Integration Theory: Lecture notes 2013

Johan Jonasson [*,†,‡]

September 2013

1 Preface

These lecture notes are written when the course in integration theory is for the first time in more than twenty years, given jointly by the the two divisions Mathematics and Mathematical Statistics. The major source is G. B. Folland: Real Analysis, Modern Techniques and Their Applications. However, the parts on probability theory are mostly taken from D. Williams: Probability with Martingales. Another source is Christer Borell's lecture notes from previous versions of this course, see

www.math.chalmers.se/Math/Grundutb/GU/MMA110/A11/

2 Introduction

This course introduces the concepts of measures, measurable functions and Lebesgue integrals. The integral used in earlier math courses is the so called Riemann integral. The Lebesgue integral will turn out to be more powerful in the sense that it allows us to define integrals of not only Riemann integrable functions, but also some functions for which the Riemann integral is not defined. Most importantly however, is that it will allow us to rigorously prove many results for which proofs of the corresponding results in the Riemann setting are usually never seen by students at the basic and intermediate level. Such results include precise conditions for when we can change order of integrals and limits, change order of integration

[*] Chalmers University of Technology
[†] Göteborg University
[‡] jonasson@chalmers.se

in multiple integrals and when we can use integration by parts. Of course, we will also prove many new results.

The concept of measurability is an advanced one, in the sense that a lot of people at first find it difficult to master; it tends to feel fundamentally more abstract than things one has encountered before. Therefore, a natural first question is why the concept is needed. To answer this, consider the following example.

Let $X = \mathbb{R}/\mathbb{Z}$, the circle of circumference 1, with addition and multiplication defined modulo 1. Suppose we want to introduce the concept of the *length* of subsets of X. A natural first assumption is that one should be able to do this so that the length is defined for *all* subsets of X. It is also extremely natural to claim that the length l, should satisfy

- $l(\emptyset) = 0$,
- $l(X) = 1$,
- $l(\cup_1^\infty)A_n = \sum_1^\infty l(A_n)$ for all disjoint $A_1, A_2, \ldots$,
- $l(A + x) = l(A)$ for all $A \subseteq X$ and $x \in X$.

However, if we insist on defining l for all subsets, this turns out to be *impossible*. Let us see why.

Partition X into equivalence classes by saying that x and y are equivalent if $x - y$ is a rational number. By the axiom of choice, there exists a set A containing exactly one element from each equivalence class. For each $q \in \mathbb{Q} \cap X$, let $A_q = A + q$. Then $\bigcup_q A_q = X$, for since for each $x \in X$, A contains an element y equivalent to x, i.e. $x \in A_{x-y}$ and $x - y \in \mathbb{Q}$.

On the other hand, the A_q's are disjoint, for if $x \in A_{q_1} \cap A_{q_2}$, then $x = y + q_1 = z + q_2$ for two elements $y, z \in A$. However, then $y - z = q_2 - q_1 \in \mathbb{Q}$, so y and z are equivalent, contradicting the construction of A.

If we could assign lengths to the A_q's, then these lengths must be equal by the fourth condition on l. On the other hand, the lengths of the A_q's must sum to 1 by the third condition. However, these two conditions are mutually exclusive.

The moral of the example is that the set A must be declared *non-measurable*; no length of A can be defined. The construction of the example is based on the axiom of choice and it can be shown that all constructions of non-measurable sets must rely on the axiom of choice.

There are even more absurd examples than this one. The famous Banach-Tarski paradox proves, using the axiom of choice, that for *any* two bounded compact sets in $\mathbb{R}^3$, the one can be divided into a *finite* number of parts which can be translated and rotated and mirrored and then put back together to form the other. For example: any grain of sand can be divided into a number of pieces that can be put back together to form a ball the size of the earth! Clearly theses pieces cannot have a well defined volume.

Examples like these call for a theory of measures and measurable sets.

3 Measures

We are going to consider measures in a very general framework: we will consider measures on a an abstract space X on we which we make no initial assumptions whatsoever. As the above example revealed, it is not always possible with meaningful measures defined on all subsets of X. Hence a concept of what classes of subsets to define a desired measure on, is needed. The two last conditions on a length measure in the above example were natural in that particular situation, but it is easy to think of other situations where neither of them is natural or even meaningful. The two first conditions however, are such that they should hold for anything that deserves to be called a measure, no matter what structure X has. Thus we keep those two conditions in mind, and ask for classes of subsets large enough to ensure that all interesting set operations on measurable sets results in a measurable set, but restrictive enough to make sure that no conflict with the basic assumptions arises. The answer is *σ-algebras.*

3.1 Algebras and σ-algebras

Definition 3.1 *Let $\mathcal{A}$ be a class of subsets of X such that*

(i) $X \in \mathcal{A}$*,*

(ii) $E^c \in \mathcal{A}$ *whenever* $E \in A$*,*

(iii) $E \cup F \in A$ *whenever* $E, F \in \mathcal{A}$*.*

Then $\mathcal{A}$ is called an algebra *(on X).*

Note that by (i) and (ii), $\emptyset = X^c \in \mathcal{A}$. Also, if $E, F \in \mathcal{A}$, then $E \cap F = (E^c \cup F^c)^c \in \mathcal{A}$ by (ii) and (iii).

Definition 3.2 *Let $\mathcal{M}$ be a class of subsets of X such that* *(i)* $X \in \mathcal{M}$,

(ii) $E^c \in \mathcal{M}$ *whenever* $E \in \mathcal{M}$,

(iii) $\bigcup_{n=1}^{\infty} E_n \in \mathcal{M}$ *whenever* $E_1, E_2, \ldots \in \mathcal{M}$.

Then $\mathcal{M}$ *is called a* σ*-algebra.*

Clearly any σ-algebra is an algebra. As above $\emptyset \in \mathcal{M}$, and analogously, if $E_1, E_2, \ldots \in \mathcal{M}$, then $\bigcap_n E_n = (\bigcup_n E_n^c)^c \in \mathcal{M}$.

A measure will always be defined on a σ-algebra. The smallest possible σ-algebra on any space X is $\{\emptyset, X\}$. The largest σ-algebra is $\mathcal{P}(X)$, the class of all subsets of X (but we have seen that meaningful measures cannot always be defined on this σ-algebra).

If $\mathcal{M}$ is a σ-algebra on X, then the pair $(X, \mathcal{M})$ is called a *measurable space* and a set $E \in \mathcal{M}$ is called $\mathcal{M}$*-measurable*.

3.2 Generated σ-algebras

Let $\mathcal{C}$ be an arbitrary class of subsets of X. We define the σ*-algebra generated by* $\mathcal{C}$ as the smallest σ-algebra containing $\mathcal{C}$, i.e.

$$\sigma(\mathcal{C}) = \bigcap \{\mathcal{F} : \mathcal{F}\ \sigma\text{-algebra},\ \mathcal{F} \supseteq \mathcal{C}\}.$$

(It is an easy exercise to show that any intersection of σ-algebras is a σ-algebra.)

The most important example is the *Borel* σ*-algebra*; if X is a topological space and $\mathcal{T}$ is the class of open sets, then the Borel σ-algebra, $\mathcal{B}(X)$, is given by

$$\mathcal{B}(X) = \sigma(\mathcal{T}).$$

Since any open set in $\mathbb{R}$ is a countable union of open intervals, it follows that

$$\mathcal{B}(\mathbb{R}) = \sigma((a,b) : a, b \in \mathbb{R}).$$

It is now easy to see (check this!) that we also have

$$\begin{aligned} \mathcal{B}(\mathbb{R}) &= \sigma([a,b) : a,b \in \mathbb{R}) = \sigma((a,b] : a,b \in \mathbb{R}) = \sigma([a,b] : a,b \in \mathbb{R}) \\ &= \sigma((-\infty,b) : b \in \mathbb{R}) = \sigma((a,\infty) : a \in \mathbb{R}). \end{aligned}$$

In integration theory, one often works with the *extended real line*, $\overline{\mathbb{R}} = [-\infty, \infty]$ and, even more, with the extended positive half-line $\overline{\mathbb{R}}_+ = [0, \infty]$. Here the arithmetics involving the points ∞ and $-\infty$ work as one would intuitively guess, and a subset is regarded as open if it is either a subset of $\mathbb{R}$ and open as such, of the form $[-\infty, a)$ or $(a, \infty]$, or the whole space. It is now straightforward to prove analogous expressions for $\mathcal{B}(\overline{\mathbb{R}})$ and $\mathcal{B}(\overline{\mathbb{R}}_+)$.

3.3 Measures

If $\mathcal{C}$ is a class of subsets of X and $\mu_0 : \mathcal{C} \to \overline{\mathbb{R}}_+$, then μ_0 is called a set function. Let $\mathcal{A}$ be an algebra. If μ_0 is a set function on $\mathcal{A}$ such that $\mu_0(\emptyset) = 0$ and $E, F \in \mathcal{A}$, $E \cap F = \emptyset$ implies $\mu_0(E \cup F) = \mu_0(E) + \mu_0(F)$, then μ_0 is said to be *additive*. If $\mu_0(\emptyset) = 0$ and μ_0 satisfies the stronger condition that $\mu_0(\bigcup_n E_n) = \sum_n \mu_0(E_n)$ whenever $E_1, E_2, \ldots \mathcal{A}$ and $\bigcup_n E_n \in \mathcal{A}$, then μ_0 is said to be *countably additive* or a *premeasure*. (Stronger since additivity follows from countable additivity by taking $E_1 = E$, $E_2 = F$ and $E_3 = E_4 = \ldots = \emptyset$.)

Definition 3.3 *Let $\mathcal{M}$ be a σ-algebra and μ a set function defined on $\mathcal{M}$. If μ is countably additive, then μ is said to be a measure.*

Let μ be a measure on the σ-algebra $\mathcal{M}$. Here are a few classifications.

- μ is said to be *finite* if $\mu(X) < \infty$.
- μ can be said to be a *probability measure* if $\mu(X) = 1$.
- μ is said to be *σ-finite* if there exist sets $E_1, E_2, \ldots \in \mathcal{M}$ such that $\bigcup_n E_n = X$ and $\mu(E_n) < \infty$ for all n.
- μ is said to be *semi-finite* if for every $E \in \mathcal{M}$ such that $\mu(E) = \infty$, there exists a set $F \subset E$ such that $0 < \mu(F) < \infty$.

The *trivial measure* is the measure μ with $\mu(E) = 0$ for all $E \in \mathcal{M}$. Clearly any probability measure is finite, any finite measure is σ-finite and every σ-finite measure is semi-finite.

Example. Let $\mu(\emptyset) = 0$ and $\mu(E) = \infty$ for any nonempty measurable E. Then μ is a measure which is not even semi-finite. $\square$

Example. Length measure on $[0, 1]$ (which, to be true, we have not defined yet) is a probability measure. Length measure on $\mathbb{R}$ is σ-finite; take e.g. $E_n = (-n, n)$. $\square$

When $\mathcal{M}$ is a σ-algebra on X and μ is a measure on $\mathcal{M}$, the triple $(X, \mathcal{M}, \mu)$ is called a *measure space*. If $\mu(X) = 1$, then we may also speak of $(X, \mathcal{M}, \mu)$ as a *probability space* and if we do that, we usually refer to $\mathcal{M}$-measurable sets as *events*.

Remark. Suppose that $\mu(X) = 1$. Then we can choose to call μ a probability measure and $(X, \mathcal{M}, \mu)$ a probability space. Whether or not we actually do that depends on the point of view we want to adopt. In many situations it is either our main purpose to model a random experiment or it is instructive or useful for some other reason to think of the points $x \in X$ as the possible outcomes of a random experiment. If this is not the case, we may instead prefer to just refer to μ as a finite measure of total mass 1.

Some general properties of measures follow. In all of these, it is assumed that $(X, \mathcal{M}, \mu)$ is a measure space.

Proposition 3.4 *(a)* $E, F \in \mathcal{M},\ E \subseteq F \Rightarrow \mu(E) \leq \mu(F)$.

(b) $E_1, E_2, \ldots \in \mathcal{M} \Rightarrow \mu(\bigcup_n E_n) \leq \sum_n \mu(E_n)$,

(c) If $\mu(X) < \infty$*, then* $\mu(E \cup F) = \mu(E) + \mu(F) - \mu(E \cap F)$,

(d) If $\mu(X) < \infty$*,* $E, F \in \mathcal{M}$ *and* $E \subseteq F$*, then* $\mu(F \setminus E) = \mu(F) - \mu(E)$.

Proof. By additivity of μ, $\mu(F) = \mu(E) + \mu(F \setminus E)$ whenever $E \subseteq F$. This proves (d) and since $\mu(F \setminus E) \geq 0$, (a) follows too. For (b), let $F_1 = E_1$ and recursively $F_n = E_n \setminus \bigcup_1^{n-1} F_j$, $n = 2, 3, \ldots$. Then the F_n's are disjoint and $\bigcup_n F_n = \bigcup_n E_n$, so by (a)

$$\mu(\bigcup_n E_n) = \sum_n \mu(F_n) \leq \sum_n \mu(E_n).$$

Finally (c) follows from

$$\mu(E \cup F) = \mu(E) + \mu(F \setminus E \cap F) = \mu(E) + \mu(F) - \mu(E \cap F)$$

by additivity and (d). □

Proposition 3.5 *(***Continuity of measures***)*

(a) If $E_1 \subseteq E_2 \subseteq \ldots$ *and* $E = \bigcup_n E_n$*, then* $\mu(E) = \lim_n \mu(E_n)$.

(b) If $F_1 \supseteq F_2 \supseteq \ldots$*,* $F = \bigcap_n F_n$ *and* $\mu(F_1) < \infty$*, then* $\mu(F) = \lim_n \mu(F_n)$.

Proof. For (a), let $A_1 = E_1$ and recursively $A_n = E_n \setminus E_{n-1}$. Then $E = \bigcup_n A_n$ and the A_n's are disjoint, so

$$\mu(E) = \sum_1^\infty \mu(A_j) = \lim_n \sum_1^n \mu(A_j) = \lim_n \mu(E_n)$$

since $E_n = \bigcup_1^n A_j$. Now (b) follows from applying (a) to $E_n = F_1 \setminus F_n$ and $E = F_1 \setminus F$ and using Proposition 3.4(d). □

Corollary 3.6 *If $\mu(N_n) = 0$ for all n, then $\mu(\bigcup_n N_n) = 0$.*

Proof. Apply e.g. Proposition 3.4(b). □

3.4 "Almost everywhere" and completeness

Let S be a proposition about points of X and suppose that $F = \{x : S(x) \text{ is false}\}$ is measurable. If $\mu(F) = 0$, then S is said to hold *almost everywhere* (with respect to μ if other measures are also under discussion), abbreviated a.e. In case μ is a probability measure, one often instead says that S holds *almost surely*, abbreviated a.s.

If S holds a.e. and T is another proposition such that $T(x)$ is true whenever S is true, then one would clearly want to think of T as also holding a.e. However this is not so in general, since even if $\mu(F) = 0$, it may be the case that some subset E of F is not measurable. If $(X, \mathcal{M}, \mu)$ is such that $E \in \mathcal{M}$ whenever $E \subset F$, $F \in \mathcal{M}$ and $\mu(F) = 0$, then the measure space is said to be *complete* and μ is said to be a complete measure.

If μ is not complete, then one can always extend the measure space, by defining the larger σ-algebra

$$\overline{\mathcal{M}} = \{E \cup F : E \in \mathcal{M}, \exists N \in \mathcal{M} : F \subset N, \mu(N) = 0\}$$

(exercise: prove that $\overline{\mathcal{M}}$ is a σ-algebra) and the measure $\overline{\mu}$ on $\overline{\mathcal{M}}$ by $\overline{\mu}(E \cup F) = \mu(E)$. Then $(X, \overline{\mathcal{M}}, \overline{\mu})$ is complete and $\overline{\mu}$ is called the completion of μ.

3.5 Dynkin's Lemma and the Uniqueness Theorem

Dynkin's Lemma will be a fundamental tool for theorem proving. It is based on the concepts of *π-systems* and *d-systems*. A π-system is a class $\mathcal{I}$ of subsets of X

that is closed under finite intersections, i.e. $E \cap F \in \mathcal{I}$ whenever $E, F \in \mathcal{I}$. The definition of a d-system follows.

Definition 3.7 *Let $\mathcal{D}$ be a class of subsets of X. Then $\mathcal{D}$ is said to be d-system if*

(a) $X \in \mathcal{D}$,

(b) $E, F \in \mathcal{D},\ E \subseteq F \Rightarrow F \setminus E \in \mathcal{D}$,

(c) $E_n \in \mathcal{D},\ E_n \uparrow E \Rightarrow E \in \mathcal{D}$.

Generated d-systems are defined analogously with generated σ-algebras:

$$d(\mathcal{C}) = \bigcap \{\mathcal{D} \supseteq \mathcal{C} : \mathcal{D}\ d\text{-system}\} S.$$

(Check that any intersection of d-systems is a d-system.)

Theorem 3.8 *Let $\mathcal{M}$ be a class of subsets of X. Then $\mathcal{M}$ is a σ-algebra if and only if it is π-system and a d-system.*

Proof. The only if-direction is obvious. The if direction follows from that $X \in \mathcal{M}$ by (a) in the definition of a d-system, $E^c = X \setminus E \in \mathcal{M}$ whenever $E \in \mathcal{M}$ by (b) and if $E_n \in \mathcal{M}$, $n = 1, 2, \ldots$, then $F_n := \bigcup_1^n E_j = (\bigcap_1^n E_j^c)^c \in \mathcal{M}$ since $\mathcal{M}$ is a π-system, so $E := \bigcup_1^\infty E_j \in \mathcal{M}$ by (c) since $F_n \uparrow E$. $\square$

Since any σ-algebra is also a d-system, it follows that $\sigma(\mathcal{C}) \supseteq d(\mathcal{C})$ for any $\mathcal{C}$. Dynkin's Lemma provides an answer to when we have equality.

Theorem 3.9 *(***Dynkin's Lemma***)*
If $\mathcal{I}$ is a π-system, then $d(\mathcal{I}) = \sigma(\mathcal{I})$.

Proof. It suffices to prove that $d(\mathcal{I}) \supseteq \sigma(\mathcal{I})$. By Theorem 3.8 it thus suffices to prove that $d(\mathcal{I})$ is a π-system. In other words, it suffices to prove that

$$\mathcal{D}_2 := \{B \in d(\mathcal{I}) : B \cap C \in d(\mathcal{I}) \text{ for all } C \in d(\mathcal{I})\}$$

equals $d(\mathcal{I})$. The proof is done in two similar steps. For step 1, define

$$\mathcal{D}_1 := \{B \in d(\mathcal{I}) : B \cap C \in d(\mathcal{I}) \text{ for all } C \in \mathcal{I}\}.$$

Since $\mathcal{I}$ is a π-system, $\mathcal{D}_1$ contains $\mathcal{I}$, so if we can show that $\mathcal{D}_1$ is a d-system, then $\mathcal{D}_1 = d(\mathcal{I})$. Part (a) in the definition of a d-system obviously holds. If $B_1, B_2 \in$

$\mathcal{D}_1$ and $B_1 \subseteq B_2$, then for any $C \in \mathcal{I}$, $(B_2 \setminus B_1) \cap C = (B_2 \cap C) \setminus (B_1 \cap C) \in d(\mathcal{I})$ since $d(\mathcal{I})$ is a d-system. Hence part (b) holds for $\mathcal{D}_1$. Finally if $B_n \in \mathcal{D}_1$ and $B_n \uparrow B$, then $B_n \cap C \uparrow B \cap C$, so $B \in \mathcal{D}_1$ since $d(\mathcal{I})$ is a d-system.

That $\mathcal{D}_1 = d(\mathcal{I})$ means that $\mathcal{D}_2 \supseteq \mathcal{I}$, so it suffices now to prove that $\mathcal{D}_2$ is a d-system, which is now done in complete analogy with step 1. (Check that you can fill this in.) $\square$

Our first application is the following uniqueness theorem for measures.

Theorem 3.10 *(***Uniqueness of finite measures***)*
Suppose that $\mathcal{I}$ is a π-system and $\mathcal{M} = \sigma(\mathcal{I})$. If μ_1 and μ_2 are two measures on $\mathcal{M}$ such that $\mu_1(X) = \mu_2(X) < \infty$ and $\mu_1(I) = \mu_2(I)$ for all $I \in \mathcal{I}$, then $\mu_1 = \mu_2$.

Proof. By Dynkin's Lemma, it suffices to prove that $\mathcal{D} := \{E \in \mathcal{M} : \mu_1(E) = \mu_2(E)\}$ is a d-system. That $X \in \mathcal{D}$ follows from the first part of the assumption. If $E, F \in \mathcal{D}$ and $E \subseteq F$, then $\mu_1(F \setminus E) = \mu_1(F) - \mu_1(E) = \mu_2(F) - \mu_2(E) = \mu_2(F \setminus E)$, so $F \setminus E \in \mathcal{D}$. Finally if $E_n \in \mathcal{D}$ and $E_n \uparrow E$, then $\mu_1(E_n) = \mu_2(E_n)$, so $\mu_1(E) = \mu_2(E)$ by the continuity of measures. $\square$

Corollary 3.11 *If two probability measures agree on $\mathcal{I}$, then they are equal.*

3.6 Borel-Cantelli's First Lemma

Definition 3.12 *Let $E_1, E_2, \ldots$ be subsets of X. Then*

$$\limsup_n E_n := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n$$

$$\liminf_n E_n := \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} E_n.$$

Note that

$$\limsup_n E_n = \{x \in X : x \in E_n \text{ for infinitely many } n\}$$

and

$$\liminf_n E_n = \{x \in X : x \in E_n \text{ for all but finitely many } n\}.$$

One sometimes writes $E_n\, i.o.$ for $\limsup_n E_n$, where $i.o.$ stands for ”infinitely often”. (There is no corresponding abbreviation for $\liminf_n E_n$.)

