in (-\infty, a]$, which is just equal to $\mathbb{P}\left(X^{-1}((-\infty, a]) \right)$. Hence we define the distribution function of the random variable X as

$$F_X(a) := \mathbb{P}\left(X^{-1}((-\infty, a])\right) = \mathbb{P}(\{\omega : X(\omega) \leq a\}).$$

Given this definition and the results we have on monotone convergence, there are some fairly obvious properties of the distribution function: (a) $F_X : \mathbb{R} \to [0, 1]$ is non-decreasing, (b) $\lim_{n\to\infty} F_X(n) = 1$, $\lim_{a\to-\infty} F_X(a) = 0$, (c) $F_X(.)$ is right continuous. To see the final property observe that

$$F_X(a) = \mathbb{P}(\{\omega : X(\omega) \leq a\}) = \lim_{n\to\infty} \mathbb{P}(\{\omega : X(\omega) \leq a + n^{-1}\}) = \lim_{n\to\infty} F_X(a + n^{-1})$$

The converse of these claims is also true. If there is a function F that satisfies the properties (a), (b), and (c) above, then we can construct a unique probability measure on $\mathbb{R}$ that has F as its distribution function. This is called the Skorokhod representation problem. These kinds of issues are important in the theory of stochastic processes for example. Suppose you have a given distribution in mind — is it possible to generate this distribution from some sort of history-dependent behaviour. Early work by Philip Strack was on this question.

7.3. Egoroff and Lusin's Theorems

In 1944 Littlewood proposed three principles that make the manipulation of sequences of functions or random variables much easier he said:

"There are three principles, roughly expressible in the following terms: Every (measurable) set is nearly a finite sum of intervals; every (measurable) function is nearly continuous; every convergent sequence of functions is nearly uniformly convergent."

Here we are going to quote more modern versions of these results that have more precision and that you can apply:

The first principle could be viewed as the regularity property that is described in Section 6.5. The second principle:

Result 11 (Lusin's Theorem) If F is a countable collection of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, (where Ω is a separable and complete metric space[1]), then for any $\delta > 0$ there is a compact subset of $K \subset \Omega$ such that every random variable $f \in F$ is continuous on K and $\mathbb{P}(\Omega \setminus K) < \delta$.

The third principle:

Result 12 (Egoroff's Theorem) If (f_n) is a sequence of random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and this sequence satisfies $f_n \to f^o$ almost surely, then for any $\delta > 0$ there exists $A \subset \Omega$ such that $\mathbb{P}(A) < \delta$ and the sequence (f_n) converges uniformly to f^o on $\Omega \setminus A$.

[1] These spaces are often called Polish spaces.

8. INDEPENDENCE

When you first learned probability, you learned that if two events were independent then you just multiplied their probabilities to find the probability of the joint event. We are now worrying about what kind of events that we are able to attach probabilities too, so we will also need to worry about what kind of events we can treat as being independent. To do this we introduce sub-σ-algebras. Suppose that $(\Omega, \mathcal{F})$ is a measurable space and let us consider a sub-collection of sets $\mathcal{G} \subset \mathcal{F}$. If the class $\mathcal{G}$ is itself a σ-algebra then it is called a sub-σ algebra.

Here is an example. Let us suppose that the state space is the set of all possible tosses of two coins: $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$. And, that we allow all possible events to be give probabilities: $\mathcal{F} = 2^{\Omega}$. Then two sub-σ-algebras of $\mathcal{F}$ are:

$$\mathcal{G}_1 := \{\Omega, \emptyset, \{(H,H),(H,T)\}, \{(T,H),(T,T)\}\} \subset \mathcal{F},$$

and

$$\mathcal{G}_2 := \{\Omega, \emptyset, \{(T,T),(H,T)\}, \{(T,H),(H,H)\}\} \subset \mathcal{F}.$$

These sub-σ-algebras are independent if the measure $\mathbb{P}$ satisfies

$$\mathbb{P}(G_1 \cap G_2) = \mathbb{P}(G_1)\mathbb{P}(G_2), \qquad \forall G_1 \in \mathcal{G}_1, G_2 \in \mathcal{G}_2.$$

Random variables X_1 and X_2 are said to be independent if the sub-σ-algebras that they generate $\sigma(X_1)$ and $\sigma(X_2)$ are also independent.

Any event $E \in \mathcal{F}$ generates a trivial sub-σ-algebra $\sigma(E) := \{\emptyset, \Omega, E, E^c\}$. Any two events E_1 and E_2 are said to be independent if and only if the σ-algebras $\sigma(E_1)$ and $\sigma(E_2)$ are independent.

EXERCISE 8: *Verify that* $\mathbb{P}(E_1)\mathbb{P}(E_2) = \mathbb{P}(E_1 \cap E_2)$ *is necessary and sufficient for* $\sigma(E_1)$ *and* $\sigma(E_2)$ *to be independent.*

To check whether two sub-σ-algebras are independent seems to require a lot of checking! In fact it is enough to check that the π-systems that generate the σ algebra's are independent:

Result 13 Let $\mathcal{I}$ and $\mathcal{J}$ by π-systems on Ω. And let $\mathcal{G} := \sigma(\mathcal{I})$ and $\mathcal{H} := \sigma(\mathcal{J})$ be the sub-σ algebras generated by these π-systems. Then $\mathcal{G}$ and $\mathcal{H}$ are independent iff

$$\mathbb{P}(I)\mathbb{P}(J) = \mathbb{P}(I \cap J), \qquad \forall I \in \mathcal{I}, J \in \mathcal{J}$$

One use of this result is a simple check for when random variables generate independent σ-algebras. For any random variable $X(.)$, the sets $\{\omega : X(\omega) \leq a\}$, for $a \in \mathbb{R}$, form a π-system. Thus this result allows us to check that random variables are independent by just checking that their distribution functions multiply, that is, if

$$\mathbb{P}(X \leq a, Y \leq b) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b), \qquad \forall a, b \in \mathbb{R}.$$

EXERCISE 9: *Define* $\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$*, where* $s > 1$*. We can now define a probability distribution on* $N \in \mathbb{N}$ *as* $\Pr(N = n) = n^{-s}/\zeta(s)$*. Let* D_p *be the event that* N *is divisible by* p*, where* p *is a prime number. Show that* $\Pr(N \in D_p) = p^{-s}$ *and that* $\Pr(N \notin D_p) = (1 - p^{-s})$*. Hence show that under this probability distribution the events* D_p *and* D_q *are independent if* p *and* q *are distinct prime numbers. Use the fact that every* $N > 1$ *is divisible by some prime number and independence to show that*

$$\zeta(s)^{-1} = \prod_{p \in \mathcal{P}} \left(1 - p^{-s}\right),$$

where $\mathcal{P}$ *is the set of all prime numbers (this does not include the number 1).*

8.1. Sequences of Independent Events — Borel Cantelli

The first major result on independent events we prove is one of the most important and is given a name to acknowledge this. Whereas the first Borel-Cantelli result gives conditions for a finite number of events to occur that do not require independence, the conditions for an infinite number of events to occur will. Recall that $\limsup F_n$ is the set of states that are in infinitely many of the F_n. So, if $\mathbb{P}(\limsup_n F_n) = 1$, then almost all states in Ω are in infinitely many of these events.

Result 14 (Second Borel-Cantelli Property) If $(F_n) \subset \mathcal{F}$ is a sequence of independent events and $\sum_{n=1}^{\infty} \mathbb{P}(F_n) = \infty$, then

$$\mathbb{P}(\limsup_n F_n) = 1.$$

Proof: We begin by getting an upper bound on the probability that only a finite number of the events occur. Let us consider the probability that for all $n \geq m$ no event F_n occurs. By independence

$$\mathbb{P}(\cap_{n \geq m} F_n^c) = \prod_{n \geq m} (1 - \mathbb{P}(F_m))$$

We can use the result that $e^x \geq 1 + x$ for all x to get an upper bound on this probability

$$\mathbb{P}(\cap_{n \geq m} F_n^c) \leq \prod_{n \geq m} e^{-\mathbb{P}(F_m)} = e^{-\sum_{n \geq m} \mathbb{P}(F_m)} = 0$$

(The last equality holds as we have assumed that $\sum_{n=1}^{\infty} \mathbb{P}(F_n) = \infty$.) As the events $\cap_{n \geq m} F_n^c$ have probability zero for each m so do their unions

$$0 = \mathbb{P}(\cup_m \cap_{n \geq m} F_n^c) = \mathbb{P}(\liminf_n F_n^c)$$

Thus we have shown that there is zero probability of only a finite number of the events occurring. But by De Morgan's Laws

$$(\liminf F_n^c)^c = (\cup_m \cap_{n \geq m} F_n^c)^c = \cap_m \cup_{n \geq m} F_n = \limsup F_n$$

Thus we have shown that the complement of $\limsup F_n$ has zero probability which is all that is required. ∎

EXERCISE 10: *To see why the independence assumption matters here, give an example of events with $\sum_{n=1}^{\infty} \mathbb{P}(F_n) = \infty$ but it is not the case an infinite number of these events occur with probability one.*

Armed with this tool we can do a lot to investigate the probability of an infinite number of events. First here's another example. Suppose we have a sequence of independent identically distributed random variables $(X_n)_{n=1}^{\infty}$. Each of these has an exponential distribution $\Pr(X_n > a) = e^{-a}$. With an infinity of independent observations we would expect there to be arbitrarily large values but can we say anything about the lim sup of the sequence X_n/n? First observe that

$$\Pr(X_n > \alpha \log n) = n^{-\alpha}$$

So if we define $F_n := \{\omega : X_n(\omega) > \alpha \log n\}$ then we know

$$\sum_{n=1}^{\infty} \Pr(F_n) = \sum_{n=1}^{\infty} n^{-\alpha} = \infty, \Leftrightarrow \alpha \leq 1.$$

by the two Borel-Cantelli Lemmas we know that $X_n(\omega) > \alpha \log n$ infinitely many times if and only if $\alpha \leq 1$. So if we compare $\frac{X_n}{\log n}$ with $\alpha \leq 1$ we know that with probability equal to 1 this ratio is bigger than 1 infinitely often. This means that $\Pr(\limsup_n \frac{X_n}{\log n} \geq 1) = 1$.

Now let us compare the fraction $\frac{X_n}{\log n}$ with something that converges to one from above, like $\alpha = 1 + k^{-1}$. We know that for all k the above sum is finite so with probability one there are only a finite number of occasions where $\frac{X_n}{\log n} > 1 + k^{-1}$ for any k. Thus the lim sup sequence $\frac{X_n}{\log n}$ is less than $1 + k^{-1}$ with probability one. As this is true for all k we have that $\Pr(\limsup_n \frac{X_n}{\log n} \leq 1) = 1$.

Combining these two steps we have shown

$$\Pr\left(\limsup_n \frac{X_n}{\log n} = 1\right) = 1.$$

Which is a pretty precise estimate about the large values of the sequence X_n! This is also an example of a phenomenon that we will investigate more in the next section: that limiting properties of independent events tend to be all or nothing kinds of things.

EXERCISE 11: *Let X be a $N(0,1)$ random variable. We know that for any $x > 0$*

$$\Pr(X > x) \leq \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Use this to show that

$$\Pr(\limsup \frac{X_n}{\sqrt{2\log n}} \leq 1) = 1.$$

Where (X_n) is a sequence of i.i.d. $N(0,1)$ random variables.

We end this section with a final technical note. It is not at all obvious that it is possible to generate an infinite sequence of independent random variables that are mutually independent. Thus far we only know how to generate one distribution function, not a whole family. We need to know that this is possible for the above to be entirely well founded. When the notion of a product measure is introduced we will be fine!

8.2. Sequences of Independent Random Variables and Kolmogorov's 0-1 Law

We will now show that a lot of events describing the limits of independent sequences have probability either zero or one. For example, if (X_n) is a sequence of independent random variables, then $\mathbb{P}(\{\sum_{m=1}^{n} X_m \text{ converges}\})$ can either equal one or zero but it cannot have any other probability.

To make this more concrete let us begin with a simple example. Suppose that (X_n) is a sequence of uniformly bounded and independent random variables. That is, there exists an M so that $\mathbb{P}(|X_n| \leq M) = 1$ for all n. We derive two new random variables: $Y_n := n^{-1}\sum_{i=1}^{n} X_i$, the sequence of averages of the X's and the $L := \limsup_n Y_n$.

There are some pretty obvious facts one can state about L. Since $\mathbb{P}(|Y_n| \leq M) = 1$ where also know (by countability) that $\mathbb{P}(|Y_n| \leq M, \forall n) = 1$. This implies that $\mathbb{P}(|L| \leq M) = 1$. But actually it is possible using the results in this section to show something much stronger: that is, there exists some $c \in [-M, M]$ such that $\mathbb{P}(L = c) = 1$! And, that we will know that either the set of states of the world where the sequence Y_n converges to a limit either has probability one, or has probability zero —no intermediate probability is possible.

The argument why this is true is based upon what occurs in the following argument. Consider the two random variables: L and X_1. These are independent of each other. If we defined

$$L' := \limsup_n \frac{1}{n} \sum_{i=2}^{n} X_i$$

it is clear L' and X_1 are independent of each other, because X_1 is independent of all the $X_2, \ldots$. But $L' = L$ (you should do the calculation to check this), so L is also independent of X_1!

It is clear that this argument also applies L and X_n for any fixed value of n. Thus the random variable L is a special kind of random variable, that is independent of what happens in finite histories. This special kind of random variable is called a tail random variable and has very special properties. Notably they are independent of themselves!

Given a general sequence (X_n) of random variables, recall that $\sigma(X_n)$ denotes the σ-algebra that is generated by X_n. Now consider the σ-algebra that is generated by the continuation $X_{n+1}, X_{n+2}, \ldots$ of the sequence. We will write this n-continuation σ-algebra as

$$\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \ldots).$$

We will define $\mathcal{T} := \cap_n \mathcal{T}_n$ to be the set of sets that are in all of these continuation sets. This is called the tail σ-algebra. Notice that the specific value of any one X_n is not measurable with respect to this σ-algebra. However, it contains a number of events we are very interested in:

$$F_1 := \{\omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\};$$

$$F_2 := \{\omega : \lim_{n\to\infty} \sum_{m=1}^{n} X_m(\omega) \text{ exists}\};$$

$$F_3 := \{\omega : \lim_{n\to\infty} n^{-1} \sum_{m=1}^{n} X_m(\omega) \text{ exists}\}.$$

Observe that changing the value of any one X_n in a sequence will not affect whether any of these events occurred or not.

