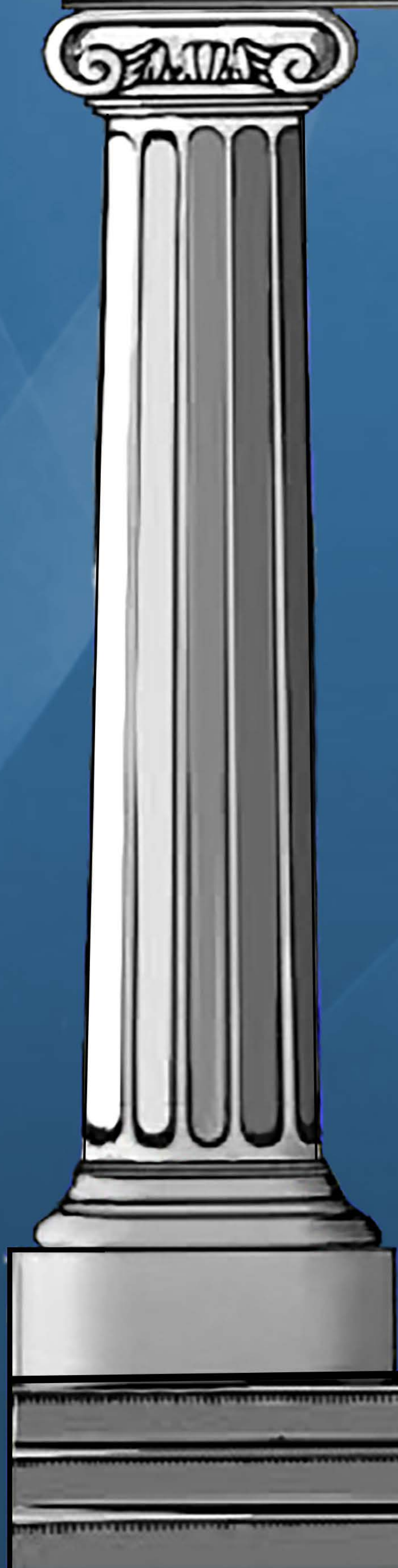
Foundations of Quantitative Finance



Book 5:

General Measure and Integration Theory

Robert R. Reitano

Foundations of Quantitative Finance: 5. General Measure and Integration Theory

Robert R. Reitano
Brandeis International Business School
Waltham, MA 02454

May, 2018

Contents

Preface

The idea for a reference book on the mathematical foundations of quantitative finance has been with me throughout my career in this field. But the urge to begin writing it didn't materialize until shortly after completing my first book, Introduction to Quantitative Finance: A Math Tool Kit, in 2010. The one goal I had for this reference book was that it would be complete and detailed in the development of the many materials one finds referenced in the various areas of quantitative finance. The one constraint I realized from the beginning was that I could not accomplish this goal, plus write a complete survey of the quantitative finance applications of these materials, in the 700 or so pages that I budgeted for myself for my first book. Little did I know at the time that this project would require a multiple of this initial page count budget even without detailed finance applications.

I was never concerned about the omission of the details on applications to quantitative finance because there are already a great many books in this area that develop these applications very well. The one shortcoming I perceived many such books to have is that they are written at a level of mathematical sophistication that requires a reader to have significant formal training in mathematics, as well as the time and energy to fill in omitted details. While such a task would provide a challenging and perhaps welcome exercise for more advanced graduate students in this field, it is likely to be less welcome to many other students and practitioners. It is also the case that quantitative finance has grown to utilize advanced mathematical theories from a number of fields. While there are also a great many very good references on these subjects, most are again written at a level that does not in my experience characterize the backgrounds of most students and practitioners of quantitative finance.

So over the past several years I have been drafting this reference book, accumulating the mathematical theories I have encountered in my work in this field, and then attempting to integrate them into a coherent collection of books that develops the necessary ideas in some detail. My target readers would be quantitatively literate to the extent of familiarity, indeed comfort, with the materials and formal developments in my first book, and sufficiently motivated to identify and then navigate the details of the materials they were attempting to master. Unfortunately, adding these details supports learning but also increases the lengths of the various developments. But this book was never intended to provide a "cover-to-cover" reading challenge, but rather to be a reference book in which one could find detailed foundational materials in a variety of areas that support current questions and further studies in quantitative finance.

Over these past years, one volume turned into two, which then became a work not likely publishable in the traditional channels given its unforgiving size and likely limited target audience. So I have instead decided to self-publish this work, converting the original chapters into stand-alone books, of which there are now nine. My goal is to finalize each book over the coming year or two.

I hope these books serve you well.

I am grateful for the support of my family: Lisa, Michael, David, and Jeffrey, as well as the support of friends and colleagues at Brandeis International Business School.

Robert R. Reitano

Brandeis International Business School

to Michael, David, and Jeffrey

Introduction

This is the fifth book in a series of several that will be self-published under the collective title of *Foundations of Quantitative Finance.* Each book in the series is intended to build from the materials in earlier books, with the first several alternating between books with a more foundational mathematical perspective, which was the case with the first and third and now this fifth book, and books which develop probability theory and some quantitative applications to finance, the focus of the second and fourth book. But while providing many of the foundational theories underlying quantitative finance, this series of books does not provide a detailed development of these financial applications. Instead this series is intended to be used as a reference work for students, researchers and practitioners of quantitative finance who already have other sources for these detailed financial applications but find that such sources are written at a level which assumes significant mathematical expertise, which if not possessed can be difficult to acquire.

Because the goal of many books in quantitative finance is to develop financial applications from an advanced point of view, it is often the case that the needed advanced foundational materials from mathematics and probability theory are introduced and summarized, but without a complete and formal development that would take the respective authors too far astray from their intended objectives. And while there are a great many excellent books on mathematics and probability theory, a number of which are cited in the references, such books typically develop materials with a eye to comprehensiveness in the subject matter, and not with an eye toward efficiently curating and developing the theory needed for applications in quantitative finance.

Thus the goal of this series is to introduce and develop in some detail a number of the foundational theories underlying quantitative finance, with topics curated from a vast mathematical and probability literature for the express purpose of supporting applications in quantitative finance. In addition, the development of these topics will be found to be at a much greater level of detail than in most advanced quantitative finance books, and certainly in more detail that most advanced mathematics and probability theory texts.

The title of this fifth book, General Measure and Integration Theory, generalizes the results of books 1 and 3 on the these respective topics. Chapter 1 begins the development with a discussion of measurable functions defined on a measure space $(X, \sigma(X), \mu)$, their properties and limits, as well as approximation results with simple functions, all reflective of similar results from book 1 in a Lebesgue measure space context. The proofs of these results are often "general" in notation only, reflecting the point made throughout book 1 that a great many of the proofs there often had nothing specifically to do with Lebesgue spaces and were generally applicable with a change of notation. This chapter also introduces the notion of monotone classes and develops the functional monotone class theorem, which proves that if a given class of functions satisfies a few, often simply verified properties, then it contains all bounded measurable functions.

A general integration theory on $(X, \sigma(X), \mu)$ is then developed in Chapter 2, largely following the outline used in book 3 in the development of the Lebesgue integral. But in addition to the various "integration to the limit" results of the bounded convergence and Lebesgue's monotone and dominated convergence theorems, this chapter also contains an additional result related to a uniformly integrable collection of functions on a finite measure space, and thus has important applications on probability spaces. This chapter then ends with an investigation into the evaluation of Lebesgue-Stieltjes integrals using Riemann sums. These integrals are defined by $\mu \equiv \mu_F$, a Borel measure induced by an increasing, right continuous function F and studied in book 1. The relationship between such integrals and the associated Riemann-Stieltjes integrals of book 3 are also discussed, where the latter integrals are defined using F as an integrator function.

Chapter 3 then investigates general results on "change of variables," starting off with the special case of Lebesgue-Stieltjes integrals where the underlying "distribution" function F is obtained from an associated density function f in the sense that:

$$\mu_F[A] = (\mathcal{L}) \int_A f(x)dx.$$

Letting $A = (a, b]$, this integral then must also equal $F(b) - F(a)$ and thus identifies the connection with F. In this context, change of variables is framed in terms of a change of measure, relating the $d\mu_F$ integrals to the associated

Lebesgue integrals. General results on change of variables under measurable transformations are then developed, followed by applications of this general result to Lebesgue integrals in two cases. The first case is when the change of variables is associated with a linear invertible transformation on $\mathbb{R}^n$, and this is then generalized to a change of variables associated with a differentiable, invertible transformation on $\mathbb{R}^n$.

The $L_p(X)$-type Banach spaces are studied in chapter 4, establishing some of their basic properties with an eye toward motivating the uniqueness of $L_2(X)$, a space that plays a prominent role in the theory of stochastic integration of book 8. Chapter 5 then turns to the topic of product space integrals and the theorems of Fubini and Tonelli, which identify conditions under which such integrals can be evaluated iteratively, or, one space variable at a time. This chapter also includes a discussion of how such results depend on the construction of the product space sigma algebra.

Applications of the Fubini-Tonelli theorems are developed in chapter 6, beginning with Lebesgue-Stieltjes integration by parts and the integrability of the convolution of integrable functions. The remainder of the chapter is focused on the Fourier transform of integrable functions and finite Borel measures, developing results related to Fourier inversion as well as a continuity theorem. Such results are key in the book 6 study of the characteristic function of a probability distribution and its applications.

The final chapter 7 focuses on various decompositions of sigma-finite measures defined on σ-finite measure spaces. The first section accumulates previously proved results to provide such a decomposition for Borel measures defined on the Lebesgue measure space on $\mathbb{R}$, and anticipates more general results to come. Following a digression into signed measures and the Hahn and Jordan decomposition theorems, the more general results developed are the Radon-Nikodým theorem and Lebesgue decomposition theorem.

Chapter 1

μ-Measurable Functions

We begin with a discussion of the definition and properties of a μ-measurable function defined on a measure space $(X, \sigma(X), \mu)$. For the definition of a general measure space $(X, \sigma(X), \mu)$ see definition 2.23 of book 1. Many of the proofs in this chapter will be omitted because the respective results from book 1 will be seen to be completely general in retrospect, after a change of notation. The reader is encouraged to provide details as an exercise.

For example it was seen in book 1 that Lebesgue measurability of a function defined on $(\mathbb{R}, \mathcal{M}_L, m)$, the Lebesgue measure space on $\mathbb{R}$, could be characterized by various equivalent properties. The following definition of a μ-**measurable function** f on $(X, \sigma(X), \mu)$ extends this equivalence. For this definition recall that for any set $A \subset \overline{\mathbb{R}}$, in the range of an extended real valued function f, the inverse f^{-1} is defined:

$$f^{-1}(A) \equiv \{x \in X | f(x) \in A\}.$$

For a discussion of the extended real numbers $\overline{\mathbb{R}}$ see section 3.1 of book 1. Finally, the statement that the set $f^{-1}(A)$ is μ-measurable means that $f^{-1}(A) \in \sigma(X)$. See also notation 1.3 below. See remark 1.2 for a justification of the equivalence of these properties.

Definition 1.1 *The* ***extended real-valued function*** $f(x) : X \rightarrow \overline{\mathbb{R}}$ *defined on the measure space* $(X, \sigma(X), \mu)$ *is said to be* μ***-measurable*** *if it is defined on a* μ*-measurable domain* $D \in \sigma(X)$*, and if any and hence all of the following conditions are satisfied:*

1. *For every real number* y*, the set* $f^{-1}((-\infty, y))$ *is* μ*-measurable.*

2. *For every real number y, the set $f^{-1}([y,\infty))$ is μ-measurable.*

3. *For every real number y, the set $f^{-1}((y,\infty))$ is μ-measurable.*

4. *For every real number y, the set $f^{-1}((-\infty,y])$ is μ-measurable.*

5. *For every Borel set $A\in\mathcal{B}(\mathbb{R})$, $f^{-1}(A)$ is μ-measurable.*

It $f(x)$ is μ-measurable then $f^{-1}(y)$ is μ-measurable for all $y\in\overline{\mathbb{R}}$.

Remark 1.2 *The book 1 proofs of the equivalence of these properties for Lebesgue measurable functions are in retrospect perfectly general because they only utilized the sigma algebra properties of $\sigma(X)$, the collection of measurable sets on the domain space. For the equivalence of 1 - 4 see proposition 3.4 of book 1, while the equivalence of 5 is proposition 3.26. Finally, the μ-measurability of $f^{-1}(y)$ is proposition 3.7.*

Notation 1.3 *There will be instances in coming chapters where we will encounter measure spaces $(X,\sigma_i(X),\mu)$ with various sigma-algebras $\sigma_i(X)$. In other words, the space X is fixed as is the measure μ, but there are various sigma algebras on which μ satisfies the definition of measure. A simple example but a common one is when $(X,\sigma(X),\mu)$ is a measure space and $\sigma_i(X)\subset\sigma(X)$ is a* ***sigma subalgebra,*** *then $(X,\sigma_i(X),\mu)$ is a measure space.*

*In this situation the notion of a μ-**measurable function** becomes ambiguous because the statement that $f^{-1}(A)$ is μ-measurable is ambiguous. When there is more than one sigma algebra on X, it is common to say that f **is** μ-**measurable relative to** $\sigma_i(X)$, or when μ is clear from the context, we say that f **is** $\sigma_i(X)$-**measurable.***

*Even more generally, if f is a mapping between measure spaces $(X,\sigma(X),\mu_X)$ and $(Y,\sigma(Y),\mu_Y)$, we could say that f **is** $\sigma(X)/\sigma(Y)$-**measurable** if $f^{-1}(A)\in\sigma(X)$ for all $A\in\sigma(Y)$. With this terminology, a μ-measurable function as defined above could be called $\sigma(X)/\mathcal{B}(\mathbb{R})$-**measurable,** but this level of formality is often not needed.*

1.1 Properties of Measurable Functions

In addition to noting the equivalence of the above defining properties, we generalize the book 1 development of the essential properties of Lebesgue measurable functions and note the applicability of the earlier proofs in this general context. For the following result recall definition 2.13 of book 1, that if the measure space $(X, \sigma(X), \mu)$ is also a topological space, the **Borel sigma algebra on** X is the smallest sigma algebra that contains the open sets of X, and denoted $\mathcal{B}(X)$.See definition 2.15 of book 1 for the definition of topology.

Proposition 1.4 *Given $f : D \rightarrow \mathbb{R}$ where $(X, \sigma(X), \mu)$ is a measure space and $D \subset X$ a μ-measurable set. Assume that X is also a topological space and that $\sigma(X)$ contains the open sets of X, and hence contains $\mathcal{B}(X)$, the Borel sigma algebra on X. If f is continuous on D, then f is a μ-measurable function on D.*

Proof. *As noted for the proof of proposition 3.13 of book 1, continuity of f means that $f^{-1}(G)$ is open in X for all G open in $\mathbb{R}$. Since the open sets in X are assumed to be measurable, this assures for example property 1 of the μ-measurable definition.* ■

Proposition 1.5 *Let $f(x)$ and $g(x)$ be **real-valued** μ-measurable functions on a common μ-measurable domain D of a measure space $(X, \sigma(X), \mu)$, and let $a, b \in \mathbb{R}$. Then the following are μ-measurable functions:*

1. $af(x) + b$,
2. $f(x) \pm g(x)$,
3. $f(x)g(x)$,
4. $f(x)/g(x)$ *on* $\{x|g(x) \neq 0\}$.

Proof. *The proof is identical to the Lebesgue measurable case in proposition 3.30 of book 1 because only the definition of μ-measurable function and the sigma algebra properties of the collection of measurable sets $\sigma(X)$ are required.* ■

The next result is an important example of when completeness is required of the measure space $(X, \sigma(X), \mu)$. As noted in definition 2.47 of book 1:

Definition 1.6 (Complete measure space) *A measure space $(X, \sigma(X), \mu)$ is **complete** if given $A \in \sigma(X)$ with $\mu(A) = 0$, then $B \in \sigma(X)$ for every $B \subset A$. Of necessity by subadditivity of μ, it is then the case that $\mu(B) = 0$ for all such B.*

As in the Lebesgue case, completeness of $(X, \sigma(X), \mu)$ is necessary for the following result.

Proposition 1.7 *Let $f(x)$ be a μ-measurable function on a* ***complete measure space*** *$(X, \sigma(X), \mu)$, and let $g(x)$ be a function with $f(x) = g(x)$ except on a set of μ-measure 0. Then $g(x)$ is μ-measurable.*

Proof. *If E is the set of μ-measure zero on which $f(x) \neq g(x)$, then*

$$\begin{aligned} \{x|g(x) < y\} &= \{x \in E|g(x) < y\} \cup \{x \notin E|g(x) < y\} \\ &= \{x \in E|g(x) < y\} \cup \{x \notin E|f(x) < y\}. \end{aligned}$$

The first set is a subset of a set of μ-measure zero and is hence μ-measurable by completeness, while the second set is the intersection of measurable $\widetilde{E}$, the complement of E, and μ-measurable $\{x|f(x) < y\}$. ■

Notation 1.8 *The notion that $f(x) = g(x)$ except on a set of μ-measure 0 is often written as $f(x) = g(x)$ μ-a.e., and read "μ almost everywhere."*

1.2 Limits of Measurable Functions

In this section we generalize many of the limiting results from book 1. See definition 3.36 of book 1 for infimum/supremum, and definition 3.42 for limits inferior and superior.

Proposition 1.9 *Given a sequence of μ-measurable functions $\{f_n(x)\}$ defined on a μ-measurable domain D of the measure space $(X, \sigma(X), \mu)$, the following functions are also μ-measurable:*

1. $\inf_n f_n(x)$,

2. $\sup_n f_n(x)$,

3. $\liminf f_n(x)$,

4. $\limsup f_n(x)$.

Proof. *The proof is again identical with the Lebesgue measurable case in proposition 3.47 of book 1 because the definitions of infimum/supremum and limits inferior/superior are identical, and only sigma algebra manipulations were needed for that proof.* ■

Corollary 1.10 *Given a sequence of μ-measurable functions $\{f_n(x)\}$ defined on μ-measurable domain D of a measure space $(X, \sigma(X), \mu)$:*

1. *If $f(x) \equiv \lim f_n(x)$ and this limit exists everywhere, then $f(x)$ is μ-measurable.*

2. *If* $\lim f_n(x)$ *exists almost everywhere and* $f(x) = \lim f_n(x)$ *almost everywhere, then* $f(x)$ *is* μ*-measurable if* $(X, \sigma(X), \mu)$ *is complete.*

Proof. *Result 1 follows from the preceding proposition since now:*

$$\lim f_n(x) = \liminf f_n(x) = \limsup f_n(x).$$

For almost everywhere convergence let E *denote the set of* μ*-measure zero on which this limit does not exist, and define* $\tilde{f}_n = f_n$ *on* $D - E$ *and* $\tilde{f}_n = 0$ *otherwise. Then* $\tilde{f}_n$ *is measurable by completeness and proposition 1.7, and* $\tilde{f}(x) \equiv \lim \tilde{f}_n(x)$ *exists everywhere and is measurable by part 1. Further, that both* $\tilde{f}$ *and* f *equal* $\lim f_n(x)$ *almost everywhere obtains that* $\tilde{f} = f$ μ*-a.e and thus* f *is measurable by completeness.* ■

We end this section with a few approximation results. The first states that pointwise convergence of measurable functions assures something more outside arbitrarily small sets, and resembles a uniform convergence result. But it does not assure uniform convergence outside a set of measure 0, nor even outside an arbitrarily small set, as discussed below.

Proposition 1.11 *Let* $\{f_n(x)\}$ *be a sequence of real-valued* μ*-measurable functions defined on a* μ*-measurable set* D *with* $\mu(D) < \infty$*, and let* $f(x)$ *be a real valued function so that* $f_n(x) \to f(x)$ *pointwise for* $x \in D$*. Then given* $\epsilon > 0$ *and* $\delta > 0$*, there is a* μ*-measurable set* $A \subset D$ *with* $\mu(A) < \delta$ *and an integer* N*, so that for all* $x \in D - A$ *and all* $n \geq N$*,*

$$|f_n(x) - f(x)| < \epsilon.$$

Proof. *Given* $\epsilon > 0$*, define*

$$G_n = \{x \mid |f_n(x) - f(x)| \geq \epsilon\},$$

and

$$D_N = \bigcup\nolimits_{n=N}^{\infty} G_n = \{x \mid |f_n(x) - f(x)| \geq \epsilon \text{ for some } n \geq N\}.$$

Then $\{D_N\}$ *is a nested sequence,* $D_{N+1} \subset D_N \subset D$*, and since* $f_n(x) \to f(x)$ *for each* $x \in D$*, it follows that for every* $x \in D$ *there is a* D_N *with* $x \notin D_N$*. Hence* $\bigcap_N D_N = \emptyset$ *and since* $\mu(D) < \infty$*, it follows by proposition 2.44 of book 1 (and the comment below remark 2.45) that* $\lim_{N\to\infty} \mu[\,D_N] \to 0$*. Thus given* $\delta > 0$ *there is an* N *with* $\mu[\,D_N] < \delta$*. Defining* $A \equiv D_N$*, then* $\mu(A) < \delta$ *and if* $x \notin A$ *it follows that* $|f_n(x) - f(x)| < \epsilon$ *for all* $n \geq N$ *by definition.* ■

Corollary 1.12 *If $(X, \sigma(X), \mu)$ is complete, the conclusion of the above proposition remains valid if $f_n(x) \to f(x)$ for each $x \in D$ outside a set of μ-measure 0.*

Proof. *Everything in the above proof remains the same except that we can now only conclude that for every $x \in D$ outside an exceptional set of measure 0, that there is an D_N with $x \notin D_N$, and hence $\bigcap_N D_N$ equals this set of measure 0. But then $\lim_{N\to\infty} m[\,D_N] \to 0$ again by proposition 2.44 of book 1and the proof follows as above.* ■

As observed in the Lebesgue case in remark 4.7 of book 1, this proposition does not imply that $f_n(x)$ converges uniformly to $f(x)$ on $D - A$ because as observed from the above proof, the set A depends on the given ϵ and δ. This result is close to but not equivalent to **Littlewood's third principle** of book 1**,** named for **J. E. Littlewood** (1885 – 1977). To improve this result to the Littlewood conclusion of "*nearly uniform convergence,*" it must be shown that A can be chosen so that $f_n(x) \to f(x)$ uniformly on $D - A$. That is, we need to find a fixed set A with $\mu(A) < \delta$ so that for any $\epsilon > 0$ there is an N such that $|f_n(x) - f(x)| < \epsilon$ for all $x \in D - A$ and all $n \geq N$. See the introduction to chapter 4 of book 1 for more on Littlewood's principles.

This next result is formally known as **Egorov's Theorem,** named for **Dmitri Fyodorovich Egorov** (1869 – 1931) and sometimes phonetically translated to **Egoroff**. It is also known as the **Severini–Egorov theorem** in recognition of the somewhat earlier and independent proof by **Carlo Severini** (1872 – 1951). The following proof is identical to the Lebesgue case in proposition 4.8 of the book 1.

Proposition 1.13 (Severini–Egorov theorem) *Let $\{f_n(x)\}$ be a sequence of μ-measurable functions defined on a μ-measurable set D with $\mu(D) < \infty$, and let $f(x)$ be a μ-measurable function so that $f_n(x) \to f(x)$ pointwise for $x \in D$. Then given $\delta > 0$ there is a μ-measurable set $A \subset D$ with $\mu(A) < \delta$, so that $f_n(x) \to f(x)$ uniformly on $D - A$. That is, for $\epsilon > 0$ there is an N so that $|f_n(x) - f(x)| < \epsilon$ for all $x \in D - A$ and $n \geq N$.*

Proof. *Given $\delta > 0$ define $\epsilon_m = 1/m$ and $\delta_m = \delta/2^m$ and apply proposition 1.11. The result is a set A_m with $\mu(A_m) < \delta_m$, and an integer N_m, so that for all $x \in D - A_m$ we have $|f_n(x) - f(x)| < \epsilon_m$ for $n \geq N_m$. Now let $A = \cup A_m$. By countable subadditivity, $\mu(A) \leq \sum \mu(A_m) = \delta$. We now show that $f_n(x) \to f(x)$ uniformly on $D - A$. Given ϵ there is an m so that $\epsilon_m < \epsilon$, and hence an N_m so that for all $x \in D - A_m$ we have $|f_n(x) - f(x)| < \epsilon_m < \epsilon$ for $n \geq N_m$. But then this statement is also true for $x \in D - A$ since $A_m \subset A$.*
■

Corollary 1.14 (Severini–Egorov theorem) *The result above remains valid if $f_n(x) \to f(x)$ μ-a.e. for $x \in D$ if $(X, \sigma(X), \mu)$ is complete.*
Proof. *Left as an exercise.* ■

Remark 1.15 *Note that the above propositions apply without the explicit need for the restriction of $\mu(D) < \infty$ in finite measure spaces, and in particular, probability spaces. In such a space we can conclude that pointwise convergence on any measurable set assures nearly uniform convergence.*

1.3 Approximating μ-Measurable Functions

We begin by generalizing the book 1 definition of a simple function, and note in advance that such functions are μ-measurable.