Let $(X, \mathcal{M}, \mu)$ be a measure space and suppose that $E_1, E_2, \ldots \in \mathcal{M}$. Since a σ-algebra is closed under countable intersections and unions, it is clear that $\limsup_n E_n$ and $\liminf_n E_n$ are then also measurable.

Lemma 3.13 *(***Borel-Cantelli's Lemma I***)*
If $\sum_{n=1}^{\infty} \mu(E_n) < \infty$, then $\mu(\limsup_n E_n) = 0$.

Proof. Write $F_m = \bigcup_{n=m}^{\infty} E_n$ and $F = \limsup_n E_n$. Then $F_n \downarrow F$. Since $\bigcup_{n=1}^{M} E_n \uparrow F_1$ it follows from the continuity of measures (from below) and the hypothesis that

$$\mu(F_1) = \lim_M \mu(\bigcup_1^M E_n) \le \lim_M \sum_1^M \mu(E_n) = \sum_1^\infty \mu(E_n) < \infty.$$

Hence the continuity of measures (from above) and the hypothesis imply that

$$\mu(F) = \lim_m \mu(F_m) \le \sum_{n=m}^{\infty} \mu(E_n) = 0.$$

□

The Borel-Cantelli Lemma is an important tool, in particular in probability theory.

Example. (The doubling strategy.)

Assume that $(X, \mathcal{M}, \mathbb{P})$ is a probability space and suppose that $E_1, E_2, \ldots$ are events such that $\mathbb{P}(E_n) = 2^{-n}$, $n = 1, 2, \ldots$. Then by the Borel-Cantelli Lemma,

$$\mathbb{P}(\limsup_n E_n) = \mathbb{P}(E_n\, i.o.) = 0.$$

One way to describe this in words is the following. Suppose we play a sequence of games such that at the n'th game we win one c.u. with probability $1 - 2^{-n}$ and lose $2^n - 1$ c.u. with probability 2^{-n}. Each game is fair in terms of expectation, but by the Borel-Cantelli Lemma, we will almost surely lose money only finitely many times. Hence, over the whole infinite sequence of games, we will almost surely win an infinite amount of money. (In practice this strategy fails, of course, since there are always some bounds that will set things up, e.g. one can only play a certain number of games in a lifetime.) □

3.7 Carathéodory's Extension Theorem

A set function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is said to be an *outer measure* if

- $\mu^*(\emptyset) = 0$,
- $\mu^*(E) \le \mu^*(F)$ whenever $E \subseteq F$,
- $\mu^*(\bigcup_{n=1}^{\infty} E_n) \le \sum_{n=1}^{\infty} \mu^*(E_n)$ for all sets $E_1, E_2, \ldots$.

If μ^* is an outer measure, then we say that a set $A \in \mathcal{P}(X)$ is *μ^*-measurable* if, for all $E \in \mathcal{P}(X)$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

By the definition of outer measure, it is immediate that the left hand side is bounded by the right hand side, so to prove that a given set A is μ^*-measurable, it suffices to show that $\mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \cap A^c)$ for arbitrary E with $\mu^*(E) < \infty$.

Theorem 3.14 *(***Carathéodory's Extension Theorem***)*
Let $\mathcal{A}$ be an algebra on X and let $\mu_0 : \mathcal{A} \to [0, \infty]$ be a countably additive set function. Then there exists a measure μ on $\sigma(\mathcal{A})$ such that $\mu(A) = \mu_0(A)$ for all $A \in \mathcal{A}$. If $\mu_0(X) < \infty$, then μ is the unique such measure.

The uniqueness part follows immediately from Theorem 3.10. The existence part will be proved via a sequence of claims. These will also reveal some other useful facts, apart from the statement of the theorem.

Claim I. Let μ^* be an outer measure and let $\mathcal{M}$ be the collection of μ^*-measurable sets. Then $\mathcal{M}$ is a σ-algebra. Moreover, the restriction of μ^* to $\mathcal{M}$ is a complete measure.

Proof. It is obvious that $X \in \mathcal{M}$. From the symmetry between A and A^c in the definition of μ^*-measurability, it is also obvious that $\mathcal{M}$ is closed under complements. It remains to show that $\mathcal{M}$ is closed under countable unions.

Suppose that $A, B \in \mathcal{M}$ and let E be an arbitrary subset of X. Then $A \cup B \in \mathcal{M}$ since

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) \\
&= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c) \\
&= \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c
\end{aligned}$$

where the last inequality follows from that $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$, so that the definition of outer measure implies that the first three terms in the middle expression bound the first term in the last expression, and that $(A \cap B)^c = A^c \cap B^c$. Moreover, if $A \cap B = \emptyset$, then $(A \cup B) \cap A = A$ and $(A \cup B) \cap A^c = B$, so the applying the definition of μ^*-measurability of A with $E = A \cup B$ gives

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

In summary $\mathcal{M}$ is closed under finite unions and μ^* is additive on $\mathcal{M}$.

Now suppose that $A_j \in \mathcal{M}$, $j = 1, 2, \ldots$ are disjoint sets. Write $B_n = \bigcup_1^n A_j$ and $B = \bigcup_1^\infty A_j$. Let E be an arbitrary subset of X. By the μ^*-measurability of A_n,

$$\begin{aligned} \mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) \\ &= \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}) \end{aligned}$$

so by induction it follows that

$$\mu^*(E \cap B_n) = \sum_1^n \mu^*(E \cap A_j).$$

Above, we proved that $\mathcal{M}$ is closed under finite unions, so $B_n \in \mathcal{M}$ for each n. Hence

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) = \sum_1^n \mu^*(E \cap A_j) + \mu^*(E \cap B_n^c) \\ &\geq \sum_1^n \mu^*(E \cap A_j) + \mu^*(E \cap B^c). \end{aligned}$$

Letting $n \to \infty$ and using the definition of outer measure, it follows that

$$\begin{aligned} \mu^*(E) &\geq \sum_1^\infty \mu^*(A_j) + \mu^*(E \cap B^c) \geq \mu^*\Big(\bigcup_1^\infty (E \cap A_j)\Big) + \mu^*(E \cap B^c) \\ &= \mu^*(E \cap B) + \mu^*(E \cap B^c) \geq \mu^*(E). \end{aligned}$$

Hence all the inequalities must be equalities and it follows that $B \in \mathcal{M}$. This proves that $\mathcal{M}$ is closed under disjoint countable unions and it is an easy exercise to show that this entails that $\mathcal{M}$ is closed under arbitrary countable unions, i.e. $\mathcal{M}$ is a σ-algebra. Moreover, taking $E = B$ gives

$$\mu^*(B) = \sum_1^\infty \mu^*(A_j)$$

proving that the restriction of μ^* to $\mathcal{M}$ is a measure. It remains to prove completeness. Assume that $N \in M$, $\mu^*(N) = 0$ and $A \subseteq N$. Then $\mu^*(A) = 0$ by the definition of outer measure. Therefore

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leq \mu^*(E)$$

proving that $A \in \mathcal{M}$. □

Next assume that μ_0 is a countably additive set function on the algebra $\mathcal{A}$. Define $\mu^* : \mathcal{P}(X) \to [0, \infty]$ by

$$\mu^*(E) = \inf\{\sum_1^\infty \mu_0(A_j) : A_j \in \mathcal{A}, \bigcup_1^\infty A_j \supseteq E\}. \tag{1}$$

Claim II. μ^* is an outer measure.

Proof. It is trivial that $\mu^*(\emptyset) = 0$ and $E \subseteq F \Rightarrow \mu^*(E) \leq \mu^*(F)$. It remains to prove countable subadditivity. Fix $\epsilon > 0$. If $E_j \in \mathcal{P}(X)$, $j = 1, 2, \ldots$, then for each j one can find $A_j(k) \in \mathcal{A}$, $k = 1, 2, \ldots$ so that $\bigcup_k A_j(k) \supseteq E_j$ and $\sum_k \mu_0(A_j(k)) \leq \mu^*(E_j) + \epsilon 2^{-j}$. Since $\bigcup_{j,k} A_j(k) \supseteq \bigcup_j E_j$, we get

$$\mu^*(\bigcup_j E_j) \leq \sum_{j,k} \mu_0(A_j(k)) \leq \sum_j \mu^*(E_j) + \epsilon$$

and since ϵ was arbitrary,

$$\mu^*(\bigcup_j E_j) \leq \sum_j \mu^*(E_j)$$

as desired. □

For the final two claims, it is assumed that μ^* is defined by (1) and $\mathcal{M}$ is the σ-algebra of μ^*-measurable sets.

Claim III. $\mu^*(E) = \mu_0(A)$ for all $E \in \mathcal{A}$.

Proof. If $E \in \mathcal{A}$, take $E_1 = A$ and $E_2 = E_3 = \ldots = \emptyset$ in the definition of μ^* to see that $\mu^*(E) \leq \mu_0(E)$. Proving the reverse inequality amounts to showing that $\mu_0(A) \leq \sum_j \mu_0(A_j)$ whenever $A_j \in \mathcal{A}$ and $\bigcup_j A_j \supseteq E$. Let $B_n = E \cap (A_n \setminus \bigcup_1^{n-1} A_j$. Then the B_n's are disjoint and $\bigcup_n B_n = E$. By the countable additivity of μ_0, it follows that

$$\mu_0(E) \sum_n \mu_0(B_n) \leq \sum_n \mu_0(A_n).$$

□

Claim IV. $\mathcal{A} \subseteq \mathcal{M}$.

Proof. Pick $A \in \mathcal{A}$ and arbitrary $E \subseteq X$ and $\epsilon > 0$. By the definition of μ^*, there exist $B_j \in \mathcal{A}$ such that $\bigcup_j B_j \supseteq E$ and $\sum_j \mu_0(B_j) < \mu^*(E) + \epsilon$. We get, by the additivity of μ_0 on $\mathcal{A}$,

$$\begin{aligned} \mu^*(E) + \epsilon &> \sum_j \mu_0(B_j \cap A) + \sum_j \mu_0(B_j \cap A^c) \\ &\geq \mu^*(E \cap A) + \mu(E \cap A^c) \end{aligned}$$

where the last equality follows from the definition of μ^*. □

Taken together, these four claims prove Carathéodory's Theorem.

3.8 The Lebesgue measure and Lebesgue-Stieltjes measures

Up to now, we have not seen any concrete examples of non-trivial measures. When X is a countable space, $X = \{x_1, x_2, \ldots\}$, then it is easy to construct such measures. Take e.g. $\mathcal{M} = \mathcal{P}(X)$, let $\{w(x_n\}_{n=1}^{\infty}$ be any collection of non-negative numbers and let μ be defined by $\mu(A) = \sum_{x \in A} w(x)$. We have also seen that for $X = (0, 1]$ and $\mathcal{M} = \mathcal{P}(X)$, no sensible length measure exists. We are now equipped with the tools needed to construct a proper length measure on $\mathbb{R}$. Since it is not possible to do this for all subsets, we have to settle for a smaller σ-algebra. Clearly sets of the form constructed in Section 2 via the axiom of choice, are "unnatural" to expect to be able to measure in terms of length. On the other hand, any sensible length measure must be able to measure the length of an interval. If we could also measure the length of any set that can be constructed from a countable number of set operations on intervals, then it is difficult enough to come up with an example of a set which would not have a length (such as the set A in

Section 2) and even harder to motivate why one would even wish to give such a set a length if doing so causes problems. This point of view is what we are going to adopt.

Now recall that the Borel σ-algebra is the σ-algebra generated by all intervals and hence, by virtue of being a σ-algebra, contains all sets we wish to assign a length to. Hence the aim is to construct a length measure on $\mathcal{B}(\mathbb{R})$. It turns out to be slightly more comfortable to restrict to $(0,1]$ and $\mathcal{B}(0,1]$. Having done so, we obviously also have length measures on $(n,n+1]$ for all $n \in \mathbb{Z}$ by translation and can extend to the whole real line by letting, for $E \in \mathcal{B}(\mathbb{R})$, defining the length of E be the sum of the lengths of $E \cap (n,n+1]$, $n \in \mathbb{Z}$.

Let $X = (0,1]$ and let $\mathcal{A}$ be the algebra consisting of finite disjoint unions of intervals of the type $(a,b]$, $0 \leq a \leq b \leq 1$. Hence any $A \in \mathcal{A}$ can be written as $\bigcup_1^n (a_j, b_j]$ for some $n \in \mathbb{Z}_+$ and the $(a_j,b_j]$'s disjoint. Define $\mu_0 : \mathcal{A} \to [0,1]$ by

$$\mu_0(\bigcup_1^n (a_j,b_j]) = \sum_1^n (b_j - a_j).$$

Clearly the length of any set in $\mathcal{A}$ must be given by $\mu_0(A)$, so we would like to extend μ to a measure on $\mathcal{B}(0,1] = \sigma(\mathcal{A})$. By Carathéodory's Extension Theorem, there is a unique such extension, *provided* that μ_0 is a countably additive set function on $\mathcal{A}$. It is trivial that $\mu_0(\emptyset) = 0$ and that μ_0 is additive, but countable additivity is not so clear. It must be proved that $\mu_0(\bigcup_1^\infty A_n) = \sum_1^\infty \mu_0(A_n)$ whenever A_1, A_2 are disjoint sets in $\mathcal{A}$ and $\bigcup_1^\infty \in \mathcal{A}$. Since μ_0 is finitely additive, we may assume without loss of generality that the A_n's and A consist of a single interval: $A_n = (a_n, b_n]$ and $A = (a,b]$.

On one hand, by finite additivity,

$$\mu_0(A) = \mu_0(A \setminus \bigcup_1^n A_j) + \mu_0(\bigcup_1^n A_j) \geq \mu_0\Big(\bigcup_1^n A_j\Big) = \sum_1^n \mu_0(A_j)$$

for every n, so letting $n \to \infty$ gives

$$\mu_0(A) \geq \sum_1^\infty \mu_0(A_j).$$

Now we focus on the reverse inequality. Fix $\epsilon > 0$. The sets $(a_n - \epsilon 2^{-n}, b_n)$ form an open cover of the set compact set $[a, b-\epsilon]$ and can hence be reduced to a finite subcover $(a_n - \epsilon 2^{-n}, b_n)$, $n = 1, \ldots, N$. Let $c_n = a_n - \epsilon 2^{-n}$ and assume without loss of generality that $c_1 \le c_2 \le \ldots \le c_N$ (otherwise just reorder). We may also assume without loss of generality that $b_1 \le b_2 \le \ldots \le b_N$, otherwise discard those intervals that are contained in one of the others; this cannot increase $\sum_1^N (b_j - a_j)$. Then, since $b_j \le a_{j+1}$ for all $i = 1, \ldots, N-1$,

$$b - \epsilon - a \le b_N - c_1 \le \sum_1^N (b_j - c_j) \le \sum_1^N (b_j - a_j + \epsilon 2^{-j}) \le \sum_1^\infty (b_j - a_j) + \epsilon.$$

Hence

$$\mu_0(A) = b - a \le \sum_1^\infty \mu_0(A_j) + 2\epsilon.$$

This establishes that μ_0 is countably additive.

Hence μ_0 extends to a unique length measure μ on $\mathcal{B}(0,1]$. This measure is known as the *Lebesgue measure* and the notation we will use for it is m. Looking back on the proof of Carathéodory's Extension Theorem, we find that for sets $E \in \mathcal{B}(0,1]$ that are not in $\mathcal{A}$, $m(E)$ is explicitly expressed in terms of μ_0 by

$$\mu^*(E) = \inf\{\sum_n \mu_0(A_n) : A_n \in \mathcal{A}, \bigcup_1^\infty A_n \supseteq E\} \tag{2}$$

and m the restriction of the outer measure μ^* to $\mathcal{B}(0,1]$. Moreover, we recall that μ_0 actually extends to a complete measure on the σ-algebra $\mathcal{M}$ of μ^*-measurable sets. This σ-algebra contains $\mathcal{A}$ and hence $\mathcal{B}(0,1]$, but nothing says that it could not be larger. Indeed, it turns out that $\mathcal{M}$ equals the completion of $\mathcal{B}(0,1]$ with respect to m and that this σ-algebra is strictly larger than the Borel σ-algebra. The larger σ-algebra $\mathcal{M}$ is called the *Lebesgue σ-algebra*, denoted $\mathcal{L}(0,1]$. Since this extension comes at no extra cost, it will be assumed throughout that the Lebesgue measure is the complete measure defined on $\mathcal{L}(0,1]$, unless otherwise stated.

The construction of the Lebesgue measure can easily be generalized in the following way. Let $F : \mathbb{R} \to \mathbb{R}$ be a non-decreasing right-continuous function. Redefine the μ_0 above by

$$\mu_{0,F}(\bigcup_1^n A_j) = \sum_1^n (F(b_j) - F(a_j)).$$

An analogous argument shows that μ_0 is countably additive on $\mathcal{A}$ and hence extends to a unique measure μ_F on $\mathcal{B}(\mathbb{R})$. For sets $E \in \mathcal{B}(\mathbb{R}) \setminus \mathcal{A}$, (2) becomes

$$\mu_F^*(E) = \inf\{\sum_n \mu_{0,F}(A_n) : A_n \in \mathcal{A}, \bigcup_1^\infty A_n \supseteq E\} \tag{3}$$

and μ_F the restriction of μ_F^* to $\mathcal{B}(\mathbb{R})$. As for the Lebesgue measure, the σ-algebra $\mathcal{M}_F$ of μ_F^*-measurable sets is strictly larger than $\mathcal{B}(\mathbb{R})$ and the restriction of μ_F^* to $\mathcal{M}_F$ coincides with the completion of μ_F. In analogy with the Lebesque measure, we will henceforth take the notation μ_F to denote this completion unless otherwise stated. The measure μ_F thus constructed is called the *Lebesgue-Stieltjes measure* associated to F.

From (3) it follows (exercise!) that a Lebesgue-Stieltjes measure satisfies the following regularity properties, called outer regularity and inner regularity respectively.

Proposition 3.15 *For all $E \in \mathcal{M}_F$,*

$$\begin{aligned}\mu_F(E) &= \inf\{\mu_F(U) : U \text{ open}, U \supseteq E\} \\ &= \sup\{\mu_F(K) : K \text{ compact}, K \subseteq E\}.\end{aligned}$$

Another property in the same vein is the following.

Proposition 3.16 *For all $E \in \mathcal{M}_F$ and $\epsilon > 0$, there exists a set A, which is a finite union of open intervals, such that*

$$\mu_F(A \Delta E) < \epsilon.$$

3.9 The Cantor Set

For any $x \in \mathbb{R}$, we have $m(\{x\}) = 0$, so for any countable subset $E \subseteq \mathbb{R}$, $m(E) = 0$. Does the reverse implication also hold? I.e. are countable sets the only ones to have Lebesgue measure 0? The answer is no. The most well-known example is the *Cantor set*. It is constructed the following way. Let for $n = 1, 2, \ldots$,

$$D_n = \bigcup_{j=0}^{3^{n-1}} \Big((3j+1)3^{-n}, (3j+2)3^{-n}\Big).$$

Let $C_1 = [0, 1] \setminus D_1$ and recursively $C_n = C_{n-1} \setminus D_n$. Let $C = \bigcap_1^\infty C_n$. The set C is the Cantor set.

In words, the process is the following. Start with the closed unit interval with the open mid third removed; this is C_1. From the two closed intervals that make up C_1, remove from each of them the open mid third to get C_2. Now C_2 is the union of four closed intervals. Remove from each of these the open mid third to get C_3, etc. The Cantor set is the limiting set of this process. Clearly $m(C_n) = (2/3)^n$, so by the continuity of measures $m(C) = 0$.

On the other hand, C has the same cardinality as $(0, 1]$. To see this, write each number $x \in [0, 1]$ by its trinary expansion:

$$x = \sum_{n=1}^{\infty} a_n(x) 3^{-n}$$

where $a_n(x) \in \{0, 1, 2\}$. The expansion is unique for all x except those that are of the type $x = j3^{-n}$, $j \in \mathbb{Z}_+$, for which one can either choose an expansion ending with an infinite sequence of 0's or one ending with an infinite sequence of 2's. In such cases, we pick the latter expansion. Then

$$C = \{x \in \{0, 1\} : a_n(x) \in \{0, 2\} \text{ for each } n\}.$$

Hence, by mapping each 2 to 1, we see that C is in a 1-1-correspondence with the set of all binary expansions $\sum_1^\infty b_n 2^{-n}$, i.e. with $(0, 1]$.

4 Measurable functions / random variables

Let $(X, \mathcal{M}, \mu)$ be a measure space and let $(Y, \mathcal{N})$ be a measurable space.

Definition 4.1 *A function $f : X \to Y$ is said to be* $(\mathcal{M}, \mathcal{N})$-measurable *if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{N}$.*

So f is $(\mathcal{M}, \mathcal{N})$-measurable if $\{x \in X : f(x) \in A\}$ is $\mathcal{M}$-measurable whenever A is $\mathcal{N}$-measurable. In words, this could be phrased as that f is measurable if statements that "make sense" in terms of the values of f also "make sense" in terms of the values of x. See the probabilistic interpretation of this in the example below.

When one of the σ-algebras is understood, we may speak of f as simply $\mathcal{M}$-measurable or $\mathcal{N}$-measurable and if $\mathcal{M}$ and $\mathcal{N}$ are both understood, we may speak of f as simply measurable. If $(X, \mathcal{M}, \mu)$ is a probability space, an $(\mathcal{M}, \mathcal{N})$-measurable function is usually called a (Y-valued) *random variable*.