EXERCISE 12: *Prove that F_1 is $\mathcal{T}$ measurable (Hint Recall Result 10).*

There is an amazing feature of all the events in the tail σ-algebra they must have either probability zero or probability one. The reason is that if an event A is in the tail σ-algebra it has to be independent from itself. That means that $\Pr(A \cap A) = \Pr(A)\Pr(A)$. Hence we get that $\Pr(A)^2 = \Pr(A)$ and the only way this can be true is if the event is has either zero probability one. You might be worried about the idea of being independent of itself, but this is is generally true of probability one events. If A is a probability one event then it is independent of all other events because $\Pr(A \cap B) = \Pr(B)$ and $\Pr(A)\Pr(B) = 1\Pr(B)$!

Result 15 (Kolmogorov's 0-1 Law) Let (X_n) be a sequence of independent random variables and $\mathcal{T}$ the tail σ-algebra, then $F \in \mathcal{T}$ implies $\mathbb{P}(F) \in \{0, 1\}$.

Proof: Define the σ-algebra generated by the first n r.v's as $\mathcal{S}_n := \sigma(X_1, \ldots, X_n)$.

1. $\mathcal{S}_n$ and $\mathcal{T}_n$ are independent: From our previous results we know that the events of the form

$$I := \{\omega : X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n\}, \qquad (x_1, \ldots, x_n) \in (\mathbb{R} \cup \{\infty\})^n$$

are a π-system for $\mathcal{S}_n$. Similar events J for $X_{n+1}, \ldots, X_{n+r}$ with $r \in \mathbb{N}$ are a π system for $\mathcal{T}_n$. The assumption that the X_n's are independent mean we multiply the probabilities of the I and J events to get the probabilities of the joint event. Result 13 then proves this claim.

2. $\mathcal{S}_n$ and $\mathcal{T}$ are independent: This follows because $\mathcal{T} \subset \mathcal{T}_n$ and $\mathcal{T}_n$ and $\mathcal{S}_n$ are independent.

Now we define $\mathcal{S}_\infty := \sigma(X_1, X_2, ...)$.

3. $\mathcal{S}_\infty$ and $\mathcal{T}$ are independent: Define the class of sets $\mathcal{K}_\infty := \cup_n \mathcal{S}_n$. This class is not necessarily a σ-algebra although it is the union of them. But it is a π-system — taking intersections of $S \in \mathcal{S}_n$ and $S' \in \mathcal{S}_{n+m}$ is in this class because $\mathcal{S}_n \subset \mathcal{S}_{n+m}$. Furthermore, the π-system $\mathcal{K}_\infty$ generates the σ-algebra $\mathcal{S}_\infty := \sigma(X_1, X_2, ...)$, because the variable X_n is measurable with respect to $\mathcal{S}_n$.

To show that $\mathcal{S}_\infty$ and $\mathcal{T}$ are independent it is sufficient to show that K_∞ and $\mathcal{T}$ are independent because of Result 13. But step 2 shows us that $\mathcal{T}$ and $\mathcal{S}_n$ are independent for all n so $\mathcal{T}$ and $K_\infty = \cup_m \mathcal{S}_n$ are independent.

4. Finally: We have that $\mathcal{S}_\infty$ and $\mathcal{T}$ are independent but also we know that $\mathcal{T} \subset \mathcal{S}_\infty$. So this means that $\mathcal{T}$ is independent of itself. Thus is $F \in \mathcal{T}$ then

$$\mathbb{P}(F) = \mathbb{P}(F \cap F) = \mathbb{P}(F)\mathbb{P}(F)$$

so $\mathbb{P}(F) \in \{0, 1\}$. ■

Now let us return to the sequence of uniformly bounded independent random variables at the start of this section and prove the claim we made there. That is, there exists $c \in [-M, M]$ such that $\mathbb{P}(L = c) = 1$.

Proof: Let $q \in [-M, M]$ be a rational number and define the event $F_q := \{L \geq q\}$. This is a tail event, so by the result above it has probability either zero or 1.

Consider the set of q's with the property that $\mathbb{P}(F_q) = 1$, take q^* to be the supremum of this set. As there are a countable number of such q's we know that the the probability of L being bigger than all $q \geq q^*$ is also one. Thus $\mathbb{P}(L \geq q^*) = 1$.

We also know that $\mathbb{P}(L \geq q) = 0$ for all $q \leq q^*$ or that $\mathbb{P}(L \leq q) = 1$ for all $q \leq q^*$. and repeat the above argument to show that $\mathbb{P}(L \leq q^*) = 1$. ■

EXERCISE 13: *Let (X_n) be a sequence of independent real-valued random variables converging almost surely to X. Show that X is almost surely constant.*

EXERCISE 14: *Let (X_n) be a sequence of (not necessarily identically distributed) random variables. This sequence is said to converge completely to X if*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty, \qquad \forall \varepsilon > 0$$

Show that if the X_n are independent then complete convergence is equivalent to almost sure convergence.

9. Integration and Expectation

In this section we will develop the ideas of integration relative to a measure. We have already studied the definitions of the Riemann and Lebesgue integrals a little, so some of this section is revision. The text persists in using strange notation for integration given on the right below, here I will use the more standard notation on the left below:

$$\int_{s\in A} f(s)\mu(ds) := \mu(f;A); \qquad \int f(s)\mu(ds) := \mu(f).$$

We start with a measure space (S,Σ,μ) and let $m\Sigma$ be the set of measurable functions $f : S \to \mathbb{R}$. As $\mathbb{R}$ is endowed with the Lebesgue measure, we can also say $m\Sigma$ is the set of Lebesgue measurable functions. (Revisit Section 7.1 to find the definition of a measurable function — the most important example of a measurable function is a random variable.) Sometimes we only want to consider positive-valued Lebesgue measurable functions we will use $m\Sigma_+$ to denote the set of these.

9.1. *Integrals of Simple Functions*

For some $A \in \Sigma$, let $I_A : S \to \mathbb{R}$ denote the indicator function

$$I_A(s) := \begin{cases} 1, & s \in A; \\ 0, & s \notin A. \end{cases}$$

We begin by defining the integral for this case as being the measure attached to the set A, that is:

$$\int I_A(s)\mu(ds) := \mu(A), \qquad \text{or} \qquad \int I_A \, d\mu := \mu(A).$$

Using linearity this also allows us to write $\int \alpha I_A \, d\mu := \alpha\mu(A)$ for all $\alpha > 0$.

We will define $h \in m\Sigma_+$ to be a simple function if it can be written as a finite weighted sum of indicator functions that is if $h = \sum_{k=1}^m \alpha_k I_{A_k}$, where $\alpha_k > 0$

$$\int h \, d\mu = \int \sum_{k=1}^m \alpha_k I_{A_k} \, d\mu = \sum_{k=1}^m \alpha_k \mu(A_k). \qquad \text{(Simple Functions)}$$

One would need to check that this definition makes sense because there are an infinity of different ways of decomposing a step function like h above into indicator sets. For example, if $A = B \cup C$ and $B \cap C = \emptyset$ then we can write $I_A = I_B + I_C$. Thus for our definition of integration to make sense we would need to check that

$$\int I_A \, d\mu = \int I_B + I_C \, d\mu$$
$$\int I_A \, d\mu = \int I_B \, d\mu + \int I_C \, d\mu$$
$$\mu(A) = \mu(B) + \mu(C)$$

the bottom line is true as the measure is additive.

EXERCISE 15: *Show that if* $f = \sum_{i=1}^{n} a_i I_{A_i} = \sum_{i=1}^{m} b_i I_{B_i}$ *where* $a_i, b_i \geq 0$ *and* A_i, B_i *are measurable*

$$\sum_{i=1}^{n} a_i \mu(A_i) = \sum_{i=1}^{m} b_i \mu(B_i).$$

Now we can define the integral of an arbitrary measurable positive function $f \in m\Sigma_+$. This is something we've already seen several times, it is the supremum of the integral of all simple functions that lie below f

Definition 8 *If* $f \in m\Sigma_+$, *then we define*

$$\int f \, d\mu := \sup_{h \in H(f)} \int h \, d\mu, \qquad \text{where} \qquad H(f) := \{h = \sum_{k=1}^{m} \alpha_k I_{A_k} : f \geq h\}.$$

The difference here is that the measurable functions are not just the intervals of the Riemann integral. It is quite possible that there are points x that are given a measure $\mu(x)$ that are also part of the simple functions in the supremum. That is why we are now able to integrate countable sets, like the rationals, and get the answer zero. A simple but important consequence of this is

EXERCISE 16: *If* $f \in m\Sigma_+$ *and* $\int f \, d\mu = 0$, *show that* $\mu(\{s : f(s) > 0\}) = 0$ *using the fact that* $\{s : f(s) > 0\} = \cup_n \{f(s) : s > 1/n\}$ *so if* $\mu(\{s : f(s) > 0\}) > 0$ *there must exist an* n *such that* $\mu(\{s : f(s) > 1/n\}) > 0$.

This result indicates that a lot of the theory of integration relies on the properties of monotone sequences of sets that are discussed in Result 1. Here is a second monotonicity result that underlies most of the results on integration and expectation.

Result 16 (Monotone Convergence) If (f_n) is a sequence of non-negative measurable functions on the measure space (S, Σ, μ) satisfying $f_n \leq f_{n+1}$ for all n, such that $f_n \to f^o$, then

$$\int f_n \, d\mu \leq \int f_{n+1} \, d\mu \to \int f^o \, d\mu.$$

Proof: First we need to show that $f_n \leq f_{n+1}$ implies $\int f_n \, d\mu \leq \int f_{n+1} \, d\mu$. This follows from the definition above. Any simple function that is bounded above by f_n must also be bounded above by f_{n+1} so the set $H(f_n)$ increases as n increases. This means the supremum cannot go down. As $\int f_n \, d\mu \leq \int f_{n+1} \, d\mu$ we have an increasing sequence of numbers so they have a (possibly infinite) limit. Let $L := \lim_{n\to\infty} \int f_n \, d\mu$. Our job is to show $L = \int f^o \, d\mu$.

First observe that as $f_n \leq f^o$ we know that $\int f_n \, d\mu \leq \int f^o \, d\mu$ for all n (for the same reason as $\int f_n \, d\mu \leq \int f_{n+1} \, d\mu$). This implies that $L \leq \int f^o \, d\mu$, as $\int f^o \, d\mu$ is an upper bound on the sequence $(\int f_n \, d\mu)_n$. Now we will show that $L \geq \int f^o \, d\mu$ and hence equality must hold.

Let h be a non-negative simple function dominated by f^o: $h \in \{\sum_{k=1}^m a_k I_{A_k} : f^o \geq \sum_{k=1}^m a_k I_{A_k}\}$. Let E_n be the places where f_n is bigger than $\alpha \in (0,1)$ times this h:

$$E_n := \{s : f_n(s) \geq \alpha h(s)\}.$$

As $f_{n+1} \geq f_n$ these sets get bigger $E_1 \subset E_2 \subset \dots$. Also as $\alpha < 1$ and f_n converges monotonically to f^o and $s \leq f^o$ these sets must eventually include all of S as $n \to \infty$ ($h \leq f^o$ so $\alpha h < f^o$). Thus $S = \cup_n E_n$.

Now we will prove that

$$\int h \, d\mu = \lim_{n\to\infty} \int_{E_n} h \, d\mu.$$

As h is a simple function $h = \sum_k a_k I_{A_k}$, and we can write

$$\int_{E_n} h \, d\mu = \sum_k a_k \mu(A_k \cap E_n).$$

The sequence $(A_k \cap E_n)_n$ of measurable sets is non-decreasing (as E_n is) and $\cup_n(A_k \cap E_n) = A_k \cap S$ (as $S = \cup E_n$). By Result 1 $\lim_{n\to\infty} \mu(A_k \cap E_n) = \mu(\cup_n(A_k \cap E_n)) = \mu(A_k \cap S)$ for every k. So,

$$\begin{aligned}
\lim_{n\to\infty} \int_{E_n} h \, d\mu &= \lim_{n\to\infty} \sum_k a_k \mu(A_k \cap E_n) \\
&= \sum_k a_k \lim_{n\to\infty} \mu(A_k \cap E_n). \\
&= \sum_k a_k \mu(A_k \cap S) = \sum_k a_k \mu(A_k). \\
&= \int h \, d\mu.
\end{aligned}$$

Now we will use this compare the integral of αh with $L := \lim_{n\to\infty} \int f_n \, d\mu$.

$$\begin{aligned}
\alpha \int h \, d\mu &= \lim_{n\to\infty} \alpha \int_{E_n} h \, d\mu \\
&\leq \limsup_{n\to\infty} \int_{E_n} f_n \, d\mu \\
&\leq \lim_{n\to\infty} \int f_n \, d\mu = L
\end{aligned}$$

(The limsup appears because on E_n $f_n \geq \alpha h$ but we don't know that the limit exists when we only integrate over E_n. However in the next step — increases the domain and we do know that this limit exists.)

Hence we have shown that $\int h \, d\mu \le L/\alpha$ for every simple function h dominated by f^o. But recalling the definition of $\int f^o \, d\mu$

$$\int f^o \, d\mu = \sup_{h \in H(f^o)} \int h \, d\mu \le L/\alpha, \qquad \forall \alpha \in (0,1).$$

As $\alpha \to 1$ we get $\int f^o \, d\mu \le L$ which is what we wanted to show. ■

9.2. *Fatou's Lemmas*

These are a bunch of results about taking limits inside integrals. Suppose that (f_n) is a sequence of non-negative measurable functions on the measure space (S, Σ, μ), then

$$\int \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu.$$

Here's an example of why this is true. Let μ be Lebesgue measure on $S := (0,1]$ and let $f_n(s)$ be the function that is equal to n on the interval $(0, 1/n]$ and is equal to zero on $(1/n, 1]$. Then $\liminf f_n = 0$, but $\int f_n \, d\mu = 1$ for all n, so $\liminf_{n\to\infty} \int f_n \, d\mu = 1$. The proof uses the monotone convergence theorem. Observe that

$$\liminf f_n = \lim_{n\to\infty} \underbrace{\inf_{m \ge n} f_m}_{:= g_n}$$

Observe that the sequence of functions g_n is increasing and converges to $\liminf f_n$, so by Monotone convergence

$$\int g_n \, d\mu \to \int \liminf f_n \, d\mu$$

But for any $m \ge n$ we have $f_m \ge g_n$ and so $\int f_m \, d\mu \ge \int g_n \, d\mu$ for all $m \ge n$. So

$$\liminf_m \int f_m \, d\mu \ge \int g_n \, d\mu \to \int \liminf f_n \, d\mu$$

Which proves the result.