Definition 1.16 *A **simple function** on $(X, \sigma(X), \mu)$ is defined by*

$$\varphi(x) = \sum\nolimits_{i=1}^{n} a_i \chi_{A_i}(x), \tag{1.1}$$

where:

1. *$\{A_i\}_{i=1}^n \subset \sigma(X)$ are disjoint μ-**measurable sets**,*
2. *$\chi_{A_i}(x)$ is the **characteristic function** or **indicator function** for A_i, defined as $\chi_{A_i}(x) = 1$ for $x \in A_i$ and 0 otherwise,*
3. *$a_i \geq 0$ for all i.*

Remark 1.17 *As will be discussed in the first section of chapter 2, we do not restrict the definition of simple function to require that $\mu(\cup_{i=1}^n A_i) < \infty$ as in the Lebesgue case. In each case the respective definition reflects the approach taken or to be taken in the development of an integration theory. But because of this generalization, we now require $a_i \geq 0$ for all i. In the forthcoming definition of the μ-integral of such $\varphi(x)$ in 2.1, this assumption avoids the potential problem of having a definition which in effect contains terms of $\pm\infty$, a problem avoided in the Lebesgue development by requiring all $\mu(A_i) < \infty$.*

It is not strictly necessary to assume that $\{A_i\}_{i=1}^n$ are disjoint, but every simple function $\varphi(x)$ can be expressed this way. This follows because the range of a simple function is finite, say $\{b_j\}_{j=1}^m$. Then $B_j \equiv \varphi^{-1}(b_j)$ is measurable, $\{B_j\}_{j=1}^m$ are disjoint and:

$$\varphi(x) = \sum\nolimits_{j=1}^{m} b_j \chi_{B_j}(x),$$

That simple functions will be useful in the general development of an integration theory is predicted by the following proposition. Note that this result does not require that D have finite measure.

Proposition 1.18 *Let $f(x)$ be a nonnegative μ-measurable function defined on a μ-measurable set D of a measure space $(X, \sigma(X), \mu)$. Then there is an increasing sequence of simple functions $\{\varphi_n(x)\}_{n=1}^{\infty}$, so that $\varphi_n(x) \to f(x)$ for all $x \in D$.*

Proof. *Given n define $N \equiv n2^n + 1$ measurable sets $\{A_j^{(n)}\}_{j=1}^{N}$ by:*

$$A_j^{(n)} = \begin{cases} \{x \in D | (j-1)2^{-n} \leq f(x) < j2^{-n}\}, & 1 \leq j \leq N-1, \\ \{x \in D | n \leq f(x)\}, & j = N. \end{cases}$$

Next define

$$\varphi_n(x) = \left\{ (j-1)2^{-n}, \ x \in A_j^{(n)}, \ 1 \leq j \leq N. \right.$$

Then $\{\varphi_n(x)\}_{j=1}^{\infty}$ is an increasing sequence of simple functions with $\varphi_n(x) \to f(x)$ for all $x \in D$. In particular, $|f(x) - \varphi_n(x)| \leq 2^{-n}$ on $\{x \in D | f(x) < n\} \equiv D - A_N^{(n)}$. ■

Remark 1.19 *The above proposition and corollaries below can be applied more generally than as stated. For example, if $f(x)$ is a μ-measurable function defined on a μ-measurable set D of a measure space $(X, \sigma(X), \mu)$, express $f(x) = f^+(x) - f^-(x)$ where $f^+(x)$ and $f^-(x)$ are nonnegative functions defined below in 2.15 and 2.16 of definition 2.36. Then there are increasing sequences of simple functions $\{\varphi_n^+(x)\}$ and $\{\varphi_n^-(x)\}$ so that $\varphi_n^+(x) \to f^+(x)$ and $\varphi_n^-(x) \to f^-(x)$ for all $x \in D$. Defining $\varphi_n(x) = \varphi_n^+(x) - \varphi_n^-(x)$, then $\varphi_n(x) \to f(x)$ for all $x \in D$. Also, $\varphi_n(x)$ is nonnegative and increasing if $f(x) \geq 0$, and negative and decreasing if $f(x) \leq 0$. In addition, defining the simple function sequence, $\{|\varphi_n(x)|\}_{j=1}^{\infty} \equiv \{\varphi_n^+(x) + \varphi_n^-(x)\}_{j=1}^{\infty}$, this sequence is increasing and $|\varphi_n(x)| \to |f(x)|$ for all $x \in D$.*

For the following result, recall:

Definition 1.20 (σ-finite) *A measure space $(X, \sigma(X), \mu)$ is* ***sigma finite,*** *or σ-**finite**, if there exists a countable collection $\{B_j\} \subset \sigma(X)$ with $\mu(B_j) < \infty$ for all j and $X = \bigcup_{j=1}^{\infty} B_j$.*

Remark 1.21 *It is common to say in the above case, that the measure μ is a **sigma finite,** or σ-**finite measure,** since the key defining property is that of the measure of such subsets.*

Corollary 1.22 *If $f(x)$ is a nonnegative μ-measurable function defined on a μ-measurable set D of a measure space $(X, \sigma(X), \mu)$, then $\{\varphi_n(x)\}_{j=1}^{\infty}$ defined in proposition 1.18 above can be constructed so that each $\varphi_n(x)$ is zero outside a set of finite measure in the following cases:*

1. $\mu(D) < \infty$,
2. $(X, \sigma(X), \mu)$ *is* σ*-finite,*
3. $f(x)$ *is* μ*-integrable on* D.

Proof. *Statement 1 needs no proof. For 2, by definition 1.20 there exists a countable collection of measurable sets $\{B_j\}_{j=1}^{\infty}$, so that $X = \cup_{j=1}^{\infty} B_j$ and $\mu(B_j) < \infty$ for all j. Without loss of generality we can assume that this collection is nested, $B_j \subset B_{j+1}$, since given a general collection $\{B_i'\}$ we simply define $B_j = \bigcup_{i \leq j} B_i'$. Now redefine each $\varphi_n(x)$ by*

$$\varphi_n(x) = \left\{ (j-1)2^{-n}, \ x \in A_j^{(n)} \cap B_n, \ 1 \leq j \leq N. \right.$$

Finally, the last result is somewhat out of place since we have not yet even defined μ-integrability. However, given that the reader undoubtedly has an intuition of this definition from book 3, and that this proof requires little, we proceed justified by the convenience of having this result in this section. If $f(x)$ is nonnegative and integrable, then $0 \leq \varphi_n(x) \leq f(x)$ implies that each $\varphi_n(x)$ is integrable. But a simple function can be integrable if and only if it is zero outside a set of finite measure. ■

The next and last result generalizes proposition 1.18 above to allow more control over the choice of the $A_j^{(n)}$-sets in the case where the measure space is constructed as in chapters 1, 5 and 6 of book 1 using an outer measure μ_A^* defined relative to an algebra $\mathcal{A}$ and a measure on this algebra, μ_A. In these cases and developed in the referenced chapter 6, $\sigma(X)$ is the complete sigma algebra of **Carathéodory measurable sets** defined relative to this outer measure μ_A^*, and the measure μ is defined to equal μ_A^* restricted to $\sigma(X)$.

See definition 6.8 of book 1 for **semi-algebras** $\mathcal{A}$ and **algebras** $\mathcal{A}'$. Exercise 6.10 of book 1 demonstrates that the collection of finite **disjoint** unions of a semi-algebra $\mathcal{A}'$ is an algebra $\mathcal{A}$, called the **algebra generated by** $\mathcal{A}'$.

Remark 1.23 *The qualification of "disjoint" was inadvertently omitted in the statements of exercise 6.10 and proposition 6.13. However, note that the proof of the proposition assumed disjointedness.*

Proposition 1.24 *Let $f(x)$ be a nonnegative μ-measurable function defined on a μ-measurable set D of a complete measure space $(X, \sigma(X), \mu)$, where $\sigma(X)$ is the complete sigma algebra of Carathéodory measurable sets defined with respect to an outer measure μ_A^*. Assume that this outer measure is based on an algebra of sets $\mathcal{A}$ generated by a semi-algebra $\mathcal{A}'$, and a measure on this semi-algebra, μ_A. Then if:*

1. *D has finite measure, or,*

2. *$(X, \sigma(X), \mu)$ is σ-finite,*

then there is a sequence of simple functions $\{\psi_n(x)\}_{j=1}^{\infty}$ so that $\psi_n(x) \to f(x)$ for almost all $x \in D$. Further, each $\psi_n(x)$ is defined by characteristic functions of sets in $\mathcal{A}'$, and is zero outside a set of finite measure.

Proof. *If $\{\varphi_n(x)\}_{j=1}^{\infty}$ is the sequence constructed in proposition 1.18 for part 1, or in corollary 1.22 for part 2, then each $\varphi_n(x)$ is defined on $N = n2^n + 1$ finite μ-measurable $A_j^{(n)}$-sets which are disjoint by construction for each n. Recalling proposition 6.5 of book 1 (note notational change), for each n and j there exists $B_j^{(n)} \in \mathcal{A}_\sigma$, the collection of countable unions of sets in the algebra $\mathcal{A}$, so that $A_j^{(n)} \subset B_j^{(n)}$ and with N as above, $\mu(B_j^{(n)} - A_j^{(n)}) < 1/2N^2$. Now $B_j^{(n)} \in \mathcal{A}_\sigma$ implies that $B_j^{(n)} = \bigcup_{k=1}^{\infty} B_{jk}^{(n)}$ with $B_{jk}^{(n)} \in \mathcal{A}$ and hence:*

$$\bigcap\nolimits_M \left(B_j^{(n)} - \bigcup\nolimits_{k=1}^{M} B_{jk}^{(n)} \right) = \emptyset.$$

The above referenced proposition 6.5 also assures that $\mu(B_j^{(n)}) < \infty$, and thus it follows from continuity from above of μ (section 2.7.2, book 1) that there is an $M_j(n)$ so that $\mu\left(B_j^{(n)} - \cup_{k=1}^{M_j(n)} B_{jk}^{(n)}\right) < 1/2N^2$. By subadditivity:

$$\begin{aligned} & \mu\left[\left(\bigcup\nolimits_{k=1}^{M_j(n)} B_{jk}^{(n)} - A_j^{(n)}\right) \cup \left(A_j^{(n)} - \bigcup\nolimits_{k=1}^{M_j(n)} B_{jk}^{(n)}\right)\right] \\ \leq\ & \mu\left(B_j^{(n)} - A_j^{(n)}\right) + \mu\left(B_j^{(n)} - \bigcup\nolimits_{k=1}^{M_j(n)} B_{jk}^{(n)}\right) \\ <\ & 1/N^2. \end{aligned}$$

*Thus for each j, $\mu\left(A_j^{(n)} \triangle \cup_{k=1}^{M_j(n)} B_{jk}^{(n)}\right) < 1/N^2$ where $B_{jk}^{(n)} \in \mathcal{A}$ for all k. (Note: $A \triangle B$ is the **symmetric set difference** of definition 4.1 of book 1).*

Now given n, $\{\cup_{k=1}^{M_j(n)} B_{jk}^{(n)}\} \equiv \{C_j^{(n)}\}_j$ is a collection of $\mathcal{A}$-sets where $1 \leq j \leq N$, each of which is by definition a finite disjoint union of $\mathcal{A}'$-sets, say $C_j^{(n)} = \bigcup_{k=1}^{N_j} C_{jk}^{(n)}$. Then for each n the above estimate obtains:

$$\sum\nolimits_{j=1}^{N} \mu\left[A_j^{(n)} \triangle \bigcup\nolimits_{k=1}^{N_j} C_{jk}^{(n)}\right] < 1/N.$$

Now define $\psi_n(x)$ by

$$\psi_n(x) = \left\{ (j-1)2^{-n},\ x \in \bigcup\nolimits_{k=1}^{N_j} C_{jk}^{(n)},\ 1 \leq j \leq N. \right.$$

Then:

$$|f(x) - \psi_n(x)| \leq |f(x) - \varphi_n(x)| + |\varphi_n(x) - \psi_n(x)|,$$

and $|f(x) - \varphi_n(x)| \to 0$ for all x by proposition 1.18. For the second term, if $x \in A_j^{(n)}$ then since $\{C_j^{(n)}\}_j$ need not be disjoint it is possible in the worst case that $x \in \bigcup_{k=1}^{N_j} C_{lk}^{(n)}$ for many l. But by the above estimate:

$$\mu\{|\varphi_n(x) - \psi_n(x)| \neq 0\} \leq \sum\nolimits_{j=1}^{N} \mu\left[A_j^{(n)} \triangle C_j^{(n)}\right] < 1/N,$$

and thus $|\varphi_n(x) - \psi_n(x)| \to 0$ a.e. ■

Example 1.25 *If $(X, \sigma(X), \mu) = (\mathbb{R}, \mathcal{M}_L, m)$, Lebesgue measure space, or $(X, \sigma(X), \mu) = (\mathbb{R}, \mathcal{M}_{\mu_F}(\mathbb{R}), \mu_F)$, a Borel measure space, then as X is σ-finite in either case, the above proposition applies with $\mathcal{A}'$ defined as the semi-algebra of right semi-closed intervals. Thus if $f(x)$ is a nonnegative measurable function defined on a measurable set D, there is a sequence of simple functions $\{\psi_n(x)\}_{j=1}^{\infty}$, defined by characteristic functions of right semi-closed intervals, $(a_{jk}^{(n)}, b_{jk}^{(n)}] \in \mathcal{A}'$, so that $\psi_n(x) \to f(x)$ for almost all $x \in E$, and each $\psi_n(x)$ is zero outside a set of finite measure.*

1.4 The Functional Monotone Class Theorem

In this section we develop a result that will be very important in the stochastic integration theory of book 8. This result is the **monotone class theorem**, a characterizing result on sigma algebras attributed to **Paul Halmos** (1916 – 2006). This result then gives rise to the **functional monotone class theorem** which provides an expedient way to prove that a given collection of functions are in fact measurable if they satisfy a few relatively easily verifiable characteristics.

We begin with a definition of a monotone class of sets.

Definition 1.26 (Monotone class of sets) *A finite or countable collection* $\{A_j\}$ *is* ***monotone*** *if either:*

1. ***monotone increasing:*** $A_j \subset A_{j+1}$ *for all* j, *or,*
2. ***monotone decreasing:*** $A_{j+1} \subset A_j$ *for all* j.

A nonempty class of sets, M, *is a* ***monotone class*** *if given any monotone collection* $\{A_j\}$ *we have* $\lim A_j \in M$ *where:*

1. ***monotone increasing:*** $\lim A_j \equiv \bigcup_j A_j$,
2. ***monotone decreasing:*** $\lim A_j \equiv \bigcap_j A_j$.

Example 1.27 1. *Every sigma algebra is a monotone class since it is closed under all countable unions and intersections, not just unions and intersections of monotone (that is, nested) sets.*

2. *An algebra* $\mathcal{A}$ *that is a monotone class is in fact a sigma algebra since given* $\{A_j\}_{j=1}^{\infty} \subset \mathcal{A}$, *the collection* $\left\{\bigcup_{k=1}^{j} A_k\right\}_{j=1}^{\infty}$ *is monotone increasing. Hence if* $\mathcal{A}$ *is a monotone class,* $\bigcup_j \left[\bigcup_{k=1}^{j} A_k\right] = \bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$, *and so* $\mathcal{A}$ *is closed under countable unions and is thus a sigma algebra.*

3. *A monotone class need not be a sigma algebra. For example, let* $\mathcal{C} = \{A \subset \mathbb{R} | A \text{ is countable}\}$. *Then* $\mathcal{C}$ *is a monotone class but not a sigma algebra since for example,* $\mathcal{C}$ *is not closed under complements. In fact,* $\mathcal{C}$ *is not even a semi-algebra (definition 6.8 of book 1).*

Definition 1.28 *Given any collection of subsets* E *of a space* X, *define* $M(E)$ ***as the smallest monotone class that contains*** E, *and also called* ***the monotone class generated by*** E.

Remark 1.29 *This notion is well defined because the collection of all subsets of* X *is a monotone class that contains* E, *and the intersection of monotone classes is a monotone class. Hence the smallest monotone class is well defined, just as was the case for the smallest semi-algebra,* $\mathcal{A}'(E)$, *the smallest algebra,* $\mathcal{A}(E)$, *or smallest sigma algebra,* $\sigma(E)$, *that contain* E.

In these latter cases it is also common to refer to these collections as the ***semi-algebra, algebra,*** *and* ***sigma algebra generated by*** E.

The monotone class theorem was originally stated and proved in the context of rings and σ-rings of sets, an alternative approach to measure theory from that taken in this book which uses algebras and σ-algebras of sets. We state and prove this result from the latter perspective.

Proposition 1.30 (Monotone Class theorem) *If $\mathcal{A}$ is an algebra of sets, then $M(\mathcal{A}) = \sigma(\mathcal{A})$. Hence, any monotone class that contains $\mathcal{A}$ contains $\sigma(\mathcal{A})$.*

Proof. *By the above example 1, $M(\mathcal{A}) \subset \sigma(\mathcal{A})$ since $\sigma(\mathcal{A})$ is a monotone class. To complete the proof we will show that $M(\mathcal{A})$ is an algebra and hence by example 2 above $M(\mathcal{A})$ is a sigma algebra and so $\sigma(\mathcal{A}) \subset M(\mathcal{A})$.*

To show that $M(\mathcal{A})$ is an algebra requires the study of new collections of subsets. For any set A, define $C(A) \subset M(\mathcal{A})$ by:

$$C(A) = \{B \in M(\mathcal{A}) | A - B, B - A, A \cup B \text{ are in } M(\mathcal{A})\}.$$

We first show that for any A, $C(A)$ is a monotone class if it is not empty. If $\{B_j\}_{j=1}^{\infty} \subset C(A)$ is a monotone sequence then since $\{A \cup B_j\}_{j=1}^{\infty} \subset M(\mathcal{A})$ is also monotone, it follows that $\lim (A \cup B_j) \in M(\mathcal{A})$. But

$$\lim (A \cup B_j) = A \cup \lim B_j,$$

and so $A \cup \lim B_j \in M(\mathcal{A})$. A similar argument shows that $A - \lim B_j$ and $\lim B_j - A$ are also in $M(\mathcal{A})$, and thus by definition $\lim B_j \in C(A)$. So $C(A)$ is a monotone class if it is not empty, and this conclusion is true for any set A.

Now if $A \in \mathcal{A}$, note that $B \in C(A)$ for all $B \in \mathcal{A}$ since algebras are closed under finite operations. Thus $\mathcal{A} \subset \mathcal{C}(A)$ for any such $A \in \mathcal{A}$, and since $C(A)$ is a monotone class and $M(\mathcal{A})$ is the smallest monotone class containing $\mathcal{A}$, this yields that $M(\mathcal{A}) \subset C(A)$ for any $A \in \mathcal{A}$. It then follows that if $A \in M(\mathcal{A})$ and $B \in \mathcal{A}$, then $A \in C(B)$ and by symmetry $B \in C(A)$. Thus $\mathcal{A} \subset \mathcal{C}(A)$ and again as $C(A)$ is a monotone class, $M(\mathcal{A}) \subset C(A)$ for any $A \in M(\mathcal{A})$.

It now follows that $M(\mathcal{A})$ is an algebra. If $A, B \in M(\mathcal{A})$, then $A \in C(B)$ implies that $A \cup B \in M(\mathcal{A})$. Similarly, given $A \in M(\mathcal{A})$, then $A \cup \widetilde{A} \in \mathcal{A} \subset M(\mathcal{A})$, and it follows that $A \in C(A \cup \widetilde{A})$ and so $\widetilde{A} = A \cup \widetilde{A} - A \in M(\mathcal{A})$. Thus $M(\mathcal{A})$ is an algebra. ■

Example 1.31 (Uniqueness of Extensions of Measures) *In proposition 6.14 of book 1 was proved that the extension of a sigma finite measure from an algebra to a sigma algebra is effectively unique:*

Proposition 6.14 (Book 1): *Let μ_A be a sigma finite measure on an algebra $\mathcal{A}$, and μ the extension of μ_A induced by the outer measure $\mu_{\mathcal{A}}^*$. By extension is meant that $\mu(A) = \mu_A(A)$ for all $A \in \mathcal{A}$. If μ' is any other extension of μ_A, then $\mu(B) = \mu'(B)$ for all $B \in \sigma(\mathcal{A})$, where $\sigma(\mathcal{A})$ denotes the smallest sigma algebra that contains $\mathcal{A}$.*

Proof. *As an application of the monotone class theorem, we prove this result by showing that the class of sets on which $\mu = \mu'$ is a monotone class, and since this class contains the algebra $\mathcal{A}$ by assumption, it must contain $\sigma(\mathcal{A})$. By definition 1.20, if μ_A is **sigma finite** then the measure space X can be expressed as a countable union of $\mathcal{A}$-sets of finite measure:*

$$X = \cup_{j=1}^{\infty} X_j, \qquad \mu_A(X_j) < \infty.$$

As in the proof of corollary 1.22, we can assume that such sets are nested, so that $X_j \subset X_{j+1}$. By assumption, $\mu(X_j) = \mu'(X_j)$ for all j, so if it can be shown that for every j, $\mu(X_j \cap A) = \mu'(X_j \cap A)$ for all $A \in \sigma(\mathcal{A})$, then the conclusion follows by continuity from below of measures, that:

$$\mu(A) = \lim_{j\to\infty} \mu(X_j \cap A) = \lim_{j\to\infty} \mu'(X_j \cap A) = \mu'(A).$$

To show this, fix j and let $\mathcal{C} = \{A \in \sigma(\mathcal{A}) | \mu(X_j \cap A) = \mu'(X_j \cap A)\}$. Then $\mathcal{A} \subset \mathcal{C}$ by assumption, and to show that $\mathcal{C}$ is a monotone class let $\{A_k\} \subset \mathcal{C}$ be a monotone increasing sequence. By continuity from below and the definition of $\mathcal{C}$,

$$\mu(X_j \cap \lim A_j) = \lim \mu(X_j \cap A_j) = \mu'(X_j \cap \lim A_j),$$

and so $\lim A_j \in \mathcal{C}$. If this sequence is monotone decreasing, continuity from above is applicable because μ and μ' are finite measures restricted to X_j, and thus $\lim A_j \in \mathcal{C}$.

Hence $\mathcal{C}$ is a monotone class that contains $\mathcal{A}$ and it now follows from the above proposition that $\sigma(\mathcal{A}) \subset \mathcal{C}$. ■

As noted above, there is a functional counterpart to this theorem known as the **functional monotone class theorem** which allows one to conclude that if a given class of functions satisfies a few, often simply verified properties, then it contains all bounded measurable functions. This result is a useful tool and will be applied in forthcoming books.

Proposition 1.32 (Functional Monotone Class theorem) *Let $(X, \sigma(X), \mu)$ be a measure space and $\mathcal{A}'$ a semi-algebra that generates $\sigma(X)$, meaning that $\sigma(X) = \sigma(\mathcal{A}')$, the smallest sigma algebra that contains $\mathcal{A}'$. Let $\mathcal{L}$ denote a class of functions with the following properties:*

1. *$\chi_A \in \mathcal{L}$ for all $A \in \mathcal{A}'$ and $\chi_X \in \mathcal{L}$.*

2. *$\mathcal{L}$ is a vector space: If $f, g \in \mathcal{L}$ then $af + bg \in \mathcal{L}$ for all $a, b \in \mathbb{R}$.*

3. *If $f : X \to \mathbb{R}^+$ is bounded and the pointwise limit of $\{f_n\} \subset \mathcal{L}$, then $f \in \mathcal{L}$.*

Then $\mathcal{L}$ contains all bounded measurable functions defined on X.