Example. Let $(X, \mathcal{M}, \mathbb{P})$ be a probability space and suppose $Y = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Let $\xi : X \to \mathbb{R}$ be a random variable. This means that ξ is a $(\mathcal{M}, \mathcal{B}(\mathbb{R}))$-measurable function, i.e.

$$\xi^{-1}(B) = \{x \in X : \xi(x) \in B\} \in \mathcal{M}$$

whenever $B \in \mathcal{B}(\mathbb{R})$. Hence $\mathbb{P}(\xi^{-1}(B)) = \mathbb{P}(\xi \in B)$ is defined for all Borel sets B. I.e. measurability means that it makes sense to speak of the probability that ξ belongs to B for any given Borel set B. $\square$

Clearly the composition of two measurable functions is measurable. More specifically, if $(Z, \mathcal{O})$ is a third measurable space, $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable and $g : Y \to Z$ is $(\mathcal{N}, \mathcal{O})$-measurable, then, since $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$, $g \circ f$ is $(\mathcal{M}, \mathcal{O})$-measurable.

The following result is an indispensable tool for proving that a given function is measurable.

Theorem 4.2 *Let $\mathcal{E}$ be a class of subsets of Y and assume that $\mathcal{N} = \sigma(\mathcal{E})$. Then $f : X \to Y$ is measurable if and only if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{E}$.*

Proof. The only if direction is trivial. Let $\mathcal{F} = \{A \in \mathcal{N} : f^{-1}(A) \in \mathcal{M}\}$. Since $\mathcal{F} \supseteq \mathcal{E}$, it suffices to show that $\mathcal{F}$ is a σ-algebra. The key is then to recall that f^{-1} commutes as an operator with the basic set operations, i.e. $f^{-1}(A^c) = f^{-1}(A)^c$ and

$$f^{-1}\Big(\bigcup_\alpha A_\alpha\Big) = \bigcup_\alpha f^{-1}(A_\alpha),\ f^{-1}\Big(\bigcap_\alpha A_\alpha\Big) = \bigcap_\alpha f^{-1}(A_\alpha)$$

for all A and A_α and α ranging over arbitrary index sets. Hence

- $X = f^{-1}(Y)$ and $X \in \mathcal{M}$ (since $\mathcal{M}$ is a σ-algebra), so $Y \in \mathcal{F}$,
- $A \in \mathcal{F} \Rightarrow f^{-1}(A) \in \mathcal{M} \Rightarrow f^{-1}(A)^c \in \mathcal{M} \Rightarrow f^{-1}(A^c) \in \mathcal{M} \Rightarrow A^c \in \mathcal{F}$,
- $A_n \in \mathcal{F}$, $n = 1, 2, \ldots \Rightarrow f^{-1}(A_n) \in \mathcal{M} \Rightarrow \bigcup_n f^{-1}(A_n) \in \mathcal{M} \Rightarrow f^{-1}(\bigcup_n A_n) \in \mathcal{M} \Rightarrow \bigcup_n A_n \in \mathcal{F}$.

$\square$

Corollary 4.3 *If X and Y are topological spaces and $\mathcal{M}$ and $\mathcal{N}$ are the Borel σ-algebras, then any continuous function is measurable.*

Proof. Let f be continuous and let $\mathcal{T}$ be the topology (i.e. the family of open sets) of Y. By the definition of continuity, $f^{-1}(U)$ is open for all $U \in \mathcal{T}$ and hence measurable by the definition of the Borel σ-algebra on X. Since $\mathcal{B}(Y) = \sigma(\mathcal{T})$ an application of Theorem 4.2 with $\mathcal{E} = \mathcal{T}$ gives the result. $\square$

Corollary 4.4 *A map $f : X \to \overline{\mathbb{R}}$ is (Borel)-measurable in either of the following cases*

- $f^{-1}[-\infty, a] \in \mathcal{M}$ *for all* $a \in \overline{\mathbb{R}}$,
- $f^{-1}[\infty, a) \in \mathcal{M}$ *for all* $a \in \overline{\mathbb{R}}$,
- $f^{-1}[a, \infty] \in \mathcal{M}$ *for all* $a \in \overline{\mathbb{R}}$,
- $f^{-1}(a, \infty] \in \mathcal{M}$ *for all* $a \in \overline{\mathbb{R}}$.

Since either of the four classes generate $\mathcal{B}(\overline{\mathbb{R}})$, the proofs follow on mimicking the proof of Corollary 4.3. Of course analogous statements are valid if $\overline{\mathbb{R}}$ is replaced with $\mathbb{R}$, $\mathbb{R}_+$ or $\overline{\mathbb{R}_+}$.

Example. Let X be the sample space of a random experiment. Then $\xi : X \to \mathbb{R}$ is a random variable iff $\{\xi \leq a\}$ is an event for all $a \in \mathbb{R}$. This is sometimes taken as the definition of a random variable in courses which want to present the necessary fundamentals without involving unnecessary measure-theoretic detail. $\square$

Theorem 4.5 *Let $f, g : X \to \overline{\mathbb{R}}$ be measurable and $\lambda \in \mathbb{R}$ a constant. Then $f + g$, λf and fg are all measurable functions. The same is true for $1/f$ provided that $f(x) > 0$ for all $x \in X$.*

Proof. We do $f + g$ and leave the other cases as exercises. By Corollary 4.4 it suffices to show that $\{x : f(x) + g(x) < a\} \in \mathcal{M}$ for all $a \in \overline{\mathbb{R}}$. However

$$\{x : f(x) + g(x) < a\} = \bigcup_{q \in \mathbb{Q}} \Big(\{x : f(x) < q\} \cap \{x : g(x) < a - q\} \Big) \in \mathcal{M}$$

since $\mathbb{Q}$ is countable and f and g are measurable. $\square$

Theorem 4.6 *Assume that $f_1, f_2, \ldots$ are measurable. Then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$ and $\liminf_n f_n$ are measurable. Moreover, the set $\{x : \lim_n f_n(x), \text{ exists}\}$ is measurable and if $\lim_n f_n(x)$ exists for all x, then $\lim_n f_n(x)$ is a measurable function.*

Proof. That $\sup_n f_n$ is measurable follows from the observation that $\{x : \sup_n f_n(x) \leq a\} = \bigcap_n \{x : f_n(x) \leq a\}$, a countable union of measurable sets. Since constant functions are trivially measurable, we get that $\inf_n f_n = 0 - \sup_n(-f_n)$ is measurable. Since $\limsup_n f_n = \inf_m \sup_{n\geq m} f_n$ and $\liminf_n f_n = \sup_m \inf_{n\geq m} f_n$, these are then also measurable. If $\lim_n f_n(x)$ exists for all x, then $\lim_n f_n = \liminf_n f_n = \limsup_n f_n$ and is hence measurable. Finally

$$\{x : \lim_n f_n(x) \text{ exists}\} = \{x : \limsup_n(x) - \liminf_n(x) = 0\}$$

is measurable by Theorem 4.5 (since $\{0\} \in \mathcal{B}(\overline{\mathbb{R}})$). $\square$

Example. Construction of a uniform random variable.

Let $(X, \mathcal{M}, \mathbb{P}) = ([0,1], \mathcal{B}, m)$ and $\xi(x) = x$, $x \in X$. Then ξ is continuous and hence a random variable and

$$\mathbb{P}(\xi \leq a) = m\{x : \xi(x) \leq x\} = m\{x : x \leq a\} = m[0,a] = a.$$

$\square$

Example. Construction of a random variable with given distribution.

Assume that $F : \mathbb{R} \to \mathbb{R}$ is non-decreasing and right continuous with

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

We want to construct a random variable ξ so that $\mathbb{P}(\xi \leq a) = F(a)$. Recall the Lebesque-Stieltjes measure μ_F. The conditions on F imply that μ_F is a probability measure, so let $(X, \mathcal{M}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}, \mu_F)$ and $\xi(x) = x$, $x \in \mathbb{R}$. Then

$$\mathbb{P}(\xi \leq a) = \mu_F(-\infty, a] = F(a).$$

An alternative construction is the following, which is most conveniently described in the case when F is continuous and strictly increasing. Then F^{-1} exists, so we can take $(X, \mathcal{M}, \mathbb{P}) = ([0,1], \mathcal{B}, m)$ and $\xi(x) = F^{-1}(x)$ and get

$$\mathbb{P}(\xi \leq a) = m\{x : F^{-1}(x) \leq a\} = m[0, F(a)] = F(a).$$

In the general case, one can replace F^{-1} with the generalized inverse, which maps all points in $[F(x-), F(x+)]$ to x and points $y \in [0,1]$ for which $F^{-1}(\{y\})$ is an

interval, which must have the form $[c, d)$ or $[c, d]$ since F is right continuous, to c.
$\square$

Example. Construction of a sequence of uniform random variables.

Again take $(X, \mathcal{M}, \mathbb{P}) = ([0,1], \mathcal{B}, m)$. Represent each $x \in [0,1]$ with its binary expansion

$$x = \sum_1^\infty a_n(x) 2^{-n}.$$

Each $a_n(x)$ is a $\{0,1\}$-valued measurable function of x, since $a_n^{-1}(\{1\})$ is a union of 2^{n-1} intervals (of length 2^{-n}). Let $\{n_{ij}\}_{j=1}^\infty$, $i = 1, 2, \ldots$ be disjoint sequences and let

$$\xi_i(x) = \sum_{j=1}^\infty a_{n_{ij}} 2^{-j}.$$

Then ξ is measurable for each i by Theorems 4.5 and 4.6 (why do we need them both?) and clearly $\mathbb{P}(\xi_i \le a) = a$ as in the first of the previous examples. $\square$

Example. Construction of a sequence of fair coin flips.

With the same setting as in the previous example, let simply $\xi_i(x) = a_i(x)$. $\square$

We end this section with a few notes on completeness. Suppose that g is $\mathcal{M}$-measurable and that $f = g$ a.e. If μ is complete, then this implies that f is measurable. However if μ is not complete, then this may not be the case. On the other hand, by the construction of the completion $\overline{\mu}$ of μ, it is clear that f is $\overline{\mathcal{M}}$-measurable. Similarly, if μ is complete, $f_1, f_2, \ldots$ measurable and $f_n \to f$ a.e., then f is measurable. (These facts make up Proposition 2.11 in Folland.) Vice versa, if f is $\overline{\mathcal{M}}$-measurable, then there exists an $\mathcal{M}$-measurable function such that $f = g$ $\overline{\mu}$-a.e. (This last fact is Proposition 2.12 in Folland.)

4.1 Product-σ-algebras and complex measurable functions

Let $(Y, \mathcal{N})$ be a measurable space and $f : X \to Y$. Then the *σ-algebra on X generated by f* is given by

$$\sigma(f) := \sigma\{f^{-1}(A) : A \in \mathcal{N}\}.$$

In other words, $\sigma(f)$ is the smallest σ-algebra on X that makes f measurable. (In fact $\{f^{-1}(A) : A \in \mathcal{N}\}$ is a σ-algebra (prove this!), so $\sigma(f)$ equals this set.)

More generally, if $\mathcal{F}$ is a family of functions from X to Y, then

$$\sigma(\mathcal{F}) := \sigma\{f^{-1}(A) : f \in \mathcal{F},\ A \in \mathcal{N}\}.$$

Now let $(X_1, \mathcal{M}_1)$ and $(X_2, \mathcal{M}_2)$ be two measurable spaces. The *projection maps* π_1 and π_2 are given by

$$\pi_i : X_1 \times X_2 \to X_i,\ \pi_i(x_1, x_2) = x_i$$

$i = 1, 2$.

Definition 4.7 *The product σ-algebra of $\mathcal{M}_1$ and $\mathcal{M}_2$ is given by*

$$\mathcal{M}_1 \times \mathcal{M}_2 := \sigma(\pi_1, \pi_2) = \sigma\{E_1 \times E_2 : E_i \in \mathcal{M}_i, i = 1, 2\}.$$

More generally

$$\prod_1^\infty \mathcal{M}_n = \sigma\{\pi_n : n = 1, 2, \ldots\} = \sigma\{\prod_1^\infty E_n : E_n \in \mathcal{M}_n\}$$

and for a general index set I

$$\begin{aligned} \prod_{\alpha \in I} \mathcal{M}_\alpha &= \sigma\{\pi_\alpha : \alpha \in I\} \\ &= \sigma\{\prod_{\alpha \in I} E_\alpha : E_\alpha \in \mathcal{M}\alpha \text{ and } E_\alpha = X_\alpha \text{ for all but countably many } \alpha\}. \end{aligned}$$

Make sure that you understand the equalities in the definitions.

Proposition 4.8 *Let $(X, \mathcal{M})$ and $(Y_\alpha, \mathcal{N}_\alpha)$, $\alpha \in I$, be measurable spaces. A map $h = (f_\alpha)_{\alpha \in I} : X \to \prod_{\alpha \in I} Y_\alpha$ is $(\mathcal{M}, \prod_{\alpha \in I} \mathcal{N}_\alpha)$-measurable if and only if each f_α is $(\mathcal{M}, \mathcal{N}_\alpha)$-measurable.*

Proof. Since $f_\alpha = \pi_\alpha \circ h$, a composition of two measurable maps, the only if direction holds. On the other hand, if all f_α are measurable, then for any α and $A \in \mathcal{N}_\alpha$,

$$h^{-1}(\pi_\alpha^{-1}(A)) = (\pi_\alpha \circ h)^{-1}(A) = f_\alpha^{-1}(A) \in \mathcal{M}.$$

Since $\prod_\alpha \mathcal{N}_a$ is generated by π_α, $\alpha \in I$, the if direction now follows from Theorem 4.2. $\square$

Proposition 4.9 $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$.

Proof. Let $\mathcal{A} = \{(a_1, b_1) \times (a_2, b_2) : a_1, b_1, a_2, b_2 \in \mathbb{Q}\}$. Since any open set in $\mathbb{R}^2$ can be written as a countable union of sets in $\mathcal{A}$, we have $\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{A})$. By definition $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$ contains $\mathcal{A}$ and hence $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \supseteq \mathcal{B}(\mathbb{R}^2)$.

On the other hand, $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$ is generated by $\pi^{-1}(A)$, $A \in \mathcal{B}(\mathbb{R})$, $i = 1, 2$. We have $\pi_1^{-1}(A) = A \times \mathbb{R}$, so it suffices to show that $A \times \mathbb{R} \in \mathcal{B}(\mathbb{R}^2)$ for every $A \in \mathcal{B}(\mathbb{R})$. (The similar statement for π_2 is of course analogous.) Since $A \times \mathbb{R}$ is open in $\mathbb{R}^2$ whenever A is open in $\mathbb{R}$, this holds for all open A. Hence, the family $\{A \in \mathcal{B}(\mathbb{R}) : A \times \mathbb{R} \in \mathcal{B}(\mathbb{R}^2)\}$ contains all open sets, so if we can show that it is also a σ-algebra, we are done. This, however, is obvious. $\square$

Two immediate corollaries follow.

Corollary 4.10 $\mathcal{B}(\mathbb{C}) = \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$.

Corollary 4.11 *A function $f : X \to \mathbb{C}$ is $(\mathcal{M}, \mathcal{B}(\mathbb{C}))$-measurable if and only if $\Re f$ and $\Im f$ are both measurable.*

4.2 Independent random variables

In the next sections $(X, \mathcal{M}, \mathbb{P})$ will be a probability space.

Definition 4.12 *Let I be an arbitrary set and let $\mathcal{E}_\alpha$, $\alpha \in I$, be subclasses of $\mathcal{M}$.*

- *We say that $\{\mathcal{E}_\alpha\}_{\alpha \in I}$ is independent if*

  $$\mathbb{P}(\bigcap_{j \in J} E_j) = \prod_{j \in J} \mathbb{P}(E_j)$$

  for all finite $J \subseteq I$ and all $E_j \in \mathcal{E}_j$, $j \in J$.

- *The family of random variables $\{\xi_\alpha\}_{\alpha \in I}$ said to be independent if $\{\sigma(\xi_\alpha)\}_{\alpha \in I}$ is independent.*

- *The family of events $\{E_\alpha\}_{\alpha \in I}$, is said to be independent if $\{\chi_{E_\alpha}\}_{\alpha \in I}$ is independent.*

The given definition is completely general in terms of the index set I. Although having I uncountable can be useful sometimes, e.g. when defining *Gaussian white noise*, it will not be so here, so in the sequel I will be either finite or countably infinite.

Lemma 4.13 *Assume that $\mathcal{I}, \mathcal{J} \subseteq \mathcal{M}$ are two π-systems and let $\mathcal{N} = \sigma(\mathcal{I})$ and $\mathcal{O} = \sigma(\mathcal{J})$. Then $\{\mathcal{N}, \mathcal{O}\}$ is independent if and only if $\{\mathcal{I}, \mathcal{J}\}$ is independent.*

Proof. The only if direction is trivial. The if direction will be proved by a two-step procedure. First fix arbitrary $I \in \mathcal{I}$ and define two measures on $\mathcal{O}$ by, for each $B \in \mathcal{O}$, setting

$$\mu_1(B) = \mathbb{P}(I \cap B)$$

$$\mu_2(B) = \mathbb{P}(I)\mathbb{P}(B)$$

By hypothesis μ_1 and μ_2 agree on $\mathcal{J}$ and $\mu_1(X) = \mu_2(X) \leq 1 < \infty$, so by the Uniqueness Theorem for measures, $\mu_1 = \mu_2$. Next fix arbitrary $B \in \mathcal{O}$ and define two measures on $\mathcal{N}$ by setting, for each $A \in \mathcal{N}$,

$$\mu_3(A) = \mathbb{P}(A \cap B)$$

$$\mu_4(A) = \mathbb{P}(A)\mathbb{P}(B).$$

By what we just proved, μ_3 and μ_4 agree on $\mathcal{I}$. They are also finite and agree on X, and are hence equal. This proves independence. $\square$

Clearly Lemma 4.13 extends to all finite collections of π-systems and their generated σ-algebras. Since independence of an infinite family of σ-algebras is equivalent to independence of finite subfamilies, Lemma 4.13 also extends to:

Corollary 4.14 *Let $\mathcal{I}_1, \mathcal{I}_2, \ldots \subseteq \mathcal{M}$ be π-systems. If $\{\mathcal{I}_1, \mathcal{I}_2, \ldots\}$ is independent, then also $\{\sigma(\mathcal{I}_1), \sigma(\mathcal{I}_2), \ldots\}$ is independent.*

The following two examples are important. First observe the following useful fact. Let $f : X \to (Y, \mathcal{N})$ and suppose that $\mathcal{E} \subseteq \mathcal{P}(Y)$ generates $\mathcal{N}$. Then $\{f^{-1}(E) : E \in \mathcal{E}\}$ generates $\sigma(f)$; this is so since $\{E \subseteq Y : f^{-1}(E) \in \sigma(f)\}$ is a σ-algebra, by the commutativity of inverse images and basic set operations.

Example. Let ξ and η be two random variables. Then $\{\xi^{-1}(-\infty, a] : a \in \mathbb{R}\}$ and $\{\eta^{-1}(-\infty, b] : b \in \mathbb{R}\}$ are π-systems and generate $\sigma(\xi)$ and $\sigma(\eta)$ respectively. Hence by Lemma 4.13 $\{\xi, \eta\}$ is independent iff $\mathbb{P}(\xi^{-1}(-\infty, a] \cap \eta^{-1}(-\infty, b]) = \mathbb{P}(\xi^{-1}(-\infty, a])\mathbb{P}(\eta^{-1}(-\infty, b])$ for all a, b, i.e. if

$$\mathbb{P}(\xi \leq a, \eta \leq b) = \mathbb{P}(\xi \leq a)\mathbb{P}(\eta \leq b)$$

for all $a, b \in \mathbb{R}$. More generally, by Corollary 4.14, $\{\xi_1, \xi_2, \ldots\}$ is independent iff

$$\mathbb{P}(\xi_{i_1} \leq a_1, \ldots, \xi_{i_n} \leq a_n) = \prod_{k=1}^{n} \mathbb{P}(\xi_{i_k} \leq a_k)$$

for all $n = 1, 2, \ldots$, all $1 \leq i_1 < \ldots < i_n$ and all $a_1, \ldots, a_n \in \mathbb{R}$. $\square$

For trivial reasons, $\{f(\xi), g(\eta)\}$ are independent whenever ξ and η are independent. (Check that you understand why!). Analogously, if $\{\{\xi_1, \xi_2, \ldots\}, \{\eta_1, \eta_2, \ldots\}\}$ is an independent pair of families of random variables (i.e. an independent pair of $\mathbb{R}^\infty$-valued random variables; there is nothing in the above definitions that prevents us from considering random variables taking on values in an arbitrary space), then $f(\xi_1, \xi_2, \ldots)$ and $g(\eta_1, \eta_2, \ldots)$ are independent.

It is intuitively clear that if $\{\xi_1, \xi_2, \ldots\}$ is independent, then, if we extract two disjoint subfamilies, these two should make an independent pair of $\mathbb{R}^\infty$-valued random variables. The next example shows that this is indeed the case.

Example. Let $\xi_1, \xi_2, \ldots$ be independent random variables and let I and J be two disjoint index sets (i.e. $I, J \subseteq \mathbb{N}$ and $I \cap J = \emptyset$). Then $\{\{\xi_{i_1} \leq a_1, \ldots, \{\xi_{i_n} \leq a_n\} : n = 1, 2, \ldots, i_1 < \ldots < i_n, a_1, \ldots, a_n \in \mathbb{R}\}$ is a π-system that generates $\sigma(\xi_i : i \in I)$ and the analogous π-system generates $\sigma(\xi_j : j \in J)$.