As a corollary to this we have the reverse property provided we assume that there exists a measurable $g \ge f_n$ for all n, with $\int g \, d\mu < \infty$:

$$\int \limsup_{n\to\infty} f_n \, d\mu \ge \limsup_{n\to\infty} \int f_n \, d\mu$$

EXERCISE 17: *Give an example of a sequence of measurable functions where the inequality above is strict.*

9.3. *Integrable Functions and the Spaces $\mathcal{L}^p(S,\Sigma,\mu)$*

We need to define integration for positive and negative functions. For any given measurable function f we do this by separately integrating the positive and negative parts and then taking a difference. First we define the positive and negative parts of a function:

$$f^+(s) := \max\{0, f(s)\}, \qquad f^-(s) := \max\{0, -f(s)\}.$$

(Notice that $f(s) = f^+(s) - f^-(s)$ and that $|f(s)| = f^+(s) + f^-(s)$.) Then as we know how to integrate positive functions, for any arbitrary measurable function we define

$$\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu.$$

This definition immediately implies the inequality $|\int f \, d\mu| \leq \int |f| \, d\mu$.

EXERCISE 18: *Let $(S,\Sigma,\mu) := (\mathbb{N}, 2^{\mathbb{N}}, \nu)$ where $\nu(N) := |N|$ for all $N \subset \mathbb{N}$. (This is called the counting measure.) Show that for any non-negative function $g(n)$*

$$\int_{\mathbb{N}} g(n) \, d\nu = \sum_{n=1}^{\infty} g(n).$$

Let $f(n) := (-1)^n/n$ for all $n \in \mathbb{N}$. Is $\int_{\mathbb{N}} f(n) \, d\nu$ well defined?

Definition 9 *We define the space $\mathcal{L}^p(S,\Sigma,\mu)$ to be the class of all (S,Σ,μ) measurable functions for which the integral $\left|\int |f|^p \, d\mu\right|^{1/p}$ is finite.*

Notice these spaces are much larger than the spaces $\mathcal{C}(0,1)$, because measurability is a much weaker condition than continuity.

EXERCISE 19: *Calculate the $\mathcal{L}^p(S,\Sigma,\mu)$ norm for the functions in the sequence $f_n(s)$ above (the function that is equal to n on the interval $(0,1/n]$ and is equal to zero on $(1/n,1]$). How does p influence these norms as $n \to \infty$?*

Notice also that the sequence of functions in this exercise are not monotone. If they were then we could immediately apply Result 16 to help us. Nor is this sequence dominated, so the following result does not help either!

Now we will mention two very useful but quite straightforward consequences of Monotone Convergence and Fatou. There is a property that is called "Uniform Integrability" which if it holds makes a lot of these results a lot easier to establish.

Result 17 (Dominated Convergence) Suppose that $f_n(s) \to f^o(s)$ pointwise for all $s \in S$ and that there exists $g \in \mathcal{L}^1(S,\Sigma,\mu)$ such that $|f_n| \leq g$, then

$$\int f_n \, d\mu \to \int f^o \, d\mu$$

and $f^o \in \mathcal{L}^1(S, \Sigma, \mu)$.

To prove this observe that $\limsup \int |f^o - f_n|\, d\mu \leq \int \limsup |f^o - f_n|\, d\mu = \int 0 d\mu$, because the functions being integrated here are all less than $2g$ and hence that Fatou's lemma applies. The rest of the proof just requires some rearranging of absolute values.

EXERCISE 20: *Given a measure space (S, Σ, μ) where $\mu(S) < \infty$. Let $\{f_n\}$ be a sequence of integrable functions converging uniformly in S. Then show the limiting function f is integrable, and*

$$\lim_{n\to\infty} \int_S f_n d\mu = \int_S f d\mu.$$

Lemma 1 (Scheffé) *Suppose that (f_n) and f^o are non-negative elements of $\mathcal{L}^1(S, \Sigma, \mu)$ and that $f_n \to f^o$ almost everywhere. Then $\int |f^o - f_n|\, d\mu \to 0$ if and only if $\int f_n\, d\mu \to \int f^o\, d\mu$*

It is the "if" part that is difficult. Suppose $\int f_n\, d\mu \to \int f^o\, d\mu$. Consider the value of the difference in f_n and f^o in the places where $f_n < f^o$, that is the variable $(f_n - f^o)^-$. As $0 \leq (f_n - f^o)^- \leq f^o$ we can apply dominated convergence to get

$$\int (f_n - f^o)^-\, d\mu \to \int 0\, d\mu = 0.$$

Cannot do the same argument with $(f_n - f^o)^+$ but instead we do

$$\int (f_n - f^o)^+\, d\mu = \int_{\{f_n \geq f^o\}} f_n - f^o\, d\mu = \underbrace{\int f_n\, d\mu - \int f^o\, d\mu}_{\to 0} - \int_{\{f_n < f^o\}} f_n - f^o\, d\mu$$

We also have that

$$0 \leq \left| \int_{\{f_n < f^o\}} f_n - f^o\, d\mu \right| \leq \int (f_n - f^o)^-\, d\mu \to 0$$

This proves the result.

The "standard machine" talked about here is a process of establishing a property is true in some space like $\mathcal{L}^1$. The process consists of four steps

1. Show that the property holds for all indicator functions.
2. The use some version of linearity to show that it holds for all simple functions.
3. Use monotone convergence to show that it holds for all non-negative measurable functions.
4. Finally, break the function into positive and negative parts and try to use linearity to show it holds for both of these parts.

9.4. *Radon-Nikodým Theorem*

When you first learn about probability densities, like the Normal, Gamma, etc. you think of them as describing a probability measure. But actually they do something else they transform one measure into another. Consider some, $A \in \mathcal{B}$, a (Borel)-measurable subset of the real line. When you compute the integral

$$\frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2}\, dx,$$

you are actually doing two things: First, you are calculating the probability the $N(0,1)$ random variable X lies in the set A. But second you are taking the Lebesgue measure on $\mathbb{R}$ which gives the measure $\lambda(A)$ to this set A and giving it a different measure that is equal to the integral above. Thus what the density, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, does is it takes the Lebesgue measure on $\mathbb{R}$ and transforms it into another measure on $\mathbb{R}$.

$$\lambda(A) \mapsto \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2}\, dx, \qquad \forall A \in \mathcal{B}$$

You might then reasonably ask—can this always be done? That is, for any two measures μ and λ on the same measurable space (S, Σ) is it possible to find density such that

$$\lambda(F) = \int_F f\, d\mu, \qquad \forall F \in \Sigma.$$

If it is possible to do this, then we write

$$\frac{d\lambda}{d\mu} = f,$$

where the top and bottom of this derivative is a measure! The circumstances under which this is possible are described by Radon-Nikodým Theorem below. We are not going to prove this theorem in this course - but it is something you need to be aware of.

EXERCISE 21: *Given measure space (S, Σ, μ) and $f \in (m\Sigma)^+$. Define*

$$\lambda(A) = \int_A f d\mu, A \in \Sigma.$$

Show (λ) is a measure on (S, Σ).

Result 18 If μ and λ are σ-finite measures on (S, Σ) satisfying the condition

$$\mu(F) = 0 \Rightarrow \lambda(F) = 0$$

then there exists a non-negative measurable function f such that $\frac{d\lambda}{d\mu} = f$.

(σ-finite just requires that S is the countable union of sets given finite measure. The condition in the theorem is sometimes stated as λ is "absolutely continuous" with respect to μ and written as $\mu \gg \lambda$.)

9.5. *Expectation: Definitions and Properties*

When we are not just dealing with a measure space (S, Σ, μ) but a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we call an integral of a random variable an "expectation":

$$E(X) := \int X\, dP \equiv \int X(\omega)\, \mathbb{P}(d\omega), \qquad X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Notice that the assumption $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ is more than enough to ensure the integral exists as it ensures $\int |X|\, dP < \infty$.) We can make different assumptions about the space of functions/random variables $X(\omega)$ that we are taking an expectation over—it does not have to be $\mathcal{L}^1$ but it better be a subset of this if we are going to have unambiguous integrals. Later we will show that $\mathcal{L}^p \subset \mathcal{L}^1$ if $p > 1$ and consider expectations of random variables in this smaller set.

All of the limiting results we have proved above for integrals hold for expectations too. Suppose that (X_n) is a sequence of random variables in $\mathcal{L}^1$ that converge almost surely to X (in $\mathcal{L}^1$) , then

- If $0 \leq X_n \leq X_{n+1} \to X$, then $E(X_n) \leq E(X_{n+1}) \to E(X)$.
- If $X_n \geq 0$ then $E(X) \leq \liminf E(X_n)$
- If $|X_n(\omega)| \leq Y(\omega)$ and $E(Y) < \infty$ then $E(|X_n - X|) \to 0$ and $E(X_n) \to E(X)$.
- If $E(|X_n|) \to E(|X|)$ then $E(|X - X_n|) \to 0$.
- If $|X_n(\omega)| \leq K < \infty$ then $E(|X_n - X|) \to 0$.

EXERCISE 22: *Prove all these.*

9.6. *Expectation: Inequalities*

There are whole books written on probabilistic inequalities. We are going to describe some of the most simple here. Do not think that this list is in anyway complete.

9.6.1. Markov's Inequality and the Chernoff Bound

This can be written in many ways. Let us suppose that $X \in \mathcal{L}^1$, $X \geq 0$, and $a > 0$

$$\frac{E(X)}{a} \geq \mathbb{P}(\{X \geq a\}), \qquad \text{Markov's Inequality}$$

To see this notice that

$$E(X) = \int X\, dP \geq \int_{X \geq a} X\, dP \geq \int_{X \geq a} a\, dP = a\mathbb{P}(\{X \geq a\})$$

(In the book X is replaced by a non-negative function of a random variable, but it's clear this is also a non-negative random variable.)

EXERCISE 23: *Explain why Markov's inequality is false for random variables that can take negative values. Give an example of this.*

Markov's inequality gives an upper bound on the probability that the random variable takes extreme values. This is extremely useful. These kind of events need to be made very rare in many situations. Markov's inequality is the main tool to limit the probability of these "large deviations". The steps in doing this were first developed by Cramer and Chernoff and are now described. First we observe that the event $\{X \geq a\}$ is the same as the event $\{e^{sX} \geq e^{sa}\}$, if $s > 0$. Also observe that by taking an exponential we have ensured that $e^{sX} > 0$ and e^{sa} so we do not need to restrict X or a to take only positive values. This has also introduced an extra parameter s into the statement that we can use. But let us first re-write Markov's inequality in this case for all X and a:

$$\mathbb{P}(\{X \geq a\}) = \mathbb{P}(\{e^{sX} \geq e^{sa}\}) \leq \frac{E(e^{sX})}{e^{sa}} = e^{-sa}E(e^{sX}) = E(e^{s(X-a)}).$$

As s does not appear on the left hand side of this inequality we can choose s to make this upper bound as tight as possible:

$$\mathbb{P}(\{X \geq a\}) \leq \inf_{s \in \mathcal{S}} E(e^{s(X-a)})$$

(Here the set of possible values of $s \in \mathcal{S}$ is chosen so that the expectation exists. Clearly the expectation exists for $s = 0$ but whether non-zero values of s are possible is a technical question.)

Now let consider a particular value of X. Let $Y_1, Y_2, \ldots, Y_n$ be in $\mathcal{L}^\infty$, independent, identically distributed and have a zero expectation $E(Y_i) = 0$. Let $S_n := Y_1 + \cdots + Y_n$ and substitute $X = S_n$ into the above

$$\begin{aligned} \mathbb{P}(\{S_n \geq an\}) &\leq e^{-ans}E(e^{sS_n}) \\ &= e^{-ans}E(\prod_{i=1}^{n} e^{sY_i}) \\ &= e^{-ans}E(e^{sY_i})^n = \left(\frac{E(e^{sY_i})}{e^{as}}\right)^n \end{aligned}$$

Now we introduce a special concept that you may have encountered before in your probability theory. Define the function $M(s) := E(e^{sY_i})$. This is called the Moment Generating Function (mgf) of the random variable Y_i. If we substitute this into the above upper bound we get

$$\mathbb{P}(\{S_n \geq an\}) \leq e^{-n(as - \log M(s))}$$

This holds for all s for which the mgf is defined. Finally define

$$r := \sup_{s>0} as - \log M(s)$$

(For this to exist you need to ensure that the RHS here has a positive derivative in s.) This gives the very important upper bound

$$\mathbb{P}(\{S_n \geq an\}) \leq e^{-nr}. \qquad \text{Chernoff Bound}$$

In fact this is an equality not just an upper bound, but to establish this requires a much more complex argument. Notice that this is a very powerful result. It says that it is exponentially unlikely that S_n grows faster than an if it has a zero mean. This holds for any $a > 0$ (but of course the r varies with a).

EXERCISE 24: *Calculate this upper bound in the case that Y_i is a Bernoulli random variable that takes the value 1 with probability p and zero with probability $1-p$. And hence show that if $a > p$*

$$\mathbb{P}(\{S_n \geq an\}) \leq e^{-nH(a\|p)}$$

where $H(a\|p) := a \ln \frac{a}{p} + (1-a) \ln \frac{1-a}{1-p}$ is the Kulback-Liebler divergence between the $(a, 1-a)$ and $(p, 1-p)$.

Finally we note that the more common Tchebycheff/Chebyshev's inequality is just the Markov inequality applied to the non-negative random variable $(X - E(X))^2$.

EXERCISE 25: *Do this derivation of the Tchebycheff/Chebyshev's inequality,*

$$\mathbb{P}(|X - E(X)| > c) \leq \frac{Var(X)}{c^2}.$$

Make sure you choose the correct $\mathcal{L}^p$ space for your random variables to live in.

The Tchebycheff/Chebyshev's inequality, when applied to sums of independent random variables, says it is unlikely S_n is a long way from $E(S_n)$, but it does not give an exponential bound on this probability. These kind of properties are used in the "Laws of Large Numbers" that we will develop later. Notice the Chernoff bound gives more information about the rates of convergence to the limits that these laws describe.

9.6.2. Jensen's Inequality

Let G be an open interval in $\mathbb{R}$ and let $c : G \to \mathbb{R}$ be a <u>convex</u> function that satisfies

$$c(\lambda x + (1-\lambda)y) \leq \lambda c(x) + (1-\lambda)c(y), \qquad \forall \lambda \in [0,1], x < y \in G.$$

Then if X is a random variable such that: $\mathbb{P}(X \in G) = 1$, $E(|X|) < \infty$, $E(|c(X)|) < \infty$; then

$$E(c(X)) \geq c(E(X)). \qquad \text{Jensen's Inequality}$$

(Notice that to apply Jensen's inequality we need the expectation to exist and a sufficient condition for this is that X and $c(X)$ are in $\mathcal{L}^1$.)