Proof. *We first show that $\chi_A \in \mathcal{L}$ for all $A \in \mathcal{A}$, the algebra generated by $\mathcal{A}'$, and defined as the class containing the empty set and all finite disjoint unions of elements of $\mathcal{A}'$. To this end let $\mathcal{K}$ denote the class of all $A \in \sigma(X)$ so that $\chi_A \in \mathcal{L}$. Then by* 1*, $\mathcal{A}' \subset \mathcal{K}$ and $X \in \mathcal{K}$. If $A \in \mathcal{A}$ then by definition $A = \bigcup_{k=1}^n A_k$, a disjoint union of $\{A_k\}_{k=1}^n \subset \mathcal{A}'$, so by* 1 *and* 2*, $\chi_A \equiv \sum_{k=1}^j \chi_{A_k} \in \mathcal{L}$ and hence $A \in \mathcal{K}$. Also $\emptyset \in \mathcal{K}$ letting $a = b = 0$ in* 2*, so $\mathcal{A} \subset \mathcal{K}$.*

We next show as an application of the monotone class theorem that $\chi_A \in \mathcal{L}$ for all $A \in \sigma(X)$ and thus $\sigma(X) \subset \mathcal{K}$. Let $\{A_k\}_{k=1}^\infty \subset \mathcal{K}$ be a monotone sequence, and note that if $A = \lim A_j$ then $A \in \mathcal{K}$ by 3 *because $\chi_A = \lim_{j\to\infty} \chi_{A_j} \in \mathcal{L}$. Thus $\mathcal{K}$ is a monotone class that contains the algebra $\mathcal{A}$ and by proposition 1.30, $\mathcal{K}$ contains the sigma algebra $\sigma(X)$.*

If f is a bounded measurable function, write $f = f^+ - f^-$, with f^+ and f^- nonnegative and bounded and defined in 2.15 and 2.16. By 2*, $f \in \mathcal{L}$ if both f^+, $f^- \in \mathcal{L}$. Simplifying notation, assume that f is a bounded nonnegative measurable function. If f has finite range, $\{y_j\}_{j=1}^n$, then with $A_j = f^{-1}(y_j)$,*

$$f(x) = \sum\nolimits_{j=1}^n y_j \chi_{A_j}(x),$$

and so $f \in \mathcal{L}$ by 2 *because $A_j \in \sigma(X)$.*

More generally, for a bounded nonnegative measurable function f, define

$$f_n(x) = 2^{-n} \lfloor 2^n f(x) \rfloor ,$$

where $\lfloor 2^n f(x) \rfloor$ is the notation for the ***floor function,*** *or* ***greatest integer function,*** *defined as the greatest integer less than or equal to $2^n f(x)$. Then since f is nonnegative and bounded, $f_n(x)$ is finite valued and so $f_n \in \mathcal{L}$ for all n. But $\lfloor 2^n f(x) \rfloor = 2^n f(x) - \varepsilon_n(x)$ with $0 \leq \varepsilon_n(x) < 1$, and so $f_n(x) = f(x) - 2^{-n}\varepsilon_n(x)$. Consequently, $f_n(x) \to f(x)$ pointwise, and by* 3*, $f \in \mathcal{L}$.* ■

Exercise 1.33 *Show directly as an application of a variant of the* ***inclusion-exclusion principle*** *of proposition 8.8 of book* 1*, that for general $\{A_j\}_{j=1}^n \subset$*

$\mathcal{A}'$ *that* $A = \bigcup_{j=1}^{n} A_j \in \mathcal{K}$. *[Hint: Note that*

$$\chi_A = \chi_X - \prod_{j=1}^{n} \left(1 - \chi_{A_j}\right),$$

and the right hand expression can be expanded as a linear combination of χ_{B_k} *for* $B_k \in \mathcal{A}'$, *since for* $B_k \equiv \bigcap_{k=1}^{n_k} A_{j_k}$:

$$\chi_{B_k} = \prod_{k=1}^{n_k} \chi_{A_{j_k}}.$$

Then use 2.*]*

Remark 1.34 *In the section above, Approximating μ-Measurable Functions with Simple Functions, it was proved in proposition 1.18 that given a bounded nonnegative measurable function f, there exists an* ***increasing sequence of simple functions*** $\{\varphi_n(x)\}_{n=1}^{\infty}$, *defined to have finite range, so that $\varphi_n(x) \to f(x)$ pointwise. The significance of this is that the functional monotone class theorem can be stated with the more restrictive assumption:*

3'. If $f : X \to \mathbb{R}^+$ is bounded and the pointwise limit of increasing $\{f_n\} \subset \mathcal{L}$, then $f \in \mathcal{L}$.

The proof then uses increasing $\{\varphi_n(x)\}_{n=1}^{\infty}$ instead of the constructed $\{f_n(x)\}_{n=1}^{\infty}$ to demonstrate that $\mathcal{L}$ contains all bounded measurable functions on X.

Chapter 2

General Integration Theory

In this chapter we generalize the integration theory of the Lebesgue measure space $(\mathbb{R}^n,\ \mathcal{M}_L^n,\ m^n)$ of book 3, to Borel measure spaces $(\mathbb{R}, \mathcal{B}_{\mu_F}(\mathbb{R}), \mu_F)$, in which case the integrals are known as **Lebesgue-Stieltjes integrals**, and to more general measure spaces $(X, \sigma(X), \mu)$. Importantly this latter collection of spaces includes finite products of measure spaces,

$$(X, \sigma(X), \mu) \equiv (\prod_{i=1}^n X_i, \sigma\left(\prod_{i=1}^n X_i\right), \prod_{i=1}^n \mu_i),$$

defined in chapter 7 of book 1 with component spaces $\{(X_i, \sigma(X_i), \mu_i) | i = 1, ..., n\}$, in which case the integrals are known as **product measure integrals**, or **product space integrals**.

In the development of the **Lebesgue integral** of book 3, we followed the sequential steps:

1. Define the Lebesgue integral of simple functions which equal zero outside sets of finite measure.

2. Extend this definition to bounded measurable functions which again equal zero outside sets of finite measure, using "limits" of integrals of subordinate and dominant simple functions defined in step 1. It was then seen that a bounded function on a set E with $m(E) < \infty$ was Lebesgue integrable if and only if it was Lebesgue measurable.

3. Extend the definition to nonnegative Lebesgue measurable functions based on the integrals of bounded functions from step 2.

4. Extend the definition of Lebesgue integral to general measurable functions, splitting such functions into positive and negative parts and applying the results of step 3.

5. Along the way, various important limit theorems on the Lebesgue integral of sequences of functions were developed that are fundamental to the applications of the theory.

The above development reflects the conventional approach to this theory. However we could well have jumped from step 1 to step 3, defining the integral of nonnegative functions directly in terms of the integrals of simple functions, and done this without the restriction that simple functions are zero outside sets of finite measure. In a general measure space, the Lebesgue approach must in fact be modified in this way, because any restriction to functions which are zero outside a set of finite measure may create a counter-intuitive result.

Example 2.1 *Consider a measure space* $(X, \sigma(X), \mu)$ *with* $\sigma(X) = \{\emptyset, X\}$ *and* μ *defined by* $\mu(\emptyset) = 0$, $\mu(X) = \infty$. *An example of this is Lebesgue measure* $\mu = m$ *defined on the trivial sigma algebra on* $\mathbb{R}$. *Logically one expects that an integration theory will obtain* $\int_X 1 d\mu = \mu(X) = \infty$, *but we cannot derive this conclusion from the sequence of steps used above. The problem is that there are no simple functions defined to be zero outside a set of finite measure.*

Remark 2.2 *Note that this kind of result cannot occur on a* σ*-finite measure space.*

General Measure Space Integrals: For a general measure space $(X, \sigma(X), \mu)$, we must abandon the notion that a general domain $D \subset X$ with $\mu(D) = \infty$ contains measurable domains with finite measure, or that a given function $f(x)$ can be approximated with simple functions defined on such domains. Consequently, the approach take here will be to:

1. Define the μ-integrals of simple functions, defined generally without restrictions on the μ-measure of their domains.

2. Extend this definition to nonnegative μ-measurable functions, using the μ-integrals of simple functions.

3. Extend the definition of μ-integral to general μ-measurable functions, splitting such functions into positive and negative parts and applying the results of step 2.

4. Along the way, various important limit theorems on the μ-integral of sequences of functions will be developed that will be fundamental to the applications of the theory.

2.1 Integrating Simple Functions

Simple functions were defined in 1.1 of definition 1.16:

$$\varphi(x) = \sum\nolimits_{i=1}^{n} a_i \chi_{A_i}(x),$$

Next, we define the μ**-**integral of a simple function. While $a_i \geq 0$ for all i, the integral defined next need not be finite since there is no definitional restriction on the μ-measures of the disjoint sets $\{A_i\}_{i=1}^{n}$.

Definition 2.3 *Given a simple function defined in 1.1,* ***the*** μ***-integral of*** $f(x)$ *is defined as:*

$$\int \varphi(x) d\mu \equiv \sum\nolimits_{i=1}^{n} a_i \mu(A_i). \tag{2.1}$$

As noted in the book 3 developments, we must again show that this definition is well defined since a simple function can be represented with infinitely many choices of coefficients $\{a_i\}$ and sets $\{A_i\}$. In addition, the value of this integral will not be changed by redefinitions of $\varphi(x)$ on sets of μ-measure 0. The following result echoes proposition 2.15 of book 3 for the Lebesgue case, but must also address the generalization that the A_i-sets need not have finite measure.

Proposition 2.4 *With $\varphi(x)$ defined above, assume that $\varphi(x) = \psi(x)$ μ-a.e., where:*

$$\psi(x) = \sum\nolimits_{j=1}^{m} a'_j \chi_{A'_j}(x)$$

with $\{A'_j\}_{j=1}^{m}$ disjoint. Then

$$\sum\nolimits_{j=1}^{m} a'_j \mu(A'_j) = \sum\nolimits_{i=1}^{n} a_i \mu(A_i). \tag{(*)}$$

Further, if $a_i > 0$ and $a'_j > 0$ for all i, j :

$$\mu\left(\bigcup\nolimits_{j=1}^{m} A'_j - \bigcup\nolimits_{i=1}^{n} A_i\right) = \mu\left(\bigcup\nolimits_{i=1}^{n} A_i - \bigcup\nolimits_{j=1}^{m} A'_j\right) = 0.$$

Proof. *The proof is similar to that of proposition 2.15 for the Lebesgue case. For any i, j define $B_{i,j} \equiv A_i \cap A'_j$, and also define $B_{i,0} \equiv A_i \cap \left(\bigcup\nolimits_{j=1}^{m} A'_j\right)^c$*

and $B_{0,j} \equiv \left(\bigcup_{i=1}^{n} A_i\right)^c \cap A'_j$. *Then the full collection* $\{B_{i,j}\}$ *are disjoint,* $A_i = \bigcup_{j=0}^{m} B_{i,j}$ *and* $A'_j = \bigcup_{i=0}^{n} B_{i,j}$. *Now since* $\varphi(x) = \psi(x)$ μ*-a.e. it follows that for every* i *either* $\mu[B_{i,0}] = 0$ *or* $a_i = 0$, *and similarly for every* j *either* $\mu[B_{0,j}] = 0$ *or* $a'_j = 0$. *By finite additivity,* $\mu(A_i) = \sum_{i=0}^{m} \mu[B_{i,j}]$ *and* $\mu\left(A'_j\right) = \sum_{i=0}^{n} \mu[B_{i,j}]$, *and so using the previous sentence:*

$$\begin{aligned}
\sum_{i=1}^{n} a_i \mu(A_i) &= \sum_{i=1}^{n} \sum_{j=0}^{m} \mu[B_{i,j}]\, a_i = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu[B_{i,j}]\, a_i, \\
\sum_{i=1}^{m} a'_i \mu(A'_i) &= \sum_{i=1}^{n} \sum_{j=1}^{m} \mu[B_{i,j}]\, a'_j = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu[B_{i,j}]\, a'_j.
\end{aligned}$$

The identity in $(*)$ *now follows because for every* i, j *either* $\mu[B_{i,j}] > 0$ *and then of necessity* $a_i = a'_j$, *or* $\mu[B_{i,j}] = 0$.

If $a_i > 0$ *and* $a'_j > 0$ *for all* i, j *then as noted above,* $\mu[B_{i,0}] = 0$ *all* i *and* $\mu[B_{0,j}] = 0$ *all* j. *But then:*

$$0 = \mu\left[\bigcup_{i=1}^{n} B_{i,0}\right] = \mu\left[\left(\bigcup_{i=1}^{n} A_i\right) \cap \left(\bigcup_{j=1}^{m} A'_j\right)^c\right],$$

and this last expression is $\mu\left(\bigcup_{j=1}^{m} A'_j - \bigcup_{i=1}^{n} A_i\right)$. *The same applies to* $\mu\left(\bigcup_{i=1}^{n} A_i - \bigcup_{j=1}^{m} A'_j\right)$. ■

Example 2.5 **1**. *If* $(X, \sigma(X), \mu) = (\mathbb{R}, \mathcal{M}_{\mu_F}(\mathbb{R}), \mu_F)$, *a Borel measure space on* $\mathbb{R}$ *as in chapter 5 of book 1, then if* $\{(a_j, b_j]\}_{j=1}^{m} \subset \mathcal{M}_{\mu_F}(\mathbb{R})$ *is any disjoint collection of right semi-closed intervals and* $\varphi(x) = \sum_{j=1}^{m} c_j \chi_{(a_j, b_j]}(x)$, *then since* $\mu_F[(a_j, b_j]] = F(b_j) - F(a_j)$:

$$\int \varphi(x) d\mu_F = \sum_{j=1}^{m} c_j [F(b_j) - F(a_j)].$$

2. *If* $(X, \sigma(X), \mu) = (\mathbb{R}^n, \mathcal{M}^n_{\mu_F}(\mathbb{R}^n), \mu^n_F)$, *a product of Borel measure spaces:* $\{(\mathbb{R}, \mathcal{M}_{\mu_{F_i}}(\mathbb{R}), \mu_{F_i})\}_{i=1}^{n}$, *as in chapter 7 of book 1, then if* $\{\Pi_{i=1}^{n}(a_{ji}, b_{ji}]\}_{j=1}^{m} \subset \mathcal{M}_{\mu_F}(\mathbb{R}^n)$ *is any disjoint collection of right semi-closed rectangles and* $\psi(x) = \sum_{j=1}^{m} c_j \chi_{A_j}(x)$ *with* $A_j \equiv \Pi_{i=1}^{n}(a_{ji}, b_{ji}]$, *then since* $\mu^n_F[\Pi_{i=1}^{n}(a_{ji}, b_{ji}]] = \Pi_{i=1}^{n}[F_i(b_{ji}) - F_i(a_{ji})]$:

$$\int \psi(x) d\mu^n_F = \sum_{j=1}^{m} c_j \prod_{i=1}^{n} [F_i(b_{ji}) - F_i(a_{ji})].$$

3. *If* $(X, \sigma(X), \mu) = (\mathbb{R}^n, \mathcal{M}_F(\mathbb{R}^n), \mu_F)$*, a general Borel measure space as in chapter 8 of book 1, induced by a continuous from above and* n*-increasing function* $F(x_1, ..., x_n)$*, then if* $\{\Pi_{i=1}^n (a_{ji}, b_{ji}]\}_{j=1}^m \subset \mathcal{M}_F(\mathbb{R}^n)$ *is any disjoint collection of right semi-closed rectangles and* $\psi(x) = \sum_{j=1}^m c_j \chi_{A_j}(x)$ *with* $A_j \equiv \Pi_{i=1}^n (a_{ji}, b_{ji}]$*, then:*

$$\int \psi(x) d\mu_F = \sum\nolimits_{j=1}^m c_j \sum\nolimits_{x_{j_k}} sgn(x_{j_k}) F(x_{j_k}).$$

This follows since:

$$\mu_F \left[\prod\nolimits_{i=1}^n (a_{ji}, b_{ji}] \right] = \sum\nolimits_{x_{j_k}} sgn(x_{j_k}) F(x_{j_k}),$$

where for each j*, the summation over* x_{j_k} *denotes the sum over the* 2^n *vertices of* A_j*, and* $sgn(x_{j_k}) \equiv -1$ *if the number of components of* x_{j_k} *equal to* a_{ji} *is odd, and* $sgn(x_{j_k}) = 1$ *otherwise.*

Proposition 2.6 *Let* $\varphi(x)$ *and* $\psi(x)$ *be simple functions. Then:*

1. *If* $\varphi(x) \leq \psi(x)$ *except on a set of* μ*-measure* 0*, then*

$$\int \varphi(x) d\mu \leq \int \psi(x) d\mu. \tag{2.2}$$

2. *For any positive constants* a *and* b *:*

$$\int [a\varphi(x) + b\psi(x)] d\mu = a \int \varphi(x) d\mu + b \int \psi(x) d\mu. \tag{2.3}$$

Proof. *Left as an exercise. For part 2, decompose* $\varphi(x)$ *and* $\psi(x)$ *into summations as in the proof of proposition 2.4.* ■

Corollary 2.7 *The definition of the integral of a simple function is well defined even if the measurable collection* $\{A_i\}_{i=1}^n$ *are not disjoint.*

Proof. *This follows from part 2 since then* $\varphi(x) = \sum\nolimits_{j=1}^m a_j \varphi_j(x)$ *with* $\varphi_j(x) \equiv \chi_{A_j}(x)$*.* ■

Finally, we introduce the definition of the integral of simple function over a μ-measurable set E. This definition will apply in the general setting below, so is stated here in that general notational context despite the open question of existence of such integrals beyond nonnegative simple functions.

Definition 2.8 *Let E be a μ-measurable set, so $E \in \sigma(X)$, and $\chi_E(x)$ the* ***characteristic function*** *of E. If $f(x)$ a μ-measurable function, define :*

$$\int_E f(x)d\mu \equiv \int \chi_E(x)f(x)d\mu, \tag{2.4}$$

when the integral on the right exists.

Remark 2.9 *Note that in the case of a simple function that 2.4 modifies the definition in 2.1 to:*

$$\int_E \varphi(x)d\mu \equiv \sum\nolimits_{i=1}^n a_i\mu(A_i \cap E). \tag{2.5}$$

Exercise 2.10 *Show that if E_1, E_2 are disjoint and measurable and $E \equiv E_1 \cup E_2$, then for a simple function $\varphi(x)$:*

$$\int_E \varphi(x)d\mu = \int_{E_1} \varphi(x)d\mu + \int_{E_2} \varphi(x)d\mu.$$

2.2 Integrating Nonnegative Measurable Functions

Continuing with the program outlined in the introduction, we start with the definition:

Definition 2.11 *If $f(x)$ is a nonnegative μ-measurable, extended real valued function defined on a μ-measurable set E of a measure space $(X, \sigma(X), \mu)$, define the* ***the μ-integral of*** *$f(x)$* ***over*** *E by:*

$$\int_E f(x)d\mu = \sup_{\varphi \leq f} \int_E \varphi(x)d\mu. \tag{2.6}$$

Here $\varphi(x)$ is a ***simple function*** *as in definition 1.16. When this supremum is finite, $f(x)$ is said to be μ-**integrable.** When the supremum is infinite, though **not** μ-**integrable,** such a function will be said to have $\int_E f(x)d\mu = \infty$.*

Remark 2.12 *The above definition differs from that initially imposed in the second step for the Lebesgue integral, there generalizing the integral from simple functions to bounded functions. Specifically, the integral of bounded f over E with $\mu(E) < \infty$ was defined when:*

$$\inf_{\psi \geq f} \int_E \psi(x)d\mu = \sup_{\varphi \leq f} \int_E \varphi(x)d\mu. \tag{2.7}$$

When true, the integral $\int_E f(x)d\mu$ *was defined to be this common value.*

Although we do not do this here, this approach could also be applied in the current context by proving the following generalization of proposition 2.29 of book 3. While the following statement appears different from the earlier result because of the explicit requirement for completeness of $(X, \sigma(X), \mu)$*, recall that Lebesgue measure space is complete.*

Proposition 2.13 *Let* $f(x)$ *be defined and bounded on a measurable set* E *in a measure space* $(X, \sigma(X), \mu)$ *with* $\mu(E) < \infty$*. If* $f(x)$ *is* μ*-measurable, then 2.7 is satisfied. Conversely, if 2.7 is satisfied and* $(X, \sigma(X),\mu)$ *is complete, then* $f(x)$ *is* μ*-measurable.*

Proof. *Left as an exercise in transcribing the earlier proof to this context. Note that the validity of 2.7 will imply only that* $f(x)$ *is equal to a* μ*-measurable function* μ*-a.e., and hence the need for the completeness of* $(X, \sigma(X),\mu)$ *to justify an application of proposition 1.7.* ■

Recalling corollary 1.22, we document the following results on the construction of subordinate simple functions.

Proposition 2.14 *Let* $f(x)$ *be a nonnegative* μ*-measurable function defined on a* μ*-measurable set* E *of a measure space* $(X, \sigma(X),\mu)$ *which is* μ-***integrable*** *by definition 2.11. Then each* $\varphi_n(x)$ *of proposition 1.18 will automatically equal zero outside a set of finite measure.*

Proof. *Because* $\{\varphi_n(x)\}_{j=1}^{\infty}$ *is an increasing sequence and* $\varphi_n(x) \leq f(x)$*, we can conclude by definition 2.11 that for every* n*,*

$$\int_E \varphi_n(x)d\mu \leq \int_E f(x)d\mu,$$

and hence μ*-integrability of* $f(x)$ *assures that* $\int_E \varphi_n(x)d\mu < \infty$ *for all* n*. By 2.1 this implies that* $\mu\left(A_j^{(n)}\right) < \infty$ *for all* j*, and thus also that* $\mu\left(\bigcup_{j=1}^{N} A_j^{(n)}\right) < \infty$*.* ■

Corollary 2.15 *Let* $f(x)$ *be a nonnegative* μ*-measurable function defined on a* μ*-measurable set* E *of a complete measure space* $(X, \sigma(X),\mu)$*, where* $\sigma(X)$ *is the sigma algebra of Carathéodory measurable sets defined with an outer measure* μ_A^* *that is based on an algebra of sets* $\mathcal{A}$ *generated by a semi-algebra* $\mathcal{A}'$ *and a measure on this semi-algebra,* μ_A*. If* $f(x)$ *is* μ-***integrable*** *by definition 2.11, there is an increasing sequence of simple functions,* $\{\psi_n(x)\}_{j=1}^{\infty}$*, defined by characteristic functions of disjoint sets* *in $\mathcal{A}'$, so that $\psi_n(x) \to f(x)$ for almost all $x \in E$, and each $\psi_n(x)$ is zero outside a set of finite measure.*

Proof. *The construction is the same as in proposition 1.24 since μ-integrability of $f(x)$ assures that each of the $A_j^{(n)}$-sets has finite measure by proposition 2.14.* ■

While the integrability criterion in definition 2.11 is accessible, in practice it is not yet very useful for developing deeper properties of this integral because there is no apparent way to organize the potentially uncountably many simple functions this definition contemplates. For simpler properties, such as those summarized next, the above definition applies directly.

Proposition 2.16 *Given nonnegative μ-measurable functions, $f(x)$, $g(x)$ defined on E :*

1. ***Monotonicity:*** *If $f(x) \leq g(x)$ then:*
$$\int_E f(x)d\mu \leq \int_E g(x)d\mu.$$

2. ***Linearity:*** *If $a > 0$,*
$$\int_E af(x)d\mu = a\int_E f(x)d\mu.$$

3. ***Domain Decomposition:*** *If E_1, E_2 are disjoint and measurable and $E \equiv E_1 \cup E_2$, then:*
$$\int_E f(x)d\mu = \int_{E_1} f(x)d\mu + \int_{E_2} f(x)d\mu.$$

Proof. *Left as an exercise.* ■

However, the following property would seem to require some additional tools beyond that provided by the supremum definition:

4. **Additivity**: Given nonnegative μ-measurable functions $f(x)$ and $g(x)$ defined on E :
$$\int_E [f(x) + g(x)]\, d\mu = \int_E f(x)d\mu + \int_E g(x)d\mu.$$

What would make derivations easier is a limit theorem, the effect of which would be that if $\{f_n(x)\}$ is a sequence of functions with $f_n(x) \to f(x)$ in some manner, we could conclude that:

$$\int_E f(x)d\mu = \lim_{n\to\infty} \int_E f_n(x)d\mu.$$

In the Lebesgue development it was noted that the **bounded convergence theorem**, **Lebesgue's monotone convergence theorem**, and **Lebesgue's dominated convergence theorem**, each provided exactly this kind of conclusion for different categories of measurable functions, $f(x)$, based on different criteria on the given sequences.

Following the Lebesgue analysis for nonnegative measurable functions, we first state and prove **Fatou's lemma,** named for its discoverer **Pierre Fatou** (1878 – 1929), and then turn to **Lebesgue's monotone convergence theorem,** named for **Henri Léon Lebesgue (**1875 – 1941).

Remark 2.17 *Fatou's lemma does not require the given sequence* $\{f_n(x)\}$ *to converge pointwise, but provides information on the integrability of the function* $f(x) \equiv \liminf_{n\to\infty} f_n(x)$ *based on the existence of a finite limit inferior of the sequence* $\left\{\int_E f_n(x)d\mu\right\}$. *Of course if this sequence converges pointwise, then* $\liminf_{n\to\infty} f_n(x) = \lim_{n\to\infty} f_n(x)$ *and thus this result addresses the integrability of this limit function based on the same criterion.*

Fatou's lemma is sometimes stated as:

$$\int_E \left[\liminf_{n\to\infty} f_n(x)\right] d\mu \leq \liminf_{n\to\infty} \int_E f_n(x)d\mu. \tag{2.8}$$

This notation emphasizes that the result addresses the interchanging of two limiting processes, the limit inferior and the value of an integral which is defined in terms of the supremum of subordinate functions.