By the previous example, the two π-systems are independent. Hence the collections $(\xi_i : i \in I)$ and $(\xi_j : j \in J)$ are independent, by Corollary 4.14. $\square$

To relax our language a bit, let us take the statement "$\xi_1, \xi_2, \ldots$ are independent" to mean that the family $\{\xi_1, \xi_2, \ldots\}$ is independent. Note that it is actually important to spell this out, since another interpretation of the statement could have been that the random variables are all *pairwise* independent. This, however, is a much weaker statement. Consider for example the three $\{0, 1\}$-valued random variables ξ_1, ξ_2, ξ_3 given by $\mathbb{P}(\xi_1 = 0, \xi_2 = 0, \xi_3 = 1) = \mathbb{P}(\xi_1 = 0, \xi_2 = 1, \xi_3 = 0) = \mathbb{P}(\xi_1 = 1, \xi_2 = 0, \xi_3 = 0) = \mathbb{P}(\xi_1 = 1, \xi_2 = 1, \xi_3 = 1) = 1/4$, which are pairwise independent, but clearly not independent since any of them is the xor sum of the other two. Hence, in the sequel, saying that a set of random variables are independent means something stronger than saying that the same random variables are pairwise independent.

Theorem 4.15 *(***Borel-Cantelli's Second Lemma***)*

Let $E_1, E_2, \ldots$ be a sequence of independent events. If $\sum_1^\infty \mathbb{P}(E_n) = \infty$, then $\mathbb{P}(\limsup_n E_n) = 1$.

Proof. Note that

$$(\limsup_n E_n)^c = (\bigcap_m \bigcup_{n \geq m} E_n)^c = \bigcup_m \bigcap_{n \geq m} E_n^c$$

so by the continuity of measures, it suffices to show that $\mathbb{P}(\bigcap_{n\geq m} E_n^c) = 0$ for all m. This in turn follows from the following computations

$$\begin{aligned}\mathbb{P}(\bigcap_{n\geq m} E_n^c) &= \lim_r \mathbb{P}(\bigcap_m^r E_n^c) = \lim_r \prod_m^r \mathbb{P}(E_n^c) \\ &= \prod_m^\infty (1 - P(E_n)) \leq \prod_m^\infty e^{-\mathbb{P}(E_n)} = e^{-\sum_m^\infty \mathbb{P}(E_n)} = 0\end{aligned}$$

□

Example. Let $\xi_1, \xi_2, \ldots$ be independent random variables with exponential(1) distribution, i.e.

$$\mathbb{P}(\xi > x) = e^{-x},\ x \geq 0.$$

Then

$$\mathbb{P}\Big(\frac{\xi_n}{\log n} > a\Big) = e^{-a \log n} = n^{-a}.$$

Hence $\sum_1^\infty \mathbb{P}(\xi_n > a \log n)$ is finite for $a > 1$ and infinite for $a \leq 1$. By the Borel-Cantelli Lemmas, this entails that

- if $a \leq 1$, then almost surely $\xi_n > a \log n$ for infinitely many n,
- if $a > 1$, then almost surely $\xi_n > a \log n$ for only finitely many n.

□

4.3 Kolmogorov's 0-1-law

Let $\xi_1, \xi_2, \ldots$ be independent random variables. For each n, let

$$\mathcal{T}_n = \sigma(\xi_{n+1}, \xi_{n+2}, \ldots)$$

and

$$\mathcal{T} = \bigcap_n \mathcal{T}_n.$$

The σ-algebra $\mathcal{T}$ is called the *tail-σ-algebra* (w.r.t. $\xi_1, \xi_2, \ldots$). A set $E \in \mathcal{T}$ is called a *tail event* and a random variable which is $\mathcal{T}$-measurable is called a *tail function* of the ξ_n's.

A tail event does not, for any n, depend on the first n of the ξ_k's, so at a first glance it may seem that $\mathcal{T}$ should be trivial. This, however, would be the wrong impression, since $\mathcal{T}$ actually contains a lot of interesting events. E.g. the event $\{x \in X : \lim_n \xi_n(x) \text{ exists}\}$ is a tail event and $\eta = \limsup_n(\frac{1}{n})\sum_1^n \xi_k$ is a tail function; they are $\mathcal{T}_n$-measurable of every n and hence $\mathcal{T}$-measurable. Kolmogorov's 0-1-law states that the probability for a tail event must be either 0 or 1 and that any tail function must be a constant a.s.

Theorem 4.16 *(***Kolmogorov's 0-1-law***)*

Let $\xi_1, \xi_2, \ldots$ be independent random variables.

(i) If $E \in \mathcal{T}$, then $\mathbb{P}(E) \in \{0, 1\}$,

(ii) If η is $\mathcal{T}$-measurable, then there exists a constant $c \in \mathbb{R}$ such that $\eta = c$ a.s.

Proof.

(i) Let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, $n = 1, 2, \ldots$. By the above example, $\mathcal{F}_n$ and $\mathcal{T}_n$ are independent. Since $\mathcal{T} \subseteq \mathcal{T}_n$, $\mathcal{F}_n$ and $\mathcal{T}$ are independent for every n. Hence $\bigcup_n \mathcal{F}_n$ and $\mathcal{T}$ are independent. Since $\bigcup_n \mathcal{F}_n$ is a π-system, it follows that $\sigma(\bigcup_n \mathcal{F}_n)$ and $\mathcal{T}$ are independent. However $\mathcal{T} \subseteq \sigma\{\xi_1, \xi_2, \ldots) = \sigma(\bigcup_n \mathcal{F}_n)$, so $\mathcal{T}$ is independent of itself. This means that for each $E \in \mathcal{T}$,

$$\mathbb{P}(E) = \mathbb{P}(E \cap E) = \mathbb{P}(E)^2$$

which entails that $\mathbb{P}(E)$ is either 0 or 1.

(ii) For all $a \in \mathbb{R}$, $\mathbb{P}(\eta \leq a) \in \{0, 1\}$ by (i). Let $c = \inf\{a : \mathbb{P}(\eta \leq a) = 1\}$. Then

$$\mathbb{P}(\eta \leq c) = \mathbb{P}\Big(\bigcap_n \{x : \eta(x) \leq c + \frac{1}{n}\}\Big) = 1$$

and

$$\mathbb{P}(\eta < c) = \mathbb{P}\Big(\bigcup_n \{x : \eta(x) \leq c - \frac{1}{n}\}\Big) = 0.$$

□

Example. (Monkey typing Shakespeare)

Suppose that a monkey is typing uniform random keys on a laptop. There are, say, N keys on the laptop. The collected works of Shakespeare (to be abbr. CWS) comprises, say, M symbols. Let E be the event that the monkey happens to type CWS eventually. Will E occur?

If we let F be the event that the monkey types CWS infinitely many times, then by Kolmogorov's 0-1-law, $\mathbb{P}(F)$ is 0 or 1. Let F_n be the event that the monkey types CWS with the $nM+1$'th to $(n+1)M$'th symols it types. Then $\mathbb{P}(F_n) = 1/N^m$, so $\sum_n \mathbb{P}(F_n) = \infty$ and hence $\mathbb{P}(\limsup_n F_n) = 1$ by Borel-Cantelli. Hence

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

So the answer is yes, the monkey will eventually type CWS (but, of course, very much provided that it has an infinite life and can be persuaded to spend an infinite amount of time at the laptop). $\square$

Note that they key in the example was really Borel-Cantelli's Second Lemma and that the information provided by Kolmogorov's 0-1-law was only that $\mathbb{P}(F) \in \{0, 1\}$. In the next example, the 0-1-law plays a more vital role.

Example. (Percolation)

Consider the two-dimensional integer lattice, i.e. the graph obtained by placing a vertex at each integer point (n, k) in the Euclidean plane and placing an edge between (n, k) and (m, j) if either $n = m$ and $|k - j| = 1$ or $k = j$ and $|n - m| = 1$. Now remove edges at random by letting each edge be kept (or open[1]) with probability p and removed (or closed) with probability $1 - p$, independently of other edges. The resulting random graph will of course a.s. fall into (infinitely many) connected components. However, will there be an infinitely large connected component?

Let E be the event that an infinite connected component exists. Let ξ_i be the status, i.e. kept or removed, of edge number i; here assume that edges are numbered according to their distance from the origin and arbitrarily among those edges that are equally far away. Now observe that E is a tail event. This is so since the presence or absence of infinite components cannot be changed by changing the status of the first n edges no matter the value of n. (For an outcome where infinite components exist, changing a finite number of edges can change the number of such components, but never change presence/absence.) Hence, by Kolmogorov's 0-1-law, $\mathbb{P}(E)$ is 0 or 1.

Determining for what p we have $\mathbb{P}(E) = 0$ and for what p we have $\mathbb{P}(E) = 1$

[1]Percolation theory has its origins in the study of water flow through porous materials. The edges then represent microscopic channels which may or may not be open for water flow.

is a different story. This is of course a general fact about applications of Kolmogorov's 0-1-law; it tells us that a tail event has probability 0 or 1, but never tells which it is. However, knowing that $\mathbb{P}(E)$ is 0 or 1 is still very helpful since if we can also show that $\mathbb{P}(E) > 0$, then it follows immediately that $\mathbb{P}(E) = 1$.

In the percolation setting of this example, consider the probability that no vertex in the $2n \times 2n$-box centered at the origin, is part of an infinite path of kept edges. It can be shown that this probability is bounded by $n(3(1-p))^n$. (This is done by bounding the number of ways that the box can be "cut off from infinity".) This is less than 1 for large enough n if $p > 2/3$. Hence $\mathbb{P}(E) = 1$ for $p > 2/3$. On the other hand, by similar counting, it is easy to see that $\mathbb{P}(E) = 0$ for $p < 1/3$. In fact, the critical probability for when $\mathbb{P}(E)$ switches from 0 to 1 is $p = 1/2$. This a central and highly non-trivial fact of percolation theory. (When $p = 1/2$, then $\mathbb{P}(E) = 0$.) $\square$

5 Integration of nonnegative functions

Defining the Lebesgue integral is a stepwise procedure. It starts with nonnegative *simple* functions.

Definition 5.1 *A function $\phi : (X, \mathcal{M}, \mu) \to \mathbb{C}$ is said to be simple if it is of the form*

$$\phi(x) = \sum_1^n z_j \chi_{E_j}(x)$$

for some n, where $z_j \in \mathbb{C}$ and $\{E_1, \ldots, E_n\}$ is a partition of X such that $E_j \in \mathcal{M}$ for all j.

Let $L^+(X, \mathcal{M}, \mu)$ denote the set of all $\mathcal{M}$-measurable functions $f : X \to [0, \infty]$. Depending on the level of risk for confusion, we often use shorthand notations such as $L^+(X)$, $L^+(\mathcal{M})$ or simply L^+.

Definition 5.2 *Let $\phi = \sum_1^n a_j \chi_{E_j}$, $a_j \in \mathbb{R}_+$ be simple. Then the integral of ϕ with respect to μ is given by*

$$\int_X \phi(x) d\mu(x) := \sum_1^n a_j \mu(E_j).$$

Example. Let $(X, \mathcal{M}, \mu) = ([0,1], \mathcal{L}, m)$ and $\phi = \chi_{\mathbb{Q}\cap[0,1]}$. Since $\mathbb{Q} \cap [0,1]$ is countable, it is measurable, so ϕ is a simple function and $\int \phi dm = 0$. Compare this with what happens if we try to calculate the Riemann integral of this function. Since the Riemann integral is defined in terms of approximations of ϕ from above and from below by simple functions that are constant intervals, we find that the Riemann integral of ϕ is not defined. Thus, there are functions defined on an interval of the real line which the Lebesgue integral can handle, but which the Riemann integral cannot. Later, we will also see that any Riemann integrable function on an interval is Lebesgue integrable and that for such functions, the two methods give the same result. □

Alternative and/or shorthand notations for the integral are $\int_X \phi(x)\mu(dx)$, $\int \phi d\mu$ and $\int \phi$. The representation of a simple function as a finite linear combination of characteristic functions is of course not unique, but it is easy to see that different representations give the same result, so the integral is well-defined. For $A \in \mathcal{M}$, write

$$\int_A \phi d\mu := \int_X \phi \chi_A d\mu.$$

This is well-defined since $\phi\chi_A = \sum_1^n a_j \chi_{A\cap E_j} + 0 \cdot \chi_{A_c}$ is simple. A few basic facts follow.

Proposition 5.3 *Let $c \in \mathbb{R}_+$ and $\phi = \sum_1^n a_j \chi_{E_j}, \psi = \sum_1^m b_j \chi_{F_j} \in L^+$ be simple functions. Then*

(a) $\int c\phi = c \int \phi$,

(b) $\int(\phi + \psi) = \int \phi + \int \psi$,

(c) $\phi \leq \psi \Rightarrow \int \phi \leq \int \psi$,

(d) The map $A \to \int_A \phi$, $A \in \mathcal{M}$, is a measure.

Proof. Part (a) is trivial. For part (b) observe that

$$\phi + \psi = \sum_i \sum_j (a_i + b_j)\chi_{E_i \cap F_j}.$$

Hence

$$\begin{aligned}\int(\phi+\psi) &= \sum_i \sum_j (a_i + b_j)\mu(E_i \cap F_j) = \sum_i a_i \sum_j \mu(E_i \cap F_j) + \sum_j b_j \sum_i \mu(E_i \cap F_j) \\ &= \sum_i a_i \mu(E_i) + \sum_j b_j \mu(F_j) = \int \phi + \int \psi.\end{aligned}$$

For part (c) use the representations $\phi = \sum_i \sum_j a_i \chi E_i \cap F_j$ and $\psi = \sum_i \sum_j b_j \chi E_i \cap F_j$. On each $E_i \cap F_j$ we have $a_i \leq b_j$, so the result follows immediately from the definition.

To prove part (d), it must be shown that if $A_1, A_2, \ldots$ are disjoint sets in $\mathcal{M}$, then $\int_{\cup_k A_k} \phi = \sum_k \int_{A_k} \phi$. We have

$$\begin{aligned}
\int_{\cup_{k=1}^{\infty} A_k} \phi &= \sum_{j=1}^{n} a_j \mu\Big(E_j \cap (\bigcup_k A_k)\Big) = \sum_{j=1}^{n} \sum_{k=1}^{\infty} a_j \mu(E_j \cap A_k) \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^{n} a_j \mu(E_j \cap A_k) = \sum_{k=1}^{\infty} \int_{A_k} \phi
\end{aligned}$$

where the second equality is countable additivity of μ. □

The next step is to define integrals of arbitrary functions in L^+ by approximating with simple functions. The following approximation result tells us that it makes sense to do so.

Theorem 5.4 *(a) Let $f \in L^+$. There are simple functions $\phi_n \in L^+$ such that $\phi_n(x) \uparrow f(x)$ for every $x \in X$.*

(b) Let $f : X \to \mathbb{C}$ be measurable. Then there are simple functions ϕ_n such that $|\phi_1| \leq |\phi_2| \leq \ldots \leq |f|$ and $\phi_n \to f$ pointwise.

Proof. In (a), let $A_j = \{x : f(x) \in [j2^{-n}, (j+1)2^{-n})\}$, $j = 0, \ldots, n2^n - 1$ and let $A_{n2^n} = \{x : f(x) \geq n\}$. Since f is measurable, all these sets are measurable, so letting

$$\phi_n(x) = \sum_{0}^{n2^n} j 2^{-n} \chi_{A_j}(x)$$

gives ϕ_n's of the desired form.

For (b), apply the proof of (a) to all four *parts* of f; see below for definitions. □

In the light of Theorem 5.4, we make the following definition.

Definition 5.5 *Let $f \in L^+$. Then*

$$\int_X f(x) d\mu(x) := \sup \Big\{ \int_X \phi(x) d\mu(x) : 0 \leq \phi \leq f,\ \phi \textit{ simple} \Big\}.$$

For $A \in \mathcal{M}$,

$$\int_A f d\mu := \int f \chi_A d\mu.$$

It is obvious that if $c \in \mathbb{R}_+$, then $\int cf = c \int f$ and if $f \leq g$, $f, g \in L^+$, then $\int f \leq \int g$. The next result is one of the key results in integration theory.

Theorem 5.6 *(**The Monotone Convergence Theorem**)*
Assume that $f_n, f \in L^+$ and $f_n \uparrow f$ pointwise. Then $\int f_n d\mu \uparrow \int f d\mu$.

Proof. Since $\{f_n\}$ is increasing, $\{\int f_n\}$ is increasing and hence $\lim_n \int f_n$ exists (but may be equal to ∞). Since $f_n \leq f$ for all n, $\lim_n \int f_n \leq \int f$.

Now pick an arbitrary simple function $\phi \in L^+$ such that $\phi \leq f$ and an arbitrary $a \in (0, 1)$. Since $f_n \uparrow f$ pointwise, the sets $A_n := \{x : f_n(x) > a\phi(x)\}$ are increasing in n and $\bigcup_n A_n = X$. Since the map $A \to \int_A \phi$ is a measure, it follows from the continuity of measures that $\int_{A_n} \phi \uparrow \int \phi$. Therefore

$$\lim_n \int f_n \geq a \liminf_n \int_{A_n} \phi = a \int \phi.$$

Since a was arbitrary, letting $a \uparrow 1$ entails that $\lim_n \int f_n \geq \int \phi$. The result now follows from the definition of $\int f$. □

The first consequence of the MCT is that the integral is additive. Indeed, it is in fact countably additive:

Theorem 5.7 *Let $f_n \in L^+$. Then*

$$\int (\sum_1^\infty f_n) d\mu = \sum_1^\infty \int f_n d\mu.$$

Proof. First consider finite additivity. By Theorem 5.4, there are simple nonnegative functions ϕ_n and ψ_n such that $\phi_n \uparrow f_1$ and $\psi_n \uparrow f_2$ pointwise. By the MCT and Proposition 5.3,

$$\int (f_1 + f_2) = \lim_n \int (\phi_n + \psi_n) = \lim_n \int \phi_n + \lim_n \int \psi_n = \int f_1 + \int f_2.$$

Now finite additivity follows by induction. Since $\sum_1^N f_n \uparrow \sum_1^\infty f_n$ as $N \to \infty$, another application of the MCT now shows that

$$\int (\sum_1^\infty f_n) = \lim_N \int (\sum_1^N f_n) = \lim_N \sum_1^N \int f_n = \sum_1^\infty \int f_n.$$

$\square$

Corollary 5.8 *Let $f \in L^+$. Then the map $A \to \int_A f d\mu$, $A \in M$, is a measure.*

The hypothesis in the MCT is that $f_n \uparrow f$ pointwise. This can be relaxed a bit; it suffices to have $f_n \uparrow f$ a.e. To see this, first observe that if $\int f = 0$, then we can find simple $\phi_n \in L^+$ with $\phi_n \uparrow f$ pointwise and $\int \phi_n = 0$. However, since ϕ_n is simple, this trivially means that $\phi_n = 0$ a.e. Now if x is a point such that $f(x) > 0$, then $\phi_n(x) > 0$ for all sufficiently large n. Hence

$$\mu\{x : f(x) > 0\} \leq \mu\Big(\bigcup_n \{x : \phi_n(x) > 0\}\Big) = 0.$$

In summary

Proposition 5.9 *Let $f \in L^+$. Then $\int f d\mu = 0$ if and only if $f = 0$ a.e.*

Suppose now that $f_n \uparrow f$ a.e. and let $E = \{x : f_n(x) \to f(x)\}$. Then $f_n\chi_E \to f\chi_E$ pointwise so by the MCT, $\int f_n\chi_E \to \int f\chi_E$. Since $f - f\chi_E \in L^+$ and $f - f\chi_E = 0$ a.e., Proposition 5.9 implies that $\int f\chi_E = \int f$. From the same argument, $\int f_n\chi_E = \int f_n$. Putting these facts together gives $\int f_n \to \int f$. (This result is Corollary 2.17 in Folland.)

The MCT states that if $f_n \in L^+$ and $f_n \uparrow f$ a.e., then $\int f_n \to \int f$, but what about when $f_n \to f$ without being increasing in n? Does this also imply $\int f_n \to \int f$? The answer is no, as the following example shows. Let $(X, \mathcal{M}, \mu) = ([0,1], \mathcal{L}, m)$ and $f_n(x) = n\chi_{[0,1/n]}(x)$. Then $f_n \to 0$ a.e. (but not pointwise, since $f_n(0) \to \infty$), but $\int f_n = 1$ for every n.

Hence some further assumption is needed to guarantee that $f_n \to f$ a.e. entails that $\int f_n \to \int f$. Such a condition will be given in the Dominated Convergence Theorem below. Before that, we will extend the integral from nonnegative real functions to general complex functions. First however, we finish the present section with the important Fatou's Lemma and a note on σ-finiteness.

Theorem 5.10 *(**Fatou's Lemma**) If $f_n \in L^+$, $n = 1, 2, \ldots$, then*

$$\int (\liminf_n f_n) d\mu \leq \liminf_n \int f_n d\mu.$$

Proof. Note that $\inf_{n\geq m} f_n$ is increasing in m, so by the MCT,

$$\begin{aligned} \int \liminf f_n &= \int \lim_m(\inf_{n\geq m} f_n) = \lim_m \int \inf_{n\geq m} f_n \\ &\leq \lim_m \inf_{n\geq m} \int f_n = \liminf_n \int f_n \end{aligned}$$

where the inequality follows from that $\inf_{n\geq m} f_n \leq f_n$, and hence $\int \inf_{n\geq m} \leq \int f_n$, for every $n \geq m$. $\square$

An immediate consequence of Fatou's Lemma is that $\int f d\mu \leq \liminf_n \int f_n d\mu$ whenever $f_n \in L^+$ and $f_n \to f$ a.e.

The final result of this section makes the observation that if $f \in L^+$ and $\int f < \infty$, then $\mu\{x : f(x) < \infty\} = 0$, which is obvious, and that μ can always be regarded to be σ-finite as far as f is concerned: $\{x : f(x) > 0\} = \bigcup_n \{x : f(x) > 1/n\}$ is σ-finite. This is stated in Folland as Proposition 2.20. The last result extends to the conclusion that $\bigcup_n \{x : f_n(x) > 0\}$ is σ-finite whenever $\int f_n d\mu < \infty$ for all n.