Proof: We know from the analysis course that such a function $c(.)$ is continuous on G. Writing $z = \lambda x + (1-\lambda)y$ in the definition of convexity gives

$$0 \leq \lambda[c(x) - c(z)] + (1-\lambda)[c(y) - c(z)]$$

Then writing $\lambda = (y-z)/(y-x)$ gives

$$\frac{c(z) - c(x)}{z - x} \leq \frac{c(y) - c(z)}{y - z}$$

This says the slope from x to z is less than the slope from z to y. If we now let x increase towards z. The left above increases but is also bounded above, thus the limit below exists

$$\lim_{x \to z} \frac{c(z) - c(x)}{z - x} := D_{-}c(z) \leq \frac{c(y) - c(z)}{y - z} \qquad \forall y > z.$$

A similar argument says that as y approaches z from above there is a decreasing sequence that is bounded below and hence converges, that is

$$D_{-}c(z) \leq D_{+}c(z) =: \lim_{y \to z} \frac{c(y) - c(z)}{y - z}$$

We also know that these functions $D_{-}c(z)$ and $D_{+}c(z)$ are non-decreasing in z. Pick an $m \in [D_{-}c(z), D_{+}c(z)]$, this implies that

$$c(u) \geq c(z) + m(u - z) \qquad \forall u \in G. \tag{1}$$

(Why? The above can be re-written as $\frac{c(u)-c(z)}{u-z} \geq m$ if $u > z$ and $\frac{c(u)-c(z)}{u-z} \leq m$ if $u < z$. If $u > z$ then we know $\frac{c(u)-c(z)}{u-z} \geq D_{+}c(z)$ and $D_{+}c(z) \geq m$. If $u < z$ then we know $\frac{c(u)-c(z)}{u-z} \leq D_{-}c(z)$ and $D_{+}c(z) \leq m$.)

Now choose $z = E(X)$ in the inequality (1).

$$c(u) \geq c(E(X)) + m(u - E(X)).$$

Finally treat $u = X$ and take an expectation to get

$$E(c(X)) \geq c(E(X)) + m(E(X) - E(X)) = c(E(X))$$

■

EXERCISE 26: *Suppose $g(\cdot)$ is a convex function and X is a random variable with finite mean $E[X]$. Then, for any constant c, $E[g(X - E(X) + c)] \geq g(c)$.*

EXERCISE 27: ***Kullback-Leibler Divergence:*** *Given two probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega, \mathcal{F}, \mathbb{Q})$ where $\mathbb{P}$ and $\mathbb{Q}$ are discrete probability distribution satisfying that if $\mathbb{P}(x) = 0$, then $\mathbb{Q}(x) = 0$. The relative entropy from $\mathbb{P}$ to $\mathbb{Q}$ is defined to be*

$$D_{KL}(\mathbb{Q}||\mathbb{P}) = \sum_{x \in \Omega; \mathbb{Q}(x) > 0} \mathbb{Q}(x) \ln \left(\frac{\mathbb{Q}(x)}{\mathbb{P}(x)} \right).$$

Show that this relative entropy is nonnegative.
(Hint: By Jensen's inequality

$$\int_{R_{\mathbb{Q}}} \left[-\ln \left(\frac{\mathbb{P}}{\mathbb{Q}} \right) d\mathbb{Q} \right] \geq -\ln \left[\int_{R_{\mathbb{Q}}} \left(\left(\frac{\mathbb{P}}{\mathbb{Q}} \right) d\mathbb{Q} \right) \right],$$

$R_Q = \{x \in \Omega, \mathbb{Q}(x) > 0\}$.*)*

9.6.3. $\mathcal{L}^p$ Norms and Spaces

The $\mathcal{L}^p$ norm for a random variable in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is defined as

$$\|X\|_p := (E(|X|^p))^{1/p} = \left(\int |X|^p \, d\mathbb{P} \right)^{1/p}$$

We can use Jensen's inequality to show that if $X \in \mathcal{L}^r$ and $1 \leq p \leq r < \infty$ then

$$\|X\|_p \leq \|X\|_r.$$

This says that every element of the space $\mathcal{L}^r$ is automatically in $\mathcal{L}^p$, so we could write $\mathcal{L}^r \subset \mathcal{L}^p$. (However, the complete reverse condition holds for the sequence (Banach) spaces we studied last term, that is, $\ell^p \subset \ell^r$ if $r \geq p$.) The main problem here is we don't immediately know that $X \in \mathcal{L}^p$ so it is necessary to construct something that approximates X that is.

Proof: The main step in the proof is to show that X is actually in $\mathcal{L}^p$ so the other part of this inequality makes sense. We begin by first creating an sequence of artificial variables that are actually in $\mathcal{L}^1$, so we can apply Jensen's inequality to them. Define the bounded and non-negative random variable

$$Z_n := \min\{|X|, n\}^p.$$

As Z_n is bounded, it is in $\mathcal{L}^1$. So too is the random variable $Z_n^{r/p}$. If $r > p$, the function $c(x) = x^{r/p}$ is convex on $(0, \infty)$ so (by Jensen's Inequality)

$$E(Z_n^{r/p}) \geq E(Z_n)^{r/p}.$$

But we can decompose $E(Z_n^{r/p})$

$$E(Z_n^{r/p}) = E(\min\{|X|, n\}^r) \leq E(|X|^r).$$

Thus we have

$$E(Z_n)^{r/p} \leq E(Z_n^{r/p}) \leq E(|X|^r).$$

Raising both sides to the power $1/r$

$$E(Z_n)^{1/p} \leq E(|X|^r)^{1/r}.$$

This holds for all n. Observe that as n increases Z_n increases and is bounded above by $|X|^p$ thus by monotone convergence

$$E(Z_n)^{1/p} \to E(|X|^p)^{1/p} \leq E(|X|^r)^{1/r}.$$

Thus the expectation $E(|X|^p)$ exists and we have also constructed the inequality we wanted. ■

The truncation result argument combined with dominated convergence used in this proof can be used to do many things. Here is an exercise for you to practice using this trick.

EXERCISE 28: *Let $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, show that $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Show that $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ can fail if we only know that $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$*

Now we will show that the space $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ is complete and hence it is almost a Banach Space (as has a vector space space structure). The norm on this space violates one requirement $\|X\|_p = 0$ does not mean that $X = 0$. For example, the random variable $\delta_0 : \mathbb{R} \to \mathbb{R}$ which equals zero on $\mathbb{R} \setminus \{0\}$ and equals unity at zero has a zero integral but does not equal zero.

Proving completeness is something we did not do in the previous part of this course, but now we have established integration we can do this proof. First, recall that completeness requires that any Cauchy sequence in the space converges to a point in the space.

Result 19 $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ is complete.

Proof: Suppose that (X_n) is a Cauchy sequence in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$. That is, for any $\varepsilon > 0$ there exists N such that for all $n, m > N$

$$\|X_n - X_m\|_p < \varepsilon.$$

Choose a sequence $R_N \to \infty$ such that for all $n, m > R_N$

$$\|X_n - X_m\|_p < \frac{1}{2^N}.$$

Now consider the subsequence (X_{R_N}). From above we know that

$$E(|X_{R_N} - X_{R_{N+1}}|) = \|X_{R_N} - X_{R_{N+1}}\|_1 \leq \|X_{R_N} - X_{R_{N+1}}\|_p < \frac{1}{2^N}.$$

Hence we know that

$$E(\sum_N |X_{R_N} - X_{R_{N+1}}|) < \infty.$$

Hence this sum can only be infinite on a zero probability event. Thus the sum $\sum_{N=1}^{M} |X_{R_N} - X_{R_{N+1}}|$ converges almost surely. So the sum $\sum_{N=1}^{M}(X_{R_N} - X_{R_{N+1}})$ also converges almost surely. But

$$\sum_{N=1}^{M}(X_{R_N} - X_{R_{N+1}}) = X_{R_0} - X_{R_{M+1}},$$

so (X_{R_N}) converges almost surely.

Let $X^o(\omega) := \limsup_N X_{R_N}$. This is $\mathcal{F}$ measurable, from our earlier results on lim sup's. We now want to show that our Cauchy sequence converges to this. If $r > R_N$ and $M \geq N$ we know that

$$E(|X_r - X_{R_M}|^p) = (\|X_r - X_{R_M}\|_p)^p \leq \frac{1}{2^{Np}}.$$

Letting $M \to 0$ and using Fatou's lemma we get

$$E(|X_r - X^o|^p) \leq \frac{1}{2^{Np}}.$$

This says that X_o is in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$. As we can also let $N \to \infty$ this also says that $X_r \to X^o$ in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$. ■

9.6.4. Hölder Inequality

A final result on $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ is a famous generalization of the Cauchy-Schwartz inequality. Suppose that we have two different spaces $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{L}^q(\Omega, \mathcal{F}, \mathbb{P})$ where

$$\frac{1}{p} + \frac{1}{q} = 1, \qquad \Leftrightarrow \qquad p + q = pq.$$

Then the inequality states that if $f \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ and if $g \in \mathcal{L}^q(\Omega, \mathcal{F}, \mathbb{P})$, then $fg \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and

$$\|fg\|_1 \leq \int |fg|\, d\mu \leq \left(\int |f|^p\, d\mu\right)^{\frac{1}{p}} \left(\int |f|^q\, d\mu\right)^{\frac{1}{q}} = \|f\|_p \|g\|_q$$

Proof: We know that when $f = 0$ almost surely the above is always true, so we suppose that there is a set of positive measure where $f \neq 0$. Now we define a new measure that is defined using f,

$$\nu(A) := \frac{\int_A |f|^p\, d\mu}{\int |f|^p\, d\mu} := \frac{\int_A |f|^p\, d\mu}{B}$$

The normalization is justified as $\int |f|\,d\mu > 0$ as there is a set of positive measure where $f \neq 0$. We can take expectations with respect to this measure.

We now introduce a random variable

$$z(\omega) := \begin{cases} 0, & f(\omega) = 0; \\ \frac{|g(\omega)|}{|f(\omega)|^{p-1}}, & f(\omega) \neq 0. \end{cases}$$

Finally we state Jensen's Inequality for this random variable relative the measure ν. And substitute in for the above definitions. Let $\Omega' \subset \Omega$ be the set of states where $f \neq 0$

$$\begin{aligned}
\int_{\Omega'} z^q \, d\nu &\geq \left(\int_{\Omega'} z \, d\nu \right)^q \\
\int_{\Omega'} \left(\frac{|g|}{|f|^{p-1}} \right)^q \frac{|f|^p \, d\mu}{B} &\geq \left(\int_{\Omega'} \frac{|g|}{|f|^{p-1}} \frac{|f|^p \, d\mu}{B} \right)^q \\
\int_{\Omega'} \frac{|g|^q |f|^{p+q-pq} \, d\mu}{B} &\geq \left(\int \frac{|f||g| \, d\mu}{B} \right)^q \\
B^{q-1} \int_{\Omega'} |g|^q 1 \, d\mu &\geq \left(\int |f||g| \, d\mu \right)^q \\
B^{\frac{q-1}{q}} \left(\int_{\Omega'} |g|^q \, d\mu \right)^{\frac{1}{q}} &\geq \int |fg| \, d\mu \\
\left(\int |f|^p \, d\mu \right)^{\frac{1}{p}} \left(\int_{\Omega'} |g|^q \, d\mu \right)^{\frac{1}{q}} &\geq \int |fg| \, d\mu \qquad \frac{1}{p} = \frac{q-1}{q} \\
\|f\|_p \|g\|_q &\geq \|fg\|_1
\end{aligned}$$

■

EXERCISE 29: *Given a measure space (S, Σ, μ) where $\mu(S) < \infty$. Show that for $0 < r < p$,*

$$\|f\|_{L^r((S,\Sigma,\mu))} \leq \mu(S)^{\frac{p-r}{pr}} \|f\|_{L^p((S,\Sigma,\mu))}$$

which implies $L^p((S, \Sigma, \mu)) \subset L^r((S, \Sigma, \mu))$ if $\mu(S) < \infty$.

EXERCISE 30: *A random variable X with mean ν_0 is called Sub-Gaussian with parameter σ (SG(σ)) if:*

$$E(e^{\lambda(X-\nu_0)}) \leq e^{\frac{\lambda^2 \sigma^2}{2}}, \quad \forall \lambda \in \mathbb{R}.$$

Assume two random variables X_1 and X_2 defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $X_1 \in SG(\sigma_1)$ with mean ν_1 and $X_2 \in SG(\sigma_2)$ with mean ν_2 . Show that $X_1 + X_2 \in SG(\sigma_0)$ for some $\sigma_0 > 0$.

9.6.5. Inequalities in $\mathcal{L}^2$

This is the space of random variables where $E(|X|^2)$ exists, so it is not entirely surprising that we are thinking about the Cauchy-Schwartz inequality. We have already proved this once in the course, but again it is necessary to check that all the relevant random variables are in the right spaces. You have already completed the main new step in proving this in Exercise 9.6.3 by constructing a sequence Z_n of truncated variables in $\mathcal{L}^1$. (We will not repeat that here.)

Result 20 (Cauchy-Schwartz) If $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ then $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and

$$E(|XY|) \le \|X\|_2 \|Y\|_2 = \left(\int |X|^2 \, d\mathbb{P} \right)^{1/2} \left(\int |Y|^2 \, d\mathbb{P} \right)^{1/2}$$

The space $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space (we have not mentioned it is compete yet)! In this space we can define an inner product $\langle X, Y \rangle$ and the above result says that it exists for $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ because

$$\langle X, Y \rangle := E(XY) \le E(|XY|) \le \|X\|_2 \|Y\|_2$$

We also get a lot of familiar statistical properties like:

$$E[(X+Y)^2] = E(X^2) + E(Y^2) \Leftrightarrow E(XY) = 0$$

$$\text{Cov}(X, Y)^2 \le \text{Var}(X)\text{Var}(Y)$$

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \Leftrightarrow E(XY) = 0$$

10. Laws of Large Numbers

10.1. Strong Laws of Large Numbers

We will begin with the strong law of large numbers, which says (roughly) that the average of a sequence of independent random variables, with zero mean, finite fourth moment, not necessarily identically distributed converges almost surely to zero. This is the strongest mode of convergence we have, hence we call it "strong". What it says is that the averages of the sequence, which is itself a sequence of random variables, converges to zero for almost every history. Not that with high probability it is close to zero.