Proposition 2.18 (Fatou's Lemma) *If* $\{f_n(x)\}$ *is a sequence of nonnegative* μ*-measurable functions, and* $f(x) \equiv \liminf_{n\to\infty} f_n(x)$ *on a* μ*-measurable set* E, *then*

$$\int_E f(x)d\mu \leq \liminf_{n\to\infty} \int_E f_n(x)d\mu. \tag{2.9}$$

Proof. *First note that* $f(x)$ *is measurable by proposition 1.9, and nonnegative. To prove 2.9 we must show that if* $\varphi(x)$ *is a simple function with* $\varphi(x) \leq f(x)$, *then*

$$\int_E \varphi(x)d\mu \leq \liminf_{n\to\infty} \int_E f_n(x)d\mu.$$

This will then assure the result for the supremum of all such φ-integrals, and hence the result in 2.9.

To this end, first assume that for some such $\varphi(x) \leq f(x)$ that $\int_E \varphi(x)d\mu = \infty$. Then as a simple function in 1.1, there is at least one μ-measurable set $A \subset E$ with $\mu(A) = \infty$, and for which $\varphi(x) > a > 0$ for $x \in A$. Since we then have that $f(x) > a > 0$ for $x \in A$, let $B_n = \{x | f_k(x) > a$ for all $k \geq n\}$. Then each B_n is μ-measurable, and $\{B_n\}$ is an increasing sequence, $B_n \subset B_{n+1}$. Also, $A \subset \cup B_n$ since for each $x \in A$, $f(x) = \lim_{n\to\infty} f_n(x) > a$ implies that $f_k(x) > a$ for $k \geq n$ and some n. So $\mu(\cup B_n) = \infty$ and as it is an increasing sequence, $\lim \mu(B_n) = \infty$. Hence

$$\int_E f_n(x)d\mu > a\mu(B_n),$$

and it follows that $\liminf \int_E f_n(x)d\mu = \infty$.

Next, if $\varphi(x) \leq f(x)$ and $\int_E \varphi(x)d\mu < \infty$, then $\mu(A_i) < \infty$ for each of the n defining sets for φ in 1.1, and we can without loss of generality assume all $a_i > 0$. Define $A = \cup_{i\leq n} A_i$, then $A \subset E$, $\mu(A) < \infty$, and $\varphi(x) = 0$ on $E - A$. Also, for $x \in A$, $\varphi(x) \leq a \equiv \max_{i\leq n} a_i$. Given $\epsilon > 0$ define:

$$B_n = \{x \in E | f_k(x) > (1-\epsilon)\varphi(x) \text{ for all } k \geq n\}.$$

Note that equivalently:

$$B_n = \{x \in E | \inf_{k\geq n} f_k(x) > (1-\epsilon)\varphi(x)\}.$$

Then $B_n \subset B_{n+1}$ so $\{B_n\}$ is an increasing sequence of μ-measurable sets. Also $A \subset \cup B_n$ because for $x \in A$ we have by definition of limit inferior:

$$0 < \varphi(x) \leq f(x) = \sup_n \inf_{k\geq n} f_k(x),$$

so $x \in B_n$ for all $n \geq N(\epsilon)$. Thus $\{A - B_n\}$ is a decreasing sequence with limit $\emptyset$ and since $\mu(A) < \infty$, continuity from above of μ obtains that $\lim \mu(A - B_n) = 0$. For the given $\epsilon > 0$, choose n so that $\mu(A - B_k) < \epsilon$ for $k \geq n$.

Then for $k \geq n$, recalling exercise 2.10:

$$\begin{aligned}
\int_E f_k(x)d\mu &\geq \int_{B_k} f_k(x)d\mu \\
&\geq (1-\epsilon)\int_{B_k} \varphi(x)d\mu \\
&= (1-\epsilon)\int_E \varphi(x)d\mu - (1-\epsilon)\int_{E-B_k} \varphi(x)d\mu.
\end{aligned}$$

Recalling that $\varphi = 0$ *on* $E - A$, $\mu(A - B_k) < \epsilon$, *and* $\varphi(x) \le a$ *for* $x \in A$,

$$\begin{aligned}\int_E f_k(x)d\mu \ &\ge \ (1-\epsilon)\int_E \varphi(x)d\mu - \int_{A-B_k}\varphi(x)d\mu \\ &\ge \ \int_E \varphi(x)d\mu - \epsilon\left[\int_E \varphi(x)d\mu + a\right].\end{aligned}$$

Since ϵ *was arbitrary it follows that* $\int_E \varphi(x)d\mu \le \int_E f_k(x)d\mu$ *for* $k \ge n$, *and thus* $\int_E \varphi(x)d\mu \le \liminf_{k\to\infty} \int_E f_k(x)d\mu$. ■

Corollary 2.19 (Fatou's Lemma) *If* $\{f_n(x)\}$ *is a sequence of nonnegative* μ*-measurable functions, and* $f(x) \equiv \liminf_{n\to\infty} f_n(x)$ μ*-a.e on a* μ*-measurable set* E, *then 2.9 remains true if it is assumed either that* f *is* μ*-measurable or that the measure space* $(X, \sigma(X), \mu)$ *is complete.*
Proof. *Either assumption assures that* $f(x)$ *is* μ*-measurable, and otherwise the proof remains the same. As a detail, it is now true that we can only conclude that* $f(x)$ *is nonnegative* μ*-a.e, but we can replace the negative values of* $f(x)$ *by* 0 *without affecting* μ*-measurability,* μ*-a.e convergence, or the value of the integral.* ■

While Fatou's lemma provides "only" an upper bound to the value of the μ-integral of $f(x)$ over E vis-a-vis the μ-integrals of the given sequence of nonnegative functions $\{f_n(x)\}$, this result is the key ingredient for a short proof of the final result of this section on such function sequences. **Lebesgue's monotone convergence theorem,** named for **Henri Léon Lebesgue (**1875 – 1941), replaces Fatou's inequality with equality under the additional constraint that the given sequence is increasing. As was the case for Fatou's lemma, the next result does not require that the functions in the sequence be integrable, which is to say, have finite integrals.

Lebesgue's result will be generalized below to be applicable to an increasing sequence of measurable, but not necessarily nonnegative, functions. It is then known as **Beppo Levi's theorem,** named for **Beppo Levi** (1875 – 1961). For this latter result it must then be assumed that the functions in the sequence are in fact integrable and that the associated integral values have a finite supremum.

Remark 2.20 *As was the case for Fatou's lemma, this next result is sometimes expressed as*

$$\int_E \lim_{n\to\infty} f_n(x)d\mu = \lim_{n\to\infty}\int_E f_n(x)d\mu, \tag{2.10}$$

to emphasize the interchanging of two limiting processes.

Proposition 2.21 (Lebesgue's Monotone Convergence theorem) *If $\{f_n(x)\}$ is an increasing sequence of nonnegative μ-measurable functions which converge on a μ-measurable set E to a function $f(x)$, then:*

$$\int_E f(x)d\mu = \lim_{n\to\infty} \int_E f_n(x)d\mu. \tag{2.11}$$

Proof. *As a limit of nonnegative μ-measurable functions, $f(x)$ is also nonnegative and μ-measurable. From Fatou's lemma we conclude that*

$$\int_E f(x)d\mu \leq \liminf_{n\to\infty} \int_E f_n(x)d\mu.$$

But the assumption that $\{f_n(x)\}$ is an increasing sequence implies that for all n,

$$\int_E f_n(x)d\mu \leq \int_E f(x)d\mu,$$

and hence

$$\limsup_{n\to\infty} \int_E f_n(x)d\mu \leq \int_E f(x)d\mu.$$

As the limit superior cannot be smaller than the limit inferior, the above two inequalities imply that these limits are equal, and 2.11 follows. ■

Remark 2.22

1. *As noted above it is not assumed that the functions in the given sequence are integrable, nor is it concluded that f is integrable. Since $\{f_n(x)\}$ is an increasing sequence, so too is $\left\{\int_E f_n(x)d\mu\right\}$ in the sense that $\int_E f_n(x)d\mu \leq \int_E f_{n+1}(x)d\mu$ when both are finite, while if $\int_E f_n(x)d\mu = \infty$ then $\int_E f_m(x)d\mu = \infty$ for $m > n$. Thus the proof confirms that f is integrable if and only if $\int_E f_n(x)d\mu \leq K < \infty$ for all n.*

2. *Perhaps surprisingly, Lebesgue's monotone convergence theorem does not apply to decreasing function sequences. A simple example from Lebesgue integration noted in remark 2.47 of book 3 is to define $f_n(x) = \chi_{[n,\infty)}(x)$ and note that $f_n(x) \to 0$ pointwise.*

Corollary 2.23 (Lebesgue's Monotone Convergence theorem) *If $\{f_n(x)\}$ is an increasing sequence of nonnegative μ-measurable functions which converge almost everywhere on a μ-measurable set E to a function $f(x)$, then 2.11 remains true if it is assumed either that f is μ-measurable or that the measure space $(X, \sigma(X), \mu)$ is complete.*

Proof. *As for the corollary to Fatou's lemma above, either assumption assures the μ-measurability of $f(x)$, and if $f(x)$ is not nonnegative on the exceptional set of μ-measure 0, then it can be redefined to be 0 on this set without affecting μ-measurability, μ-a.e convergence, or the value of the integral.* ■

Remark 2.24 (Evaluating Integrals of Nonnegative Functions) *While the definition of the integral $\int_E f(x)d\mu$ in 2.6 potentially contemplates uncountably many simple functions in the supremum calculation, Lebesgue's monotone convergence theorem gives a more practical and useful way to evaluate the μ-integral of a nonnegative function. Specifically, if $\{f_n(x)\}$ is* ***any*** *sequence of* ***increasing μ-measurable simple functions*** *defined on a μ-measurable set E with $f_n(x) \to f(x)$ for all $x \in E$, then:*

$$\int_E f(x)d\mu = \lim_{n\to\infty} \int_E f_n(x)d\mu.$$

In theory, the μ-integral can be evaluated using any such increasing sequence of functions, but by using simple functions the integrals of the function sequence are simply evaluated by 2.1.

In the case where the measure space $(X, \sigma(X),\mu)$ is complete, the same conclusion follows if only $f_n(x) \to f(x)$ for μ-almost all $x \in E$.

That this result is useful is based on proposition 1.18 that states that for any measurable set E and nonnegative μ-measurable function $f(x)$, there is an increasing sequence of simple functions $\{f_n(x)\}$, so that $f_n(x) \to f(x)$ for all $x \in E$.

Example 2.25 *If $(X, \sigma(X), \mu) = (\mathbb{R}, \mathcal{M}_{\mu_F}(\mathbb{R}), \mu_F)$, the Borel measure space on $\mathbb{R}$ defined with $F(x) = x^2$, we evaluate $\int_E f(x)d\mu_F$ for $E = [0, 2]$ and $f(x) = x^2$. Because E is an interval it is natural to utilize a sequence of simple functions defined as step functions. If we can then show that $f(x)$ is the pointwise limit of an increasing sequence of such functions, the value of this integral can be obtained by Lebesgue's monotone convergence theorem.*

Given n *define* $\{A_i\}_{i=1}^n$ *by* $A_i = [2(i-1)/n, 2i/n]$, *and the simple function* $\varphi_n(x) \leq f(x)$ *by:*

$$\varphi_n(x) = \sum\nolimits_{i=1}^n [2(i-1)/n]^2 \chi_{A_i}(x).$$

To ensure that $\{\varphi_n(x)\}$ *is an increasing sequence, we use interval bisection, choosing* $n = 2m$ *for positive integers* m. *Then recalling that* $\mu_F[(a,b]] = F(b) - F(a)$:

$$\begin{aligned}\int_E \chi_{A_i}(x)d\mu_F &= (2i/n)^2 - (2(i-1)/n)^2 \\ &= 4[2i-1]/n^2,\end{aligned}$$

and by 2.1,

$$\begin{aligned}\int_E \varphi_n(x)d\mu_F &= 16\sum\nolimits_{i=1}^n (i-1)^2(2i-1)/n^4 \\ &= 16\sum\nolimits_{i=1}^{n-1} \left(2i^3 + i^2\right)/n^4.\end{aligned}$$

Since $\sum\nolimits_{i=1}^{n-1} i^2 = O(n^3)$, *this term can be ignored, and hence*

$$\int_E \varphi_n(x)d\mu_F = 32\sum\nolimits_{i=1}^{n-1} i^3/n^4 + O(1/n).$$

Now $\sum\nolimits_{i=1}^n i^3 = [n(n+1)]^2/4$, *and thus,*

$$\int_0^2 x^2 d\mu_F = \lim_{n\to\infty}\int_E \varphi_n(x)d\mu_F = 8.$$

With the help of Lebesgue's monotone convergence theorem, we now establish additional properties of the μ-integral of nonnegative measurable functions in the following proposition. These properties will be generalized below.

Proposition 2.26 *If* $f(x)$ *and* $g(x)$ *are nonnegative* μ*-measurable functions defined on a* μ*-measurable set* E, *then:*

1. $\int_E f(x)d\mu = 0$ *if and only if* $f(x) = 0$ μ*-a.e. on* E.

2. *For any* $a, b > 0$,

$$\int_E [af(x) + bg(x)]d\mu = a\int_E f(x)d\mu + b\int_E g(x)d\mu.$$

3. *If $f(x) = g(x)$ μ-a.e., then*

$$\int_E f(x)d\mu = \int_E g(x)d\mu.$$

4. *If $f(x) \leq g(x)$ μ-a.e., then*

$$\int_E f(x)d\mu \leq \int_E g(x)d\mu.$$

5. *If $E' \subset E$ is μ-measurable, then*

$$\int_{E'} f(x)d\mu \leq \int_E f(x)d\mu.$$

6. *If $E = A \cup B$, a union of disjoint μ-measurable sets, then*

$$\int_E f(x)d\mu = \int_A f(x)d\mu + \int_B f(x)d\mu.$$

Proof. *For 1, assume that $\int_E f(x)d\mu = 0$, since the reverse implication is apparent by 2.6 and 2.1. If $A_n = \{x \in E | f(x) \geq 1/n\}$, then $f(x) \geq \chi_{A_n}(x)/n$ obtains from 2.6 and 2.1 that $\mu(A_n) \leq n\int_E f(x)d\mu = 0$. But $\{x \in E | f(x) > 0\} = \cup A_n$, and hence this set has μ-measure 0 by countable subadditivity of μ.*

Linearity of the μ-integral in 2 is proved by recalling that this is true for simple functions by proposition 2.6, and if $\{\varphi_n(x)\}$ is an increasing sequence converging to $f(x)$, and $\{\psi_n(x)\}$ is increasing and converging to $g(x)$, then $\{a\varphi_n(x) + b\psi_n(x)\}$ is increasing and converging to $af(x) + bg(x)$. So by Lebesgue's monotone convergence theorem:

$$\begin{aligned}
\int_E [af(x) + bg(x)]d\mu &= \lim \int_E [a\varphi_n(x) + b\psi_n(x)]d\mu \\
&= \lim \left[a\int_E \varphi_n(x)d\mu + b\int_E \psi_n(x)d\mu \right] \\
&= a\int_E f(x)d\mu + b\int_E g(x)d\mu.
\end{aligned}$$

Part 3 states that $f(x) - g(x) = 0$ μ-a.e. on E, and the result follows from 1 and 2, and the same argument applies to $f(x) - g(x) \geq 0$ μ-a.e for part 4. Part 5 follows from 4 and definition 2.4, since $f(x)\chi_{E'}(x) \leq f(x)$ on E, while 6 follows from 2 since $f(x) = f(x)\chi_A(x) + f(x)\chi_B(x)$ for $x \in E$. ∎

Part 4 of this proposition provides an interesting corollary which in essence states that every nonnegative μ-integrable function can be used as a **test function** to identify other μ-integrable functions.

Corollary 2.27 *If $g(x)$ is a nonnegative μ-measurable function which is μ-integrable on a μ-measurable set E, then for any nonnegative μ-measurable function $f(x)$ with $f(x) \leq g(x)$, it follows that $f(x)$ is also μ-integrable.*

We record next two important corollaries to the Lebesgue monotone convergence theorem. The first applies to the integral of a function series, providing a condition which allows the reversal of the two limiting processes: summation and integration. The second allows the decomposition of an integral into a countable number of disjoint domains. The proofs are identical with the Lebesgue integral case in corollaries 2.54 and 2.55 of book 3.

Corollary 2.28 *If $\{f_n(x)\}$ is a sequence of nonnegative μ-measurable functions, and $f(x) = \sum_{n=1}^{\infty} f_n(x)$ on μ-measurable E, then*

$$\int_E f(x)d\mu = \sum_{n=1}^{\infty} \int_E f_n(x)d\mu. \tag{2.12}$$

Corollary 2.29 *If $f(x)$ is a nonnegative μ-integrable function on μ-measurable set E, and $E = \cup E_j$, a union of disjoint μ-measurable sets, then*

$$\int_E f(x)d\mu = \sum_{k=1}^{\infty} \int_{E_j} f(x)d\mu. \tag{2.13}$$

2.2.1 Product Space Measures Revisited

It may be recalled that in chapter 7 of book 1, it was a mighty challenge to prove that the product set function μ_0, defined on measurable rectangles in definition 7.1, could be extended to a measure on the semi-algebra $\mathcal{A}'$ of such rectangles, or the associated algebra $\mathcal{A}$ of finite disjoint unions of $\mathcal{A}'$-rectangles. Indeed, with the tools then at hand we could only prove that this product set function was finitely additive on the semi-algebra $\mathcal{A}'$, and needed to expand $\mathcal{A}'$ to the algebra $\mathcal{A}$ to address countable additivity. And even then it was necessary to assume that the component measure spaces were σ-finite, an extraneous assumption, but one needed to justify the continuity from above argument that was utilized. It was noted in that development that the tools of the current book, and in particular Lebesgue's monotone convergence theorem, would allow a relatively easy proof that μ_0 is in fact both finitely and countable additive on the semi-algebra $\mathcal{A}'$, and indeed even smaller semi-algebras.

For completeness, we recall the notation and definitions.

Definition 2.30 *Given measure spaces* $\{(X_i, \sigma(X_i), \mu_i) | i = 1, ..., n\}$*, the product space* $X = \prod_{i=1}^n X_i$ *is defined:*

$$X = \{(x_1, x_2, ..., x_n) | x_i \in X_i\}.$$

A measurable rectangle A *in* X *is a set:*

$$A = \prod_{i=1}^n A_i = \{x \in X | x_i \in A_i\},$$

where with $A_i \in \sigma(X_i)$*. We denote by* $\mathcal{A}'$ *the collection of measurable rectangles in* X*.*

On a measurable rectangle $\prod_{i=1}^n A_i$*, the product set function* μ_0 *is defined by:*

$$\mu_0(A) = \prod_{i=1}^n \mu_i(A_i).$$

In proposition 7.2 of book 1 was proved that $\mathcal{A}'$ is a semi-algebra, and in proposition 7.15 was proved that μ_0 is finitely additive on $\mathcal{A}'$. We now prove countable additivity with the aid of Lebesgue's monotone convergence theorem, and without the assumption of finite additivity.

Proposition 2.31 *The product set function* μ_0 *is countably additive on* $\mathcal{A}'$*. That is, if* $\{B_j\}_{j=1}^\infty \subset \mathcal{A}'$ *is a disjoint collection of measurable rectangles with* $\cup_j B_j \in \mathcal{A}'$*, then*

$$\mu_0\left(\bigcup_{j=1}^\infty B_j\right) = \sum_{j=1}^\infty \mu_0(B_j). \tag{2.14}$$

Proof. *Let* $B_j = \prod_{i=1}^n A_{ji}$ *where* $\{A_{ji}\}_{j=1}^\infty \subset \sigma(X_i)$ *for each* i*, and assume that* $\bigcup_{j=1}^\infty B_j = \prod_{i=1}^n A_i$ *with* $A_i \in \sigma(X_i)$*. First, note that for* $x = (x_1, x_2, ..., x_n) \in X$ *with* $x_i \in X_i$ *:*

$$\prod_{i=1}^n \chi_{A_i}(x_i) = \sum_{j=1}^\infty \prod_{i=1}^n \chi_{A_{ji}}(x_i).$$

As above $\chi_A(x_i)$ *denotes the characteristic function of* A *for* $A = A_i$ *or* $A = A_{ji}$*, and is defined as* $\chi_A(x_i) = 1$ *if* $x_i \in A$ *and* 0 *otherwise. To prove this identity, note that the left side product equals* 1 *if and only if* $x \in \prod_{i=1}^n A_i = \bigcup_{j=1}^\infty \prod_{i=1}^n A_{ji}$*, and hence* $x \in \prod_{i=1}^n A_{ji}$ *for at least one* j*. However, the disjointedness of* $\{B_j\}_{j=1}^\infty$ *assures that such* j *is unique.*

Next, fix $x_1, x_2, ..., x_{n-1}$, and consider this identity as a functional identity in x_n. Because $\chi_{A_n}(x_n)$ and each $\chi_{A_{jn}}(x_n)$ are μ_n-measurable functions, we can take μ_n-integrals of this identity to produce:

$$\int \prod_{i=1}^{n-1} \chi_{A_i}(x_i)\chi_{A_n}(x_n)d\mu_n = \int \left[\sum_{j=1}^{\infty} \left(\prod_{i=1}^{n-1} \chi_{A_{ji}}(x_i)\right) \chi_{A_{jn}}(x_n)\right] d\mu_n.$$

Now applying 2.12, noting that $\int \chi_A(x_n)d\mu_n = \mu_n(A)$ for $A = A_n$ or A_{nj}, and that $\prod_{i=1}^{n-1} \chi_{A_{ji}}(x_i)$ is a constant:

$$\begin{aligned}\int \left[\sum_{j=1}^{\infty} \left(\prod_{i=1}^{n-1} \chi_{A_{ji}}(x_i)\right) \chi_{A_{jn}}(x_n)\right] d\mu_n &= \sum_{j=1}^{\infty} \prod_{i=1}^{n-1} \chi_{A_{ji}}(x_i) \int \chi_{A_{jn}}(x_n)d\mu_n \\ &= \sum_{j=1}^{\infty} \prod_{i=1}^{n-1} \chi_{A_{ji}}(x_i)\mu_n(A_{jn}).\end{aligned}$$

Similarly:

$$\int \prod_{i=1}^{n-1} \chi_{A_i}(x_i)\chi_{A_n}(x_n)d\mu_n = \prod_{i=1}^{n-1} \chi_{A_i}(x_i) \int \chi_{A_n}(x_n)d\mu_n,$$

and combining:

$$\prod_{i=1}^{n-1} \chi_{A_i}(x_i)\mu_n(A_n) = \sum_{j=1}^{\infty} \prod_{i=1}^{n-1} \chi_{A_{ji}}(x_i)\mu_n(A_{jn}).$$

This identity is true for all fixed $x_1, x_2, ..., x_{n-1}$.