6 Integration of complex functions

Consider a function $f : (X, \mathcal{M}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Define the two functions f^+ and f^- by

$$f^+(x) = \max(f(x), 0)$$

and

$$f^-(x) = f^+(x) - f(x) = -\min(0, f(x)).$$

These two nonnegative functions are called the *positive part* and *negative part* of f respectively.

Definition 6.1 *A function $f : (X, \mathcal{M}, \mu) \to (\mathbb{R}, \mathcal{B})$ is said to be integrable if $\int f^+ d\mu < \infty$ and $\int f^- d\mu < \infty$. The integral of an integrable function f is given by*

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

A function $f : (X, \mathcal{M}, \mu) \to (\mathbb{C}, \mathcal{B})$ is said to be integrable if $\Re f$ and $\Im f$ are both integrable, and the integral of f is then given by

$$\int f d\mu = \int (\Re f) d\mu + i \int (\Im f) d\mu.$$

The integral of a complex function is well-defined since measurability of f is equivalent to measurability of its real and imaginary parts, by Corollary 4.11. It is easy to see that the integral is linear and that f is integrable iff $\int |f| < \infty$.

By $L^1(X, \mathcal{M}, \mu)$ we will mean the space of all integrable complex functions on X. Simplified notations are $L^1(X)$, $L^1(\mathcal{M})$, $L^1(\mu)$ or just L^1 when these can be used without risk of confusion. The space L^1 is, as we just observed, a complex vector space.

Proposition 6.2 *For any $f \in L^1$,*

$$\Big|\int f d\mu\Big| \leq \int |f| d\mu.$$

Proof. For real-valued f, this is just the ordinary triangle inequality:

$$\Big|\int f\Big| = \Big|\int f^+ - \int f^-\Big| \leq \int f^+ + \int f^- = \int |f|.$$

For the general case, represent complex numbers z as $z = |z|\mathrm{sgn} z$. Then $|\int f| = \alpha \int f$, where $\alpha = \overline{\mathrm{sgn}(\int f)}$. Thus

$$\begin{aligned}\Big|\int f\Big| &= \int \alpha f = \Re \int \alpha f = \int \Re(\alpha f)\\ &\leq \int |\Re(\alpha f)| \leq \int |\alpha f| = |\alpha| \int |f| = \int |f|.\end{aligned}$$

□

Proposition 6.3 *Let $f, g \in L^1$. Then*

(a) $\{x : f(x) \neq 0\}$ *is σ-finite,*

(b) $\int_E f = \int_E g$ *for all $E \in \mathcal{M}$ iff $\int |f - g| d\mu = 0$ iff $f = g$ a.e.*

Proof. Part (a) is the corresponding result for nonnegative functions applied to the four parts of f. We also saw in the previous section that $\int |f - g| = 0$ iff $|f - g| = 0$ a.e., so the second equivalence in (b) holds. For the if direction in (b): if $\int |f - g| = 0$, then for each $E \in \mathcal{M}$,

$$\Big|\int_E f - \int_E g\Big| = \Big|\int (f - g)\Big| \leq \int |f - g| = 0.$$

For the only if direction: Assume for contradiction that $\int_E (f-g) = 0$ for all $E \in \mathcal{M}$ and $\mu\{|f-g| > 0\} > 0$. Writing $f - g = u + iv$, we must then have that at least one of the four parts u^+, u^-, v^+ and v^- is nonzero on a set of positive measure. Assume without loss of generality that this holds for u^+, so that with $\mu\{x : u^+(x) > 0\} > 0$. Then with n sufficiently large and $E = \{x : u^+(x) > 1/n\}$ has $\mu(E) > 0$. Then, since $u^- = 0$ on E,

$$\Re \int_E (f-g) \geq \frac{1}{n}\mu(E) > 0$$

a contradiction. □

Remark. In the notation L^1 for the space of integrable functions, it is usually understood that the space is normed with the L^1-*norm* given by

$$\|f\|_1 := \int |f-g| d\mu.$$

There is actually a slight problem with this, since $\|f-g\|_1 = 0$ only implies that $f = g$ a.e. and not that f and g are identical functions. This is solved by defining equivalence classes of integrable functions by saying that f and g belong to the same equivalence class if they are equal a.e. Then these equivalence classes are formally taken to be the elements in L^1. Then a particular function f is not really an element of L^1, but rather a representative of its equivalence class. This distinction, however, will not cause any problems in this course.

Theorem 6.4 *(*The Dominated Convergence Theorem*)*
Assume that $f_1, f_2, \ldots \in L^1$ and $f_n \to f$ a.e. Assume also that there exists an integrable $g \in L^+$ such that $|f_n| \leq g$ for every n. Then

$$\int f_n d\mu \to \int f d\mu.$$

Strictly speaking, that $f_n \to f$ a.e. does not imply that f is measurable. If μ is complete, then measurability of f follows. If not, the at least f will be measurable after an alternation on a null set. Let us suppress this concern and simply assume that f is measurable.

Proof. Assume without loss of generality that the f_n's and f are real-valued. Then $g + f_n$ and $g - f_n$ are nonnegative by assumption. Hence Fatou's Lemma gives on one hand

$$\int g + \int f = \int (g+f) \leq \liminf_n \int (g + f_n) = \int g + \liminf_n \int f_n$$

and on the other

$$\int g - \int f = \int (g - f) \leq \liminf_n \int (g - f_n) = \int g - \limsup_n \int f_n.$$

$\square$

The DCT allows us to prove that the integral is countably additive under the right assumption.

Theorem 6.5 *Assume that $f_n \in L^1$ and $\sum_1^\infty \int |f_n| d\mu < \infty$. Then $g := \sum_1^\infty |f_n|$ is integrable and $\int (\sum_1^\infty f_n) d\mu = \sum_1^\infty \int f_n d\mu$.*

Proof. Since

$$\int g = \int \sum_1^\infty |f_n| = \sum_1^\infty \int |f_n| < \infty$$

by Theorem 5.7, g is integrable and $\sum_1^\infty |f_n(x)| < \infty$ for a.e. x, so that $\sum_1^\infty f_n(x)$ exists for a.e. x. Since $\sum_1^N |f_n| \leq g$ for every N, the DCT implies that

$$\int (\sum_1^\infty f_n) = \lim_N \int (\sum_1^N f_n) = \lim_N \sum_1^N \int f_n = \sum_1^\infty \int f_n.$$

$\square$

The next result states that the set of simple functions is dense in L^1.

Theorem 6.6 *If $\epsilon > 0$ and $f \in L^1$, then there exists a simple function $\phi = \sum_1^m a_j \chi_{E_j}$, $a_j \in \mathbb{C}$, such that*

$$\|f - \phi\|_1 < \epsilon.$$

If μ is a Lebesgue-Stieltjes measure on $\mathbb{R}$, then the E_j's can be taken to be open intervals. Moreover, there exists a continuous function g such that $\|f - g\|_1 < \epsilon$.

Proof. By Theorem 5.4(b), there are simple functions ϕ_k such that $\phi_k \to f$ pointwise and $|\phi_k| \leq |f|$ for every k. Then $|\phi_k - f| \leq 2|f|$ so by the DCT,

$$\int |\phi_k - f| \to 0.$$

Now take $\phi = \phi_k$ for sufficiently large k.

Assume now that μ is a Lebesgue-Stieltjes measure and ϕ as in the statement of the theorem. We have, if the a_j's are nonzero,

$$\mu(E_j) = \frac{1}{|a_j|}\int |\phi| \leq \frac{1}{|a_j|}\int |f| < \infty.$$

Hence, by Proposition 3.16, there exists a set A_j which is the finite union of open intervals, such that $\mu(A_j \Delta E_j) < \epsilon/(m|a_j|)$. Let $\psi = \sum_1^m a_j \chi_{A_j}$. Then $\int |\phi - \psi| < \epsilon$. The final assertion follows from that the characteristic function $\chi_{(a,b)}$ of an open interval can be arbitrarily well approximated by the continuous function which is 0 outside (a, b), 1 on $[a + \delta, b - \delta]$ and linear on the remaining pieces. $\square$

Consider two measurable spaces $(X_1, \mathcal{M}_1)$ and $(X_2, \mathcal{M}_2)$. For sets $E \in \mathcal{M}_1 \times \mathcal{M}_2$, Fix $x_2 \in X_2$ and define the set $E_{x_2} = \{x_1 \in X_1 : (x_1, x_2) \in E\}$. Let $\mathcal{F}$ be the family of sets in $E \in \mathcal{M}_1 \times \mathcal{M}_2$ such that $E_{x_2} \in \mathcal{M}_1$. Then $\mathcal{F}$ contains all sets of the form $E_1 \times E_2$, $E_j \in \mathcal{M}_j$, by the definition of product-σ-algebras. It is also easy to see that $\mathcal{F}$ is a σ-algebra. Hence $\mathcal{F} = \mathcal{M}_1 \times \mathcal{M}_2$, i.e. $E_{x_2} \in \mathcal{M}_1$ for every $E \in \mathcal{M}_1 \times \mathcal{M}_2$ and every $x_2 \in X_2$. A consequence of this is that if $f : X_1 \times X_2 \to Y$ is $(\mathcal{M}_1 \times \mathcal{M}_2, \mathcal{N})$-measurable and we let $f_{x_2}(x_1) = f(x_1, x_2)$, then for $B \in \mathcal{N}$, $(f_{x_1}^{-1})(B) = f^{-1}(B)_{x_2} \in \mathcal{M}_1$, i.e. f_{x_2} is $(\mathcal{M}_1, \mathcal{N})$-measurable. Hence the following statements are well-defined.

Theorem 6.7 *Let $a, b \in \mathbb{R}$, $a < b$ and let $f : X \times [a, b] \to \mathbb{C}$ be $(\mathcal{M} \times \mathcal{B}[a, b], \mathcal{B}(\mathbb{C}))$-measurable. Assume that $f(\cdot, t)$ is integrable for each $t \in [a, b]$ and let $F(t) = \int_X f(x, t)d\mu(x)$.*

- *(a) If there exists a $g \in L^1(\mu)$ such that $|f(x, t)| \leq g(x)$ for all (x, t) and $\lim_{t\to t_0} f(x, t) = f(x, t_0)$ for every x, then $\lim_{t\to t_0} F(t) = F(t_0)$. Consequently, if f is continuous, then so is F.*

- *(b) If f is partially differentiable w.r.t. t and there exists a $g \in L^1(\mu)$ such that $|(\partial f/\partial t)(x, t)| \leq g(x)$ for all (x, t). Then*

$$F'(t) = \int \frac{\partial f}{\partial t}(x, t)d\mu(x).$$

Proof. Pick arbitrary t_n converging to t_0, let $h_n(x) = f(x, t_n)$ and $h(x) = f(x, t)$ and use the DCT on h_n and h. This gives (a). For (b), let instead $h_n(x) = (f(x, t_n) - f(x, t))/(t_n - t)$ and $h(x) = (\partial f/\partial t)(x, t)$. Then $h_n \to h$ pointwise and the result follows on applying the DCT to h_n and h; this can be done since $|h_n(x)| \leq \sup_t |(\partial f/\partial t)(x,t)| \leq g$, by the Mean Value Theorem and the hypothesis. $\square$

We are now going to see that any Riemann integrable function on a closed interval $[a,b]$ is also Lebesgue integrable and that the results of the two integrals are the same. The setting is thus that $(X, \mathcal{M}, \mu) = ([a,b], \mathcal{L}, m)$. Let f be defined on X and bounded. Let $P = \{t_0, t_1, \ldots, t_n\}$, $a = t_0 < t_1 < \ldots, t_n = b$, be an arbitrary finite set of points in $[a,b]$. Let

$$m_j = m_j(P) = \inf_{t \in [t_{j-1}, t_j]} f(t),\ M_j = M_j(P) = \sup_{t \in [t_{j-1}, t_j]} f(t),$$

$$s_p f = \sum_1^n m_j(t_j - t_{j-1}),\ S_P f = \sum_1^n M_j(t_j - t_{j-1})$$

and

$$\underline{I}(f) = \sup_P s_P f,\ \overline{I}(f) = \inf_P S_P f.$$

Then f is said to be Riemann integrable if $\underline{I}(f) = \overline{I}(f)$ and $\int_a^b f(x)dx$ is defined as the common value of the two.

For a given P, let $g_P = \sum_1^n m_j \chi_{(t_{j-1}, t_j]}$ and $G_P = \sum_1^n M_j \chi_{(t_{j-1}, t_j]}$. If f is Riemann integrable, there are sets P_k such that $P_1 \subseteq P_2 \subseteq \ldots$ and $s_{P_k} \uparrow \int_a^b f(x)dx$ and $S_{P_k} \downarrow \int_a^b f(x)dx$ as $k \to \infty$. Since g_{P_k} and G_{P_k} are increasing and decreasing respectively, there are limiting functions g and G satisfying $g \leq f \leq G$. Since g_{P_k} and G_{p_k} are obviously Lebesgue measurable, so are g and G. By the DCT, $\int g dm = \int G dm = \int_a^b f(x)dx$, Hence $\int (G - g) dm = 0$, so $G = g$ a.e. which entails $f = G$ a.e. Since the Lebesgue measure is complete on $\mathcal{L}$, f is Lebesgue measurable and we get $\int f dm = \int_a^b f(x)dx$.

These results are summarized in Folland in Theorem 2.28. The results clearly extend to improper integrals and to multiple integrals of functions on $\mathbb{R}^n$.

6.1 Expectation

Let $(X, \mathcal{M}, \mathbb{P})$ be a probability space and let $\xi : (X, \mathcal{M}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ be a random variable. If ξ is integrable, then the *expectation* of ξ is

$$\mathbb{E}\xi := \int_X \xi dP.$$

For $A \in \mathcal{B}$, let

$$\mathbb{P}_\xi(A) = \mathbb{P}\{x : \xi(x) \in A\}.$$

Then $\mathbb{P}_\xi$ is a probability measure on $\mathcal{B}$. The next result shows that the expectation can be computed by integration with respect to $\mathbb{P}_\xi$.

Theorem 6.8 *(**The law of the unconscious statistician**) Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function and assume that $h \circ \xi$ is integrable. Then*

$$\mathbb{E}h(\xi) = \int_{\mathbb{R}} h(t) d\mathbb{P}_\xi(t).$$

Proof. Assume first that $h = \chi_B$ for a $B \in \mathcal{B}$. Then $h \circ \xi = h \circ \chi_B = \chi_{\{x:\xi(x)\in B\}}$, so

$$\mathbb{E}h(\xi) = \mathbb{P}\{\xi \in B\} = \mathbb{P}_\xi(B) = \int_{\mathbb{R}} \chi_B d\mathbb{P}_\xi.$$

By linearity of integrals, the result now holds for all simple functions h. By the MCT, the result then extends to all nonnegative h and finally to all h by linerity. □

A corresponding result can be shown for measurable functions on any σ-finite space.

6.2 The Monotone Class Theorem

The version of the Monotone Class Theorem presented here is slightly different from, and more efficient than, the one in Folland. It is an extension of Dynkin's Lemma and will allow us to make conclusions for all measurable functions on the basis of the corresponding conclusion for characteristic functions of the sets of a generating π-system.

Definition 6.9 *Let $\mathcal{H}$ be a class of functions defined on the space X. Then $\mathcal{H}$ is said to be a monotone class if*

(i) $\mathcal{H}$ *is a complex vector space,*

(ii) $f \equiv 1 \Rightarrow f \in \mathcal{H}$,

(iii) $f_n \in \mathcal{H},\ f_n \geq 0,\ f_n \uparrow f,\ f$ *bounded* $\Rightarrow f \in \mathcal{H}$.

Theorem 6.10 *(***The Monotone Class Theorem***) Let $\mathcal{H}$ be a monotone class of functions on X. Let $\mathcal{I} \subseteq \mathcal{P}(X)$ be a π-system and assume that $\chi_I \in \mathcal{H}$ for all $I \in \mathcal{I}$. Then $\mathcal{H}$ contains all bounded complex $\sigma(\mathcal{I})$-measurable functions.*

Proof. Let $\mathcal{D} = \{A \in \sigma(\mathcal{I}) : \chi_A \in \mathcal{H}\}$. By the conditions on a monotone class, $\mathcal{D}$ is a d-system. Hence $\chi_A \in \mathcal{H}$ for all $A \in \sigma(\mathcal{I})$ by Dynkin's Lemma. Since $\mathcal{H}$ is a vector space, $\mathcal{H}$ then contains all simple functions. If f is nonnegative and $\sigma(\mathcal{I})$-measurable, then let $\phi_n \uparrow f$ for simple ϕ_n. By (iii), $f \in \mathcal{H}$. Finally $\mathcal{H}$ now must contain all bounded $\sigma(\mathcal{I})$-measurable functions, by (i). □

6.3 Product measures

Let $(X_1, \mathcal{M}_1, \mu_1)$ and $(X_2, \mathcal{M}_2, \mu_2)$ be two measure spaces. Recall that

$$\mathcal{M}_1 \times \mathcal{M}_2 = \sigma(\mathcal{I})$$

where $\mathcal{I} = \{E_1 \times E_2 : E_j \in M_j\}$. Observe that $\mathcal{I}$ is a π-system. Let $\mathcal{A}$ be the family of finite disjoint unions of elements in $\mathcal{I}$. Then $\mathcal{A}$ is an algebra. This is not immediately clear, but follows from the observation that $(E \times F)^c = (E^c \times X_2) \cup (E \times F^c)$. Obviously $\sigma(\mathcal{A}) = \mathcal{M}_1 \times \mathcal{M}_2$. For a given set $\bigcup_1^n (E_k \times F_k) \in \mathcal{A}$, let

$$\nu\Big(\bigcup_1^n (E_k \times F_k)\Big) = \sum_1^n \mu_1(E_k)\mu_2(F_k).$$

Claim. ν is countably additive on $\mathcal{A}$.

Proof. It suffices to show that if $E_n \times F_n$, $n = 1, 2, \ldots$ are disjoint and $\bigcup_n (E_n \times F_n) = E \times F$, then $\nu(E \times F) = \sum_n \nu(E_n \times F_n)$. We do this in two steps. First fix an arbitrary $x_2 \in X_2$. Then

$$\begin{aligned}
\mu_1(E)\chi_F(x_2) &= \chi_F(x_2) \int_{X_1} \chi_E(x_1) d\mu_1(x_1) = \int_{X_1} \chi_E(x_1)\chi_F(x_2) d\mu_1(x_1) \\
&= \sum_n \int_{X_1} \chi_{E_n}(x_1)\chi_{F_n}(x_2) d\mu_1(x_1) \\
&= \sum_n \chi_{F_n}(x_2) \int_{X_1} \chi_{E_n}(x_1) d\mu_1(x_1) \\
&= \sum_n \chi_{F_n}(x_2)\mu_1(E_n),
\end{aligned}$$

where the second equality follows from the MCT and that $\chi_E(x_1)\chi_F(x_2) = \sum_n \chi_{E_n}(x_1)\chi_{F_n}(x_2)$. The second step is now the following computation.

$$\begin{aligned}
\mu_1(E)\mu_2(F) &= \int_{X_2} \mu_1(E)\chi_F(x_2)d\mu_2(x_2) \\
&= \sum_n \mu_1(E_n) \int_{X_2} \chi_{F_n}(x_2)d\mu_2(x_2) \\
&= \sum_n \mu_1(E_n)\mu_2(F_n),
\end{aligned}$$

where the second equality is the MCT and step 1. □

By Carathéodory's Extension Theorem, ν extends to a measure μ on $\mathcal{M}_1 \times \mathcal{M}_2$. The standard notation for this measure is $\mu = \mu_1 \times \mu_2$.

The construction of the product measure obviously extends to a finite product of measure spaces. It also works for a countable number of spaces after modifying the π-system $\mathcal{I}$ to $\mathcal{I} = \{\prod_1^n E_i \times \prod i > nX_i\}$. This extension is most natural when the μ_i's are probability measures.

Recalling how a measure is constructed from a countably additive set function on an algebra via an outer measure, we find that

$$(\mu_1 \times \mu_2)(A) = \inf\{\sum_1^\infty \mu_1(E_j)\mu_2(F_j) : E_j \in \mathcal{M}_1, F_j \in \mathcal{M}_2\}.$$

Applying this to the two-dimensional Lebesque measure, some useful analogs of approximation results for one-dimensional Lebesque measure follow. Let m be the two-dimensional Lebesgue measure. Then

$$\begin{aligned}
m(A) &= \inf\{m(U) : U \supseteq A : U \text{ open}\} \\
&= \sup\{m(K) : K \subseteq A, K \text{ compact}\}.
\end{aligned}$$

This is part (a) of Folland's Theorem 2.40. We will also need Theorem 2.40(c) which states that for any set E with $m(E) < \infty$ and $\epsilon > 0$, one can find a set A, which is a finite union of rectangles, such that $m(A\Delta E) < \epsilon$. This result is analogous to Proposition 3.16 as is its proof.

By mimicking the proof of Theorem 6.6 one also gets

Theorem 6.11 *If $f \in L^1(m)$ and $\epsilon > 0$, then there exists a simple function $\phi = \sum_1^m a_j \chi_{R_j}$, where the R_j's are rectangles, such that*

$$\int |f - \phi| dm < \epsilon.$$

There is also a continuous function $g : \mathbb{R}^2 \to \mathbb{R}$ with bounded support, such that

$$\int |f - g| dm < \epsilon.$$

Of course, these results extend to Lebesgue measure and Lebesgue measurable functions on $\mathbb{R}^n$ for arbitrary $n = 2, 3, 4, \ldots$.