Result 21 (Strong Law of Large Numbers) Suppose that (X_n) is a sequence of independent random variables such that: $E(X_n) = 0$ for all n, $(X_n) \subset \mathcal{L}^4(\Omega, \mathcal{F}, \mathbb{P})$, there exists $K > 0$ such that $\forall n \; E(|X|_n^4) < K$, then

$$Y_n := \frac{1}{n} \sum_{i=1}^{n} X_i \to 0, \qquad \mathbb{P}\text{-almost surely.}$$

Proof: First we use independence to argue that all fourth moments that contain odd numbers of powers are equal to zero

$$0 = 0E(X_j^3) = E(X_i X_j^3), \qquad 0 = 0.0E(X_j^2) = E(X_i X_k X_j^2), \qquad 0 = 0.0.0.0 = E(X_i X_k X_j X_h),$$

where $i \neq j \neq k \neq h$. Hence if we compute the expectation of Y_n^4 we get

$$E(Y_n^4) = \frac{1}{n^4}\sum_{i=1}^{n} E(X_i^4) + \frac{1}{n^4}\binom{4}{2}\sum_{i<j\leq n} E(X_i^2 X_j^2)$$

We have assumed that $E(X_i^4) < K$, but if we write this as $E\left((X_i^2)^2\right) < K$ then take the second square outside we have by Jensen's inequality $K > E\left((X_i^2)^2\right) \geq E\left(X_i^2\right)^2$. Also, by independence $E(X_i^2 X_j^2) = E(X_i^2)E(X_j^2)$. If the upper bounds on $E(X_i^4)$ and $K^{1/2} > E(X_i^2)$ are substituted into $E(Y_n^4)$ we get

$$E(Y_n^4) < \frac{K}{n^3} + \frac{6K}{n^4}\frac{1}{2}n(n-1) < \frac{7K}{n^2}$$

Thus the sequence of non-negative random variables Y_n^4 have an expectation that converges to zero. Now we just use Markov's inequality $\Pr(Y_n^4 > a) < E(Y_n^4)/a$. Choosing $a = \varepsilon^2$ and substituting from above gives

$$\Pr(Y_n^2 > \varepsilon) = \Pr(Y_n^4 > \varepsilon^2) < \frac{E(Y_n^4)}{\varepsilon^2} < \frac{7K}{\varepsilon^2 n^2}$$

The probabilities on the right are finite when summed over n. So by the first Borel-Cantelli property, for any $\varepsilon > 0$ we have that $Y_n^2 > \varepsilon$ only a finite number of times. Thus $\limsup Y_n^2 < \varepsilon$ with probability one. Taking the intersection of the countable sequence of probability one events $\limsup Y_n^2 < \frac{\varepsilon}{2^k}$ implies that $\limsup Y_n^2 = 0$. ∎

EXERCISE 31: *Why does* $\limsup Y_n^2 = 0$ *almost surely imply* $Y_n \to 0$ *almost surely?*

This version of the SLLN requires that the fourth moment exists, but does not require identically distributed X_n. There are many other versions of this law with different requirements. Here's another

Result 22 (Another SLLN Result) Suppose that (X_n) is a sequence of independent and identically distributed random variables such that: $E(X_n) = 0$ for all n and $(X_n) \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, then

$$Y_n := \frac{1}{n}\sum_{i=1}^{n} X_i \to 0, \qquad \mathbb{P}\text{-almost surely.}$$

10.2. The Weak Laws of Large Numbers in $\mathcal{L}^2$ and $\mathcal{L}^1$

If we have a sequence (X_n) of random variables in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ we know that $E(X_n^2)$ exists so we can define their variance $\mathrm{Var}(X_n) := E((X - E(X))^2)$. Armed with this we can use Chebyshev's inequality (that you proved earlier) to get the "Weak Law of Large Numbers" or (WLLN). Why is it called "weak"? The reason is the kind of convergence that the result describes. If you think of Y_n as a sequence of random variables and μ as another (constant) random variable, then this result says the random variable Y_n is close to μ on a lot of the states. (It is like measuring the distance between two functions by integration.) This kind of convergence is called "convergence in probability".

Definition 10 *The sequence of random variables (X_n) are said to converge in probability to X, $X_n \to_P X$ if* $\lim_{n\to\infty} \Pr(|X - X_n| \geq \varepsilon) = 0$ *for all $\varepsilon > 0$.*

EXERCISE 32: *Does convergence in probability imply almost sure convergence? Does almost sure convergence imply convergence in probability? Justify your answers with examples.*

Result 23 (WLLN.1) If $(X_n) \subset \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ is a sequence of independent random variables that satisfies $E(X_n) = \mu$ and $\mathrm{Var}(X_n) \leq \sigma^2$ for all $n \in \mathbb{N}$. If $Y_n := \frac{1}{n}\sum_{i=1}^n X_i$, then for every $\varepsilon > 0$ and $n \in \mathbb{N}$

$$\mathbb{P}(|Y_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

so $\lim_{n\to\infty} \mathbb{P}(|Y_n - \mu| \geq \varepsilon) = 0$.

EXERCISE 33: *Do this proof. Apply Chebychev's inequality to the random variable Y_n. It has the mean μ and*

$$\mathrm{Var}(Y_n) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) \leq \frac{\sigma^2}{n}.$$

Here's a little more challenging version of the result that does not use Chebychev.

Result 24 (WLLN.2) If $(X_n) \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ is a sequence of independent and identically distributed random variables that satisfies $E(X_n) = \mu$ for all $n \in \mathbb{N}$. If $Y_n := \frac{1}{n}\sum_{i=1}^n X_i$, then for every $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - \mu| \geq \varepsilon) = 0.$$

Proof: We will do the proof for the case $\mu = 0$. Fix $\varepsilon > 0$ and $\delta > 0$. Now we for $m \in \mathbb{N}$ we define two censored random variables X_n^{+m} to be the places where $|X_n| > m$ and X_n^{-m} to be the places where $|X_n| \leq m$, that is:

$$X_n^{+m} := \mathbb{1}_{\{|X_n|>m\}} \times X_n, \qquad X_n^{-m} := X_n \times \mathbb{1}_{\{|X_n|\leq m\}}.$$

Both of the sequences of random variables $(|X_n^{+m}|)_{m=1}^{\infty}$ and $(X_n^{-m})_{m=1}^{\infty}$ are dominated by X_n. By Result 17 we have that $E(X_n^{-m}) \to \mu = 0$ and $E(|X_n^{+m}|) \to 0$ as $m \to \infty$. Now we can use these new random variables to re-write Y_n:

$$Y_n := \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} X_i^{+m} + \frac{1}{n}\sum_{i=1}^{n} X_i^{-m} := Y_n^{+(m)} + Y_n^{-(m)}$$

(Notice that the $Y_n^{+(m)}$ and $Y_n^{-(m)}$ are not censored random variables.) We will decompose $|Y_n - 0| = |Y_n - \mu|$ as below.

$$|Y_n - 0| \leq \underbrace{|E(X_1^{-m}) - 0|}_{E(X_n^{-m})\to 0} + \underbrace{|Y_n^{-(m)} - E(X_1^{-m})|}_{WLLN1} + |Y_n^{+(m)}| \tag{2}$$

We will now bound the terms in this expression. We will deal with the first and last term. Choose m so large that $E(|X_n^{+m}|) < \varepsilon\delta/4$, which is possible as this expectation converges to zero as m increases. And m so large that $|E(X_1^{-m}) - 0| < \varepsilon/4$ which is possible as $E(X_1^{-m}) \to 0$. Then we have

$$E(|Y_n^{+(m)}|) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^{+m}) \leq \frac{1}{n}\sum_{i=1}^{n} E(|X_i^{+m}|) \leq \frac{\varepsilon\delta}{4}.$$

Applying Markov's inequality then gives

$$\mathbb{P}\left(|Y_n^{+(m)}| > \frac{\varepsilon}{2}\right) < \frac{\delta}{2}.$$

To bound the middle term notice that the X_i^{-m} variables are bounded so they are in $\mathcal{L}^2$. Thus for any value of m the previous WLLN applies. Thus there exists an N' such that for all $n > N'$

$$\mathbb{P}\left(\ |Y_n^{-m} - E(X_i^{-m})| \geq \frac{\varepsilon}{4} \ \right) < \frac{\delta}{2}$$

If (2) is bigger than ε then the last two terms have got to contribute at least $(3/4)\varepsilon$. needs to be bigger than $\varepsilon/2$ if they sum to something bigger than ε. This can occur with probability at most δ. This now completes the proof. ∎

EXERCISE 34: *Fill in the details of this last step in the proof.*

EXERCISE 35: *Let (F_n) be a sequence of independent events. Let $\bar{p}_n := \frac{1}{n}\sum_{m=1}^{n}\mathbb{P}(F_m)$ and let X_n be a random variable that counts the number of the events $F_1, \ldots, F_n$ have occurred. Show that*

$$n^{-1}X_n - \bar{p}_n \to_P 0.$$

EXERCISE 36: *Suppose that the sequence of random variables (X_n) is m-dependent. That is, terms more than m apart are independent of each other. Suppose that $E(X_n) = 0$ for all n and that there exists an $M > 0$ such that $|X_n| < M$ for all n. Show that $S_n/n \to 0$, where $S_n := X_1 + \cdots + X_n$, with probability one.*

An important consequence of the weak law is the result that a distribution on $[0, \infty)$ is uniquely determined by its moment generating function. We will state this result without a proof.

Result 25 If μ and ν are probability measures on $[0, \infty)$ and

$$\int_0^\infty e^{-sx}\,\mu(dx) = \int_0^\infty e^{-sx}\,\nu(dx), \qquad \forall s \geq s_0 \geq 0$$

then $\mu = \nu$.

10.3. Maximal Inequalities

In this section we report one other kind of result on sequences of random variables that are also important and seem naturally to fit in this part of the notes.

Let (X_n) be a sequence of independent, zero-mean, random variables. By adding these together we can define a random walk:

$$S_n := X_1 + \cdots + X_n.$$

What the first inequality result below does, is consider this random walk over the first n periods. It takes the values $S_1, S_2, \ldots, S_n$. It asks the question what is the probability that at one of these places the random walk leaves the interval $(-\alpha, \alpha)$? It shows that this probability is big only if S_n has a large variance.

Result 26 If (X_n) is a sequence of independent random variables with zero mean and finite variance, then for all $\alpha > 0$

$$\mathbb{P}\left(\max_{1\leq m\leq n} |S_m| \geq \alpha\right) \leq \frac{\mathrm{Var}(S_n)}{\alpha^2}$$

Proof: Define the event A_k to be the event that $|S_m| < \alpha$ for all $m < k$ but $|S_k| \geq \alpha$. So the sequence is in A_k if it first leaves the interval $(-\alpha, \alpha)$ at time k. Notice these events are disjoint provide a partition of the event $\max_{1\leq m\leq n} |S_m| \geq \alpha$.

Now we do a simple calculation

$$\begin{aligned} E(S_n^2) &\geq \sum_{m \leq n} \int_{A_m} S_n^2 \, d\mathbb{P} \\ &= \sum_{m \leq n} \int_{A_m} \left(S_m^2 + 2S_m(S_n - S_m) + (S_n - S_m)^2 \right) d\mathbb{P} \\ &= \sum_{m \leq n} \int_{A_m} \left(S_m^2 + 2S_m(S_n - S_m) \right) d\mathbb{P} \end{aligned}$$

Consider the second term in this integral $\int_{A_m} S_m(S_n - S_m) \, d\mathbb{P}$ on the event A_m the term S_m is fixed and only S_n varies. As the increments of the random walk have zero mean on average $S_n = S_m$ on the event A_m. Thus this second term in the integral is zero and we have

$$E(S_n^2) \geq \sum_{m \leq n} \int_{A_m} S_m^2 \, d\mathbb{P} \geq \alpha^2 \sum_{m \leq n} \mathbb{P}(A_m) = \alpha^2 \mathbb{P} \left(\max_{1 \leq m \leq n} |S_m| \geq \alpha \right)$$

S_n has a zero expectation, so $\text{Var}(S_n) = E(S_n^2)$ and we are done. ■

EXERCISE 37: *Calculate this upper bound for the cases: (1) X_n is a random variable that takes values $+1$ and -1 with probability $1/2$, (2) X_n has a standard normal, $N(0, 1)$ distribution.*

11. PRODUCT MEASURES AND FUBINI'S THEOREM

This is a short technical section. It states the results required when we want to build up a measure on a sequence space from the measures on the elements of the sequence. I think the explanation in **?** of this material is clearer than the text. Here's a simple idea of the problem. Suppose you have two spaces X and Y and you have three σ-algebras: $\mathcal{X}$ on X, $\mathcal{Y}$ on Y, and $\mathcal{Z}$ on the product space $X \times Y$. This gives us three different measurable spaces: $(X, \mathcal{X})$, $(Y, \mathcal{Y})$, and $(X \times Y, \mathcal{Z})$. When is it the case that we can think of the σ-algebra $\mathcal{Z}$ as being consistent with the σ-algebras $\mathcal{X}$ and $\mathcal{Y}$?

What we could try to do is to use the σ-algebras $\mathcal{X}$ and $\mathcal{Y}$ to build a σ-algebra on the product space $X \times Y$ and then compare it with $\mathcal{Z}$. To build the new σ-algebra you could proceed as follows: First pick a set $A \in \mathcal{X}$ and a set $B \in \mathcal{Y}$ and then create the "measurable rectangle" $A \times B \subset X \times Y$. (Notice that this isn't really a rectangle unless A and B are both intervals, which is certainly possible if $\mathcal{X}$ and $\mathcal{Y}$ are Borel.) These measurable rectangles form a π-system.

EXERCISE 38: *Prove that the measurable rectangles form a π system.*

Then we can define the σ-algebra $\mathcal{X} \times \mathcal{Y}$ on $X \times Y$ to be the σ-algebra that is generated by the measurable rectangles. This set is much larger than just the rectangles.

Result 27 If $F \in \mathcal{X} \times \mathcal{Y}$ then for all $x \in X$ the <u>section</u> set $\{y : (x,y) \in F\}$ is in $\mathcal{Y}$. And, for each y the <u>section</u> set $\{x : (x,y) \in F\}$ is in $\mathcal{X}$. If f is measurable with respect to $\mathcal{X} \times \mathcal{Y}$ then for each x the function $f(x,.)$ is measurable with respect to $\mathcal{Y}$ and the function $f(.,y)$ is measurable with respect to $\mathcal{X}$

11.1. Product Measures

We have $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ as two measurable spaces and now we are going to endow them with σ-finite measures μ and ν. So we have measure spaces $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$.

What we are going to do is to show that we can use these set functions $\mu : \mathcal{X} \to \mathbb{R}$ and $\nu : \mathcal{Y} \to \mathbb{R}$ to create a new set-valued map $\pi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ on the product σ-algebra. It is easy to see that if $A \times B$ is a measurable rectangle then we can define

$$\pi(A \times B) := \mu(A)\nu(B).$$

What is difficult is that we will need to define π for sets that are not measurable rectangles, so we will need to be able to somehow generalize this formula to arbitrary sets $F \in \mathcal{X} \times \mathcal{Y}$. The previous result helps here. It says that if $F \in \mathcal{X} \times \mathcal{Y}$ then we can slice it up into "sections" $\{y : (x,y) \in F\}$ one for each $x \in F$, or into sections $\{x : (x,y) \in F\}$ one for each $y \in F$. Each of these slices are measurable in the appropriate space. So if we can take the measure given to these sections and average them some how, then we will be able to give a measure to an arbitrary $F \in \mathcal{X} \times \mathcal{Y}$. But the big question is, can we do that in a consistent way? That is if we average over the X sections we will get one answer and if we average over the Y sections it might be possible that we get a different answer.