We now fix $x_1, x_2, ..., x_{n-2}$ and repeat the derivation to conclude that

$$\prod_{i=1}^{n-2} \chi_{A_i}(x_i)\mu_{n-1}(A_{n-1})\mu_n(A_n) = \sum_{j=1}^{\infty} \prod_{i=1}^{n-2} \chi_{A_{ji}}(x_i)\mu_{n-1}(A_{j,n-1})\mu_n(A_{jn}).$$

Continuing in this way obtains:

$$\prod_{i=1}^{n} \mu_i(A_i) = \sum_{j=1}^{\infty} \prod_{i=1}^{n} \mu_i(A_{ji}),$$

which is 2.14 by the definition of μ_0. ■

Corollary 2.32 *The product measure μ_0 is finitely additive on $\mathcal{A}'$.*
Proof. *Simply choose $B_j = \emptyset$ for all but finitely many j in the above proposition, and note that by definition, $\mu_0(\emptyset) = 0$.* ■

Recall corollary 7.3 of book 1 which defined $\mathcal{A}'(\mathcal{A}_i')$, respectively $\mathcal{A}'(\mathcal{A}_i)$:

Definition 2.33 *$\mathcal{A}'(\ )$ denotes the collection of measurable rectangles in X, defined by $A = \Pi_{i=1}^{n} A_i$ with:*

1. $\mathcal{A}'(\mathcal{A}'_i) : A_i \in \mathcal{A}'_i$, *where* $\mathcal{A}'_i \subset \sigma(X_i)$ *is a semi-algebra, or,*

2. $\mathcal{A}'(\mathcal{A}_i) : A_i \in \mathcal{A}_i$, *where* $\mathcal{A}_i \subset \sigma(X_i)$ *is an algebra.*

It was then proved in this corollary that each collection of rectangles is again a semi-algebra, and by definition with transparent notation:

$$\mathcal{A}'(\mathcal{A}'_i) \subsetneqq \mathcal{A}'(\mathcal{A}_i) \subsetneqq \mathcal{A}'(\sigma(X_i)).$$

Using the same proof as above we have:

Proposition 2.34 *The product measure* μ_0 *is countably additive on* $\mathcal{A}'(\mathcal{A}'_i)$, *respectively* $\mathcal{A}'(\mathcal{A}_i)$, *and hence also finitely additive.*

Remark 2.35 *As noted in section 7.3.2 of book 1, the assumption of proposition 7.18 there, that the given component measure spaces were* σ*-finite, was only needed due the approach taken to prove countable additivity on* $\mathcal{A}$*, the algebra of all finite unions of disjoint elements from* $\mathcal{A}'(\sigma(X_i))$*. Specifically, this proof used continuity from above of measures (see section 2.7.2 of book 1). As can be seen from the above proof, countable additivity of* μ_0 *can be proved on the respective* $\mathcal{A}'$ *collections without this* σ*-finiteness assumption, and then extended to countably additive measures* μ_A *on the respective algebras.*

That said, the σ*-finite assumption is still needed to assure uniqueness of the extension from a measure* μ_A *defined on the algebra* $\mathcal{A}$*, to a measure* μ *defined on* $\sigma(\mathcal{A})$*, the smallest sigma algebra that contains* $\mathcal{A}$*. This is proposition 6.14 of book 1.*

2.3 Integrating General Measurable Functions

The final step in the μ-integration sequence is to extend the definition from nonnegative to general μ-measurable functions, and this is relatively easy to do. First we recall the definition and formulas from book 3.

Definition 2.36 *Given* $f(x)$*, the* ***positive part of*** $f(x)$*, denoted* $f^+(x)$*, is defined by:*

$$f^+(x) = \max\{f(x), 0\}, \tag{2.15}$$

and the ***negative part of*** $f(x)$*, denoted* $f^-(x)$*, is defined by:*

$$f^-(x) = \max\{-f(x), 0\}. \tag{2.16}$$

Both the positive and negative part of a function are nonnegative functions, so we can apply the previous section's results to either part. The original function and its absolute value can then be recovered from these component functions:

$$f(x) = f^+(x) - f^-(x), \qquad |f(x)| = f^+(x) + f^-(x), \tag{2.17}$$

and these will provide a basis for the definitions of $\int_E f(x)dx$ and $\int_E |f(x)|\,dx$.

If $f(x)$ is μ-measurable, then by proposition 1.9 so too is $f^+(x)$ and $f^-(x)$, and hence so too $|f(x)|$. For example, if $A \equiv \{x | f(x) > 0\}$ then:

$$f^+(x) = f(x)\chi_A(x),$$

and $\chi_A(x)$ is μ-measurable because $A \in \sigma(X)$. Recalling the Lebesgue development of section 2.5 of book 3, the following will be no surprise.

Definition 2.37 *A μ-measurable, extended real valued function $f(x)$ is said to be μ-**integrable** over a μ-measurable set E if both $f^+(x)$ and $f^-(x)$ are integrable over E, and in this case we define:*

$$\int_E f(x)d\mu = \int_E f^+(x)d\mu - \int_E f^-(x)d\mu. \tag{2.18}$$

In this case $|f(x)|$ is also μ-integrable over E and we define:

$$\int_E |f(x)|\,d\mu = \int_E f^+(x)d\mu + \int_E f^-(x)d\mu. \tag{2.19}$$

*If one of the functions $f^+(x)$ and $f^-(x)$ is μ-integrable and one is not, then $f(x)$ is said to be **not** μ-**integrable** although it is then common to define $\int\limits_E f(x)d\mu = \infty$ or $\int\limits_E f(x)d\mu = -\infty$ as appropriate. If neither function is μ-integrable, then $f(x)$ is said to be **not** μ-**integrable,** and $\int\limits_E f(x)dx$, which formally equals the expression $\infty - \infty$, is undefined. But in this case we can again say $\int\limits_E |f(x)|\,d\mu = \infty$.*

Identical to the case for nonnegative Lebesgue integrable functions of example 2.42 of book 3:

Proposition 2.38 *If $f(x)$ is μ-integrable, then*

$$\mu(\{x|f(x) = \pm\infty\}) = 0.$$

Proof. *If $E = \{x|f(x) = \infty\}$, $E' = \{x|f(x) = -\infty\}$ and $\mu(E \cup E') > 0$, then at least one of E or E' has positive μ-measure and hence at least one of $f^+(x)$ and $f^-(x)$ would not be integrable, and thus $f(x)$ would not be integrable.* ■

But also as in the Lebesgue case, $\mu(E) = \mu(E') = 0$ does not assure μ-integrability as the following examples demonstrates.

Example 2.39 *Recall example 2.25 above with $(X, \sigma(X), \mu) = (\mathbb{R}, \mathcal{M}_{\mu_F}(\mathbb{R}), \mu_F)$, the Borel measure space on $\mathbb{R}$ defined with $F(x) = x^2$, but now we attempt to evaluate $\int_E f(x)d\mu_F$ for $E = [0, \infty)$ and $f(x) = x^2$. Because E is an interval we again utilize a sequence of simple functions defined to be step functions. Since $f(x)$ is nonnegative, recall that if it can be shown that $f(x)$ is the pointwise limit of an increasing sequence of simple functions, the integral of $f(x)$ can be evaluated by Lebesgue's monotone convergence theorem.*

Given n define a partition of $E_N \equiv [0, N]$ by $\{A_i\}_{i=1}^n$ with $A_i = [N(i-1)/n, Ni/n]$, and a simple function with $\varphi_n(x) \leq f(x)$ defined on E_N by:

$$\varphi_n(x) = \sum\nolimits_{i=1}^{n} [N(i-1)/n]^2 \chi_{A_i}(x).$$

To ensure that $\{\varphi_n(x)\}$ is an increasing sequence we again use interval bisection, meaning $n = 2m$ for positive integers m. Then

$$\int_{E_N} \chi_{A_i}(x)d\mu_F = N^2 [2i-1]/n^2,$$

and then by 2.1 have that

$$\int_{E_N} \varphi_n(x)d\mu_F = N^4 \sum\nolimits_{i=1}^{n-1} \left(2i^3 + i^2\right)/n^4,$$

and by proposition 2.26:

$$\int_0^\infty x^2 d\mu_F \geq \int_{E_N} \varphi_n(x)d\mu_F = N^4/2.$$

Hence, x^2 is not μ_F-integrable on E.

We summarize the essential properties of the μ-integral in the following.

Proposition 2.40 *If $f(x)$ and $g(x)$ are μ-integrable functions defined on a μ-measurable set E, then:*

1. *For any a,*
$$\int_E af(x)d\mu = a\int_E f(x)d\mu.$$

2. *Arbitrarily defining $f(x)+g(x)$ on the set of μ-measure 0 for which this sum is $\infty-\infty$ or $-\infty+\infty$:*
$$\int_E [f(x)+g(x)]d\mu = \int_E f(x)dx + \int_E g(x)d\mu.$$

3. *If $f(x)=g(x)$ a.e., then*
$$\int_E f(x)d\mu = \int_E g(x)d\mu.$$

4. *If $f(x)\leq g(x)$ a.e., then*
$$\int_E f(x)d\mu \leq \int_E g(x)d\mu.$$

5. *If measurable $E'\subset E$, then*
$$\int_E f(x)d\mu \leq \int_{E'} f(x)d\mu.$$

6. *If $E=\bigcup_i E_i$, a union of disjoint μ-measurable sets, then*
$$\int_E f(x)d\mu = \sum\nolimits_i \int_{E_i} f(x)d\mu. \tag{2.20}$$

7. *The **triangle inequality:***
$$\left|\int_E f(x)d\mu\right| \leq \int_E |f(x)|\, d\mu. \tag{2.21}$$

Proof. *The proof is left as an exercise, applying the results for nonnegative functions.* ■

2.4 Integration to the Limit

For nonnegative measurable functions, **Fatou's lemma** and **Lebesgue's Monotone Convergence theorem** of the above section provide important results on the relationship between the limit of the integrals of a function sequence and the integral of the pointwise limit function. In this section, additional results relating to "integration to the limit" are developed. The first result is a generalization of Lebesgue's monotone convergence theorem known as **Beppo Levi's theorem,** and named for **Beppo Levi** (1875 – 1961). In contrast to the earlier result, for this generalization it must be assumed that the functions in the sequence are in fact integrable and that the associated integral values have finite supremum. But this is then sufficient to guarantee that the limiting function is integrable, and to specify the value of its integral.

Remark 2.41 *As was the case for earlier results, this result and those below are sometimes expressed as*

$$\int_E \lim_{n\to\infty} f_n(x)d\mu = \lim_{n\to\infty} \int_E f_n(x)d\mu,$$

to emphasize the interchanging of two limiting processes.

Proposition 2.42 (Beppo Levi's theorem) *Let $\{f_n(x)\}$ is an increasing sequence of μ-measurable functions which converge pointwise on a μ-measurable set E to a function $f(x)$. Assume that $\int_E f_n(x)d\mu \leq K < \infty$ for all n. Then f is integrable, and:*

$$\int_E f(x)d\mu = \lim_{n\to\infty} \int_E f_n(x)d\mu. \tag{2.22}$$

Proof. *Consider the nonnegative, increasing function sequence, $\{f_n(x) - f_1(x)\}_{n\geq 2}$ with pointwise limit $f(x) - f_1(x)$. By 2.11 of Lebesgue's monotone convergence theorem and the assumption of integrability of f_n :*

$$\int_E [f(x) - f_1(x)]\, d\mu = \lim_{n\to\infty} \int_E [f_n(x) - f_1(x)]\, d\mu = \lim_{n\to\infty} \int_E f_n(x)d\mu - \int_E f_1(x)d\mu.$$

The sequence of integrals on the right is increasing and bounded by K, and so $f(x) - f_1(x)$ is integrable. The integrability of f and the identity in 2.22 now follow by addition, since $\int_E f_1(x)d\mu$ is finite. ■

The next result is a cornerstone limit theorem for function sequences, **Lebesgue's Dominated Convergence theorem** named for **Henri Léon Lebesgue** (1875 – 1941), which parallels the result of Lebesgue integration theory. After noting a few corollaries to this result, one of which is the **Bounded Convergence theorem,** a new integration to the limit result is developed that reflects the notion of uniform integrability introduced in definition 3.63 of book 4. .

Proposition 2.43 (Lebesgue's Dominated Convergence theorem) *Let $\{f_n(x)\}$ be a pointwise convergent sequence of μ-measurable functions on a μ-measurable set E, with $f(x) \equiv \lim_{n\to\infty} f_n(x)$, and assume that there is a μ-integrable function $g(x)$ so that:*

$$|f_n(x)| \leq g(x), \text{ all } n.$$

Then $f(x)$ is integrable on E and

$$\int_E f(x)d\mu = \lim_{n\to\infty} \int_E f_n(x)d\mu. \tag{2.23}$$

Further,

$$\int_E |f_n(x) - f(x)|\, d\mu \to 0, \text{ as } n \to \infty. \tag{2.24}$$

Proof. *Note that the limit function $f(x)$ is μ-measurable by proposition 1.9, and since $|f_n(x)| \leq g(x)$ implies that $|f(x)| \leq g(x)$, all $f_n(x)$ and the limit function $f(x)$ are μ-integrable. Now if 2.10 is demonstrated, the triangle inequality in 2.21 obtains:*

$$\left|\int_E [f_n(x) - f(x)]d\mu\right| \leq \int_E |f_n(x) - f(x)|\, d\mu,$$

and thus

$$\int_E [f_n(x) - f(x)]d\mu \to 0.$$

Since $f(x)$ is μ-integrable, 2.23 is proved by addition of finite $\int_E f(x)d\mu$.

For 2.10, it follows from Fatou's lemma applied to nonnegative $g(x) + f_n(x)$:

$$\int_E [g(x) + f(x)]\, d\mu \leq \liminf_{n\to\infty} \int_E [g(x) + f_n(x)]\, d\mu,$$

noting that $f(x) = \liminf_{n\to\infty} f_n(x)$ *by definition. Subtracting finite* $\int_E g(x)d\mu$*:*

$$\int_E f(x)d\mu \leq \liminf_{n\to\infty} \int_E f_n(x)d\mu.$$

Next, apply Fatou's lemma to nonnegative $g(x) - f_n(x)$ *:*

$$\int_E [g(x) - f(x)]\, d\mu \leq \liminf_{n\to\infty} \int_E [g(x) - f_n(x)]\, d\mu,$$

and subtracting finite $\int_E g(x)d\mu$ *:*

$$\int_E f(x)d\mu \geq -\liminf_{n\to\infty} \int_E [-f_n(x)]\, d\mu = \limsup_{n\to\infty} \int_E f_n(x)d\mu.$$

Combining:

$$\limsup_{n\to\infty} \int_E f_n(x)d\mu \leq \int_E f(x)d\mu \leq \liminf_{n\to\infty} \int_E f_n(x)d\mu,$$

and these inequalities yield 2.23 since they imply the limits superior and inferior agree. ■

Remark 2.44 *For the following results which involve* μ*-a.e convergence,* $f_n(x) \to f(x)$*, there are two approaches to an affirmative conclusion:*

1. *Assume that* $(X, \sigma(X), \mu)$ *is complete, and then the measurability of* $f(x)$ *is part of the conclusion, along with its integrability and the limiting result in 2.23;*

2. *For general* $(X, \sigma(X), \mu)$ *the measurability of* $f(x)$ *must be assumed, and then the conclusion is the integrability of* $f(x)$ *and the limit result in 2.23.*

For simplicity, the following results are stated consistent with approach 1.

Corollary 2.45 (Lebesgue's Dominated Convergence theorem) *Let* $\{f_n(x)\}$ *be a sequence of* μ*-measurable functions on a complete measure space* $(X, \sigma(X), \mu)$ *with* $f(x) \equiv \lim_{n\to\infty} f_n(x)$ μ*-a.e on a* μ*-measurable set* E*. If for all* n*,* $|f_n(x)| \leq g(x)$ μ*-a.e for* μ*-integrable* $g(x)$*, then the conclusions of Lebesgue's dominated convergence theorem remain true.*

Proof. *As for the corollaries to the Fatou's lemma and Lebesgue's monotone convergence theorem above, completeness assures the μ-measurability of $f(x)$. Then $f(x)$ can be arbitrarily redefined on the exceptional sets of μ-measure 0 on which it is not bounded by $g(x)$, nor is equal to the limit of the $f_n(x)$-series, without affecting μ-measurability, μ-a.e convergence, or the value of the integral.* ■

Proposition 2.46 (Bounded Convergence theorem) *Let $\{f_n(x)\}$ be a sequence of μ-measurable functions with $f(x) \equiv \lim_{n\to\infty} f_n(x)$ on a μ-measurable set E with $\mu[E] < \infty$. If $|f_n(x)| \leq M < \infty$ on E, then the conclusions of Lebesgue's dominated convergence theorem remain true.*

If $f(x) \equiv \lim_{n\to\infty} f_n(x)$ μ-a.e and $|f_n(x)| \leq M < \infty$ μ-a.e on E, then the conclusions of Lebesgue's dominated convergence theorem remain true if the measure space $(X, \sigma(X), \mu)$ is complete.

Proof. *Let $g(x) = M\chi_E(x)$ above.* ■

Remark 2.47 (Evaluating Integrals of Functions) *The definition of the integral $\int_E f(x)d\mu$ in 2.18, as was the case for the definition in 2.6, contemplates potentially uncountably many simple functions in the supremum calculation. Analogous to the case of nonnegative functions addressed in remark 2.24, Lebesgue's dominated convergence theorem gives a more practical and useful way to evaluate the μ-integral of a general function.*

*Specifically, if $\{f_n(x)\}$ is a sequence of μ-**measurable simple functions** defined on a μ-measurable set E with $|f_n(x)| < g(x)$ for some μ-integrable $g(x)$ and $f_n(x) \to f(x)$ for $x \in E$, then*

$$\int_E f(x)dx = \lim_{n\to\infty} \int_E f_n(x)dx.$$

In theory, the Lebesgue integral can be evaluated using any such sequence of functions, but by using simple functions the integrals in the sequence are then easily evaluated by 2.1.

When $(X, \sigma(X),\mu)$ is complete, the same conclusion follows if only $f_n(x) \to f(x)$ for μ-almost all $x \in E$.

This result is useful because of remark 1.19 that shows that proposition 1.18 can be applied to derive that given any μ-measurable function $f(x)$, there is a sequence of simple functions, $\{\varphi_n(x)\}$, so that $\varphi_n(x) \to f(x)$ for all $x \in E$, and $\{|\varphi_n(x)|\}$ is an increasing sequence with $|\varphi_n(x)| \to |f(x)|$. By Lebesgue's monotone convergence theorem,

$$\int_E |f(x)|\, dx = \lim_{n\to\infty} \int_E |\varphi_n(x)|\, dx,$$

and hence if $\int_E |\varphi_n(x)|\,dx$ *is bounded in* n *then this limit exists and* $f(x)$ *is* μ*-integrable. If* $\int_E |\varphi_n(x)|\,dx$ *is unbounded in* n *then* $f(x)$ *is not* μ*-integrable.*

To then evaluate $\int_E f(x)dx$ *when finite also requires that* $|\varphi_n(x)| \leq g(x)$ *for some* μ*-integrable* $g(x)$.

We next record important corollaries to the Lebesgue dominated convergence theorem, generalizing the earlier results for nonnegative functions. We leave the proofs as exercises. We state the first result in the general case of almost everywhere convergence, thereby necessitating the assumption of completeness.

Corollary 2.48 *Let* $\{h_j(x)\}$ *be a sequence of* μ*-measurable functions on a* μ*-measurable set* E *in a complete measure space* $(X, \sigma(X), \mu)$, *and assume that* $f(x) \equiv \sum_{j=1}^{\infty} h_j(x)$ *converges except on a set of* μ*-measure* 0. *If there is a* μ*-measurable function* $g(x)$, *integrable on* E, *so that for all* n :

$$\left|\sum\nolimits_{j=1}^{n} h_j(x)\right| \leq g(x), \ \ \mu\text{-a.e.},$$

then $f(x)$ *is integrable on* E *and*

$$\int_E f(x)d\mu = \sum\nolimits_{j=1}^{\infty} \int_E h_j(x)d\mu. \tag{2.25}$$

Further as $n \to \infty$,

$$\int_E \left|\sum\nolimits_{j=n}^{\infty} h_j(x)\right| d\mu \to 0.$$

Corollary 2.49 *If* $f(x)$ *is a* μ*-integrable function on* μ*-measurable set* E *and* $E = \bigcup_j E_j$, *a union of disjoint* μ*-measurable sets, then:*

$$\int_E f(x)d\mu = \sum\nolimits_{j=1}^{\infty} \int_{E_j} f(x)d\mu. \tag{2.26}$$

2.4.1 Uniform Integrability Convergence theorem

The final limit theorem is "new" in the sense that it does not generalize a result of Lebesgue integration from book 3. There is a simple reason for this, and that is that this result is applicable in finite measure spaces such as probability spaces, and thus the Lebesgue measure space does not qualify. This new result is sometimes useful when the function sequence converges pointwise or pointwise μ-a.e., but does not satisfy one of the conditions addressed in the propositions noted above and summarized below.

Each of the prior results is applicable to a sequence of measurable functions $\{f_n\}_{n=1}^{\infty}$ when $f_n \to f$ pointwise, as well as for μ-a.e. convergence with the additional assumption either that f is measurable or that $(X, \sigma(X), \mu)$ is complete. The additional requirements are then:

1. If $\{f_n(x)\}$ is nonnegative and monotonically increasing, this allows the application of **Lebesgue's monotone convergence theorem**;

2. If $\{f_n(x)\}$ is bounded by a constant and the measure space is finite, $\mu[X] < \infty$, this allows the application of the **bounded convergence theorem**;

3. If $\{f_n(x)\}$ is absolutely bounded by an integrable function $g(x)$, this allows the application of **Lebesgue's dominated convergence theorem**.

The next result uses the notion of **uniform integrability** introduced in definition 3.63 of book 4, but here adapted to the current notation:

Definition 2.50 *Given a finite measure space $(X, \sigma(X), \mu)$, a sequence of measurable functions $\{f_n\}_{n=1}^{\infty}$ is said to be **uniformly integrable** if:*

$$\lim_{N\to\infty} \sup_n \int_{|f_n|\geq N} |f_n(x)|\, d\mu = 0. \tag{2.27}$$

Remark 2.51 *As noted above the following result requires that $(X, \sigma(X), \mu)$ be a **finite measure space,** $\mu(X) < \infty$. Then the assumption of uniform integrability assures that the functions in the sequence are in fact integrable, and with integrals that are uniformly bounded. Indeed, choosing N so that $\sup_n \int_{|f_n|\geq N} |f_n(x)|\, d\mu \leq 1$ say:*

$$\int_X |f_n(x)|\, d\mu \leq \int_{|f_n|\leq N} |f_n(x)|\, d\mu + \sup_n \int_{|f_m|\geq N} |f_n(x)| \leq N\mu(X) + 1.$$

If $f_n \to f$ μ-a.e., this will be enough below to assure that f is also integrable, and that the integrals converge.

If X is not a finite measure space then 2.27 does not assure that $\left\{\int_X |f_n(x)|\, d\mu\right\}$ is bounded, nor that f is integrable if $f_n \to f$, μ-a.e. This follows because every uniformly bounded sequence of functions, $|f_n(x)| \leq M$ say, satisfies 2.27. For example, on $X = [1, \infty)$ as a Lebesgue measure space, let $f_n(x) = x^{-1-1/n}$, and $f(x) = x^{-1}$. Then $\{f_n\}$ are uniformly bounded and satisfy 2.27, $f_n \to f$ everywhere, yet $\int_X |f_n(x)|\, dm = n$ and f is not integrable. If $f_n(x) = 1 + 1/n$ and $f(x) = 1$, then again $\{f_n\}$ satisfy 2.27 and $f_n \to f$ everywhere, but none of these functions are integrable.

Hence while the notion of uniform integrability has important applications in probability spaces or more generally finite measure spaces, this is not the case in more general measure spaces.