Example. Construction of a sequence of independent random variables. For each $n = 1, 2, \ldots$, let $(X_n, \mathcal{M}_n, \mathbb{P}_n) = ([0,1], \mathcal{L}, m)$ and let $\xi : X_n \to \mathbb{R}$ be a random variable with desired distribution, constructed as in earlier examples. Let $(X, \mathcal{M}, \mathbb{P}) = (\prod_1^\infty X_n, \prod_1^\infty \mathcal{M}_n, \prod_1^\infty \mathbb{P}_n)$ and set $\eta_n(x_1, x_2, \ldots) = \xi_n(x_n)$. Then, by the construction of product measure, letting $E_n = \{x_n \in X_n : \xi_n(x) \in B\}$ and $E_j = X_j$ for $j \neq n$,

$$\mathbb{P}(\eta_n \in B) = \mathbb{P}(\prod_1^\infty E_j) = \mathbb{P}_n(\xi_n \in B).$$

Similarly it follows that the η_n's are independent on the π-system consisting of sets of the form $\prod_1^\infty \{x : \eta_n(x) \in (-\infty, b_n)\}$ with $b_n = \infty$ for all but finitely many n and hence on the whole of $\sigma(\eta_1, \eta_2, \ldots)$ by Corollary 4.14. (In fact, we made precisely this observation in the example following Corollary 4.14.) $\Box$

The next question in focus will be when it is possible to change the order of integration for a double integral. First, however, some work is required to establish that the question makes sense. Let $(X_j, \mathcal{M}_j, \mu_j)$, $j = 1, 2$, be *finite* measure spaces and let $(X, \mathcal{M}, \mu)$ be the product space. Let f be a complex- or $\mathbb{R}_+$-valued Borel function on X.

Lemma 6.12 *For every $x_1 \in X_1$ and $x_2 \in X_2$, $f(\cdot, x_2)$ and $f(x_1, \cdot)$ are $\mathcal{M}_1$-measurable and $\mathcal{M}_2$-measurable respectively.*

Proof. Obviously it suffices to check the first statement. Let

$$\mathcal{H} = \{f : f(\cdot, x_2) \text{ is } \mathcal{M}_1\text{-measurable}\}.$$

Let

$$\mathcal{I} = \{E \times F : E \in \mathcal{M}_1,\ F \in \mathcal{M}_2\}.$$

Then $\mathcal{I}$ is a π-system that generates $\mathcal{M}$. Pick $E \times F \in \mathcal{I}$, let $f = \chi_{E\times F}$ and $g = f(\cdot, x_2)$. Pick a set $B \in \mathcal{B}(\mathbb{C})$. Suppose $1 \in B$ and $0 \notin B$. If $x_2 \in F$, then $g^{-1}(B) = E$ and if $x_2 \notin F$, then $g^{-1}(B) = \emptyset$. Hence $g^{-1}(B)$ and $g^{-1}(B^c)$ are measurable. In case B contains both 0 and 1, $g^{-1}(B) = X_1$. We conclude that g is measurable. Since $\mathcal{H}$ is a monotone class, $\mathcal{H}$ contains all bounded functions. Since limits of measurable functions are measurable, approximating by a sequence of simple functions now shows that $\mathcal{H}$ contains all $\mathcal{M}$-measurable functions as desired. $\square$

By this lemma, the two functions $g : X_1 \to \mathbb{C}$ and $h : X_2 \to \mathbb{C}$ given by

$$g(x_1) = \int_{X_2} f(x_1, x_2) d\mu_2(x_2)$$

$$h(x_2) = \int_{X_1} f(x_1, x_2) d\mu_1(x_1)$$

are well-defined.

Lemma 6.13 *The functions g and h are measurable.*

Proof. Let $\mathcal{H}$ be the family of f's such that the corresponding g and h are measurable. If $f \equiv 1$, then $g = \mu_2(X_2)$ and $h = \mu_1(X_1)$ are finite constants and hence measurable. Since measurability is closed under linear operations, $\mathcal{H}$ is a vector space. If $f_n \uparrow f$, $f_n \geq 0$ and f is bounded, then $f \in \mathcal{H}$ by the MCT. Thus $\mathcal{H}$ is a monotone class.

Now if $f = \chi_{E\times F}$, $E \in \mathcal{M}_1$, $F \in \mathcal{M}_2$, then $g \equiv 0$ if $x_1 \notin E$ and $g \equiv \mu_2(F)$ if $x_1 \in E$, which is clearly measurable in either case. Hence $\mathcal{H}$ contains all bounded functions by the Monotone Class Theorem. Now extend to all f by the MCT and linearity. $\square$

By this Lemma 6.13, it makes sense to define

$$\int_{X_1} \Big(\int_{X_2} f(x_1, x_2) d\mu_2(x_2) \Big) d\mu_1(x_1)$$

and

$$\int_{X_2} \Big(\int_{X_1} f(x_1, x_2) d\mu_1(x_1) \Big) d\mu_2(x_2).$$

However, are they equal? Also, how do they relate to $\int_{X_1\times X_2} f d(\mu_1 \times \mu_2)$?

Theorem 6.14 *(***Tonelli's Theorem***) If $f \in L^+(X, \mathcal{M}, \mu)$, then*

$$\begin{aligned}\int_X f d\mu &= \int_{X_1} \Big(\int_{X_2} f(x_1, x_2) d\mu_2(x_2) \Big) d\mu_1(x_1) \\ &= \int_{X_2} \Big(\int_{X_1} f(x_1, x_2) d\mu_1(x_1) \Big) d\mu_2(x_2).\end{aligned}$$

Proof. Let $\mathcal{H}$ be the class of bounded f for which the statement holds. Then $\chi_{E\times F} \in \mathcal{H}$ for every $E \times F \in \mathcal{M}$, since all three expressions are then equal to $\mu_1(E)\mu_2(F)$. Taking $E = X_1$ and $F = X_2$ shows that $1 \in \mathcal{H}$. Since μ_1 and μ_2 are finite, $\mathcal{H}$ is a complex vector space. By the MCT, it now follows that $\mathcal{H}$ is a monotone class. Hence $\mathcal{H}$ contains all bounded functions, by the Monotone Class Theorem. The proof is now completed via another appeal to the MCT. □

For $f \in L^1(X, \mathcal{M}, \mu)$, Tonelli's Theorem together and linearity of integrals show:

Theorem 6.15 *(***Fubini's Theorem***) If $f \in L^1(X, \mathcal{M}, \mu)$, then*

$$\begin{aligned}\int_X f d\mu &= \int_{X_1} \Big(\int_{X_2} f(x_1, x_2) d\mu_2(x_2) \Big) d\mu_1(x_1) \\ &= \int_{X_2} \Big(\int_{X_1} f(x_1, x_2) d\mu_1(x_1) \Big) d\mu_2(x_2).\end{aligned}$$

It is very useful to note that in order to check that a given function f is integrable with respect to the product measure, one can use Tonelli's Theorem on $|f|$ to do the integration in the most convenient order and check if the resulting integral is finite.

By countable additivity, Tonelli's and Fubini's Theorem's extend to σ-finite measure spaces. However, they do not extend beyond that. Consider for example $X_1 = X_2 = [0, 1]$, $\mathcal{M}_1 = \mathcal{M}_2 = \mathcal{B}[0, 1]$, $\mu_1 = m$ and μ_2 counting measure (i.e. $\mu_2(F)$ is the number of points on F, so that μ_2 is infinite for all infinite sets). Note that μ_2 is not σ-finite. Let A be the diagonal, i.e. $A = \{(x, x) : x \in [0, 1]\}$. (Why does $A \in \mathcal{B} \times \mathcal{B}$?) Then

$$\int_{X_1} \Big(\int_{X_2} \chi_A(x_1, x_2) d\mu_2(x_2) \Big) d\mu_1(x_1) = 1$$

since the inner integral is constantly 1, whereas

$$\int_{X_2} \Big(\int_{X_2} \chi_A(x_1, x_2) d\mu_1(x_1) \Big) d\mu_2(x_2) = 0$$

since the inner integral is constantly 0 in this case. (Exercise: What is $\int_X f d\mu$?)

In Fubini's Theorem, also the integrability condition is necessary. For an example that demonstrates this, let X_1 and X_2 both be the set of natural numbers and μ_1 and μ_2 both counting measure. Let A be the diagonal $\{(k,k) : k = 1, 2, \ldots\}$ and B the off-diagonal $\{(k, k+1), k = 1, 2, \ldots\}$. Letting $f = \chi_A - \chi_B$, we get $\int_{X_1} \int_{X_2} f d\mu_2 d\mu_1 = 0$ and $\int_{X_2} \int_{X_1} d\mu_1 d\mu_2 = 1$, whereas $\int_X f d\mu$ is undefined.

7 Signed measures

Let $(X, \mathcal{M})$ be a measurable space and let $\nu : X \to \overline{\mathbb{R}}$.

Definition 7.1 *The function ν is said to be a signed measure if*

- $\nu(\emptyset) = 0$,
- *ν assumes at most one of the values ∞ and $-\infty$,*
- *$\nu(\bigcup_1^\infty E_n) = \sum_1^\infty \nu(E_n)$ whenever $E_n \in \mathcal{M}$ are disjoint and the sum converges absolutely if $\nu(\bigcup_1^\infty E_n)$ is finite.*

Sometimes when we speak of a measure in a context where also some signed measure appears, we will refer to the measure as a *positive measure* to make the distinction clear.

Example. If μ_1 and μ_2 are two measures on $\mathcal{M}$ and at least one of them is finite, then $\mu_1 - \mu_2$ is a signed measure. □

Example. If f is real-valued and $\mathcal{M}$-measurable and at least one of f^+ and f^- is integrable, then

$$\nu(E) = \int_E f d\mu$$

defines a signed measure. A function of this kind is called an *extended integrable* function. □

Proposition 7.2 *Let ν be a signed measure. If $E_n \uparrow E$, then $\nu(E_n) \to \mu(E)$. If $E_n \downarrow E$ and $\nu(E_1)$ is finite, then $\nu(E_n) \to \nu(E)$.*

Proof. Let $F_n = E_n \setminus E_{n-1}$ so that the F_n's are disjoint and $E = \bigcup_1^\infty F_n$. Then, exactly is in the positive measure case,

$$\nu(E) = \sum_1^\infty \nu(F_n) = \lim_N \sum_1^N \nu(F_n) = \lim_N \nu(E_N).$$

The second part also goes through exactly as for positive measures. $\square$

7.1 Jordan-Hahn Decompositions

Definition 7.3 *Let ν be a signed measure. A set E is said to be a positive set for ν if $\nu(F) \geq 0$ whenever F is measurable and $F \subseteq E$. A negative set is defined analogously. If E is both positive and negative for ν, then E is said to be a null set for ν.*

It is obvious from the definition that any subset of a positive/negative set is positive/negative. It is also clear that the union and the intersection of two positive/negative sets are positive/negative.

Lemma 7.4 *Let $P_1, P_2, \ldots$ be positive sets for the signed measure nu. Then $P = \bigcup_1^\infty P_n$ is also positive.*

Proof. Let $Q_1 = P_1$ and $Q_n = P_n \setminus \bigcup_1^{n-1} P_j$, so that the Q_n's are disjoint and $\bigcup_1^\infty Q_n = P$. Then each Q_n is positive, so for any $E \subseteq P$, $\nu(E \cap Q_n) > 0$. Hence

$$\nu(E) = \sum_1^\infty \nu(E \cap Q_n) \geq 0$$

by countable additivity of ν. $\square$

The next result states that given a signed measure, the space can be partitioned into a positive and a negative part, in an essentially unique way.

Theorem 7.5 *(**The Hahn Decomposition Theorem**) Let ν be a signed measure on $(X, \mathcal{M})$. Then the is a positive set P and a negative set N such that $X = P \cup N$. If P' and N' are two other such sets, then $P \Delta P'$ and $N \Delta N'$ are null for ν.*

Proof. Assume without loss of generality that ν does not assume the value $+\infty$. Let $m = \sup\{\nu(E) : E \text{ positive}\}$. Pick a sequence $\{P_j\}$ of positive sets such that $\nu(P_j) \to m$. Since positivity is closed under finite unions, we may assume that the P_j's are increasing. Let $P = \bigcup_1^\infty P_j$. By Lemma 7.4, $\nu(P_j) \to \nu(P)$, so $\nu(P) = m$.

Let $N = X \setminus P$. We claim that N is negative. Assume for contradiction that N is not negative. Observe that there can be no positive subset E of N with $\nu(E) > 0$, since that would imply that $P \cup E$ is positive and $\nu(P \cup E) = \nu(P) + \nu(E) > m$, contradicting the definition of m. Hence there must be an $E \subseteq N$ with $\nu(E) > 0$, but E not positive. This means that there is an $F \subset E$ with $\nu(F) < 0$. This implies that $\nu(E \setminus F) > \nu(E)$. Iterating this observation will lead to the desired contradiction.

Let n_1 be the smallest positive integer such that there exists an $A_1 \subset N$ with $\nu(A_1) > 1/n_1$. Pick such an A_1. Since A_1 is not positive, we can let n_2 be the smallest positive integer such that there exists and $A_2 \subset A_1$ with $\nu(A_2) > \nu(A_1) + 1/n_2$. Pick such an A_2. Since $\nu(A_2) > 0$ and $A_2 \subset N$, A_2 is not positive. Iterate the procedure to produce smallest possible integers $n_3, n_4, \ldots$ and $A_3, A_4, \ldots$ with $\nu(A_k) > \sum_1^k n_j^{-1}$. Let $A = \bigcap_1^\infty A_n$. Recall our assumption that ν does not take on the value $+\infty$. Consequently $\nu(A) < \infty$. Hence Proposition 7.2 implies that $\nu(A_n) \to \nu(A)$ so that

$$\sum_{j=1}^{\infty} \frac{1}{n_j} < \nu(A) < \infty.$$

From this it follows in particular that $\lim_j n_j = \infty$. However $A \subset N$, so A is not positive. Thus there exists a positive integer n and a $B \subset A$ such that $\nu(B) > \nu(A) + 1/n$. Since $n_j \to \infty$, $n_j > n$ for large enough j. Thus $\nu(B) > \nu(A) + 1/n > \nu(A_j) + 1/n$ and $B \subset A \subset A_j$. This contradicts the choice of n_j as the *smallest* integer n for which such a B exists.

Finally if $P' \cup N'$ is another partition into a positive and a negative set, then $P \setminus P' \subseteq P \cap N'$ and is hence null. Analogously $P' \setminus P$, $N \setminus N'$ and $N' \setminus N$ are null. $\square$

A partition of the space X into the sets P and N, as in the Hahn Decomposition Theorem, is called a *Hahn decomposition* (with respect to ν).

Definition 7.6 *If ν_1 and ν_2 are two signed measures on $(X, \mathcal{M})$, then they are said to be mutually singular (or just singular) if there exist $E, F \in \mathcal{M}$ such that $E \cup F = \mathcal{X}$, E is null for ν_2 and F is null for ν_1.*

In words, ν_1 and ν_2 are singular if they live on disjoint parts of X. When ν_1 and ν_2 are singular, this is denoted by $\nu_1 \perp \nu_2$. It follows from the Hahn Decomposition Theorem that any signed measure ν can be written as the difference of two positive measures. These are mutually singular and unique. This is summarized in the following result.

Theorem 7.7 *(***The Jordan Decomposition Theorem***) Let ν be a signed measure on $(X, \mathcal{M})$. Then there exist two unique positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$.*

Proof. Let $X = P \cup N$ be a Hahn decomposition with respect to ν and let

$$\nu^+(E) = \nu(E \cap P)$$

$$\nu^-(E) = -\nu(E \cap N),$$

$E \in \mathcal{M}$. Then ν^+ and ν^- are positive, singular and $\nu = \nu^+ - \nu^-$. It remains to prove uniqueness. Assume that ν can also be written as $\nu = \mu^+ - \mu^-$ for two other positive singular measures μ^+ and μ^-. Then there are disjoint sets $E, F \in M$ such that $E \cup F = X$ and $\mu^+(F) = \mu^-(E) = 0$. Hence $E \cup F$ is another Hahn decomposition of X and hence $P \Delta E$ is null for ν. Therefore, for any $A \in \mathcal{M}$,

$$\mu^+(A) = \mu^+(A \cap E) = \nu(A \cap E) = \nu(A \cap P) = \nu^+(A).$$

Thus $\mu^+ = \nu^+$ and analogously $\mu^- = \nu^-$. □

A decomposition of a signed measure in this way is called a *Jordan decomposition* or a *Jordan-Hahn decomposition*. The measures ν^+ and ν^- are called the *positive variation* of ν and the *negative variation* of ν respectively. The *total variation* of μ is the measure $|\nu| := \nu^+ + \nu^-$. The integral with respect to the signed measure ν is given by

$$\int f d\nu = \int f d\nu^+ - \int f d\nu^-, \ f \in L^1(|\nu|).$$

We say that ν is *finite* if $|\nu|$ is finite and we say that ν is *σ-finite* if $|\nu|$ is σ-finite.

7.2 The Lebesgue-Radon-Nikodym Theorem

Let ν be a signed measure and μ a positive measure on $(X, \mathcal{M})$.

Definition 7.8 *If $\nu(E) = 0$ whenever $E \in \mathcal{M}$ and $\mu(E) = 0$, then ν is said to be absolutely continuous with respect to μ, denoted $\nu \ll \mu$.*

Immediate consequences of the definition are that $\nu \ll \mu$ iff ($\nu^+ \ll \mu$ and $\nu^- \ll \mu$) iff $|\nu| \ll \mu$ and that ($\nu \ll \mu$ and $\nu \perp \mu$) iff $\nu = 0$.

Example. Let $\xi : (\mathcal{X}, \mathcal{M}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable. Recall the measure $\mathbb{P}_\xi$ on $\mathcal{B}$ given by $\mathbb{P}_\xi(B) = \mathbb{P}\{\xi \in B\}$. If $\mathbb{P}_\xi \ll m$, then ξ is said to be a *continuous random variable*. □

The classical Radon-Nikodym Theorem states that whenever $\nu \ll \mu$, there exists an $\mathcal{M}$-measurable function f such that

$$\nu(E) = \int_E f d\mu, \; E \in \mathcal{M},$$

provided that μ and ν are σ-finite. The Lebesgue-Radon-Nikodym Theorem (LRNT) provides even more information. Before that, a preparatory lemma is required.

Lemma 7.9 *Assume that ν and μ are two finite measures on $(X, \mathcal{M})$. Then either $\nu \perp \mu$ or there exists $\epsilon > 0$ and $E \in \mathcal{M}$ such that $\mu(E) > 0$ and E is positive for $\nu - \epsilon\mu$.*

Proof. Let $P_n \cup \mathcal{N}_n$ be a Hahn decomposition for $\nu - n^{-1}\mu$, $n = 1, 2, \ldots$. Write $P = \bigcup_n P_n$ and $N = \bigcap_n N_n$, so that $P_n \uparrow P$ and $N_n \downarrow N$. Since N is negative for $\nu - n^{-1}\mu$, $\nu(N) \leq n^{-1}\mu(N)$ for all n and since μ is finite, this implies that $\nu(N) = 0$. If $\mu(P) = 0$, then $\mu \perp \nu$. If $\mu(P) > 0$, then $\mu(P_k) > 0$ for some k by continuity of measures. Now take $E = P_k$ and $\epsilon = 1/k$. □

Theorem 7.10 *(**The Lebesgue-Radon-Nikodym Theorem**) Let ν be a signed measure and μ a positive measure on $(X, \mathcal{M})$, both σ-finite. Then*

(a) there exist unique σ-finite signed measures λ and ρ such that

$$\lambda \perp \mu, \;\; \rho \ll \mu, \;\; \nu = \lambda + \rho$$

(b) there exists an extended μ-integrable function f such that

$$\rho(E) = \int_E f d\mu$$

for all $E \in \mathcal{M}$. If g is another such function, then $f = g$ μ-a.e.

Proof. We do this for ν, μ finite positive measures; the extensions are straightforward. The uniqueness parts are left for exercises (or reading in Folland).

Let $\mathcal{F}$ be the set of $\mathcal{M}$-measurable nonnegative functions f such that $\int_E f d\mu \leq \nu(E)$ for all $E \in \mathcal{M}$. Then $\mathcal{F}$ is nonempty (since at least $0 \in \mathcal{F}$) and $\mathcal{F}$ is closed under finite maxima, since if $f, g \in \mathcal{F}$, then

$$\int_E f \vee g \, d\mu = \int_{E \cap A} f d\mu + \int_{E \cap A^c} g d\mu \leq \nu(E \cap A) = \nu(E \cap A^c) = \nu(A)$$

where $A = \{x : f(x) \geq g(x)\}$. Let $a = \sup\{\int f d\mu : f \in \mathcal{F}\}$. Note that $a \leq \nu(X) < \infty$. Pick $f_n \in \mathcal{F}$ such that $\int f_n d\mu \to a$. Letting $g_n = \max(f_1, \ldots, f_n)$ we get that $g_n \uparrow g := \sup_n f_n$ pointwise, so that the MCT implies that

$$\int g d\mu = \lim_n \int g_n d\mu = a.$$

The MCT, applied to $g_n \chi_E$ for each $E \in \mathcal{M}$, also implies that $g \in \mathcal{F}$. Hence the set function λ defined by

$$\lambda(E) = \nu(E) - \int_E g d\mu$$

is a positive measure. Set $\rho(E) = \int_E g d\mu$. Then we are done if we can prove that λ and μ are singular. If not, Lemma 7.9 implies that we can find E with $\mu(E) > 0$ and $\epsilon > 0$ such that $\lambda \geq \epsilon \mu$ on E. However then for any $F \in \mathcal{M}$,

$$\int_F (g + \epsilon \chi_E) d\mu = \int_F g d\mu + \epsilon \mu(F \cap E) \leq \int_F g d\mu + \lambda(F) = \nu(F),$$

i.e. $g + \epsilon \chi_E \in \mathcal{F}$, a contradiction. □

Writing $\nu = \lambda + \rho$ with $\lambda \perp \mu$ and $\rho \ll \mu$ is called the *Lebesgue decomposition* of ν.