Here is an example of what might go wrong. Suppose that $X = Y = [0,1]$. But let the measure μ on X be the Lebesgue measure and let the measure ν on Y be the counting measure (that just counts the number of elements in the set). Consider the diagonal set $D = \{(x,x) \in [0,1]^2 : 0 \le x \le 1\}$. If we take the ν measure of the section $\{y : (x,y) \in D\}$ we get $\nu(\{y : (x,y) \in D\}) = 1$ because there is one point on each section — the diagonal. Then if we integrate these slices $\nu(\{y : (x,y) \in D\})$ for all $x \in [0,1]$ over Lebesgue measure we integrate 1 over $[0,1]$ and get the answer that the product measure should give measure 1 to the diagonal set D. However, if we take the μ measure of the section $\{x : (x,y) \in D\}$ we get $\mu(\{x : (x,y) \in D\}) = 0$ because there is one point on each diagonal and this has Lebesgue measure zero. Thus every y-slice of D is given measure zero. Adding these up using the counting measure we still get the answer zero. Thus using this approach, the product measure should give measure 0 to the diagonal set D!

What goes wrong with the above example is that the measures are not σ-finite. The first result below says that we can add up slices and get consistent way of describing a product measure provided the measures are σ-finite.

Result 28 If $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ are σ-finite measure spaces then we can define a σ-finite measure $\pi = \mu \times \nu$ on $(X \times Y, \mathcal{X} \times \mathcal{Y})$ as follows

$$\pi(F) := \int_X \nu(\{y : (x, y) \in F\})\, d\mu, \qquad \text{and} \qquad \pi(F) := \int_Y \mu(\{y : (x, y) \in F\})\, d\nu.$$

if it is the unique measure for which $\pi(A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{X}$ and $B \in \mathcal{Y}$.

Instead of giving measures to sets, which are integrals of indicator functions, Fubini's theorem talks about when we can integrate functions, slice by slice and get a consistent answer.

Result 29 (Fubini) If $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ are σ-finite, then the set function $\pi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is called the product measure of $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ and for any bounded non-negative $\mathcal{X} \times \mathcal{Y}$ measurable function f

$$\int f\, d\pi = \int I_f(x)\, d\mu = \int J_f(y)\, d\nu$$

where

$$I_f(x) := \int_Y f(x, y) d\nu, \qquad I_f(y) := \int_X f(x, y) d\mu,$$

11.2. *Sequences of Random Variables and Stochastic Processes*

This kind of approach can be generalized to any countable product of the measurable spaces $\{(S_n, \Sigma_n)\}_{n=1}^{\infty}$. That is, we could take a countable intersection to get a simple rectangle $F_1 \times F_2 \times \ldots$. Thus for stochastic processes in discrete time, the analysis above is all that is required.

However, when we want to consider continuous-time stochastic processes more is required. The above works to describe measures on a finite collections of time periods, but what we need to do is to find some way of extending these distributions on finite numbers of time periods to an uncountable product of measurable spaces. A more complicated argument is necessary—this is called Kolmogorov's Existence/Extension Theorem. To be able to do this it is necessary that the consistency conditions on the probabilities are placed on the finite dimensional rectangles.

12. Conditional Expectation

We are used to thinking about conditional expectations as a way of predicting the value of a random variable when we already have partial information on it. Thus we tend to think of conditional expectations as point estimates that use the conditional probabilities (derived from Bayes' rule) to calculate an expectation. In virtually all cases when dealing with non-zero probability events this is correct. However, this is not how conditional expectations are defined.

Formally conditional expectations are *not* defined like this. In fact they are defined as random variables that satisfy certain measurability conditions. This may seem very strange at first, and takes a bit of getting used to. Let us begin with a simple two dimensional example of how this works. Suppose we have a two-dimensional set of states of the world $\omega \in \{1,2,3\}^2 := \Omega$ and two random variables (X, Y) defined on these states. These random variables are are measurable with respect to sub-σ-algebras of Ω where X and Y can each take 3 values. (We use the rows and the columns of the table below to describe these sub-σ algebras.) The probabilities $\mathbb{P}$ of the states of the world $\omega \in \{1,2,3\}^2 := \Omega$ are given in the table below

	$X=1$	$X=2$	$X=3$
$Y=1$	$\frac{1}{7}$	$\frac{1}{7}$	0
$Y=2$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$
$Y=3$	0	$\frac{1}{7}$	$\frac{1}{7}$

If we wanted to calculate the conditional expectation of Y given X we would do the three calculations:

$$E(Y|X=1) = \frac{1+2}{2} = 1.5, \qquad E(Y|X=2) = \frac{1+2+3}{3} = 2, \qquad E(Y|X=3) = \frac{3+2}{2} = 2.5.$$

Now suppose we use these numbers to define a new random variable, called Z, on $\omega \in \{1,2,3\}^2 := \Omega$ as follows

Z	$X=1$	$X=2$	$X=3$
$Y=1$	1.5	2	2.5
$Y=2$	1.5	2	2.5
$Y=3$	1.5	2	2.5

Notice several things: (1) This random variable is defined for every state of the world, whereas the conditional expectations we first calculated are only defined on a sub-σ-algebra — the X values. (2) This new random variable is still measurable with respect to the σ-algebra generated by the values of X, this is because it is constant on states that have the same X value or are in the same set as the σ algebra induced by X. (3) If $\mathcal{G}$ is the σ-algebra generated by X then taking expectations over Z and of X on the sets in $\mathcal{G}$ gives the same answer.

$$\int_G Z\, d\mathbb{P} = \int_G Y\, d\mathbb{P}, \qquad \forall G \in \mathcal{G}.$$

For an example of this last statement, suppose that G is the event that $X \leq 2$. Then

$$\int_G Z\, d\mathbb{P} = \frac{1}{7}(1.5+1.5) + \frac{1}{7}(2+2+2)$$

and
$$\int_G Y \, d\mathbb{P} = \frac{1}{7}(1+2) + \frac{1}{7}(1+2+3)$$
Thus the random variable Z integrates up to give the conditional expectation.

Hence we will now define conditional expectations as taken with respect to σ algebras and we will think of them as being random variables.

Definition 11 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ and let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $E(|X|) < \infty$. Suppose there is a random variable Z such that*

1. *Z is $\mathcal{G}$ measurable,*
2. *$E(|Z|) < \infty$,*
3. $$\int_G X \, d\mathbb{P} = \int_G Z \, d\mathbb{P}, \qquad \forall G \in \mathcal{G};$$

then Z is called a version of the conditional expectation of X given $\mathcal{G}$ and is written as $E(X|\mathcal{G})$.

Thus the conditional expectation $E(X|\mathcal{G})$ is a random variable and for each $\omega \in \Omega$ we have that $E(X|\mathcal{G})(\omega) = Z(\omega)$.

EXERCISE 39: *Let the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given. The event $G \in \mathcal{F}$ is used to define a simple sub σ-algebra $\mathcal{G}' = \{\emptyset, \Omega, G, G^c\}$. And let X be a random variable in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Describe the random variable $E(X|\mathcal{G}')$ in the two cases: $\mathbb{P}(G) > 0$ and $\mathbb{P}(G) = 0$.*

EXERCISE 40: *Let the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given. And let X be a random variable in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that we have two nested sub-σ-algebras $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$. Show that*
$$E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{G}), \qquad E(E(X|\mathcal{H})|\mathcal{G}) = E(X|\mathcal{G}).$$

The main result we have in this section is that

Result 30 The conditional expectation of $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ given $\mathcal{G}$, or $E(X|\mathcal{G})$ exists and is unique up to a set of zero measure.

(There may be many versions of a conditional expectation but these different random variables are almost surely equal.)

Proof: Step 1 Almost sure uniqueness: Suppose that $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and that Y and $\tilde{Y}$ are both versions of $E(X|\mathcal{G})$. Then integrating over all Ω gives that $Y, \tilde{Y} \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and we know that

$$\int_G (Y - \tilde{Y})\, d\mathbb{P} = 0, \qquad \forall G \in \mathcal{G}.$$

If Y is not almost surely equal to $\tilde{Y}$ there must exist a positive probability set where they differ, so suppose $\mathbb{P}(Y > \tilde{Y}) > 0$. The sequence of events $\{\omega : Y > \tilde{Y} + 1/n\}$ increases and converges to the event $Y > \tilde{Y}$, so one of the events in this sequence has positive probability. But the set $Y - \tilde{Y} > n^{-1}$ is in $\mathcal{G}$. Thus we have

$$\int_{\{Y - \tilde{Y} > n^{-1}\}} Y - \tilde{Y}\, d\mathbb{P} = 0$$

from the definition of conditional expectations, but also

$$\int_{\{Y - \tilde{Y} > n^{-1}\}} Y - \tilde{Y}\, d\mathbb{P} > n^{-1}\mathbb{P}(Y - \tilde{Y} > n^{-1}) > 0$$

by construction — a contradiction.

Step 2 Existence of the Random Variable $E(X|\mathcal{G})$ if $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$: First we define the subset of $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ of $\mathcal{G}$-measurable functions, that is, $\mathcal{K} := \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. This is a closed linear subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. $\mathcal{K}$ is complete in the $\mathcal{L}^2$ norm. This means we can find a random variable Y in $\mathcal{K}$ that is closest to X and is orthogonal to it:

$$E((X - Y)^2) = \inf\{E((X - Z)^2) : Z \in \mathcal{K}\}$$

and

$$\int (X - Y)Z\, d\mathbb{P} = 0, \qquad \forall Z \in \mathcal{K}.$$

(This is a result we omitted when we did expectations.) Suppose that $Z = \mathbb{1}_G$, that is it is restricted to the set of indicator functions for some set $G \in \mathcal{G}$. Then the above says

$$\int_G X - Y\, d\mathbb{P} = 0, \qquad \forall G \in \mathcal{G}.$$

Thus this Y satisfies the properties required of $E(X|\mathcal{G})$.

Step 3 Extending this to $\mathcal{L}^1$. First you split X into its positive and negative parts $X = X^+ - X^-$. Then you approximate these parts by bounded random variables that are therefore also in $\mathcal{L}^2$ and apply the above results using monotone convergence.

■

EXERCISE 41: *If $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$ and $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. If $\int_G X\, d\mathbb{P} = \int_G Y\, d\mathbb{P}$ for every G in a π-system, containing Ω, that generates $\mathcal{G}$ then $\int_G X\, d\mathbb{P} = \int_G Y\, d\mathbb{P}$ for every $G \in \mathcal{G}$.*

12.1. A Problem and its Resolution

If we are dealing with random variables that take finite (or countable) numbers of values and there are no zero probability events then the above is all you need to know about conditional expectations.

To understand the problem it is easiest to think about a continuous bi-variate distribution $(x, y) \in [0, 1]^2$ with the pdf $f(x, y)$. Suppose that we wanted to think about the conditional distributions of y given some information on x. Then applying property (3) of Definition 11 to the event $x \le a$ requires that

$$\int_{\{x \le a\}} Z f(x, y)\, dx\, dy = \int_{\{x \le a\}} y f(x, y)\, dx\, dy$$

Or substituting $E(y|x \le a)$ for Z in the above

$$\int_{\{x \le a\}} E(y|x \le a) f(x, y)\, dx\, dy = \int_{\{x \le a\}} y f(x, y)\, dx\, dy$$

We get the formula that we usually use:

$$E(y|x \le a) = \frac{\int_{\{x \le a\}} y f(x, y)\, dx\, dy}{\int_{\{x \le a\}} f(x, y)\, dx\, dy}$$

(Here I have done a bit of a cheat and instead of treating $E(y|x \le a)$ as a random variable, I have just made it a number that I am able to take outside the integral — sorry.)

However, suppose instead of conditioning on a positive probability event $x \le a$ we conditioned on the zero probability event $x = a$. Then property (3) of Definition 11 applied to the event $G = \{x = a\}$ requires that

$$\int_{\{x = a\}} Z f(x, y)\, dx\, dy = \int_{\{x = a\}} y f(x, y)\, dx\, dy$$

Because the event $\{x = a\}$ has zero measure, both sides of this equality are zero — we can proceed no further! Thus our definition of conditional expectations does not tell us anything about $E(y|x = a)$. This is often something we need to consider, so how do we proceed?

What we actually need to do is to find a version of the conditional expectation $E(y|\mathcal{G})$ (where $\mathcal{G} := \sigma(x)$ is the sub σ-algebra generated by the x observation) such that we can define sensible conditional probabilities even for probability zero events. These objects actually require strictly more information that just the joint probability measure of x and y. What we need to do is define something that has a range of different names: a *probability kernel*, a *random measure* or a *disintegration*.

Suppose that we are dealing with the probability space $([0,1]^2, \mathcal{F}, \mathbb{P})$ and that $\mathcal{G} = \sigma(x)$ the sub-σ algebra of $\mathcal{F}$ generated by observing x. Then a probability kernel is a function $\mu : [0,1]^2 \times \mathcal{F} \to \mathbb{R}_+$ such that for all $F \in \mathcal{F}$

$$\mu((x,y),F) \qquad \text{is } \mathcal{G}\text{-measurable in } (x,y).$$

This means that we could actually write $\mu((x,y),F) \equiv \mu(x,F)$ for this σ-algebra. And for all x

$$\mu(x,F) \qquad \text{is a probability measure on } \mathcal{F}.$$

This says that for all x the function $\mu(x,.)$ defines a probability measure on $\mathcal{F}$. What this kernel does for us is solve the zero probability problem for conditional probabilities directly, by telling us for every x a complete conditional probability measure over $\mathcal{F}$. Thus it is a bit like a Markov transition matrix, for each current state x we have a conditional distribution over new states $\mu(x,.)$. It is also a family of probability measures indexed by the x value. But because x itself is random, it describes a random probability measure. Notice that the probability kernel does place a restriction on how probabilities vary. For a fixed event $F \in \mathcal{F}$ we ascribed these events the probabilities $\mu(x,F)$ — this function is required to be measurable with respect to the x σ-algebra.

Obviously, once we have the probability kernel we can integrate over the marginal probabilities for the x's to work out the conditional probabilities of events on y for non-zero conditioning events. So we can build up the conditional expectations in the definition from the kernel μ. Thus the definition above does discipline what kernels we can choose and there is a question about whether there is a kernel that is consistent with every conditional expectation? Before we move on to that, note that this process of breaking down a joint distribution into a kernel is called "disintegration" in modern probability theory.

A version of the conditional expectation $E(y|\mathcal{G})$ that also defines a kernel is called a regular conditional distribution of y given $\mathcal{G}$. There is a major result which will not be proved here, that says that provided the two variables x and y are real and the underlying σ-algebras are Borel. Then, there always exists a such a regular conditional distribution. More generally, if x and y live in Polish spaces (complete, separable, metric space) you can also do this. However, for more general random variables this will fail.