Proposition 2.52 (Uniform Integrability Convergence theorem) *Let $\{f_n(x)\}$ be a **uniformly integrable** sequence of functions on a finite measure space $(X, \sigma(X), \mu)$ with $f_n(x) \to f(x)$ μ-a.e. Assume either that f is μ-measurable or that $(X, \sigma(X), \mu)$ is complete. Then f is μ-integrable, and*

$$\int_X f_n(x) d\mu \to \int_X f(x) d\mu. \tag{2.28}$$

Proof. *By either assumption f is μ-measurable, and thus by Fatou's lemma and the above remark there exists N so that:*

$$\int_X |f(x)|\, d\mu \leq \liminf \int_X |f_n(x)|\, d\mu \leq N\mu(X) + 1,$$

and thus $f(x)$ is integrable. Given arbitrary $N \in \mathbb{R}$ let $f_n^{(N)} = f_n$ for $|f_n(x)| < N$ and $f_n^{(N)} = 0$ otherwise, and similarly $f^{(N)} = f$ for $|f(x)| < N$ and $f^{(N)} = 0$ otherwise. To ensure that $f_n^{(N)} \to f^{(N)}$ μ-a.e. it must be verified that $\mu\{|f| = N\} = 0$. This is because it is possible that $|f_n(x_0)| < N$ for all n yet $|f_n(x_0)| \to |f(x_0)| = N$, and then $f_n^{(N)}(x_0) \nrightarrow f^{(N)}(x_0)$. However by integrability of f, $\mu\{|f| = N\} = 0$ for all but countably many N, and so with countably many exceptions in N, $f_n^{(N)} \to f^{(N)}$ μ-a.e. The bounded convergence theorem now applies to assure that for all such N :

$$\int_X f_n^{(N)}(x) d\mu \to \int_X f^{(N)}(x) d\mu. \tag{(*)}$$

Now:

$$\int_X f_n(x) d\mu = \int_X f_n^{(N)}(x) d\mu + \int_{|f_n| \geq N} f_n(x) d\mu,$$

and similarly

$$\int_X f(x)d\mu = \int_X f^{(N)}(x)d\mu + \int_{|f|\geq N} f(x)d\mu.$$

Thus,

$$\begin{aligned}\left|\int_X f_n(x)d\mu - \int_X f(x)d\mu\right| \leq & \left|\int_X f_n^{(N)}(x)d\mu - \int_X f^{(N)}(x)d\mu\right| \\ & + \int_{|f_n|\geq N} |f_n(x)|\, d\mu + \int_{|f|\geq N} |f(x)|\, d\mu.\end{aligned}$$

By the bounded convergence result in $(*)$ *it follows that for all such* N :

$$\begin{aligned}\limsup_n \left|\int_X f_n(x)d\mu - \int_X f(x)d\mu\right| \leq & \limsup_n \int_{|f_n|\geq N} |f_n(x)|\, d\mu + \int_{|f|\geq N} |f(x)|\, d\mu \\ \leq & \sup_n \int_{|f_n|\geq N} |f_n(x)|\, d\mu + \int_{|f|\geq N} |f(x)|\, d\mu.\end{aligned}$$

Letting $N \to \infty$, *avoiding the countably many values noted above, it follows that the right hand side converges to* 0 *by uniform integrability of* $\{f_n(x)\}$ *and the integrability of* $f(x)$. ■

2.5 Lebesgue-Stieltjes Integrals by Riemann Sums

Important special cases of the general integrals above are provided by the Borel measure space $(\mathbb{R}, \mathcal{M}_{\mu_F}(\mathbb{R}), \mu_F)$ and the n-dimensional counterpart $(\mathbb{R}^n, \mathcal{M}_{\mu_F}(\mathbb{R}^n), \mu_F)$. These are often referred to as **Lebesgue-Stieltjes measure spaces** and μ_F **the Lebesgue-Stieltjes measure induced by the function** F. It is named for **Henri Lebesgue** (1875 – 1941) and **Thomas Stieltjes** (1856 – 1894). Integration in these measure spaces is then referred to as **Lebesgue-Stieltjes integration,** and the associated integrals are called **Lebesgue-Stieltjes integrals.**

The 1-dimensional Borel measure space $(\mathbb{R}, \mathcal{M}_{\mu_F}(\mathbb{R}), \mu_F)$ is developed in chapter 5 of book 1, in which case every Borel measure μ can be identified with a **monotonically increasing, right continuous** function $F(x)$, and conversely, any such function induces a Borel measure μ_F. In this latter construction, μ_F is the unique measure on the Borel sigma algebra $\mathcal{B}(\mathbb{R})$ induced by the set function μ_0 defined on the semi-algebra $\mathcal{A}'$ of right semi-closed intervals $(a, b]$ by:

$$\mu_0((a, b]) \equiv F(b) - F(a).$$

This set function is then extended to a measure $\mu_{\mathcal{A}}$ on the algebra $\mathcal{A}$ generated by $\mathcal{A}'$, which then gives rise to the outer measure $\mu^*_{\mathcal{A}}$ of definition 5.15 of book 1 and defined on all subsets of $\mathbb{R}$. The sigma algebra $\mathcal{M}_{\mu_F}(\mathbb{R})$ of the final measure space $(\mathbb{R}, \mathcal{M}_{\mu_F}(\mathbb{R}), \mu_F)$ is defined to contain all sets which are **Carathéodory measurable** by definition 5.18 of book 1, and named for **Constantin Carathéodory** (1873 – 1950). This sigma algebra is complete, $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}_{\mu_F}(\mathbb{R})$, and $\mu_F \equiv \mu^*_{\mathcal{A}}$ on $\mathcal{M}_{\mu_F}(\mathbb{R})$. Further, μ_F is an extension of μ_0 in the sense that $\mu_F = \mu_0$ on $\mathcal{A}'$.

The n-dimensional measure space $(\mathbb{R}^n, \mathcal{M}_{\mu_F}(\mathbb{R}^n), \mu_F)$ is developed in chapter 8 of book 1 with largely the same program. Again functions $F(x)$ defined on $\mathbb{R}^n$ and Borel measures on $\mathbb{R}^n$ can be identified, though the development here was slightly less general. Specifically, every **finite** Borel measure μ can be identified with an n**-increasing and continuous from above function** $F(x)$, and any such function gives rise to a (not-necessarily finite) Borel measure μ_F. Recalling these notions from book 1:

Definition 2.53 *A function $F(x)$ defined on $\mathbb{R}^n$ is:*

1. ***Continuous from above*** *at $x = (x_1, x_2, ..., x_n)$ if given $\{x^{(m)}\} \subset \mathbb{R}^n$ with $x_i^{(m)} \geq x_i$ for all i and $x^{(m)} \to x$ as $m \to \infty$:*

$$F(x) = \lim_{m\to\infty} F(x^{(m)}). \tag{2.29}$$

2. *F is n-**increasing** if given any bounded right semi-closed rectangle $A \equiv \prod_{i=1}^n (a_i, b_i]$:*

$$\Delta F \equiv \sum\nolimits_x sgn(x)F(x) \geq 0, \tag{2.30}$$

where each $x = (x_1, ..., x_n)$ in this summation is one of the 2^n vertices of A, so each $x_i = a_i$ or $x_i = b_i$, and $sgn(x)$ is defined as -1 if the number of a_i-components of x is odd, and $+1$ otherwise.

Given such F, a set function μ_0 is defined on any **bounded** right semi-closed rectangle $\prod_{i=1}^n (a_i, b_i]$ by:

$$\mu_0 \left[\prod\nolimits_{i=1}^n (a_i, b_i]\right] \equiv \sum\nolimits_x sgn(x)F(x), \tag{2.31}$$

and then this set function is extended uniquely to a measure μ_F on $\mathcal{B}(\mathbb{R}^n)$, and to a complete sigma algebra $\mathcal{M}_{\mu_F}(\mathbb{R}^n)$ with $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{M}_{\mu_F}(\mathbb{R}^n)$, and once again μ_F extends μ_0.

2.5.1 Lebesgue-Stieltjes Integrals on $\mathbb{R}$

For a continuous and thus Borel measurable function $g(x)$, we can approximate the Lebesgue-Stieltjes integral $\int_a^b gd\mu_F$ by Riemann sums as follows. Let $\Delta_n = \{x_i\}_{i=0}^n$ be a partition of $[a,b]$ with $a = x_0 < x_1 \cdots < x_n = b$, and define the step function $g_n(x)$ on $[a,b]$ by:

$$g_n(x) = \sum\nolimits_{i=0}^{n-1} g(x_i')\chi_{(x_i,x_{i+1}]}(x).$$

As usual, $\chi_{(x_i,x_{i+1}]}(x)$ denotes the characteristic function of $(x_i, x_{i+1}]$, defined to equal 1 on this interval and 0 elsewhere, and $x_i' \in [x_i, x_{i+1}]$ is arbitrary. For continuous g, $g_n(x) \to g(x)$ pointwise as $n \to \infty$ if the partitions' **mesh size** $\delta_n \equiv \max_{1\leq i\leq n}\{x_i - x_{i-1}\} \to 0$. This follows since for all $x \in [a,b]$:

$$|g(x) - g_n(x)| \leq \sup\nolimits_J \left|g(x) - g(x')\right|,$$

where $x, x' \in J \equiv [x_i, x_{i+1}]$ is a subinterval of length δ_n. By uniform continuity of g on $[a,b]$, it then follows that $g_n(x) \to g(x)$ for all x as $\delta_n \to 0$.

As $g_n(x)$ is a simple function, we have by 2.1:

$$\int_a^b g_n d\mu_F = \sum\nolimits_{i=0}^{n-1} g(x_i')\left[F(x_{i+1}) - F(x_i)\right]. \tag{2.32}$$

since $\mu_F((x_i, x_{i+1}]) = F(x_{i+1}) - F(x_i)$ by the above discussion.

Now if $\int_a^b g_n d\mu_F$ converges to a limit as $n \to \infty$, this limit must be independent of the choice of $\{x_i'\}_{i=0}^{n-1}$. This again follows by uniform continuity of $g(x)$ because if $\{x_i''\}_{i=0}^{n-1}$ is another selection:

$$\begin{aligned}&\left|\sum\nolimits_{i=0}^{n-1} g(x_i')\left[F(x_{i+1}) - F(x_i)\right] - \sum\nolimits_{i=0}^{n-1} g(x_i'')\left[F(x_{i+1}) - F(x_i)\right]\right| \\ \leq\ &\sup\nolimits_J \left|g(x') - g(x'')\right| \sum\nolimits_{i=0}^{n-1} \left|F(x_{i+1}) - F(x_i)\right|.\end{aligned}$$

As above x', $x'' \in J \equiv [x_i, x_{i+1}]$, a subinterval of length δ_n. If F is increasing this summation equals $F(b) - F(a)$, and so this difference converges to 0 as $\delta_n \to 0$ since g is uniformly continuous.

We now prove the main result of this section for increasing $F(x)$.

Proposition 2.54 (Lebesgue-Stieltjes Integrals by Riemann Sums) *Let $F(x)$ be a right continuous increasing function and $g(x)$ a continuous function. Then for any sequence of partitions of $[a,b]$, $\Delta_n = \{x_i\}_{i=0}^n$ with $\delta_n \equiv \max_{1\leq i\leq n}\{x_i - x_{i-1}\} \to 0$, and arbitrary $\{x_i'\}_{i=0}^{n-1}$ with $x_i' \in [x_i, x_{i+1}]$:*

$$\sum\nolimits_{i=0}^{n-1} g(x_i')\left[F(x_{i+1}) - F(x_i)\right] \to \int_a^b gd\mu_F. \tag{2.33}$$

Proof. *With uniqueness of the limit addressed above, we only need to prove existence. It follows by continuity of $g(x)$ that $|g| \leq K$ on $[a,b]$, and thus for increasing F :*

$$\int_a^b |g| \, d\mu_F \leq K \left[F(b) - F(a)\right],$$

and $g(x)$ is μ_F-integrable. Next, define $x_i' \in [x_i, x_{i+1}]$ so that with $J_i \equiv [x_i, x_{i+1}]$, $g(x_i') = \inf_{x \in J_i} g(x)$. Note that x_i' exists by continuity of $g(x)$ and compactness of J_i. Let $g_n(x)$ be defined as above by:

$$g_n(x) = \sum\nolimits_{i=0}^{n-1} g(x_i')\chi_{(x_i, x_{i+1}]}(x).$$

Then for all x, $g_n(x) \leq g(x)$ and $g_n(x) \to g(x)$ pointwise if $\delta_n \to 0$. Hence by Lebesgue's dominated convergence theorem:

$$\int_a^b g_n d\mu_F \to \int_a^b g d\mu_F,$$

and the result follows by 2.32. ■

Remark 2.55 *It should be noted that if $F(x)$ is right continuous and of* ***bounded variation*** *(see definition 3.23 of book 3), then the decomposition $F(x) = F_1(x) - F_2(x)$ into increasing functions is provided by proposition 3.27 of book 3. The above discussion then assures that $\sum\nolimits_{i=0}^{n-1} g(x_i') \left[F(x_{i+1}) - F(x_i)\right]$ converges as $\delta_n \to 0$ for arbitrary $\{x_i'\}_{i=0}^{n-1}$, and that this limit is independent of $\{x_i'\}_{i=0}^{n-1}$. We must be hesitant to declare however that this sequence converges to $\int_a^b g d\mu_F$ since we have not defined such integrals, nor indeed such measures. If such a "measure" existed, it would no longer be the case that $\mu_F((x_i, x_{i+1}]) = F(x_{i+1}) - F(x_i) \geq 0$ since F need not be increasing. Thus μ_F would be called a* ***signed measure,*** *a notion which is studied in the last chapter of this book.*

Lebesgue-Stieltjes vs. Riemann-Stieltjes Integrals

To induce a Borel measure μ_F on $\mathbb{R}$ it is necessary and sufficient for the function $F(x)$ to be increasing and right continuous by chapter 5 of book 1. To attempt to define an associated **Lebesgue-Stieltjes integral** $\int_a^b g d\mu_F$ above, the function $g(x)$ must be μ_F-measurable. For the **Riemann-Stieltjes integrals** of chapter 4 of book 3, denoted $\int_a^b g dF$, one existence result is that if the integrator function $F(x)$ is increasing and $g(x)$ continuous then this integral is well-defined (see proposition 4.19 of book 3). That the Lebesgue-Stieltjes and Riemann-Stieltjes integrals can agree under certain assumptions generalizes the equality of Riemann and Lebesgue integrals in certain cases as noted in propositions 2.31 and 2.64 of book 3.

These earlier results were quite general. For the current context we settle for a more limited result.

Proposition 2.56 (Lebesgue-Stieltjes vs. Riemann-Stieltjes Integrals) *If $F(x)$ is increasing and right continuous and $g(x)$ is continuous on $[a,b]$, then the associated Riemann-Stieltjes and Lebesgue-Stieltjes integrals agree:*

$$\int_a^b g d\mu_F = \int_a^b g dF. \tag{2.34}$$

Proof. *By proposition 2.54 above and proposition 4.20 of book 3, each integral equals the limit of the same Riemann sums.* ■

Remark 2.57 *By either remark 4.21 or exercise 4.22 of book 3, $\int_a^b g dF$ also exists if $g(x)$ is continuous and $F(x)$ is of* ***bounded variation****. This integral is again the limit of $\int_a^b g_n dF$ as $n \to \infty$, and this limit is independent of $\{x_i'\}_{i=0}^{n-1}$. In addition, proposition 4.27 of book 3 provides the relationship between such Riemann-Stieltjes integrals with bounded variation integrators and Riemann-Stieltjes integrals defined with respect to the increasing component function integrators, where $F(x) = F_1(x) - F_2(x)$ as noted in remark 2.55. However, we have not developed the definitional counterpart for $\int_a^b g d\mu_F$ in the case of general bounded variation F.*

2.5.2 Lebesgue-Stieltjes Integrals on $\mathbb{R}^n$

This section will follow the template for $n = 1$ almost identically, with appropriate generalizations of notation and references to results from books 1 and 3. In this context we address integrals:

$$\int_R g d\mu_F$$

for appropriate F, where $R = \prod_{i=1}^n (a_i, b_i]$, a bounded, right semi-closed rectangle in $\mathbb{R}^n$, and where $g(x)$ is continuous on the closed set $\bar{R} = \prod_{i=1}^n [a_i, b_i]$.

To partition R into subrectangles we introduce some notation from section 4.2 of book 3. Let each interval $(a_j, b_j]$ be partitioned:

$$a_j = x_{j,0} < x_{j,1} < \cdots < x_{j,m_j-1} < x_{j,m_j} = b_j,$$

with **mesh size** μ of this collection of partitions defined by:

$$\mu \equiv \max\{x_{j,i} - x_{j,i-1}\},$$

where the maximum is defined over $1 \leq i \leq m_j$ and $1 \leq j \leq n$. These interval partitions lead to a partition of R into disjoint right semi-closed rectangles:

$$\bigcup\nolimits_{J \in I} R_J = R,$$

where $R_J \equiv \prod_{j=1}^n R_{j,i_j}$ with $R_{j,i_j} \equiv (x_{j,i_j-1}, x_{j,i_j}]$ and $I = \{(i_1, i_2, ..., i_n)|1 \leq i_j \leq m_j\}$. The index set I identifies the $\prod_{j=1}^n m_j$ disjoint rectangles that are defined by this partition. For notational convenience in this section we will assume that all $m_j = m$.

To approximate $g(x)$ on R by a step function, first note that ΔF defined as in 2.30 is finitely additive over this partition. That is, if $\{R_J\}_{J \in I}$ is the collection of disjoint right semi-closed rectangles defined above, then

$$\Delta F = \sum\nolimits_{J \in I} \Delta F_J, \tag{2.35}$$

where ΔF_J denotes ΔF applied to R_J. This result is proposition 8.11 of book 1.

With this notation, define the step functions $g_m(x)$ on R by:

$$g_m(x) = \sum\nolimits_{J \in I} g(x'_J)\chi_{R_J}(x),$$

where the subrectangles $\{R_J\}$ are defined as above with all $m_j = m$, and $x'_J \in \bar{R}_J$ the closure of R_J. For continuous g, $g_m(x) \to g(x)$ pointwise as $m \to \infty$ if the partitions's mesh size $\mu \equiv \max\{x_{j,i} - x_{j,i-1}\} \to 0$. This follows since if $x \in \bar{R}_J$:

$$|g(x) - g_m(x)| \leq \sup\nolimits_{\bar{R}_J} \left|g(x) - g(x')\right|,$$

and thus $g_m(x) \to g(x)$ as $\mu \to 0$ by uniform continuity of g on $\bar{R}$.

As $g_m(x)$ is a simple function, we have by 2.1:

$$\int_R g_m d\mu_F = \sum\nolimits_{J \in I} g(x'_J)\Delta F_J, \tag{2.36}$$

since $\mu_F\left[R_J\right]=\Delta F_J$ by 2.31 and the remarks that followed there.

As before for $n=1$, if $\int_R g_m d\mu_F$ converges to a limit as $m\to\infty$, this limit must be independent of $\{x'_J\}_{J\in I}$. This again follows by uniform continuity of $g(x)$ because if $\{x''_J\}_{J\in I}$ is another selection:

$$\left|\sum\nolimits_{J\in I} g(x'_J)\Delta F_J-\sum\nolimits_{J\in I} g(x''_J)\Delta F_J\right|\leq\sup\nolimits_{\bar{R}_J}\left|g(x')-g(x'')\right|\sum\nolimits_{J\in I}\left|\Delta F_J\right|.$$

As F is n-increasing this summation equals ΔF by 2.35 and so this difference converges to 0 by uniform continuity of $g(x)$.

Proposition 2.58 (Lebesgue-Stieltjes Integrals by Riemann Sums) *Let $F(x)$ be a continuous from above and n-increasing function on $\mathbb{R}^n$, and $g(x)$ a continuous function on the closure $\bar{R}$ of the right semi-closed rectangle R. Then for any sequence of partitions of $R=\bigcup_{J\in I}R_J$ with $\mu\equiv\max\{x_{j,i}-x_{j,i-1}\}\to 0$ as $m\to\infty$, and arbitrary $\{x'_J\}_{J\in I}$ with $x'_J\in\bar{R}_J$:*

$$\sum\nolimits_{J\in I} g(x'_J)\Delta F_J\to\int_R g d\mu_F. \tag{2.37}$$

Proof. *Uniqueness of this limit is addressed above, so only existence is left to prove. By uniform continuity $|g|\leq K$ on $\bar{R}$ and thus:*

$$\int_R |g|\,d\mu_F\leq K\int_R d\mu_F=K\Delta F,$$

so $g(x)$ is μ_F-integrable. Next, define $x'_J\in\bar{R}_J$ so that $g(x'_J)=\inf_{\bar{R}_J}g(x)$ and note that x'_J exists by continuity of $g(x)$ and compactness of $\bar{R}_J$. Then with such $x'_J\in\bar{R}_J$ define $g_m(x)$ by:

$$g_m(x)=\sum\nolimits_{J\in I} g(x'_J)\Delta F_J.$$

Then for all x, $g_m(x)\leq g(x)$ and $g_m(x)\to g(x)$ pointwise as $m\to\infty$. Hence by Lebesgue's dominated convergence theorem:

$$\int_R g_m d\mu_F\to\int_R g d\mu_F,$$

and the result follows by 2.36. ■

Lebesgue-Stieltjes vs. Riemann-Stieltjes Integrals

To induce a Borel measure μ_F on $\mathbb{R}^n$ it is sufficient for the function $F(x)$ to be n-increasing and continuous from above by section 8.2.2 of book 1. In that book's section 8.2.1 it is also seen that these conditions are necessary for finite Borel measures. To define an associated **Lebesgue-Stieltjes integral** $\int_R gd\mu_F$ above, the function $g(x)$ must be μ_F-measurable. For $\int_R gdF$, the **Riemann-Stieltjes integral** of book 3, one existence result provides that if the integrator function $F(x)$ is n-increasing and $g(x)$ continuous then this integral is well-defined (see proposition 4.67 of book 3). That the Lebesgue-Stieltjes and Riemann-Stieltjes integrals can agree under certain assumptions again generalizes the equality of Riemann and Lebesgue integrals in certain cases as noted in propositions 2.31 and 2.64 of book 3.

These earlier results were quite general. For the current context we again settle for a more limited result.

Proposition 2.59 (Lebesgue-Stieltjes vs. Riemann-Stieltjes Integrals) *If $F(x)$ is n-increasing and continuous from above and $g(x)$ is continuous on $\bar{R}$ the closure of $R = \prod_{i=1}^{n}(a_i, b_i]$, then the associated Riemann-Stieltjes and Lebesgue-Stieltjes integrals agree:*

$$\int_R gd\mu_F = \int_R gdF. \tag{2.38}$$

Proof. *By proposition 2.58 above and corollary 4.68 of book 3, each integral equals the limit of the same Riemann sums.* ■

Chapter 3

Change of Variables Results

It is a common exercise in Riemann integration to use "change of variables" or "method of substitution" in a Riemann integral to simplify its evaluation.

Example 3.1 *Let* $h(x) = x\exp(-x^2)$ *and assume we wish to evaluate* $\int_0^\infty h(x)dx$*. The familiar "method of substitution" or "change of variables" approach requires one to define a new "variable"* $y \equiv g(x) = x^2$*, and then by defining* $dy \equiv g'(x)dx = 2xdx$ *one obtains:*

$$\begin{aligned}\int_0^\infty x\exp(-x^2)dx &= 0.5\int_{g(0)}^{g(\infty)}\exp(-y)dy \\ &= 0.5.\end{aligned}$$

Within the Riemann context, this manipulation can be interpreted as a ***trompe l'oeil****, French for "deceive the eye," the goal of which is to simplify the application of the fundamental theorem of calculus (proposition 3.1 of book 3) by making it easier to identify the integrand as a derivative. In other words, if we can identify a function* $H(x)$ *with* $H'(x) = x\exp(-x^2)$*, then this result states that:*

$$\int_0^\infty x\exp(-x^2)dx = H(\infty) - H(0),$$

assuming $H(\infty) \equiv \lim_{x\to\infty} H(x)$ *exists.*

The above identification of $H(x)$ *is based on an application of the formula for the derivative of a composite function, that with* $H(x) = f(g(x))$ *:*

$$H'(x) = f'(g(x))\,g'(x).$$

So formally, this method seeks to find functions f and g so that,

$$x\exp(-x^2) = f'(g(x))\, g'(x). \tag{(*)}$$

As x is a multiple of the derivative of x^2, we choose $g(x) = x^2$ and $f'(y) = 0.5\exp(-y)$, and then:

$$x\exp(-x^2) = f'(g(x))\, g'(x).$$

Thus $f(y) = -0.5\exp(-y)$

$$\begin{aligned}\int_0^\infty x\exp(-x^2)dx &= -\int_0^\infty \left[0.5x\exp(-x^2)\right]' dx \\ &= 0.5.\end{aligned}$$

From a practical point of view, the mechanical technique of substitution is simpler than the more formal approach of solving the equation in $(*)$ for f and g. This is because the objective of the substitution is to help the eye see an underlying function (here $f'(y) = 0.5\exp(-y)$), with a simple antiderivative (here $f(y) = -0.5\exp(-y) + C$). From a pedagogical point of view, substitution also supports trial and error in determining the most appropriate variable $y = g(x)$, and also supports sequential substitutions to ultimately reduce a complicated integrand to something manageable, if indeed this is possible.

In this chapter we will see that this change of variable formula also has an interpretation from a measure-theoretic perspective, and this has important applications in probability theory in book 6. We begin with the special case of evaluating certain Lebesgue-Stieltjes integrals by a change of measure. We then turn to a general investigation into the transformation of measures between measure spaces and the implications for integrals in the domain and range spaces, and then to various applications of these results to change of variables in Lebesgue integrals.

3.1 Change of Measure in Lebesgue-Stieltjes Integrals

This section provides a simple approach for evaluating a Lebesgue-Stieltjes integral in terms of a related Lebesgue integral when the Borel measure μ is of a special type, and in particular, a type often found in probability theory. A more general result, also needed for probability theory, will be addressed in the next section.

Let $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$ be a Borel measure space and assume that there is a nonnegative Lebesgue measurable function $f(x)$ defined on $\mathbb{R}^n$ so that for any Borel set $A \in \mathcal{B}(\mathbb{R}^n)$:

$$\mu(A) = (\mathcal{L}) \int_A f(x)dx. \tag{3.1}$$

Here $x \equiv (x_1, ..., x_n)$ and dx with the $(\mathcal{L})$-qualifier denotes Lebesgue measure, which is also denoted dm^n.