In case $\nu \ll \mu$, the LRNT gives states that $\nu(E) = \int_E f d\mu$ for all $E \in \mathcal{M}$, i.e. the Radon-Nikodym Theorem. It is common to write $f = d\nu/d\mu$, the reason of course being that the notation in itself suggests the property that defines the function f, namely that $\int_E (d\nu/d\mu) d\mu = \int_E d\nu$ for all E. "Multiplying" by $d\mu$, one also writes $d\nu = f d\mu$. The function $d\nu/d\mu$ is called the *Radon-Nikodym derivative* of ν with respect to μ.

Note that the LRNT works fine even if it is assumed ν is a signed measure; just Jordan decompose μ and use the LRNT on μ^+ and μ^-.

The most important applications of the LRNT are the Fundamental Theorem of Calculus and the Integration by Parts formula for Lebesgue integrals. We will come back to those shortly. Another fundamental application is the concept of conditional expectation in probability theory.

Example. (**Conditional Expectation**) Let $f \in L^1(X, \mathcal{M}, \mu)$, μ σ-finite. Define $\nu(E) = \int_E f d\mu$, $E \in \mathcal{M}$. Then ν is a finite signed measure such that $\nu \ll \mu$. Now let $\mathcal{N}$ be a sub-σ-algebra of $\mathcal{M}$. Then obviously $\nu|_{\mathcal{N}} \ll \mu|_{\mathcal{N}}$. By the LRNT, this entails that there exists a function $g \in L^1(X, \mathcal{N}, \mu|_{\mathcal{N}})$ such that

$$\nu(E) = \int_E g d\mu$$

for all $E \in \mathcal{N}$, i.e.

$$\int_E f d\mu = \int_E g d\mu$$

for all $E \in \mathcal{N}$. This provides the base for the definition of conditional expectation, as follows.

Let $(X, \mathcal{M}, \mathbb{P})$ be a probability space and ξ and η integrable random variables. We would like to find a sensible, proper definition of the conditional expectation $\mathbb{E}[\xi|\eta]$. Clearly, writing $\psi = \mathbb{E}[\xi|\eta]$, ψ should be a random variable which is a function of η. In other words, ψ should be a $\sigma(\eta)$-measurable function. Now, it is intuitively fairly clear that the conditional expectation of ξ given an event A should satisfy

$$\mathbb{E}[\xi|A] = \frac{\mathbb{E}[\xi\chi_A]}{\mathbb{P}(A)} = \frac{\int_A \xi d\mathbb{P}}{\mathbb{P}(A)}$$

for any A such that $\mathbb{P}(A) > 0$. Since $\psi = \mathbb{E}[\xi|\eta]$ should equal $\mathbb{E}[\xi|\eta \in B]$ if we are told that $\eta \in B$ (for some $B \in \mathcal{B}(\mathbb{R})$) and no more, we should have

$$\int_{\{\eta \in B\}} \psi d\mathbb{P} = \int_{\{\eta \in B\}} \xi d\mathbb{P}$$

for all $B \in \mathcal{B}$, i.e.

$$\int_A \psi d\mathbb{P} = \int_A \xi d\mathbb{P}$$

for all $A \in \sigma(\eta)$. This is the criterion that is used for the formal definition.

Definition 7.11 *Let $\mathcal{N}$ be a sub-σ-algebra of $\mathcal{M}$ and ξ an integrable random variable. Then ψ is said to be (a version of) a conditional expectation of ξ given*

$\mathcal{N}$ if ψ is $\mathcal{N}$-measurable and

$$\int_A \psi d\mathbb{P} = \int_A \xi d\mathbb{P}$$

for all $A \in \mathcal{N}$.

By the above observations, the existence of such a ψ follows from the LRNT. Note that two versions of the conditional expectation must be equal a.s. (exercise). $\square$

Here are a few more results on the validity of the $d\nu/d\mu$-notation.

Proposition 7.12 *Assume that μ, ν and λ are σ-finite measures, $\nu \ll \mu$ and $\mu \ll \lambda$.*

(a) If $g \in L^1(\nu)$, then $g(d\nu/d\mu) \in L^1(\mu)$ and

$$\int g d\nu = \int g \frac{d\nu}{d\mu} d\mu.$$

(b)

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$$

λ-a.e.

Proof.

(a) If $g = \chi_E$, $E \in \mathcal{M}$, then

$$\int g\frac{d\nu}{d\mu}d\mu = \int_E \frac{d\nu}{d\mu}d\mu = \nu(E) = \int_E d\nu = \int g d\nu.$$

Now use linearity of the integrals prove the result for simple functions, then the MCT for nonnegative functions and then linearity again for general g.

(b) Pick $E \in \mathcal{M}$ arbitrarily, let $g = \chi_E(d\nu/d\mu)$ and plug this into (a), letting μ and λ play the rôle of ν and μ respectively. Doing so gives

$$\int_E \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}d\lambda = \int_E \frac{d\nu}{d\mu}d\mu = \nu(E) = \int_E \frac{d\nu}{d\lambda}d\lambda,$$

where the first equality is by (a) and the other two by definition. By Proposition 6.3, this proves (b).

□

Example. If $\nu \ll \mu$ and $\mu \ll \nu$, then $(d\nu/d\mu)(d\mu/d\nu) = 1$ almost everywhere with respect to any of the two measures. □

7.3 Complex measures

Let $(X, \mathcal{M})$ be a measurable space. A set function $\nu : \mathcal{M} \to \mathbb{C}$ is said to be a *complex measure* if it can be written as

$$\nu = \nu_r + i\nu_i$$

where ν_r and ν_i are finite signed measures. We let $L^1(\nu) = L^1(\nu_r) \cap L^1(\nu_i)$ and for $f \in L^1(\nu)$, we define

$$\int f d\nu = \int f d\nu_r + i \int f d\nu_i.$$

For two complex measures ν and μ, we write $\nu \perp \mu$ if $\nu_j \perp \mu_k$ for all four combinations of $i, j \in \{r, i\}$. If μ is a positive measure, we write $\nu \ll \mu$ if $\nu_r \ll \mu$ and $\nu_i \ll \mu$. The Lebesgue-Radon-Nikodym Theorem now goes through unchanged if the signed measure ν is replaced with a complex measure.

The *total variation* of the complex measure ν is given by

$$|\nu|(E) = \sup\{\sum_1^\infty |\nu(F_n)| : F_1, F_2, \dots \text{ disjoint and } \bigcup_1^\infty F_n = E\}.$$

It is fairly easy to show that $|\nu|$ is a finite measure. It is obvious that $\nu \ll |\nu|$ and that for positive measures we have $\mu\nu \ll \mu$ iff $|\nu| \ll \mu$.

Proposition 7.13 *Let $f = d\nu/d|\nu|$. Then $|f| = 1$ $|\nu|$-a.e.*

Proof. On one hand

$$\Big|\int_E f d|\mu|\Big| = |\nu(E)| \leq |\nu|(E) = \int_E 1 d|\nu|$$

for all $E \in \mathcal{M}$, so $|f| \leq 1$ a.e. On the other hand, if $|f| < 1$ on a set of positive measure, then by continuity of measures and separability of $\mathbb{C}$, there must be an

$n \in \mathbf{N}$ and a $z \in \mathbb{C}$ with $|z| < 1 - 2/n$ such that $f \in B_{1/n}(z)$ on a set of positive measure. Let $E = \{x : f(x) \in B_{1/n}(z)\}$ for such n and z. Then for all $F \subseteq E$,

$$|\nu(F)| = \Big| \int_F f d|\nu| \Big| \leq \int_F |f| d|\nu| \leq (1 - \frac{1}{n})|\nu|(F).$$

Hence for all disjoint $F_1, F_2, \ldots$ whose union is E, we get

$$\sum_1^\infty |\nu(F_n)| \leq (1 - \frac{1}{n})|\nu|(F),$$

contradicting the definition of $|\nu|$. □

A few immediate consequences of the definition of the total variation and the above proposition conclude this section.

- If $f = d\mu/d|\mu|$, then $\int_E |f| d|\mu| = |\mu|(E)$ for all $E \in M$. More generally, if $\nu \ll \mu$ for a positive measure μ and $f = d\nu/d\mu$, then $|f| = d|\nu|/d\mu$ μ-a.e.
- If ν_1 and ν_2 are two complex measures, then $|\nu_1 + \nu_2| \leq |\nu_1| + |\nu_2|$.

7.4 Differentiation in $\mathbb{R}^n$

In this section, we are going to have $(X, \mathcal{M}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), m)$ for some $n = 1, 2, \ldots$ throughout.

Suppose that ν is a σ-finite signed measure satisfying $\nu \ll m$. By the Radon-Nikodym Theorem, $f = d\nu/dm$ exists and satisfies

$$\int_E f(x)dx = \nu(E)$$

for all $E \in \mathcal{B}$.

Let

$$F(x) = \lim_{r \to 0} \frac{\nu(B_r(x))}{m(B_r(x))} = \lim_{r \to 0} \frac{\int_{B_r(x))} f(t)dt}{m(B_r(x))},$$

provided that the limit exists, i.e. F is the limit of the average value of f on $B_r(x)$, when it exists. Intuitively, one would expect that $F = f$ a.e. Is this true? This question will be the focus of our attention in this section. Define

$$A_r f(x) = \frac{\int_{B_r(x))} f(t)dt}{m(B_r(x))},$$

so that $F(x) = \lim_{r \to 0} A_r f(x)$ when $f = d\nu/dm$. We define $A_r f(x)$ for all functions f for which the definition makes sense, i.e. for $f \in L^1_{loc}$ where L^1_{loc} is the space of all *locally integrable functions*, i.e. all functions g for which $\int_K |g(x)|dx < \infty$ for all compact K. (Note that L^1_{loc} is precisely the space of functions g for which $\nu(E) = \int_E g(x)dx$ defines a σ-finite measure.)

Lemma 7.14 *Let $\mathcal{C}$ be a family of open balls in $\mathbb{R}^n$ and let U be the union of all the sets in $\mathcal{C}$. Then, for any $c < m(U)$, there are disjoint sets $B_1, \ldots, B_k \in \mathcal{C}$ such that $\sum_1^k m(B_j) > 3^{-n}c$.*

Proof. Since m is inner regular, by (4), there exists a compact set $K \subset U$ such that $m(K) > c$. Since $\mathcal{C}$ is an open cover of K, there are $A_1, \ldots, A_l \in \mathcal{C}$ such that $\bigcup_1^l A_j \supset K$. Let B_1 be the largest of the A_j's (in terms of radius; if there is more than one ball with the largest radius, then choose arbitrarily). Next let B_2 be the largest of the remaing A_j's that does not intersect B_1. Then let B_3 be the largest of the now remaining A_j that does not intersect B_1 or B_2. Keep on doing this recursively until no A_j remains that does not intersect any of the chosen B_j's. Let k be the index of the last A_j chosen by this procedure.

Suppose that A_i is one of the A_j's that was not chosen. Then there is a smallest index j such that $A_i \cap B_j \neq \emptyset$. We must then have that the radius of A_i is at most as large as the radius of B_j, since otherwise A_i would itself have been chosen at step j or earlier. This means that $A_i \subseteq B_j^*$, where B_j^* is the ball centered at the same point as B_j and with three times the radius of B_j.

Repeating this argument for all A_j's that were not chosen shows that $K \subset \bigcup_1^k B_j^*$. Since $m(B_j^*) = 3^n m(B_j)$ we get

$$c < m(K) < \sum_1^k m(B_j^*) = 3^n \sum_1^k m(B_j).$$

□

Lemma 7.15 *The function $A_r f(x)$ is continuous in r and x.*

Proof. Let $c = m(B_1(0))$ so that $m(B_r(x)) = cr^n$. Hence

$$A_r f(x) = c^{-1} r^{-n} \int_{B_r(x)} f(t)dt,$$

so that it suffices to check that $\int_{B_r(x)} f(t)dt$ is continuous in (x, r). If $(x, r) \to (x_0, r_0)$, then $\chi_{B_r(x)} \to \chi_{B_{r_0}(x_0)}$ pointwise, except on a subset of the boundary of $B_{r_0}(x_0)$, a null-set. Also, for x close enough to x_0, all these characteristic functions are bounded by $\chi_{B_{r_0+1}(x_0)}$ which is an integrable function. Since f is locally integrable, it now follows from the DCT that

$$\int_{B_r(x)} f(t)dt \to \int_{B_{r_0}(x_0)} f(t)dt$$

as desired. □

Next we define the *Hardy-Littlewood maximal function*, $Hf(x)$.

Definition 7.16 *For $f \in L^1$, let*

$$Hf(x) = \sup_{r>0} A_r|f|(x),\ x \in \mathbb{R}^n.$$

Theorem 7.17 *(**The Maximal Theorem**) For $f \in L^1$ and $a > 0$, let $E_a^f = \{x \in \mathbb{R}^n : Hf(x) > a\}$. Then, for all f and a,*

$$m(E_a^f) \le \frac{3^n}{a} \int |f(t)|dt.$$

Proof. Fix f and a. If $E_a^f = \emptyset$, the result is trivial, so assume otherwise. Then, for each $x \in E_a^f$, pick $r_x > 0$ so that $A_r|f|(x) > a$. By Lemma 7.14, we can find $x_1, \dots, x_k \in E_a^f$ so that the $B_j := B_{r_{x_j}}(x_j)$'s are disjoint and $\sum_1^k m(B_j) > 3^{-n} m(E_a^f)$. However

$$\int_{B_j} |f(t)|dt = m(B_j) A_{r_{x_j}}|f|(x_j) > a m(B_j)$$

so

$$3^{-n} m(E_a^f) < \sum_1^k m(B_j) < \frac{1}{a} \sum_1^k \int_{B_j} |f(t)|dt \le \frac{1}{a} \int |f(t)|dt.$$

□

We are now ready to show that the limit as $r \to 0$ of $A_r f(x)$ is indeed $f(x)$ for any locally integrable f.

Theorem 7.18 *If $f \in L^1_{loc}$, then for a.e. $x \in \mathbb{R}^n$,*

$$\lim_{r\to 0} A_r f(x) = f(x).$$

Proof. It suffices to prove the result for $f \in [-N, N]^n$ for arbitrarily fixed N and hence we may assume without loss of generality that $f \in L^1$. Then, for any $\epsilon > 0$ by Thorem 6.11, there exists a continuous integrable function g such that

$$\int |f(t) - g(t)|dt < \epsilon.$$

Since g is continuous, there is for each x and each $\delta > 0$, an $r > 0$ such that $|g(t) - g(x)| < \delta$ whenever $|t - x| < r$. For such an r we have

$$|A_r g(x) - g(x)| = \frac{|\int_{B_r(x)}(g(t) - g(x))dt|}{m(B_r(x))} < \delta.$$

Hence $A_r g(x) \to g(x)$ as $r \to 0$. From this, it follows that

$$\begin{aligned}
\limsup_{r\to 0} |A_r f(x) - f(x)| &\leq \limsup_{r\to 0} \Big| A_r(f(x) - g(x)) \\
&+ (A_r g(x) - g(x)) + (g(x) - f(x)) \Big| \\
&\leq H(f-g)(x) + |f-g|(x),
\end{aligned}$$

by the triangle inequality and that the middle term of the second expression vanishes by the above. For $a > 0$, let $E_a = \{x : \limsup_{r\to 0} |A_r f(x) - f(x)| > a\}$. We want to show that $m(E_a) = 0$ for every a. Let $F_a = \{x : |f(x) - g(x)| > a\}$. By the above inequality,

$$E_a \subseteq F_{a/2} \cup \{x : H(f-g)(x) > a/2\}.$$

By the Maximal Theorem, the measure of the second set on the right hand side is bounded by $2 \cdot 3^n \int |f(t) - g(t)|dt < 2 \cdot 3^n \epsilon$. Also, by Markov's inequality,

$$m(E_a) \leq \frac{2}{a} \int |f(t) - g(t)|dt < \frac{2}{a}\epsilon.$$

Hence $m(E_a) < (2(1 + 3^n)/a)\epsilon$ and since ϵ was arbitrary, we are done. $\square$

Note that by applying Theorem 7.18 to the function $g(t) = |f(t) - f(x)|$, we find that also the following slightly stronger statement holds:

$$\lim_{r\to 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(t) - f(x)| dt = 0.$$

The result can be generalized a bit further by replacing the balls $B_r(x)$ by more general sets. A family of sets $\{E_r\}_{r>0}$ is said to *shrink nicely* (or E_r shrinks nicely) to x if $E_r \subseteq B_r(x)$ for all r and there is an $a > 0$, independent of r, such that $m(E_r) > am(B_r(x))$ for all r. It is now easy to see that

$$\lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} |f(t) - f(x)| = 0$$

whenever E_r shrinks nicely to x. As a special case of this, consider a signed measure ν on $\mathcal{B}(\mathbb{R}^n)$ such that $|\nu|(K) < \infty$ for all compact K and $\nu \ll m$. Letting $f = d\nu/dm$, we get that $f \in L^1_{loc}$ and hence

$$\lim_{r\to 0} \frac{\nu(E_r)}{m(E_r)} = f(x) \tag{4}$$

for almost every x, whenever E_r shrinks nicely to x. In fact, this holds even if ν is not absolutely continuous w.r.t. m. By the LRNT, one can write

$$\nu(E) = \lambda(E) + \int_E f d\mu, \ E \in \mathcal{B}(\mathbb{R}^n)$$

where $\lambda \perp m$ and $f = d\nu/dm$. Using that λ lives on a space of m-measure 0, one can show that (4) still holds. (Then, of course, if x is a point for which $\lambda\{x\} > 0$, this point must belong to the exceptional null-set where (4) is false.)

Theorem 7.19 *Let $F : \mathbb{R} \to \mathbb{R}$ be nondecreasing and right continuous. Then the set of points where F is not continuous is countable and F is differentiable a.e.*

Proof. Since

$$\sum_{x\in[-N,N]} (F(x+) - F(x-)) = F(N) - F(-N) < \infty,$$

the first assertion follows. Since $F(x + h) - F(x)$ equals $\mu_F(x, x + h]$ for $h > 0$ and $-\mu_F(x+h, x]$ for $h < 0$ and the sets $(x, x+h]$ and $(x+h, x]$ shrink nicely to x, the second statement now follows from (4). (In fact, it suffices with the $\| \ll m$ version of (4). Why?) $\square$

7.5 Bounded variation

In this section, we will investigate find the precise conditions for and the proofs of two profoundly essential results for integrals, namely the Fundamental Theorem of Calculus and the Integration by Parts Theorem. Let $F : \mathbb{R} \to \mathbb{C}$.

Definition 7.20 *The total variation of F, denoted T_F is the function given by*

$$T_F(x) = \sup\{\sum_1^n |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, -\infty < x_0 < x_1 < \ldots < x_n = x\}.$$

Note that adding an extra x_j on the right hand side of the definition of T_F only serves to increase $\sum_j |F(x_j) - F(x_{j-1}|$ for that particular set of x_j's. This means that when estimating $T_F(b)$ we may always assume that a given point $a < b$ is one of the x_j's if that is helpful. One consequence is that

$$T_F(b) - T_F(a) = \sup\{\sum_1^n |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, a = x_0 < x_1 < \ldots < x_n = x\}.$$

If $\lim_{x\to\infty} T_F(x) < \infty$,we say that F is of *bounded variation*. Let BV denote the space of functions $F : \mathbb{R} \to \mathbb{C}$ of bounded variation. By $BV[a, b]$, we denote space of F's defined on $[a, b]$ for which $T_F(b) - T_F(a) < \infty$. A function in $BV[a, b]$ is said to be of bounded variation on $[a, b]$. Here a few observations.

- If $F \in BV$, then the restriction to $[a, b]$ of F is in $BV[a, b]$.
- If $F \in BV[a, b]$, then the extension of F given by $F(x) = F(a)$ for $x < a$ and $F(x) = F(b)$ for $x > b$, is in BV.
- BV is a complex vector space.
- If F is differentiable and F' is bounded, then by the Mean Value Theorem, $T_F(b) - T_F(a) \leq (b - a)\sup_t F(t) < \infty$, and hence $F \in BV[a, b]$ for all $-\infty < a < b < \infty$.

Lemma 7.21 *If F is real-valued and $F \in BV$, then $T_F - F$ and $T_F + F$ are nondecreasing.*

Proof. Pick y arbitrarily, pick $\epsilon > 0$ and pick $x < y$. Pick $x_0 < x_1 < \ldots < x_n = x$ so that $\sum_1^n |F(x_j) - F(X_{j-1})| > T_F(x) - \epsilon$. Then

$$\begin{aligned} T_F(y) + F(y) &\geq \sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)| + F(y) \\ &\geq \sum_1^n |F(x_j) - F(x_{j-1})| + F(x) \\ &> T_F(x) - \epsilon + F(x) \end{aligned}$$

Since ϵ was arbitrary, it follows that $T_F + F$ is nondecreasing. The other part is analogous. □

Theorem 7.22 *(a)* $F \in BV$ *iff* $\Re F, \Im F \in BV$,

(b) The real-valued function F is in BV iff F can be written as the difference between two bounded nondecreasing functions.

(c) If $F \in BV$ is real-valued, then $F(x+)$ and $F(x-)$ exist for all x and $F(\pm\infty)$ both exists.

(d) If $F \in BV$, then the set of points where F is discontinuous is countable.

(e) If $F \in BV$ is real-valued and right continuous, then F is differentiable a.e.