13. MARTINGALES

A filtration $\{\mathcal{F}_n\}$ is a sequence of increasing sub-σ-algebras: $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$. One of our earliest results said that the limit algebra $\mathcal{F}_\infty := \sigma(\cup \mathcal{F}_n))$ also, then is well defined.

We think of the filtration as describing a state of information, so: in a repeated game an agent's filtration describes what they know at each point in time or in a dynamic optimization the filtration describes the history of the state variables. Importantly, though, the filtration describes the information available at the end of a period — that is after all the randomness in a period has been resolved. If you want to think about what we know "before" a period begins we need to think about $\mathcal{F}_{n-1}$.

A sequence of random variables $X = (X_n)$ is said to be *adapted* to the filtration $\{\mathcal{F}_n\}$ if X_n is $\mathcal{F}_n$-measurable for all n. Thus observing $\mathcal{F}_n$ is at least as informative as observing $(X_1, \dots, X_n)$. Sometimes we want to think about what a player knows at the beginning of a period before all the uncertainty has been revealed. There are two ways of doing this: First to think about a process (Y_n) which is adapted to $\{\mathcal{F}_n\}$ but where Y_{n-1} is chosen in period n. Second, to think about *previsible* processes (Z_n) where z_n is measurable with respect to $\mathcal{F}_{n-1}$. (The second is what is done in the book, but the first approach is what we are used to doing in economics.) Here I will adopt the first approach—thus the formula's in the text talk have pre-visible actions C_n I will replace these with B_{n-1}'s. We will call $X = (X_n)$ a process or a stochastic process.

Definition 12 *A sequence of random variables $(X_n) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ that is adapted to the filtration $\{\mathcal{F}_n\}$ is called a martingale with respect to $\{\mathcal{F}_n\}$ if*

$$E(X_{n+1}|\mathcal{F}_n) = X_n, \qquad \forall n.$$

If $E(X_{n+1}|\mathcal{F}_n) \leq X_n$ for all n it is called a supermartingale.

Taking an unconditional expectation of both sides of the definition above, and applying the properties of conditional expectations, implies that: the expected value of X_n is constant for martingales

$$E(X_{n+1}) = E(E(X_{n+1}|\mathcal{F}_n)) = E(X_n), \qquad \forall n;$$

and decreasing for supermartingales

$$E(X_{n+1}) = E(E(X_{n+1}|\mathcal{F}_n)) \leq E(X_n), \qquad \forall n.$$

We can think of unbiased random walks as martingales, but they are a much more general class of process than a random walk. At each point in the process X_n the conditional distribution of X_{n+1} can depend on the entire history of the process up to X_n. The only requirement is that this conditional distribution has a mean equal to X_n.

EXERCISE 42: *: [Pólya's Urn] Consider an urn that starts with a white and black ball. In each period a ball is sampled from the urn randomly. If the sampled ball is black it is replaced and another black ball is added. If the sampled ball is white it is replaced* *and another white ball is added. Let B_n be the number of black balls in the urn at the nth period and let $R_n = \frac{B_n}{n+1}$ be the fraction of black balls in the urn. Claim R_n is a martingale with respect to the filtration induced by the random variables $(B_1, \ldots, B_n)$ to verify this we need to check that:*

$$E(R_{n+1}|R_n, \ldots, R_1) = E(R_{n+1}|R_n) = R_n \frac{B_n + 1}{n+2} + (1 - R_n)\frac{B_n}{n+2}$$

As an exercise do this calculation.

One of the most important examples of a martingale is how an estimate of a random variable evolves as more information is acquired. Suppose that $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and there is a filtration $\{\mathcal{F}_n\}$ of $\mathcal{F}$. Then, given the information in $\mathcal{F}_n$ the best estimate of ξ is $E(\xi|\mathcal{F}_n)$. Let us define $M_n := E(\xi|\mathcal{F}_n)$, this random variable is definitely measurable with respect to $\mathcal{F}_n$ so the process (M_n) is adapted to the filtration. If we calculate $E(M_{n+1}|\mathcal{F}_n)$, then the properties of expectations ensure the martingale condition holds true.

$$E(M_{n+1}|\mathcal{F}_n) = E(E(\xi|\mathcal{F}_{n+1})|\mathcal{F}_n) = E(\xi|\mathcal{F}_n) = M_n$$

13.1. *Martingale Convergence*

The reason martingales are so useful is that we have some great results on their convergence. Moreover, the convergence is of the strongest possible kind — almost-sure convergence.

As a warm up result we will prove a convergence result in $\mathcal{L}^2$.

Result 31 Let the process (X_n) be a martingale with respect to the filtration $\{\mathcal{F}_n\}$. Suppose that $E(X_n^2) < K$ for all n. Then there exists $Z \in \mathcal{L}^2$ such that $\lim_{n\to\infty} E((X_n - Z)^2) = 0$.

Proof: Let $X_0 = 0$ and define the martingale differences $Y_n = X_n - X_{n-1}$. Notice that $\sum_{k=1}^n Y_n = X_n - X_0 = X_n$. Also notice that if $n > m$

$$\begin{aligned} E(Y_n Y_m) =& E(E(Y_n Y_m \mid \mathcal{F}_{n-1})) \\ =& E\left(Y_m E(X_n - X_{n-1} \mid \mathcal{F}_{n-1})\right) \\ =& E\left(Y_m [E(X_n \mid \mathcal{F}_{n-1}) - X_{n-1}]\right) = 0. \end{aligned}$$

Thus the Y_n's are uncorrelated. Using this fact we get that

$$K > E(X_n^2) = E\left([\sum_{k=1}^n Y_k]^2\right) = \sum_{k=1}^n E[Y_k^2]$$

This holds for all n so $K > \sum_{k=1}^{\infty} E[Y_k^2]$. Thus the sequence $E[Y_k^2] \to 0$. This implies that (X_n) is a Cauchy sequence in $\mathcal{L}^2$. As this space is complete there exists a limit Z. ■

EXERCISE 43: *Show that* (X_n) *is a Cauchy sequence in* $\mathcal{L}^2$.

Now we move on to proving that uniformly bounded martingales in $\mathcal{L}^1$ converge almost surely. The first step is a pretty intuitive property. Suppose you at the start of time n and you are considering an asset that pays off $X_n - X_{n-1}$ today, where (X_n) is a supermartingale process. The supermartingale property means that on average the asset loses money, given what you know at the start of the period: $0 \geq E(X_n|\mathcal{F}_{n-1}) - X_{n-1}$. Consider an investment strategy allows you either to buy one unit of the asset or to stay out of the market, we will write B_{k-1} as the amount you invest in period k. The success of such a strategy obviously depends on what you know. (If you could cheat the system and only buy in periods when the the returns were positive, then you could make money. But, that would require you to know X_n, or $\mathcal{F}_n$, when you decide to invest.) Thus we will restrict you to only invest based in period k on what you know at the start of the period k that is B_{k-1} which is measurable with respect to $\mathcal{F}_{k-1}$. The profits from the investment strategy are:

$$Y_n := \sum_{k=1}^{n} B_{k-1}(X_k - X_{k-1}), \qquad B_{k-1} \in \{0, 1\}.$$

(In the text B_{k-1} is replace by C_k, but C_k is previsible so only measurable with respect to $\mathcal{F}_{k-1}$. As we are only studying discrete-parameter martingales, this is simple. In continuous time the notion of previsibility is essential to write this down.) The first Lemma says that you cannot expect to profit from investing in this loss making asset, if your investment strategy only depends on what you know the start of the period.

Lemma 2 *Suppose that* (X_n) *is a supermartingale w.r.t. the filtration* $\{\mathcal{F}_n\}$ *and* $X_0 = 0$. *If* (B_n) *is a sequence of* $\{0,1\}$*-valued random variables adapted to* $\{\mathcal{F}_n\}$, *then* $Y_n := \sum_{k=1}^{n} B_{k-1}(X_k - X_{k-1})$ *is a supermartingale w.r.t.* $\{\mathcal{F}_n\}$ *and* $E(Y_n) \leq 0$.

Proof: We begin by calculating the expected value of winnings next period given what we know today:

$$\begin{aligned}
E(Y_{n+1}|\mathcal{F}_n) &= E\left(\sum_{k=1}^{n+1} B_{k-1}(X_k - X_{k-1}) \middle| \mathcal{F}_n \right) \\
&= E\left(Y_n + B_n(X_{n+1} - X_n) \middle| \mathcal{F}_n \right) \\
&= Y_n + B_n[E\left(X_{n+1} \middle| \mathcal{F}_n \right) - X_n] \\
&= Y_n + B_n[\leq 0] \\
&\leq Y_n
\end{aligned}$$

Repeatedly taking unconditional expectations gives $E(Y_n) \leq 0$. ■

EXERCISE 44: *Suppose that:* $(X_n)_{n=1}^N$ *is a supermartingale and that w.r.t. the filtration* $\{\mathcal{F}_n\}_{n=1}^N$, $X_n \in [-M, M]$ *almost surely and* $X_0 = 0$. *Consider the policy of investing in the asset until the first time that* $X_k > c > 0$ *and then stopping, or otherwise continuing to invest until the last period* N. *Calculate a lower bound on the expected profit from this policy. Hence show that the probability of* $X_k > c > 0$ *for some* $k \leq n$ *is at most* $M/(M+c)$.

We now consider a very specific investment strategy. Fix two points $a < b$. You do not invest until you first hit a period, k say, where $X_k < a$. In the next period you choose $B_k^* = 1$ and invest. You keep investing until you first hit a period where $X_{k'} > b$. At this point you stop investing. You do not invest again until $X_{k''} < a$ again at which point the whole process repeats itself. Thus the investment strategy has two states "Invest" (I) or "Not invest" (NI). You switch from I to NI in the period after $X_n > b$ and you switch from NI to I in the period after $X_n < a$. This strategy is clearly uses information that is available to the investor at the start of the period. We will define $U_n^{a,b}$ to be the number of times this strategy has switched from I to NI over the time periods $1, 2, \ldots, n$. (In the jargon these are called the "upcrossings" of the interval $[a, b]$ by the supermartingale X_n.)

Lemma 3 *Suppose that* (X_n) *is a supermartingale w.r.t. the filtration* $\{\mathcal{F}_n\}$ *and* $X_0 = 0$. *Then*

$$E(U_n^{a,b}) \leq \frac{E((X_n - a)^-)}{b - a}$$

where $(X_n - a)^- := -\min\{0, X_n - a\}$.

Proof: The profit from this strategy is

$$Y_n := \sum_{k=1}^{n} B_{k-1}^*(X_k - X_{k-1}).$$

Over an completed investment phase we know that $B_k < a$ and $B_{k'} > b$ and this strategy makes the profit

$$1\left[(X_{k+1} - X_k) + \cdots + (X_{k'} - X_{k'-1})\right] = 1(X_{k'} - X_k) > b - a.$$

If there is an uncompleted investment phase we know that $X_k < a$ but nothing about the terminal value so

$$1\left[(X_{k+1} - X_k) + \cdots + (X_n - X_{n-1})\right] = 1(X_n - X_k) > X_n - a > \min\{0, X_n - a\}.$$

Thus a lower bound on the profits from this strategy is

$$Y_n \geq U_n^{a,b}(b - a) + \min\{0, X_n - a\}.$$

We know (from the previous lemma) that $E(Y_n) \leq 0$ so

$$E(U_n^{a,b}(b-a) + \min\{0, X_n - a\}) \leq 0.$$

Rearranging this gives the result. ∎

Armed with this lemma we can prove a the almost sure convergence result.

Theorem 1 *Suppose that (X_n) is a supermartingale w.r.t. the filtration $\{\mathcal{F}_n\}$ and $X_0 = 0$. And that there exists K such that $E(|X_n|) < K$ for all n. Then, there exists $X^\infty \in \mathcal{L}^1$ such that $X_n \to X^\infty$ almost surely.*

Proof: Let $U_\infty^{a,b} = \lim_{n\to\infty} U_n^{a,b}$. (This limit exists as it is a monotone increasing sequence.) Also notice that

$$E(U_\infty^{a,b}) = \lim_{n\to\infty} E(U_n^{a,b}) \leq \lim_{n\to\infty} \frac{E((X_n - a)^-)}{b-a} \leq \frac{K + |a|}{b-a} < \infty$$

Thus the probability that $U_\infty^{a,b}$ is infinitely large is zero, because otherwise this expectation would also be infinite. So

$$\mathbb{P}(U_\infty^{a,b} < \infty) = 1, \qquad \forall a < b.$$

If $\limsup X_n > b$ and $\liminf X_n < a$ then $U_\infty^{a,b} = \infty$, so the above says that

$$\mathbb{P}(\limsup X_n > b > a > \liminf X_n) = 0, \qquad \forall a < b.$$

Now consider the set of states of the world $\omega \in \Omega$ such that X_n does not converge to $\limsup X_n$. Such a set of states would be included in the countable union

$$\bigcup_{a,b\in\mathcal{Q}} \{\omega : \limsup X_n > b > a > \liminf X_n\}.$$

But we have already shown that each set in this (countable) union has zero probability. Thus it is probability zero the X_n does not converge to $\limsup X_n := X^\infty$—almost sure convergence.

We now need to show that this limit is in $\mathcal{L}^1$. Consider

$$E(|X^\infty|) = E(\liminf |X_n|) \underbrace{\leq}_{\text{Fatou}} \liminf E(|X_n|) < K.$$

∎

13.2. Stopping Times and Optional Stopping

In general we want to consider an observer or optimizer who can choose how long to allow a martingale, or a more general stochastic process, to continue. A simple stopping rule that could always be implemented is to allow (X_n) to continue until $n = T$ and then observe the value of the process at that time X_T. But there are other kinds of rules that the observer could implement, for example stop the process the first time it takes a value less than zero or at time T or if it's always positive. Stop the process if there is a run of three negative values. Stop the process if the total gambling losses are greater than or equal to 100 otherwise continue. These are all examples of feasible strategies that could be played. In each of these cases the decision to stop at time n depends on information that is $\mathcal{F}_n$. But the time where the process stops is a random variable. This general class of random times are called Stopping Times. The requirement that the stopping time is measurable with respect to $\mathcal{F}_n$ is natural and prevents stopping that requires foresight. An examples of random times, that are not stopping times: stop if the next period's value is negative.

Definition 13 *A map $\tau : \Omega \to \mathbb{N}$ is called a stopping time if $\{\omega : \tau(\omega) = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.*

(In continuous time it is more natural to use the definition $\{\omega : \tau(\omega) \leq n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.) You have already studied stopping times when you considered investment strategies (B_n) that were adapted $\{0, 1\}$ processes. By investing, $B_k = 1$, until some condition was true you would generate X_n at that condition. Also, notice that there is no requirement that stopping occurs. For example, it might be possible that no negative value of X_n occurs so that the process continues forever. It is useful to consider stopping times that are finite, so if τ is a stopping time we will define $\tau \wedge n$ to denote the minimum of τ and n, where $n \in \mathbb{N}$.