Example 3.2 *Given nonnegative Lebesgue integrable $f(x)$ for $n = 1$ define $F(x)$ by:*

$$F(x) = (\mathcal{L}) \int_{-\infty}^{x} f(y)dy.$$

Then F is increasing since $f \geq 0$, and continuous and thus right continuous by proposition 3.33 of book 3. Hence the set function μ_F defined by:

$$\mu_F((a, b]) \equiv F(b) - F(a) = (\mathcal{L}) \int_a^b f(y)dy,$$

is well defined on the semi-algebra of right semi-closed intervals $\mathcal{A}'$ and can be uniquely extended to a measure on $\mathcal{B}(\mathbb{R})$ as in chapter 5 of book 1. Since $\mu_F = \mu$ on this semi-algebra, with μ given by 3.1, it follows that $\mu_F = \mu$ on $\mathcal{B}(\mathbb{R})$ and thus μ_F is given by 3.1 for all $A \in \mathcal{B}(\mathbb{R})$.

This example extends to a result for general n in that given nonnegative Lebesgue integrable f, define:

$$F(x_1, ..., x_n) = (\mathcal{L}) \int_{A(x_1,...,x_n)} f(y_1, ..., y_n)dy,$$

where $A(x_1, ..., x_n) \equiv \prod_{i=1}^{n}(-\infty, x_i]$. That such F is continuous from above and n-increasing by definition 2.53 is left as an exercise. As in section 8.2.2 of book 1 a set function μ_F can be defined on a bounded right semi-closed rectangle $R = \prod_{i=1}^{n}(a_i, b_i]$ as noted in 2.31:

$$\mu_F \left[\prod\nolimits_{i=1}^{n}(a_i, b_i]\right] \equiv \sum\nolimits_x sgn(x)F(x).$$

Recall that each $x = (x_1, ..., x_n)$ in this summation is one of the 2^n vertices of A, so each $x_i = a_i$ or $x_i = b_i$, and $sgn(x)$ is defined as -1 if the number of a_i-components of x is odd, and $+1$ otherwise. The conclusion that μ_F extends to a measure on $\mathcal{B}(\mathbb{R}^n)$ and that $\mu_F = \mu$ on $\mathcal{B}(\mathbb{R}^n)$ with μ given by 3.1, then follows from section 8.2.2 of book 1.

Generalizing this example, when f is nonnegative and Lebesgue **integrable,** the set function μ defined in 3.1 defines a measure on $\mathcal{B}(\mathbb{R}^n)$ by 2.13 of corollary 2.29, a result also proved for Lebesgue integrals in corollary 2.55 of book 3. More generally we have the following.

Proposition 3.3 *If $f(x)$ is a nonnegative Lebesgue measurable function on $\mathbb{R}^n$, then the set function μ in 3.1 defines a measure on $\mathcal{B}(\mathbb{R}^n)$.*
Proof. *Countable additivity is the only thing to prove, so assume $\{B_j\} \subset \mathcal{B}(\mathbb{R}^n)$ are disjoint and let $B = \bigcup_{j=1}^{\infty} B_j$. If $f(x)$ is integrable over B then countable additivity follows from corollary 2.29 as noted above. So assume that B is so defined and $f(x)$ is not integrable over B, which since nonnegative implies that $\mu(B) = \infty$. Now if there exists j with $f(x)$ also not integrable over B_j, so $\mu(B_j) = \infty$, then countable additivity is satisfied and we are done.*

So assume that $f(x)$ is integrable over all B_j, and thus $\mu(B_j) < \infty$ for all j. Define $C_n \equiv \bigcup_{j=1}^{n} B_j$ and $f_n(x) \equiv f(x)\chi_{C_n}(x)$. Then $f_n(x)$ is nonnegative, and Lebesgue integrable by 6 of proposition 2.40 with:

$$\int f_n(x)dx \equiv \int_{C_n} f(x)dx = \sum_{j=1}^{n} \int_{B_j} f(x)dx = \sum_{j=1}^{n} \mu(B_j).$$

Also, $\{f_n\}$ is a monotone sequence of nonnegative functions with $f_n(x) \to f(x)\chi_B(x)$. Thus by Lebesgue's monotone convergence theorem:

$$\int f_n(x)dx \to \int_B f(x)dx.$$

Combining:

$$\sum_{j=1}^{n} \mu(B_j) \to \int_B f(x)dx = \infty,$$

proving countable additivity in this case. ■

Remark 3.4 *Note that using the same proof, it is seen that μ also defines a measure on the complete Lebesgue sigma algebra $\mathcal{M}_L^n(\mathbb{R}^n)$, since the integral in 3.1 is well-defined for $A \in \mathcal{M}_L^n(\mathbb{R}^n)$. Recall that $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{M}_L^n(\mathbb{R}^n)$. This observation then extends to proposition 3.6 below, in that the results there are valid for Lebesgue measurable g. However, for results below on change of variables in Lebesgue integrals we will focus on Borel measurable g for reasons to be clarified there.*

Let $g(x)$ be a measurable function defined on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$:

$$g : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), m),$$

so $g^{-1}(A) \in \mathcal{B}(\mathbb{R}^n)$ for all $A \in \mathcal{B}(\mathbb{R})$. As $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$ is a measure space and g is measurable, the μ-integral of $g(x)$ over any Borel set $A \in \mathcal{B}(\mathbb{R}^n)$:

$$\int_A g(x)d\mu,$$

is definable as a Lebesgue-Stieltjes integral by definition 2.37.

The following result relates the μ-integrability of $g(x)$ and the value of its Lebesgue-Stieltjes integral to the Lebesgue integrability of $f(x)g(x)$ and the value of the associated Lebesgue integral.

Remark 3.5 *It should be noted that the next result substantially generalizes the result of proposition 4.75 of book 3, which required continuity of $f(x)$ and $g(x)$ to be able to utilize Riemann and Riemann-Stieltjes integrals. On the other hand, the earlier result also applied to $f(x)$ that was not nonnegative, as long as the associated $F(x)$ was of Vitali bounded variation (see definition 4.57 of book 3).*

Proposition 3.6 *Given a nonnegative Lebesgue measurable function $f : \mathbb{R}^n \to \mathbb{R}$ and μ defined on $\mathcal{B}(\mathbb{R}^n)$ as in 3.1. Then for any Borel measurable function $g : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), m^n) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$:*

$$\int_{\mathbb{R}^n} g(x)d\mu = (\mathcal{L}) \int_{\mathbb{R}^n} g(x)f(x)dx, \tag{3.2}$$

although both integrals may be infinite.

In addition, g is μ-integrable if and only if $f(x)g(x)$ is Lebesgue integrable, and when integrable,

$$\int_A g(x)d\mu = (\mathcal{L}) \int_A g(x)f(x)dx, \tag{3.3}$$

for all $A \in \mathcal{B}(\mathbb{R}^n)$.

Proof. *First note that if this proposition is proved for nonnegative $g(x)$ then the general result follows by applying the special result to $g^+(x)$ and $g^-(x)$, the positive and negative parts of $g(x)$, and applying 2.17.*

Now given $A \in \mathcal{B}(\mathbb{R}^n)$, let $g(x) = \chi_A(x)$. Then 3.2 is satisfied because

$$\int g(x)d\mu = \int_A d\mu = \mu(A),$$

and also:

$$(\mathcal{L}) \int_{\mathbb{R}} g(x)f(x)dx = (\mathcal{L}) \int_A f(x)dx = \mu(A).$$

If $g(x)$ is a simple function, then 3.2 again follows by linearity of both integrals.

For general nonnegative $g(x)$, let $\{g_n(x)\}$ be an increasing sequence of nonnegative simple functions given by proposition 1.18 with $g_n(x) \to g(x)$ for all x. Since f is nonnegative it follows that $\{g_n(x)f(x)\}$ is an increasing sequence of nonnegative Lebesgue measurable functions and $g_n(x)f(x) \to g(x)f(x)$. By Lebesgue's monotone convergence theorem:

$$\int g_n(x)d\mu \to \int g(x)d\mu,$$

$$\int g_n(x)f(x)dx \to \int g(x)f(x)dx,$$

and 3.2 follows since this identity is satisfied for all $g_n(x)$.

The conclusion of 3.2 now applies to general measurable $g(x)$ as note above, and to $|g(x)|$ for a general measurable function:

$$\int |g(x)|\, d\mu = (\mathcal{L}) \int f(x)\, |g(x)|\, dx.$$

Hence $g(x)$ is μ-integrable if and only if $f(x)g(x)$ is Lebesgue integrable.

Finally, if $g(x)$ is μ-integrable and $A \in \mathcal{B}(\mathbb{R}^n)$, then so too is $g(x)\chi_A(x)$ μ-integrable. Hence applying 3.2 to $g(x)\chi_A(x)$ over $\mathbb{R}^n$ obtains 3.3. ■

Example 3.7 (Expectations of Random Variables 1) *Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_F)$ be a Borel measure space induced by an increasing, right continuous and bounded function $F(x)$ as in the construction of section 5.2 of book 1. If F is absolutely continuous then $F'(x) = f(x)$ exists almost everywhere by proposition 3.58 of book 3, and assuming that $F(-\infty) = 0$ it follows by proposition 3.61 of book 3 that:*

$$F(x) = (\mathcal{L}) \int_{-\infty}^{x} f(y)dy.$$

It then follows as in example 3.2 above that μ_F is given as in 3.1 for all $A \in \mathcal{B}(\mathbb{R})$.

If $g : (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_F) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ is μ_F-measurable, then the Lebesgue-Stieltjes integral of $g(x)$ over $A \in \mathcal{B}(\mathbb{R})$ can be evaluated as a Lebesgue integral as in 3.3.

$$\int_A g(x)d\mu_F = (\mathcal{L}) \int_A g(x)f(x)dx.$$

In addition, $g(x)$ will be μ_F-integrable if and only if $g(x)f(x)$ is Lebesgue integrable.

If μ_F is a probability measure and thus $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_F)$ is a probability space, such a function $g(x)$ is called a ***random variable*** *and often denoted by an upper case letter such as X, Y, etc.. Then by 3.7 of book 4 the* ***expectation of*** *Y, denoted $E[Y]$, is defined:*

$$E[Y] \equiv \int_{\mathbb{R}} Y d\mu_F.$$

Thus by 3.3, when F is absolutely continuous:

$$E[Y] = (\mathcal{L}) \int_{\mathbb{R}} Y(x) F'(x) dx.$$

For a more general result in this direction, see example 3.17 below which utilizes the more general results on transformation of measures. See also section 3.1.2 of book 4 for a summary of these results.

The result of proposition 3.6 can be generalized beyond Lebesgue and the special Borel measures on $\mathbb{R}^n$ defined by 3.1. Specifically, if $(X, \sigma(X),$ $\mu)$ is a measure space and $f(x)$ a nonnegative μ-measurable function, $f:$ $(X, \sigma(X), \mu) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, then

$$\upsilon(A) \equiv \int_A f(x) d\mu \tag{3.4}$$

is a well defined, though not necessarily finite, set function for $A \in \sigma(X)$. Using the same proof as proposition 3.3 obtains that the set function υ is again a measure on $(X, \sigma(X))$.

Hence $(X, \sigma(X),\ \upsilon)$ is a measure space and if $g(x)$ is a ν-measurable function on X, then $\int_A g(x) d\upsilon$ is well defined though not necessarily finite for all $A \in \sigma(X)$, and the following is a straightforward generalization of the above result.

Proposition 3.8 *Given a nonnegative μ-measurable function $f : (X, \sigma(X), \mu) \rightarrow$ $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, and υ defined on $\sigma(X)$ as in 3.4. Then for any ν-measurable function $g : X \rightarrow \mathbb{R}$:*

$$\int_X g(x) d\upsilon = \int_X g(x) f(x) d\mu, \tag{3.5}$$

although both integrals may be infinite.

In addition, $g(x)$ is υ-integrable if and only if $f(x)g(x)$ is μ-integrable, and when integrable,

$$\int_A g(x)d\upsilon = \int_A g(x)f(x)d\mu, \tag{3.6}$$

for any $A \in \sigma(X)$.

Proof. *Other than the change in notation, the proof is identical to that above.* ■

3.2 Transformations and Change of Variables

3.2.1 Measures Induced by Transformations

Given sets X and X', a **transformation** $T : X \to X'$, and sometimes called a **map** or **mapping,** is simply a rule under which $Tx \in X'$ for each $x \in X$. Since this is virtually unusable as a general notion, transformations of interest almost always have additional properties. As an example, a **measurable transformation** is defined as follows. For this definition, recall that if $A' \subset X'$:

$$T^{-1}(A') = \{x \in X | Tx \in A'\}. \tag{3.7}$$

Definition 3.9 (Measurable transformations and Induced measures) *Given measure spaces $(X, \sigma(X), \mu)$ and $(X', \sigma(X'), \mu')$, a transformation, $T : X \to X'$ is **measurable** and sometimes $\sigma(X)/\sigma(X')$-**measurable** if*

$$T^{-1}\left[\sigma(X')\right] \subset \sigma(X),$$

meaning

$$T^{-1}(A') \in \sigma(X) \text{ for all } A' \in \sigma(X'). \tag{3.8}$$

*If $(X', \sigma(X'), \mu') = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu')$ for some Borel measure μ', then T is called a **measurable function**.*

*A measurable transformation induces a new measure μ_T on the range space X', and hence induces a new measure space denoted $(X', \sigma(X'), \mu_T)$. The measure μ_T is called **the measure induced by** T and defined on $A' \in \sigma(X')$ by*

$$\mu_T(A') = \mu\left[T^{-1}(A')\right]. \tag{3.9}$$

Since 3.9 only provides a set function definition, it must be checked that μ_T so defined is indeed a measure on $(X', \sigma(X'))$.

Proposition 3.10 *The set function μ_T defined in 3.9 is a measure on $(X', \sigma(X'))$.*

Proof. *That μ_T is a well-defined set function on $\sigma(X')$ follows from the measurability of T, while the property $\mu_T(A') \geq 0$ for all such A' is satisfied because μ is a measure. Since $\mu_T(\emptyset) = 0$, the last property to check is countable additivity, and to this end let $\{A'_j\}_{j=1}^{\infty} \subset \sigma(X')$ be disjoint. Then $\{T^{-1}(A'_j)\}_{j=1}^{\infty} \subset \sigma(X)$ are also disjoint by definition of a transformation, and since $T^{-1}\left(\bigcup_{j=1}^{\infty} A'_j\right) = \bigcup_{j=1}^{\infty} T^{-1}\left(A'_j\right)$ it follows that:*

$$\mu_T\left[\bigcup_{j=1}^{\infty} A'_j\right] = \mu\left[\bigcup_{j=1}^{\infty} T^{-1}(A'_j)\right] = \sum_{j=1}^{\infty} \mu_T\left[A'_j\right].$$

■

The measure μ_T is defined on a set $A' \in \sigma(X')$ to equal the μ-measure of the pre-image set $T^{-1}(A')$. This may well sound like a familiar idea, as we have encountered this already in earlier books.

Example 3.11 *Let $(\mathcal{S}, \mathcal{E}, \mu)$ be a probability space and X a* ***random variable (r.v.)****:*

$$X : \mathcal{S} \longrightarrow \mathbb{R}.$$

As defined in definition 3.1 of book 2, this means that given any bounded or unbounded interval, $\langle a, b\rangle \subset \mathbb{R}$, where $\langle a, b\rangle$ denotes that this interval may be open, closed or semi-closed:

$$X^{-1}(\langle a, b\rangle) \in \mathcal{E}.$$

It was assigned as exercise 3.3 of book 2 to show that if $X^{-1}((a, b)) \in \mathcal{E}$ for all open intervals (a, b), then $X^{-1}(A) \in \mathcal{E}$ for every Borel set $A \in \mathcal{B}(\mathbb{R})$. Hence by the above definition, a random variable as defined in book 2 is a ***measurable function*** *between $(\mathcal{S}, \mathcal{E}, \mu)$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$.*

The ***distribution function (d.f.)*** *or* ***cumulative distribution function (c.d.f.)*** *associated with X, denoted F or F_X, was then defined on $\mathbb{R}$ by:*

$$F(x) \equiv \mu[X^{-1}(-\infty, x]].$$

This definition is identical with that of μ_X as given in 3.9 for sets $A' = (-\infty, x]$. In other words:

$$F(x) = \mu_X\left[(-\infty, x]\right],$$

and thus by finite additivity, for any right semi-closed interval $(a,b]$:

$$\mu_X\left[(a,b]\right] = F(b) - F(a).$$

The random variable X *thus induces the measure space* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ *by the above definition.*

Proposition 3.60 of book 1 assures that such F *as defined above is an increasing, right continuous function, and thus in chapter 5 of book 2 these distribution functions provided the first step in defining an induced Borel measure* μ_F *on the range space* $\mathbb{R}$. *By this construction, any random variable* $X : \mathcal{S} \longrightarrow \mathbb{R}$ *induces a measure on the range space* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, *producing* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_F)$.

Since $\mu_F = \mu_X$ *on* $\mathcal{A}'$, *the semi-algebra of right semi-closed intervals* $(a,b]$, *this identity extends by proposition 6.14 of book 1 to the smallest sigma algebra that contains* $\mathcal{A}'$, *which is* $\mathcal{B}(\mathbb{R})$. *Thus except for notation:*

$$\mu_X \equiv \mu_F.$$

Although μ_T is defined on the same space and sigma algebra as was the original measure μ', the induced measure and original measure can be closely related or quite different.

Example 3.12 **1**. *Let* X_1 *be defined on a probability space,* $X_1 : (\mathcal{S}, \mathcal{E}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, *with range* $\mathbb{N} = \{0, 1, 2, 3...\}$ *such that for any* $j \in \mathbb{N}$:

$$\mu\left[X_1^{-1}(j)\right] = e^{-\lambda}\lambda^j/j!$$

for some $\lambda > 0$. *Hence*

$$F_1(x) = \mu[X_1^{-1}(-\infty, x]] = e^{-\lambda}\sum\nolimits_{j \leq x} \lambda^j/j!.$$

The induced measure μ_{X_1} *on* $\mathbb{R}$ *is the* ***Poisson probability measure*** μ_P *introduced in section 1.3 of book 2, and so in the various notations:*

$$\mu_{F_1} \equiv \mu_{X_1} = \mu_P.$$

This induced measure μ_{X_1} *is quite different from the original Lebesgue measure in that* m *assigns measure* 0 *to the set* $\mathbb{N}$ *on which* μ_{X_1} *assigns all of its measure. In chapter 7 we will say that* m *and* μ_{X_1} *are* ***mutually singular***, *denoted* $m \perp \mu_{X_1}$, *and define this to mean that there is a Borel set* A *so that* $m(A) = \mu_{X_1}(\tilde{A}) = 0$. *In this example,* $A = \mathbb{N}$ *and* $\tilde{A} = \mathbb{R} - \mathbb{N}$.

2. *Let X_2 be defined on a probability space, $X_2 : (\mathcal{S}, \mathcal{E}, \mu) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, with range $\mathbb{R}^+ \equiv \{x \in \mathbb{R} | x \geq 0\}$ such that for $x \geq 0$:*

$$F_2(x) = \mu[X_2^{-1}(-\infty, x]] = 1 - e^{-\lambda x}$$

*for some $\lambda > 0$. The induced measure μ_{X_2} on $\mathbb{R}$ is the **exponential probability measure** μ_E introduced in section 1.3 of book 2, and so in the various notations:*

$$\mu_{F_2} \equiv \mu_{X_2} = \mu_E.$$

In this case the induced measure μ_{X_2} is closely related to the original Lebesgue measure m. In particular, with the associated density function $f_E(x)$ defined for $x \geq 0$ by

$$f_E(x) = \lambda e^{-\lambda x}$$

and 0 elsewhere, we have that for $A \in \mathcal{B}(\mathbb{R})$,

$$\mu_{X_2}(A) = (\mathcal{L}) \int_A f_E(x) dx.$$

*Hence for any set $A \in \mathcal{B}(\mathbb{R})$ with $m(A) = 0$, it must be the case that $\mu_{X_2}(A) = 0$. In chapter 7 we will say that μ_{X_2} is **absolutely continuous** with respect to m, denoted $\mu_{X_1} \ll m$, where this notation is meant to suggests that if $m(A) = 0$ on a Borel set A, then $\mu_{X_1}(A) = 0$.*

3.2.2 General Change of Variables Under Transformations

The goal of this section is to relate integrals defined on the induced measure space $(X', \sigma(X'), \mu_T)$ with related integrals defined on the domain space $(X, \sigma(X), \mu)$. Intuitively, such integrals ought to be related because the μ_T-measures of sets in $\sigma(X')$ are defined in terms of the μ-measures of pre-image sets in $\sigma(X)$. Specifically, given general measure spaces $(X, \sigma(X), \mu)$ and $(X', \sigma(X'), \mu')$, a **measurable transformation** $T : X \rightarrow X'$ induces the measure μ_T on the range space X' defined on $A' \in \sigma(X')$ by 3.9:

$$\mu_T(A') = \mu\left[T^{-1}(A')\right].$$

Now consider a measurable function g defined on $(X', \sigma(X'), \mu')$ with range in $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$. By definition g is also measurable on $(X', \sigma(X'), \mu_T)$

since the definition of measurability depends only on the sigma algebra $\sigma(X')$. Diagrammatically:

$$(X, \sigma(X), \mu) \quad \xrightarrow{T} \quad (X', \sigma(X'), \mu_T) \quad \xrightarrow{g} \quad (\mathbb{R}, \mathcal{B}(\mathbb{R}), m).$$

As $(X', \sigma(X'), \mu_T)$ is a measure space and g is a measurable function,

$$\int_{A'} g(x')d\mu_T$$

is definable for any measurable set $A' \in \sigma(X')$ by the development in chapter 2, though such integrals need not be finite.

Measurability of g as defined on $(X', \sigma(X'), \mu_T)$ ensures that the composite function $g \circ T$ is measurable as a function defined on $(X, \sigma(X), \mu)$. This follows because if $B \in \mathcal{B}(\mathbb{R})$, then

$$[g \circ T]^{-1}(B) = T^{-1}\left[g^{-1}(B)\right],$$

and $g^{-1}(B) \in \sigma(X')$ ensures that $T^{-1}\left[g^{-1}(A)\right] \in \sigma(X)$ by the measurability of T. Hence as $(X, \sigma(X), \mu)$ is a measure space and $g \circ T \equiv g(T)$ is a measurable function,

$$\int_A g(Tx)d\mu,$$

is again definable for any measurable set $A \in \sigma(X)$.

The goal of the next proposition is to relate the value of these integrals.

Remark 3.13 *Note that although the μ'-integral $\int g(x')d\mu'$ can also be defined on the original range measure space $(X', \sigma(X'), \mu')$, this proposition is silent on its value. Instead because the range space X' is endowed with a new measure μ_T which provides a link to the measure in the domain measure space $(X, \sigma(X), \mu)$, this proposition can make a statement about the relative values of integrals on the given measure spaces. It is this linkage between the measures μ and μ_T that allows the following results.*

Proposition 3.14 *Let $T : (X, \sigma(X), \mu) \to (X', \sigma(X'), \mu')$ be a measurable transformation and μ_T defined on the range space by 3.9. Then for any nonnegative measurable function $g : X' \to \mathbb{R}$:*

$$\int_X g(Tx)d\mu = \int_{X'} g(x')d\mu_T, \tag{3.10}$$

though both integrals may be infinite.

More generally, a measurable function g is μ_T-integrable if and only if $g(Tx)$ is μ-integrable, and when integrable:

$$\int_{T^{-1}A'} g(Tx)d\mu = \int_{A'} g(x')d\mu_T \tag{3.11}$$

for any $A' \in \sigma(X')$.

Proof. *Given $A' \in \sigma(X')$, let $g(x') = \chi_{A'}(x')$. Then $\int_{X'} g(x')d\mu_T = \mu_T[A']$, and since $g(Tx) = \chi_{T^{-1}A'}(x)$ it follows that $\int_X g(Tx)d\mu = \mu\left[T^{-1}(A')\right]$. By 3.9 this proves 3.10 for characteristic functions of sets in $\sigma(X')$, and also for all simple functions on X' by linearity of the integral. For general non-negative $g(x')$, let $\{g_n(x')\}$ be an increasing sequence of nonnegative simple functions given by proposition 1.18 so that $g_n(x') \to g(x')$ for all x'. Then also $\{g_n(Tx)\}$ is an increasing sequence of simple functions defined on X with $g_n(Tx) \to g(Tx)$. So by Lebesgue's monotone convergence theorem:*

$$\int_X g_n(Tx)d\mu \to \int_X g(Tx)d\mu,$$

$$\int_{X'} g_n(x')d\mu_T \to \int_{X'} g(x')d\mu_T,$$

and the result follows since 3.10 is satisfied for this sequence of simple functions.

Applying this conclusion to $|g(x')|$ for a general measurable function, we conclude that

$$\int_X |g(Tx)|\, d\mu = \int_{X'} \left|g(x')\right| d\mu_T,$$

and hence $g(x')$ is μ_T-integrable if and only if $g(Tx)$ is μ-integrable. This is 3.11 with $A' = X'$.