Proof. Parts (c), (d) and (e) follow from (a), (b) and Theorem 7.19, so it suffices to prove (a) and (b). Part (a) is obvious, so it remains to prove (b). The if-direction follows from the third and fourth notes above. For the only if-direction, write

$$F = \frac{1}{2}(T_F + F) + \frac{1}{2}(T_F - F),$$

which is by Lemma 7.21 the difference of two increasing functions, which are bounded since $F \in BV$. □

Let $F \in BV$. If F is real-valued, then writing, as in (b) of the above Theorem, $F = F_1 - F_2$, where F_1 and F_2 are nondecreasing and bounded is called to decompose F in its positive and negative variations. If f is complex-valued, we can write $F = F_1 - F_2 + i(G_1 - G_2)$, where the F_i's and G_is are the positive/negative variations of $\Re F$ and $\Im F$ respectively.

Denote by NBV the space of $F \in BV$ such that $F(-\infty) = 0$ and F is right continuous. For an $F \in NBV$, the functions F_1, F_2, G_1 and G_2 are all right continuous. Hence we can define the complex measure μ_F given by $\mu_F = \mu_{F_1} - \mu_{F_2} + i(\mu_{G_1} - \mu_{G_2})$.

Proposition 7.23 *If $f \in NBV$, then $F' \in L^1(m)$. Moreover $\mu_F \perp m$ iff $F' = 0$ a.e. and $\mu_F \ll m$ iff $F(x) = \int_{-\infty}^{x} F'(t)dt$*

Note. Theorem 7.22(e) guarantees that $F'(x)$ exists for almost every x, so the present proposition should be read with the understanding that F' is extended by defining it arbitrarily on the exceptional null-set.

Proof. By the definition of derivative, $F'(x) = \lim_{r \to 0}(\mu_F(E_r)/m(E_r))$, where $E_r = (x, x+r]$ for $r > 0$ and $E_r = (x+r, x)$ for $r < 0$. By the observations following Theorem 7.18, $F'(x) = d\mu_F/dm$ a.e. By the LRNT, this entails that

$$F(x) = \lambda(-\infty, x] + \int_{-\infty}^{x} F'(t)dt$$

where $\lambda \perp m$ and F' must be in $L^1(m)$ since F must be bounded by virtue of being of bounded variation. □

One part of Proposition 7.25 is that the Fundamental Theorem of Calculus holds for $F \in NBV$ defined on the whole real line, such that $\mu_F \ll m$. Can the latter criterion be stated in a way which is in a more direct way in terms of F itself? The answer is yes:

Definition 7.24 *A function $F : \mathbb{R} \to \mathbb{C}$ is said to be absolutely continuous if for all $\epsilon > 0$ there exists a $\delta > 0$ such that $\sum_1^n |F(b_j) - F(a_j)| < \epsilon$ whenever $a_1 < b_1 < a_2 < \ldots, b_n$ and $\sum_1^n (b_j - a_j) < \delta$.*

Note that absolute continuity is stronger than uniform continuity (and thus stronger than continuity), since uniform continuity follows from taking $n = 1$ in the definition of absolute continuity. We say that F is absolutely continuous on $[a, b]$ if it satisfies the definition restricted to $a \leq a_j, b_j \leq b$.

Example. If F is differentiable everywhere and F' is bounded, then by the Mean Value Theorem, then $|F(b_j) - F(a_j)| \leq \max_x F'(x)(b_j - a_j)$ for any a_j, b_j, so F is absolutely continuous. □

Proposition 7.25 *If $f \in NBV$, then F is absolutely continuous iff $\mu_F \ll m$.*

Proof. If $\mu_F \ll m$, then we claim that for each $\epsilon > 0$ there is a $\delta > 0$ such that $\mu_F(E) < \epsilon$ whenever $m(E) < \delta$. It suffices to prove the claim for positive μ_F. Suppose for contradiction the there are E_k such that $m(E_k) < 2^{-k}$ but $\mu_F(E_k) \geq \epsilon$. By Borel-Cantelli, $m(\limsup_k E_k) = 0$. However, for each n, $\mu_F(\cup_n^\infty E_k) \geq \epsilon$. Since $F \in NBV$, μ_F is finite, so it follows from continuity of measures that $\mu_F(\limsup_k E_k) \geq \epsilon$, contradicting that $\mu_F \ll m$.

For the only-if direction, pick E so that $m(E) = 0$, pick $\epsilon > 0$ and a corresponding δ according to the definition of absolute continuity. By outer regularity of m and μ_F there are open sets $U_1 \supseteq U_2 \supseteq \ldots \supseteq E$ such that $m(U_1) < \delta$ and $\mu_F(E_j) \downarrow \mu_F(E)$. Each U_j can be written as a countable union of intervals:

$$U_j = \bigcup_k (a_j^k, b_j^k).$$

It follows from the absolute continuity of F, since $\sum_k (b_j^k - a_j^k) < \delta$ for each j, that

$$\begin{aligned}
|\mu_F(U_j)| &\leq \lim_n \sum_{k=1}^n |\mu_F(a_j^k, b_j^k)| \\
&\leq \lim_n \sum_{k=1}^n |F(b_j^k) - F(a_j^k)| \leq \epsilon.
\end{aligned}$$

Hence $\mu_F(E) = 0$, proving that $\mu_F \ll m$. □

Remark. It may come as a surprise that continuity of F is not sufficient for $\mu_F \ll m$. However, consider the Cantor set C on $[0, 1]$. As in Section 2, represent each number $x \in [0, 1]$ by its trinary expansion

$$x = \sum_1^\infty a_n(x) 3^{-n},$$

$a_n(x) \in \{0, 1, 2\}$. For $x \in C$, let $b_n(x) = a_n(x)/2$ (recall that $a_n(x) \in \{0, 2\}$ whenever $x \in C$). Let $F(x) = \sum_1^\infty b_n(x) 2^{-n}$. Extend F to a function on $[0, 1]$ by letting $F(x) = \sup\{F(c) : c \in C, c \leq x\}$. Then F is a.e. constant, in the sense that for any $x \notin C$, there is an open interval containing x on which F is constant.

Nevertheless, $F(0) = 0$ and $F(1) = 1$. Since F is increasing and $F[0,1] = [0,1]$, F is continuous. The measure μ_F however, is concentrated on C. Thus $\mu_F \perp m$, despite F being continuous. The function F is known as the *Cantor function*.

So, by Proposition 7.25, for functions $F \in NBV$, absolute continuity of F implies that $F(x) = \int_{-\infty}^{x} F'(t)dt$. For F defined on an interval $[a,b]$ (or $F(x) = F(a)$, $x < a$ and $F(x) = F(b)$, $x > b$, things are even a bit better.

Lemma 7.26 *If F is absolutely continuous on $[a,b]$, then $F \in BV[a,b]$.*

Proof. Take $\epsilon = 1$ in the definition of absolute continuity of F and pick δ accordingly. Let $N = \lfloor (b-a)/\delta \rfloor + 1$. For any given $a = x_0 < x_1 < \ldots < x_n = b$, group the intervals $(x_{j-1}, x_j]$ into N groups such that the total length of each group is less than δ; this can be done by the choice of δ, at least after adding some extra x_j's. Hence the sum of the $|F(x_j) - F(x_{j-1})|$'s over each group is bounded by 1, so

$$\sum_{1}^{n} |F(x_j) - F(x_{j-1})| \le N.$$

Since the x_j's were arbitrary, this shows that $T_F(b) \le N$, in particular $F \in BV[a,b]$. □

Summing up, we get

Theorem 7.27 *(**The Fundamental Theorem of Calculus**) Let $-\infty < a < b < \infty$. Then $F : [a,b] \to \mathbb{C}$ is absolutely continuous iff $f \in BV[a,b]$, F is differentiable a.e., $F' \in L^1([a,b], \mathcal{L}, m)$ and*

$$F(x) = \int_a^x F'(t)dt$$

for every $x \in [a,b]$.

Next we consider integration by parts. For $F \in NBV$, write $\int_E f dF$ for $\int_E f d\mu_F$.

Theorem 7.28 *(**Integration by Parts**) Let $F, G \in NBV$ and assume that G is continuous. Let $-\infty < a < b < \infty$. Then*

$$\int_{(a,b]} F dG + \int_{(a,b]} G dF = F(b)G(b) - F(a)G(a).$$

Proof. By Theorem 7.22 parts (a) and (b), it suffices to do this for F and G increasing. Let $\Omega = \{(x, y) : a < x \leq y \leq b\}$. By Tonelli, we have on one hand that

$$\begin{aligned}(\mu_F \times \mu_G)(\Omega) &= \int_{(a,b]} \int_{(a,y]} dF(x)dG(y) \\ &= \int_{(a,b]} (F(y) - F(a))dG(y) \\ &= \int_{(a,b]} FdG - F(a)(G(b) - G(a)).\end{aligned}$$

and on the other hand

$$\begin{aligned}(\mu_F \times \mu_G)(\Omega) &= \int_{(a,b]} \int_{[x,b]} dG(y)dF(x) \\ &= \int_{(a,b]} (G(b) - G(x))dF(x) \\ &= G(b)(F(b) - F(a)) - \int_{(a,b]} GdF\end{aligned}$$

where the second equality requires that G is continuous. Equating the two expressions gives the result. $\square$

8 The law of large numbers

This section is devoted to proving the strong version of the Law of Large Numbers. Of course, there is a probability space $(X, \mathcal{M}, \mathbb{P})$ underlying all statements made. We begin with a fundamental observation.

Proposition 8.1 *Let ξ and η be independent integrable random variables. Then*

$$\mathbb{E}[\xi\eta] = \mathbb{E}[\xi]\mathbb{E}[\eta].$$

Proof. If ξ and η are simple functions, then the result follows directly from the definition of independence and easy algebraic manipulation. If ξ and η are positive, then let sequences of simple functions increase to ξ and η respectively.

Choose the sequences so that the simple functions are $\sigma(\xi)$- and $\sigma(\eta)$-measurable respectively (which is what one gets if one uses the basic construction of such simple functions). Then all functions of the first sequence are independent of all functions of the second sequence, by being functions of independent random variables. The result now follows for positive functions. Finally the general result follows from linearity of integrals. □

The weak Law of Large Numbers is very easy to prove and goes as follows. Here and in the sequel the abbreviation "iid" stands for "independent and identically distributed". Also, for a sequence of real numbers $x_1, x_2, \ldots$, the quantity $\overline{x}_n$ denotes the average of the first n x_j:s, i.e. $\overline{x}_n = n^{-1}\sum_1^n x_j$.

Theorem 8.2 *(***Weak Law of Large Numbers***) Assume that $\xi_1, \xi_2, \ldots$ are iid random variables such that $\mathbb{E}[\xi_1] = 0$ and $\mathbb{E}[\xi_1^2] = M_2 < \infty$. Then for any $\epsilon > 0$,*

$$\lim_n \mathbb{P}(|\overline{\xi}_n| > \epsilon) = 0.$$

Proof. By the above proposition, $\mathbb{E}[\overline{\xi}_n^2] = n^{-1}\mathbb{E}[\xi_1^2]$. Hence, by Markov's inequality,

$$\mathbb{P}(|\overline{\xi}_n| > \epsilon) = \mathbb{P}(\overline{\xi}_n^2 > \epsilon^2) \leq \frac{M_2}{n\epsilon^2},$$

which tends to 0 as $n \to \infty$. □

Obviously, if $\mathbb{E}[\xi_1] = v \neq 0$, then applying the result to $\xi_j - v$ gives that $\mathbb{P}(|\overline{\xi}_n - v| > \epsilon) \to 0$. The strong law will make away with the assumption of finite second moment and also prove that $\overline{\xi}_n \to 0$ a.s., which is clearly a stronger result in both aspects. As for the weak law, it is obviously sufficient to consider the case $\mathbb{E}[\xi_1] = 0$.

The strong law has a reputation of having a very involved proof. This is not entirely correct. Granted, compared to the weak law it is involved, but compared to other fundamental mathematical results it is certainly not. Here we will present the "elementary proof"; the other standard proof uses martingale theory, which is not a topic of this course.

Let us begin with a short and elegant proof of a.s. convergence under the assumption of bounded fourth moment. The proof of the full strong law does not rely on this result, so we may regard it as a side track. On the other hand, it is more general in that it does not assume iid random variables, only that they are independent and have the same expectation.

Theorem 8.3 *(***Law of Large Number under Bounded 4'th Moment (LLN(4))***)* *Let $\xi_1, \xi_2, \ldots$ be independent random variables such that $\mathbb{E}[\xi_j] = 0$ for all j and such that there exists $M_4 < \infty$ such that $\mathbb{E}[\xi_j^4] \le M_4$ for all j. Then $\lim_n \overline{\xi}_n = 0$ a.s.*

Proof. Let $S_n = \sum_1^n \xi_j$. Then

$$\mathbb{E}[S_n^4] = \sum_1^n \mathbb{E}[\xi_j^4] + 6 \sum \sum_{i<j} \mathbb{E}[\xi_i^2]\mathbb{E}[\xi_j^2]$$

since the other terms of the expansion of S_n^4 have expectation 0 by assumption and the above proposition. Now suppose η is an integrable positive random variable and let $v := \mathbb{E}[\eta]$. Then $0 \le \int (\eta - v)^2 d\mathbb{P} = \int \eta^2 - 2v \int \eta + v^2 = E[\eta^2] - \mathbb{E}[\eta]^2$. Apply this on $\eta = \xi_j^2$ to get that $\mathbb{E}[\xi_j^2] \le \mathbb{E}[\xi_j^4]^{1/2} \le M_4^{1/2}$. Hence

$$\mathbb{E}[S_n^4] \le \left(n + 6\binom{n}{2}\right) M_4 \le 3n^2 M_4.$$

Therefore $\mathbb{E}[(S_n/n)^4] \le 3n^{-2}M_4$, so

$$\mathbb{E}\Big[\sum_1^\infty \Big(\frac{S_n}{n}\Big)^4\Big] < \infty$$

which in particular entails that $(S_n/n)^4 \to 0$ a.s. □

Lemma 8.4 *Let $\xi_1, \xi_2, \ldots$ be independent random variables with $\mathbb{E}[\xi_j] = 0$ and $\sum_1^\infty \mathbb{E}[\xi_j^2] < \infty$. Then $\sum_1^n \xi_j$ converges as $n \to \infty$ a.s.*

Proof. Let $M := \sum_1^\infty \mathbb{E}[\xi_j^2]$. Let $S_n = \sum_1^n \xi_j$. Fix two rational numbers $a < b$ and let U_n be the number of *up-crossings* of (a, b) if $S_1, \ldots, S_n$, i.e.

$$U_n = \max\{k : \exists s_1 < t_1 < s_2 < \ldots < t_k \le n : \forall 1 \le j \le k : S_{s_j} \le a, S_{t_j} \ge b\}.$$

Define the $0/1$-random variables $C_1, C_2, \ldots$ by taking $C_1 = 1$ if $a > 0$ and $C_1 = 0$ otherwise and then recursively

$$C_n = \chi_{\{C_{n-1}=1, S_{n-1}<b\} \cup \{C_{n-1}=0, S_{n-1}\le a\}}.$$

Let $T_n = \sum_1^n C_j \xi_j$. Since each C_n is $\sigma(\xi_1, \ldots, \xi_{n-1})$-measurable, C_n and ξ_n are independent and hence $\mathbb{E}[T_n] = 0$. However

$$T_n \geq (b-a)U_n - (S_n - a)^-$$

so the expectation of the right hand side is at most 0. Hence

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[|S_n - a|]}{b-a} \leq \frac{|a| + \mathbb{E}[S_n^2]^{1/2}}{b-a} \leq \frac{|a| + M^{1/2}}{b-a}.$$

Letting $U_\infty = \lim_n U_n$, the MCT gives $\mathbb{E}[U_\infty] < \infty$, so that $U_\infty < \infty$ a.s. By countable additivity of measures, this holds simultaneously for all rational a and b. Hence the sequence $\{S_n\}$ a.s. has only finitely many up-crossings of all nonempty intervals, which means that either S_n converges or $|S_n| \to \infty$. In either case $\lim_n |S_n|$ exists, but may be infinite. However, by Fatou's Lemma,

$$\mathbb{E}[\lim_n |S_n|] \leq \liminf_n \mathbb{E}[|S_n|] \leq \liminf_n \mathbb{E}[S_n^2]^{1/2} = \liminf_n \sum_1^n \mathbb{E}[\xi_j^2] \leq M^{1/2},$$

where the last equality follows from independence and the final inequality by assumption. Hence $\lim_n |S_n| < \infty$ a.s. $\square$

Lemma 8.5 *(***Césàro's Lemma***) Suppose that* $v_1, v_2, \ldots$ *is a sequence of real numbers such that* $\lim_n v_n = v_\infty$. *Then* $\lim_n \overline{v}_n = v_\infty$.

Proof. Fix N so large that $n > N \Rightarrow |v_n - v_\infty| < \epsilon$. Then for $n > N$,

$$\overline{v}_n > \frac{1}{n}\sum_1^N v_j + \frac{n-N}{n}(v_\infty - \epsilon) \to v_\infty - \epsilon$$

and

$$\overline{v}_n < \frac{1}{n}\sum_1^N v_j + \frac{n-N}{n}(v_\infty + \epsilon) \to v_\infty + \epsilon$$

as $n \to \infty$. $\square$

Lemma 8.6 *(***Kronecker's Lemma***) Suppose* $x_1, x_2, \ldots$ *are real numbers such that* $\sum_1^n (x_j/j)$ *converges as* $n \to \infty$. *Then* $\lim_n \overline{x}_n = 0$.

Proof. Let $v_n = \sum_1^n (x_j/j)$ and $v_\infty = \lim_n v_n$. With this notation, we get

$$\sum_1^n x_j = \sum_1^n j\frac{x_j}{j} = \sum_1^n j(v_j - v_{j-1}) = nv_n - \sum_1^n v_{j-1}.$$

Hence

$$\overline{x}_n = v_n - \frac{1}{n}\sum_1^n v_{j-1} \to 0$$

by Césàro's Lemma. □

The next step is the strong law under a mild variance restriction.

Theorem 8.7 *(***Law of Large Numbers under Variance Restriction (LLN(V))***) Let $\psi_1, \psi_2, \ldots$ be independent random variables with $\mathbb{E}[\psi_j] = 0$ for all j and $\sum_1^\infty (\mathbb{E}[\psi_j^2]/n^2) < \infty$. Then $\lim_n \overline{\psi}_n = 0$ a.s.*

Proof. By Kronecker's Lemma, it sufffices to prove that $\sum_1^n (\psi_j/j)$ converges as $n \to \infty$ a.s. This in turn follows from Lemma 8.4 on taking $\xi_n = \psi_n/n$. □

Lemma 8.8 *(***Kolmogorov's Truncation Lemma (KTL)***) Let $\xi_1, \xi_2, \ldots$ be iid random variables with $\mathbb{E}[\xi_j] = 0$ for all j. Let $\eta_j = \xi_j \chi_{\{|\xi_j| < j\}}$. Then*

(a) $\lim_n \mathbb{E}[\eta_n] = 0$,

(b) $\mathbb{P}(\limsup_n \{x : \xi_n(x) \neq \eta_n(x)\}) = 0$,

(c) $\sum_1^\infty (\mathbb{E}[\eta_j^2]/n^2) < \infty$.

Proof. Since η_n has the same distribution as $\xi_1 \chi_{\{|\xi_1| < n\}}$ which converges pointwise to ξ_1, it follows by the DCT using $|\xi_1|$ as a dominating L^1 function, that

$$\mathbb{E}[\eta_n] \to \mathbb{E}[\xi_1] = 0.$$

This proves (a). For (b):

$$\begin{aligned} \sum_1^\infty \mathbb{P}(\eta_j \neq \xi_j) &= \sum_1^\infty \mathbb{P}(|\xi_1| \geq n) \\ &= \mathbb{E}\Big[\sum_1^\infty \chi_{\{|\xi_1| \geq n\}}\Big] \\ &\leq \mathbb{E}\Big[|\xi_1|\Big] < \infty, \end{aligned}$$

where the second equality follows from the MCT. Hence (b) follows from Borel-Cantelli's Lemma. For (c):

$$\begin{aligned}
\sum_1^\infty \frac{\mathbb{E}[\eta_n^2]}{n^2} &= \sum_1^\infty \frac{\mathbb{E}[\xi_1^2 \chi_{\{|\xi_1|<n\}}]}{n^2} \\
&= \mathbb{E}\Big[\xi_1^2 \sum_1^\infty \frac{\chi_{\{|\xi_1|<n\}}}{n^2}\Big] \\
&= \mathbb{E}\Big[\xi_1^2 \sum_{n=\lfloor|\xi_1|\rfloor+1}^\infty \frac{1}{n^2}\Big] \\
&\leq 3\mathbb{E}\Big[|\xi_1|\Big] < \infty,
\end{aligned}$$

where the second equality follows from the MCT. □

Theorem 8.9 *(***The Law of Large Numbers***) Let $\xi_1, \xi_2, \ldots$ be iid random variables with $\mathbb{E}[\xi_1] = 0$. Then*

$$\lim_n \overline{\xi}_n = 0$$

almost surely.

Proof. Let $\eta_n = \xi_n \chi_{\{|_{1_n}|<n\}}$. By (c) of KTL and LLN(V), almost surely,

$$\lim_n \frac{1}{n} \sum_1^n (\eta_j - \mathbb{E}[\eta_j]) = 0.$$

By KTL (a) $\mathbb{E}[\eta_n] \to 0$, so by Césàro's Lemma, $n^{-1} \sum_1^n \mathbb{E}[\eta_j] \to 0$. Hence almost surely,

$$\lim_n \overline{\eta}_n = 0.$$

Finally, by KTL (b), almost surely $\eta_n \neq \xi_n$ for only finitely many n, so

$$\lim_n \overline{\xi}_n = 0$$

almost surely. □