EXERCISE 45: *Show that if τ is a stopping time then so is $\tau \wedge n$.*

An important property is that if (X_n) is a (super) martingale with respect to $\{\mathcal{F}_n\}$, then so too is the stopped process, so that for

$$\begin{aligned} E(X_0) &= E(X_{\tau \wedge n}), && \text{martingales,} \\ E(X_0) &\geq E(X_{\tau \wedge n}), && \text{supermartingales.} \end{aligned}$$

EXERCISE 46: *Prove this by considering a strategy that invests in every period until the stopping time, or n is reached, and then stops investing for ever. You also need to use Lemma 2.*

We would also like to know when the equalities above can be written without the imposition of a finiteness condition on the stopping time, that is when does $E(X_0) = E(X_\tau)$? This fails because the stopping time τ may not be finite — so some probability in the expectation gets lost and does not contribute to the expectation. (This may not be a problem for supermartingales, however.) As an example consider a symmetric random walk that starts at the origin $X_0 = 0$. This is a martingale, so $E(X_n) = 0$ for all n. Suppose we consider stopping the process when it hits 1. Because the symmetric random walk goes everywhere with probability 1, we know that the random walk hits 1 with probability 1. So $E(X_\tau) = 1$ and $E(X_0) = 0$ so the above inequality does not seem to hold. What is going wrong here? Suppose we stop in finite time we have

$$0 = E(X_0) = E(X_{\tau\wedge n}) = \Pr(\tau \le n)1 + (1 - \Pr(\tau \le n))E(X_n|\tau > n)$$

We know that as n increases $\Pr(\tau \le n) \to 1$, but the problem is that $E(X_n|\tau > n) \to -\infty$. This makes the left and right side equal and prevents $E(X_\tau)$ and $E(X_0)$ from being equal. Thus it is necessary to bound X_n to stop this occurring and at the same time having $\Pr(\tau \le n) \to 1$ as n increases. The following result summarizes the necessary conditions.

Result 32 (Optional Stopping) Suppose that (X_n) is a martingale/supermartingale, X_0 is integrable, and τ is a stopping time. Then, X_τ is integrable and

$$E(X_0) = E(X_\tau) \qquad \text{(martingale)} \qquad E(X_0) \ge E(X_\tau) \qquad \text{(super martingale)}$$

if one of the following conditions holds:

1. The exists N such that $\tau(\omega) \le N$ for all $\omega \in \Omega$.
2. $|X_n(\omega)| \le K$ for all n and all ω and τ is almost surely finite.
3. $E(\tau) < \infty$ and $|X_n(\omega) - X_{n-1}(\omega)| < K$ for all n, ω.

Proof: We will do the proof for supermartingales. If (X_n) is a martingale, both X_n and $-X_n$ are supermartingales and the result follows in this case. First note that $X_{\tau\wedge n}$ is integrable. This follows as $E(|X_{n\wedge\tau}|) \le E(\max_n |X_n|) < \infty$.

Part 1: As $E(X_0 - X_{n\wedge\tau}) \ge 0$ for all n, take $n = N$.

Part 2: The conditions ensure that $X_0 - X_{n\wedge\tau} \to X_0 - X_\tau$ almost surely. Then (as X_n is bounded) applying bounded convergence gives $\lim_{n\to\infty} E(X_0 - X_{n\wedge\tau}) \to E(X_0 - X_\tau) \ge 0$.

Part 3: Since

$$|X_{n\wedge\tau} - X_0| \le \sum_{k=1}^{n\wedge\tau} |X_k - X_{k-1}| \le K\tau$$

And $E(\tau) < \infty$. We can use dominated convergence to take the limit as $n \to \infty$. ■

EXERCISE 47: *Do (as an example) exercise 10.6 from the text to understand how useful this property is.*

EXERCISE 48: *Explain why the random-walk example, that precedes the statement of this result, does not satisfy any of the above three cases.*

It is useful to know that a stopping time is finite, that is $E(\tau) < \infty$. Properties like the Borel-Cantelli can be used to establish this. Another way of getting $E(\tau) < \infty$ is the following. Suppose you know that when you get to any period n there is a probability of at least $\varepsilon > 0$ of stopping in the next K periods, $\mathbb{P}(\tau < n + K|\mathcal{F}_n) > \varepsilon$. Then the probability of surviving mK periods without stopping is at most $(1-\varepsilon)^m$. This implies that :$\mathbb{P}(\tau > mK) < (1-\varepsilon)^m$, $\mathbb{P}(\tau = \infty) = 0$, and that $E(\tau) < \infty$.

EXERCISE 49: *Use the above argument to get finite upper bound on $E(\tau)$ if $\mathbb{P}(\tau < n + K|\mathcal{F}_n) > \varepsilon$ for all n.*

13.2.1. De Moivre's Martingale

Now we will consider a slightly different random walk and show how martingales can be used to calculate stopping times in a very appealing way. Suppose we start at $X_0 = k$ and study a random walk on $\{0, 1, \ldots, N\}$ that moves up $+1$ with probability p and down -1 with probability $q = 1 - p$. Thus we are considering $Y_n = \sum_{k=0}^{n} X_k$ where X_k is i.i.d. on $\{-1, +1\}$. We will stop the process when it first hits 0 or N, so we don't need to worry above what happens at the boundary in this definition.

Clearly Y_n describes the position of the random walk at time n. We claim that $(q/p)^{Y_n}$ is a martingale.

$$E\left(\left(\frac{q}{p}\right)^{Y_{n+1}} \mid Y_n\right) = p\left(\frac{q}{p}\right)^{Y_n+1} + q\left(\frac{q}{p}\right)^{Y_n-1} = \frac{q^{Y_n+1} + pq^{Y_n}}{p^{Y_n}} = \left(\frac{q}{p}\right)^{Y_n}$$

Thus if the process started at $Y_0 = k$ we know

$$\left(\frac{q}{p}\right)^k = E\left(\left(\frac{q}{p}\right)^{Y_n \wedge \tau}\right)$$

where τ is the stopping time for the first time of hitting 0 or N. As the process is bounded and the random walk hits the extremes in finite time, we know that $(q/p)^\tau$ is also a martingale. Thus

$$\left(\frac{q}{p}\right)^k = E\left(\left(\frac{q}{p}\right)^\tau\right) = \left(\frac{q}{p}\right)^0 \Pr(\tau = 0) + \left(\frac{q}{p}\right)^N (1 - \Pr(\tau = 0)).$$

Hence (if $p \neq q$) we can solve

$$\Pr(\tau = 0) = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}$$

EXERCISE 50: *Take the limit of this as $p/q \to 1$.*

14. UNIFORM INTEGRABILITY

This is a concept that can be studied independently of Martingales. Let us begin by thinking about why convergence in probability does not ensure convergence of expectations. Consider the sequence of binary random variables (X_n): $X_n = 0$ with probability $\frac{n-1}{n}$ and $X_n = n$ with probability $1/n$. Clearly X_n converges in probability to zero but $E(X_n) = 1$ for all n, so the limit of the expectations is different from the expectation of the limit. To prevent issues like this you could impose uniform bounds on the sequence random variables, like $|X_n| < K$ for all n. This can rule out important applications (such as the random walk). The condition below turns out to be enough:

Definition 14 *The sequence (X_n) is said to be uniformly integrable if*

$$\sup_n E(|X_n| \mathbb{1}_{\{|X_n| \geq a\}}) \to 0 \qquad \text{as } a \to \infty.$$

What this does is prevent weird stuff happening to the expectations with too high a probability.

EXERCISE 51: *Why is the example that begins this section not uniformly integrable? Why is the stopped symmetric random walk example in the previous section not uniformly integrable?*

EXERCISE 52: *Show that if (X_n) is uniformly integrable, then there exists K such that $E(|X_n|) < K$ for all n.*

As you might need to check whether a sequence is uniformly integrable, one way of doing this is to verify one of the following sufficient conditions:

Result 33 If the sequence of random variables (X_n) satisfies: (a) For all n, some $p > 1$ and some $K > 0$, $E(|X_n|^p) < K$. (b) There exists a random variable $Y > |X_n|$ such that $E(|Y|) < \infty$. Then it is uniformly integrable.

Proof: (a) requires the observation that if $X_n > a > 1$ then $(X_n/a)^p > (X_n/a)$ and so $a^{1-p}X_n^p > X_n$. Hence

$$E(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}) \leq E(|a^{1-p}X_n^p|\mathbb{1}_{\{|X_n|\geq a\}}) \leq a^{1-p}E(|X_n^p|) \leq a^{1-p}K$$

Letting $a \to \infty$ gives the result. Part (b) uses the fact that the probability that $|Y| > a$ tends to zero as $a \to \infty$ if the expectation exists. This gives the required properties for the upper bound on $E(|X_n|)$. ■

We will now state a result without proof that is useful in summarizing the properties of uniform integrability.

Result 34 Suppose that (X_n) is a sequence of random variables that converge in probability to X, then the following statements are equivalent:

1. (X_n) is uniformly integrable.
2. $E(|X_n|) < \infty$ for all n, $E(|X|) < \infty$ and $E(|X_n - X|) \to 0$.
3. $E(|X_n|) < \infty$ for all n, $E(|X_n|) \to E(|X|) < \infty$.

14.1. Optional Stopping Again

We know that $E(X_0) = E(X_{\tau\wedge n})$ for all stopping times τ and time periods n. We want to show that $E(X_0) = E(X_\tau)$ so a natural route to achieve this is to try to establish that the sequence of random variables $(X_{\tau\wedge n})_n$ is uniformly integrable so that $\lim_{n\to\infty} E(X_{\tau\wedge n}) = E(X_\tau)$. The following result illustrates a way of doing this.

Result 35 Suppose that (X_n) is a martingale with respect to the filtration $\{\mathcal{F}_n\}$ and that τ is a stopping time. Then $E(X_\tau) = E(X_0)$ if

1. $\Pr(\tau < \infty) = 1$, $E(\tau) < \infty$, and
2. There exists c such that

$$E(|X_{n+1} - X_n| \mid \mathcal{F}_n) \leq c, \qquad \forall n < \tau.$$

Proof: To show that $\lim_{n\to\infty} E(X_{\tau\wedge n}) = E(X_\tau)$ it is sufficient to show that $X_{n\wedge\tau}$ is uniformly integrable (by the previous result). Let $Z_n := |X_n - X_{n-1}|$ and $W := Z_1 + \cdots + Z_\tau$. By the triangle inequality $|X_{\tau\wedge n} - X_0| \leq W$ for all n so $|X_{\tau\wedge n}| < |X_0| + W$. By our second sufficient condition for uniform integrability it is sufficient to show that $E(W) < \infty$. We can decompose W into a sum

$$W = \sum_{n=1}^{\infty} Z_n \mathbb{1}_{\{\tau\geq n\}}$$

Considering the expectation of a term in this sum we can bound it by using condition (b) in the result

$$E(Z_n \mathbb{1}_{\{\tau \geq n\}} | \mathcal{F}_{n-1}) = \mathbb{1}_{\{\tau \geq n\}} E(Z_n | \mathcal{F}_{n-1}) \leq c \mathbb{1}_{\{\tau \geq n\}}$$

(Note this uses the fact that one of the things you know at the end of period $n-1$ is that you haven't stopped yet. So if you do stop it will be after period $n-1$. Now taking an unconditional expectation of both sides above we get that

$$E(Z_n \mathbb{1}_{\{\tau \geq n\}}) \leq c \Pr(\tau \geq n)$$

Now taking an expectation of both sides of the definition of W we get that $E(W) < cE(\tau)$ which is finite by assertion. This completes the proof of uniform integrability. ■

Now we will look at some more applications of the optional stopping theorem.

14.1.1. Wald's Identity

This can be used to calculate the sum of randomly many i.i.d. random variables. Consider, for example, a website experiencing i.i.d. queries each day and trying to work out how many queries it will have to process before it shuts down.

Result 36 (X_n) is a sequence of i.i.d. random variables such that $\mu = E(X_n)$ exists. Suppose that τ is a stopping time with respect to the filtration $\{\mathcal{F}_n\}$ generated by the process (X_n) and that $E(\tau)$ is finite. Then $E(X_1 + X_2 + \cdots + X_\tau) = \mu E(\tau)$.

Proof: Define $Y_n := X_1 + \cdots + X_n - n\mu$. This obviously adapted to the filtration $\{\mathcal{F}_n\}$. It is also a martingale with respect to this filtration as

$$E(Y_{n+1} | \mathcal{F}_n) = X_1 + \cdots + X_n + E(X_{n+1} | \mathcal{F}_n) - (n+1)\mu = Y_n.$$

Also notice that this says

$$E(|Y_{n+1} - Y_n| \mid \mathcal{F}_n) = E(|X_{n+1} - \mu| \mid \mathcal{F}_n) \overbrace{=}^{\text{indep}} E(|X_{n+1} - \mu|) = E(|X_1 - \mu|) < \infty$$

Hence the conditions for UI are satisfied and $E(Y_\tau) = E(Y_0) = 0$. Substitution then proves the claim. ■

This property does not just hold for the first moment but all moments — if they exist— as the following exercise shows.

EXERCISE 53: *Suppose that (X_n) is a sequence of i.i.d. random variables such that $E(X_n^2)$ exists. Suppose that τ is a stopping time with respect to the filtration $\{\mathcal{F}_n\}$*

generated by the process (X_n) and that $E(\tau)$ is finite. Show that $Var(X_1 + X_2 + \cdots + X_\tau) = \sigma^2 E(\tau)$.

There is an extension of Wald's identity that applies to the moment generating function of $X_1 + X_2 + \cdots + X_\tau$.

14.2. Doob Decomposition

It seems like martingales are very special kinds of processes. The following result says that all adapted processes in $\mathcal{L}^1$ can be written as a martingale plus some other factors that are more deterministic in nature, by taking conditional expectations many problems can be reduced to martingale problems

Result 37 Suppose that $(X_n) \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ is an adapted process, then

$$X_n = X_0 + M_n + A_{n-1}$$

where M_n is a martingale with respect to the filtration $\{\mathcal{F}_n\}$ and A_{n-1} is adapted to $\mathcal{F}_{n-1}$ (a previsible process).

Proof: Define

$$A_{n-1} := \sum_{k=1}^{n} E(X_k - X_{k-1} | \mathcal{F}_{k-1})$$

and define

$$M_n := X_n - X_0 - A_{n-1}.$$

EXERCISE 54: *Show that this A_{n-1} is $\mathcal{F}_{n-1}$ adapted and this M_n is a martingale.*

This exercise completes the proof. ■