If $g(x)$ is μ_T-integrable then so too is $g(x')\chi_{A'}(x')$ for $A' \in \sigma(X')$. The validity of 3.11 over X' implies that

$$\int_X g(Tx)\chi_{A'}(Tx)d\mu = \int_{X'} g(x')\chi_{A'}(x')d\mu_T,$$

while noting that $\chi_{A'}(Tx) = \chi_{T^{-1}A'}(x)$ completes the proof. ■

Remark 3.15 *The reader is encouraged to look closely at 3.3 and 3.11. Note that when converting from a Lebesgue-Stieltjes integral to a Lebesgue integral in 3.3, that the domain of integration $A \in \mathcal{B}(\mathbb{R})$ is not changed. The same is true of 3.6 for $A \in \sigma(X)$. For a general integral induced by a* *transformation T as in 3.11, the domain of integration in the range space must be transformed into its pre-image for the integral in the domain space.*

It is tempting to rewrite 3.11 in terms of μ-integrals over sets $A \in \sigma(X)$ in the domain space and μ_T-integrals over $T(A)$, as this initially appears to be merely a change of notation. But as will be seen in the example of the next section, while $T^{-1}A' \in \sigma(X)$ for all $A' \in \sigma(X')$, there may well exist $A \in \sigma(X)$ which cannot be so represented. This is because measurability of T means that the sigma algebra $T^{-1}[\sigma(X')] \subset \sigma(X)$, but it need not be the case that $T^{-1}[\sigma(X')] = \sigma(X)$. See example 3.187 below.

Given this remark, the above result provides the following corollary.

Corollary 3.16 *A measurable function $g(x')$ is μ_T-integrable if and only if $g(Tx)$ is μ-integrable, and when integrable,*

$$\int_A g(Tx)d\mu = \int_{T(A)} g(x')d\mu_T \tag{3.12}$$

for any $A \in T^{-1}[\sigma(X')]$.

Proof. *If $A \in T^{-1}[\sigma(X')]$, then by definition $A = T^{-1}[A']$ for some $A' \in \sigma(X')$.* ■

Example 3.17 (Expectations of Random Variables 2) *If X is a random variable defined on a probability space, $X : (\mathcal{S},\mathcal{E},\lambda) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, then as noted in example 3.7 above, the induced measure λ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which takes the place of μ_T in the current notation, is the Borel measure $\lambda_X \equiv \mu_F$ where F is the distribution function of X. Hence if $g(x)$ is a measurable function defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, or equivalently $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_F)$ since measurability on depends on the sigma algebra $\mathcal{B}(\mathbb{R})$, 3.10 becomes:*

$$\int_{\mathcal{S}} g(X)d\lambda = \int_{\mathbb{R}} g(x)d\mu_F.$$

Thus the "expectation" of $g(X)$, defined in 3.7 of book 4 as a λ-integral on $\mathcal{S}$ and denoted $E[g(x)]$, can be evaluated as a Lebesgue-Stieltjes integral on $\mathbb{R}$ with the Borel measure μ_F induced by the distribution function F. If F has an associated density function f, this Lebesgue-Stieltjes integral can then be evaluated as discussed in example 3.7. These transformations of $E[g(x)]$, initially defined as an integral on $(\mathcal{S},\mathcal{E},\lambda)$, were introduced in section 3.1.2 of book 4.

3.3 Special Cases of Change of Variables

3.3.1 Method of Substitution in Riemann and Lebesgue Integrals on $\mathbb{R}$

We first revisit example 3.1 above, applying the results from propositions 3.6 and 3.14 to provide a measure-theoretic framework for this familiar procedure.

Example 3.18 (Continuation) 1. ***False Start:*** *Recall that the goal was to evaluate* $\int_0^\infty x\exp(-x^2)dx$ *as a Riemann integral. By proposition 2.64 of book 3, since* $x\exp(-x^2)$ *is absolutely Riemann integrable it is also Lebesgue integrable and the integrals agree. Thus this can be considered as an exercise to evaluate* $(\mathcal{L})\int_0^\infty x\exp(-x^2)dx$. *The "method of substitution" or "change of variables approach" begins by defining a new variable:* $y \equiv x^2$. *This substitution can be interpreted as a continuous and thus measurable transformation:*

$$T : (\mathbb{R}, \mathcal{B}(\mathbb{R}), m) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), m_T),$$

where $Tx = x^2$, $\mathbb{R}^+ = [0, \infty)$ *and* $\mathcal{B}(\mathbb{R}^+) \equiv \mathcal{B}(\mathbb{R}) \cap [0, \infty)$.

The induced measure in the range space m_T *is given by 3.9 and we first investigate* m_T *on the semi-algebra of right semi-closed intervals,* $\mathcal{A}' \subset \mathcal{B}(\mathbb{R}^+)$, *which includes* $\{[0, b]\}$ *for* $b > 0$. *Given* $(a, b] \in \mathcal{A}'$, *since* $T^{-1}(a, b] = (\sqrt{a}, \sqrt{b}] \cup (-\sqrt{b}, -\sqrt{a}]$ *we obtain by 3.9:*

$$\begin{aligned} m_T((a, b]) &\equiv m\left(T^{-1}(a, b]\right) \\ &= 2\left(\sqrt{b} - \sqrt{a}\right), \end{aligned}$$

and this formula also works for $[0, b]$. *Diagraming in the notation of the above proposition,*

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), m) \quad \xrightarrow{T} \quad (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), m_T) \quad \xrightarrow{g} \quad (\mathbb{R}, \mathcal{B}(\mathbb{R}), m),$$

where g *is defined so that* $g(T(x)) = x\exp(-x^2)$.

Hence as a function on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), m_T)$, *the range space of* T :

$$g(x') = \sqrt{x'}\exp(-x'),$$

and 3.11 becomes:

$$\int_{T^{-1}A'} x\exp(-x^2)dm = \int_{A'} \sqrt{x'}\exp(-x')dm_T.$$

Unfortunately there is no $A' \in \mathcal{B}(\mathbb{R}^+)$ *with* $T^{-1}A' = [0,\infty)$*. More generally, this is an example noted in remark 3.15 for which* $T^{-1}\left[\mathcal{B}(\mathbb{R}^+)\right] \subsetneqq \mathcal{B}(\mathbb{R})$*.*

2. ***Corrected Attempt:*** *By redefining the domain of* T *:*

$$T : (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), m) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), m_T),$$

now $m_T((a,b]) = \sqrt{b} - \sqrt{a}$ *for* $(a,b] \in \mathcal{A}'$*, and applying 3.11 with* $A' = [0,\infty)$*:*

$$\int_0^\infty x \exp(-x^2) dm = \int_0^\infty \sqrt{x'} \exp(-x') dm_T.$$

The integrand on the right seems equally intractable as that on the left, in fact perhaps more so since a somewhat familiar Lebesgue integral has been replaced by a Lebesgue-Stieltjes integral with Borel measure m_T*. But contemplating this Borel measure, it is verified that it is of the special type in 3.1 and addressed by proposition 3.6. Namely, with* $f(x) = 0.5x^{-1/2}$ *for* $x > 0$ *and* 0 *otherwise, we have that for any set in* $\mathcal{A}'$*,*

$$m_T((a,b]) = (\mathcal{L}) \int_a^b f(x) dx,$$

and similarly for $m_T([0,b])$*.*

Hence, proposition 3.6 allows the above Lebesgue-Stieltjes integral to be restated in terms of a Lebesgue integral in which the original integrand is multiplied by $f(x)$*. The domain of integration is kept the same as seen in 3.3:*

$$\int_0^\infty \sqrt{x'} \exp(-x') dm_T = (\mathcal{L}) \int_0^\infty 0.5 \exp(-x') dx'.$$

Putting these results together, we have with the benefit of two measure-theoretic change of variable formulas, that:

$$(\mathcal{L}) \int_0^\infty x \exp(-x^2) dx = (\mathcal{L}) \int_0^\infty (0.5) \exp(-x') dx'.$$

As both $x \exp(-x^2)$ *and* $0.5 \exp(-x')$ *are continuous and Lebesgue integrable, these integrals again equal the corresponding Riemann integrals. Hence:*

$$(\mathcal{R}) \int_0^\infty x \exp(-x^2) dx = (\mathcal{R}) \int_0^\infty (0.5) \exp(-x') dx'.$$

The following proposition generalizes this example. It requires the assumption of absolute continuity for g^{-1}, a notion introduced in definition 3.54 of book 3. The details of the proof are assigned as an exercise in applying the approach of the above example.

Proposition 3.19 *Let $f(x)$ be a Lebesgue integrable function on a finite or infinite interval $I = [a, b]$. Then given a differentiable function $g : \mathbb{R} \to \mathbb{R}$ with $g'(x) \neq 0$ and g^{-1} absolutely continuous, then for $(a, b] \subset Rng[g]$, the range of g :*

$$(\mathcal{L}) \int_a^b f(y)dy = (\mathcal{L}) \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(x))g'(x)dx. \tag{3.13}$$

Proof. *Given such g, define the continuous and thus measurable transformation $g : (\mathbb{R}, \mathcal{B}(\mathbb{R}), m) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$. Then with $y = g(x)$ it follows from 3.11 that*

$$(\mathcal{L}) \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(x))g'(x)dx = \int_a^b f(y)g'(g^{-1}(y))dm_g.$$

The induced measure m_g is defined as in 3.9 by $m_g((a, b]) = g^{-1}(b) - g^{-1}(a)$ since g is continuous and $(a, b] \subset Rng[g]$. By absolute continuity and the fundamental theorem in proposition 3.61 of book 3:

$$m_g((a, b]) = \int_a^b \frac{dg^{-1}}{dy} dy.$$

But $\frac{dg^{-1}}{dy} = 1/g'(g^{-1}(y))$, and since $g'(x) \neq 0$ 3.13 now follows from 3.3. ■

Remark 3.20 *If f is continuous and Riemann integrable, and also absolutely integrable when the interval is infinite, the above proposition provides a Riemann integration result if $g'(x)$ is continuous. This follows from propositions 2.31 and 2.64 of book 3.*

3.3.2 Change of Variables for Linear Transformations on $\mathbb{R}^n$

In this section, we provide a detailed result in the special case where T is an invertible linear transformation, $T : \mathbb{R}^n \to \mathbb{R}^n$. This result will be generalized to continuously differentiable and invertible transformations on $\mathbb{R}^n$ in the coming section, Change of Variables for Differentiable Transformations on $\mathbb{R}^n$.

Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be an invertible linear transformation and denote the standard matrix representation of this transformation as $A = (a_{ij})_{i,j=1}^n$:

$$T(x) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

The "vector" x is conventionally identified with a column vector in linear algebra, and thus with $x^t \equiv (x_1, x_2, ..., x_n)$ denoting the the **transpose** of x as a row vector:

$$T(x) = Ax^t \equiv \left(\sum\nolimits_{j=1}^n a_{1j}x_j, \sum\nolimits_{j=1}^n a_{2j}x_j, \cdots, \sum\nolimits_{j=1}^n a_{nj}x_j \right)^t. \tag{3.14}$$

Notation 3.21 *For most expressions involving vectors, there is no reason to "identify" vectors as column vectors. In other words, stating that $x^t \equiv (x_1, x_2, ..., x_n)$ or $x \equiv (x_1, x_2, ..., x_n)$ conveys exactly the same information, and it is common to drop the transpose designation unless required. For expressions involving matrices this notation is considered essential. If A is an $n \times n$ matrix, then xAx is not defined although it could be argued that this expression is uniquely defined as x^tAx. On the other hand, xx is also not defined, nor is there a unique interpretation since both x^tx and xx^t are well defined and very different.*

Define a Lebesgue measure μ on $\mathcal{B}(\mathbb{R}^n)$ or on the complete Lebesgue sigma algebra $\mathcal{M}_L^n(\mathbb{R}^n)$ by:

$$\mu(B) = |\det(A)|\, m^n(B). \tag{3.15}$$

Here m^n is Lebesgue measure and $\det(A)$ is the determinant of the matrix induced by T. Note that $\det(A) \neq 0$ since this matrix is invertible. Next, consider the transformation between measure spaces:

$$T : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_T),$$

where μ_T is the measure induced by T. That T is a measurable transformation follows by noting that both T and T^{-1} are continuous, and hence both $T(G)$ and $T^{-1}(G)$ are open sets for all open G. Since $\mathcal{B}(\mathbb{R}^n)$ is the smallest sigma algebra that contains the open sets, measurability follows.

Proposition 3.22 *The measure induced by T in 3.14 satisfies: $\mu_T = m^n$, Lebesgue measure.*
Proof. *For $B \in \mathcal{B}(\mathbb{R}^n)$:*

$$\mu_T(B) \equiv \mu\left[T^{-1}(B)\right] \equiv |\det(A)|\, m^n(T^{-1}(B)).$$

If $B = \prod_{j=1}^n (0, b_j]$ is a right semi-closed rectangle and $B_j^t = (0, ..., 0, b_j, 0, ..., 0)) \in \mathbb{R}^n$ with jth component equal to $b_j > 0$, then $B = \left\{\sum_{j=1}^n \lambda_j B_j | 0 < \lambda_j \leq 1\right\}$. Because T^{-1} is also a linear transformation:

$$T^{-1}(B) = \left\{\sum\nolimits_{j=1}^n \lambda_j A^{-1}(B_j) | 0 < \lambda_j \leq 1\right\}.$$

Invertibility of T implies that $\{A^{-1}(B_j)\}_{j=1}^n$ are linearly independent so $T^{-1}(B)$ is a parallelepiped in $\mathbb{R}^n$ and thus its Lebesgue measure equals ordinary Euclidean volume. Since A^{-1} is a linear transformation, the Euclidean volume of a linearly transformed rectangle satisfies:

$$m^n(T^{-1}(B)) = \left|\det(A^{-1})\right| m^n(B).$$

Finally, $\left|\det(A^{-1})\right| = |\det(A)|^{-1}$, and the result follows.

The identity that $\mu_T = m^n$ then applies to the semi-algebra $\mathcal{A}'$ of all right semi-closed rectangles, $\left\{\prod_{j=1}^n (a_j, b_j]\right\}$, since given B of this form, linearity of T^{-1} obtains:

$$T^{-1}(B) = \left\{A^{-1}(a_1, a_2, ..., a_n)^t + \sum\nolimits_{j=1}^n \lambda_j A^{-1}(\tilde{B}_j^t) | 0 < \lambda_j \leq 1\right\},$$

where $\tilde{B}_j^t = (0, ..., 0, b_j - a_j, 0, ..., 0))$. Now since m^n is translation invariant, again $\mu_T(B) = m^n(B)$.

The transition of this identity from this semi-algebra to the algebra $\mathcal{A}$ of all disjoint unions of $\mathcal{A}'$-sets and then to the sigma algebra $\mathcal{B}(\mathbb{R}^n)$ follows by the extension process and uniqueness result developed in section 6.2 of book 1. ■

Now since $\mu_T = m^n$ we have from proposition 3.14 the following proposition and corollary. These results will be generalized below to continuously differentiable invertible transformations $T : \mathbb{R}^n \to \mathbb{R}^n$, where the absolute value of the **Jacobian determinant of** T will take the place of $|\det(A)|$ below.

Remark 3.23 *Recalling remark 3.4, the following proposition is restricted to Borel measurable g, and does not attempt to generalize to Lebesgue measurable functions. The reason for this echoes the discussion surrounding proposition 3.33 of book 1 on the measurability of composite functions, While originally presented as a discussion on $\mathbb{R}$, it is easy to appreciate that the discussion does not simplify in $\mathbb{R}^n$.*

Looking to 3.16, since both are Lebesgue integrals one needs to assure that the integrands are appropriately measurable. If g is Borel measurable, then since T is continuous it follows that $g(T)$ is Borel measurable and both integrals are definable. If g is merely Lebesgue measurable, then Lebesgue measurability of $g(T)$ requires that $T^{-1}\left(\mathcal{M}_L^n(\mathbb{R}^n)\right) \subset \mathcal{M}_L^n(\mathbb{R}^n)$. For the current linear T, this seems like a reasonable result though it appears quite challenging to prove. In any event it is not worth the effort since for the next result on continuously differentiable T such a conclusion would certainly appear elusive, if even true.

Hence in this section and the next we focus on Borel measurable g, though the domain of integration can be more generally defined. But note that for corollary 3.25, the domain of integration is again restricted to $B \in \mathcal{B}(\mathbb{R}^n)$. The reader is invited to determine why this proof does not readily apply to $B \in \mathcal{M}_L^n(\mathbb{R}^n)$.

Proposition 3.24 *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation, and $|\det(A)|$ the absolute value of the determinant of the matrix implied by T. Then for any nonnegative Borel measurable function $g : \mathbb{R}^n \to \mathbb{R}$:*

$$\int_{\mathbb{R}^n} g(Tx)\,|\det(A)|\,dm^n = \int_{\mathbb{R}^n} g(y)dm^n, \tag{3.16}$$

though both integrals may be infinite.

More generally, a Borel measurable function g is Lebesgue integrable if and only if $g(Tx)\,|\det(A)|$ is Lebesgue integrable, and when integrable:

$$\int_{T^{-1}B} g(Tx)\,|\det(A)|\,dm^n = \int_B g(y)dm^n \tag{3.17}$$

for all $B \in \mathcal{M}_L^n(\mathbb{R}^n)$.

Proof. As $\mu \equiv |\det(A)|\,m^n$ and the induced $\mu_T = m^n$, the result follows from proposition 3.14 by substitution. ∎

Corollary 3.25 *If $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, then for all $B \in \mathcal{B}(\mathbb{R}^n)$:*

$$m^n\left[T\left(B\right)\right] = |\det(A)|\,m^n\left[B\right]. \tag{3.18}$$

Proof. *Letting* $g(x) = 1$, *3.17 implies that for any* $B' \in \mathcal{B}(\mathbb{R}^n)$, $m^n[B'] = |\det(A)|\, m^n\left[T^{-1}(B')\right]$. *As* T *and* T^{-1} *are continuous,* $B' \in \mathcal{B}(\mathbb{R}^n)$ *if and only if* $T^{-1}(B') \in \mathcal{B}(\mathbb{R}^n)$, *so given* B, *let* $B' = T(B)$ *in this identity.* ■

Remark 3.26 *It may seem odd that in contrast to the expression in 3.17, that* $g'(x)$ *in 3.13 appears without absolute values since* $g(x) = ax$ *is a linear transformation on* $\mathbb{R}$ *to which either result applies. In 3.17 we would have* $|\det(A)| = |a|$, *while in 3.13,* $g'(x) = a$. *Below, we generalize 3.17 to continuously differentiable transformations and again an absolute value will be seen in the term that generalizes the role of* $\det(T)$, *so the same problem is apparently observed for general* $g(x)$.

The explanation for this is that over any interval $[c, d]$ *for which* $g'(x) < 0$, *it would follow that* $g^{-1}(d) < g^{-1}(c)$ *and hence:*

$$\int_{g^{-1}(c)}^{g^{-1}(d)} f(g(x))g'(x)dx = \int_{g^{-1}(d)}^{g^{-1}(c)} f(g(x))\left|g'(x)\right| dx = \int_{g^{-1}([c,d])} f(g(x))\left|g'(x)\right| dx.$$

For subintervals on which $g'(x) > 0$, $g^{-1}(c) < g^{-1}(d)$ *and so*

$$\int_{g^{-1}(c)}^{g^{-1}(d)} f(g(x))g'(x)dx = \int_{g^{-1}(c)}^{g^{-1}(d)} f(g(x))\left|g'(x)\right| dx = \int_{g^{-1}([c,d])} f(g(x))\left|g'(x)\right| dx.$$

Hence in any case, integrating with $|g'(x)|$ *over the interval* $g^{-1}([c, d])$ *as in 3.17 produces the same result as integrating with* $g'(x)$ *from* $g^{-1}(c)$ *to* $g^{-1}(d)$ *as in 3.13.*

3.3.3 Change of Variables for Differentiable Transformations on $\mathbb{R}^n$

In this section we generalize 3.13 and 3.17 to multivariate Lebesgue integrals and invertible, continuously differentiable transformations. Recall the set-up in the section General Change of Variables Under Transformations, here modified to:

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu) \xrightarrow{T} (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), m^n) \xrightarrow{g} (\mathbb{R}, \mathcal{B}(\mathbb{R}), m).$$

Thus T is a transformation between an n-dimensional Borel measure space with measure μ to be determined below, and the Lebesgue measure space.

For $x = (x_1, x_2, ..., x_n)$,

$$T(x) = (t_1(x), t_2(x), \cdots, t_n(x)),$$

where $t_i : \mathbb{R}^n \rightarrow \mathbb{R}$ for all i. We will say that T is continuous, differentiable, continuously differentiable, etc., if the component functions $\{t_i\}_{i=1}^n$ have such properties. As noted above, $T(x)$ is by convention deemed to be a column vector in $\mathbb{R}^n$ and is therefore expressed as the transpose of the given row vector. When this transformation has differentiable component functions, the **Jacobian matrix** associated with T, denoted $T'(x)$, is defined as:

$$T'(x) \equiv \begin{pmatrix} \frac{\partial t_1}{\partial x_1} & \frac{\partial t_1}{\partial x_2} & \cdots & \frac{\partial t_1}{\partial x_n} \\ \frac{\partial t_2}{\partial x_1} & \frac{\partial t_2}{\partial x_2} & \cdots & \frac{\partial t_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial t_n}{\partial x_1} & \frac{\partial t_n}{\partial x_2} & \cdots & \frac{\partial t_n}{\partial x_n} \end{pmatrix}. \tag{3.19}$$

The **Jacobian determinant** associated with T, denoted $\det(T'(x))$, is defined as the determinant of $T'(x)$. This matrix and its determinant are named for **Carl Gustav Jacob Jacobi** (1804 – 1851), an early developer of determinants and their applications in analysis.

Notation 3.27 *The Jacobian matrix is denoted in many ways:*

$$T'(x) \equiv \frac{\partial(t_1, t_2, ..., t_n)}{\partial(x_1, x_2, ..., x_n)} = \left(\frac{\partial(t_1, t_2, ..., t_n)}{\partial(x_1, x_2, ..., x_n)}\right) = \left(\frac{\partial T}{\partial x}\right),$$

and the Jacobian determinant

$$\det T'(x) \equiv \left|\frac{\partial(t_1, t_2, ..., t_n)}{\partial(x_1, x_2, ..., x_n)}\right|.$$

Absolute values are commonly used with a matrix to denote determinant, but it can be ambiguous in the context of the current development where we will want to use the absolute value of the Jacobian determinant, $|\det T'(x)|$, *so we will avoid this alternative notation.*

Remark 3.28 *We do not review multivariate calculus in any detail in this book, but there are two key properties of the Jacobian matrix which are used below. The reader can find such results in Edwards (1973) in the references and elsewhere.*

1. ***Linear Approximations:*** *If* $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ *is continuously differentiable at* x, *then*

$$T(y) = T(x) + T'(x)(y - x) + o(|y - x|),$$

where $|y-x|$ is the standard Euclidean norm on $\mathbb{R}^n$. In the notation of the next chapter, this is also called the l_2-norm of $x \equiv (x_1, x_2, ..., x_n)$ and defined by:

$$|(x_1, x_2, ..., x_n)| = \sqrt{\sum_{j=1}^{n} x_j^2}. \tag{3.20}$$

*The "**little** o" error term means that*

$$\frac{o(|y-x|)}{|y-x|} \to 0 \text{ as } |y-x| \to 0.$$

Thus the Jacobian matrix serves the same role in linearly approximating T as does $f'(x)$ in the usual Taylor series approximation for continuously differentiable functions:

$$f(y) = f(x) + f'(x)(y-x) + o(|y-x|),$$

where here the absolute value norm on $\mathbb{R}$ is identical with the l_2-norm defined above.

As it turns out, the error in the linear approximation to T can be bounded in terms of the errors in the component functions, $\{t_i(x)\}_{i=1}^n$. The next proposition identifies this connection and will be applied below. For notational simplicity, we denote $\frac{\partial t_i}{\partial x_j}$ by $\partial_j t_i$.

Proposition 3.29 *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable transformation with $T(x) = (t_1(x), t_2(x), \cdots, t_n(x))$, and $A = \prod_{i=1}^n [a_i, b_i]$ a closed rectangle in $\mathbb{R}^n$. If for all i, j:*

$$|\partial_j t_i(y) - \partial_j t_i(x)| \leq \beta$$

for all $x, y \in A$, then

$$|T(y) - T(x) - T'(x)(y-x)| \leq \beta n^2 |y-x| \tag{3.21}$$

for all $x, y \in A$.

Proof. *First note that if $x \equiv (x_1, x_2, ..., x_n)$, then $|x| \leq \sum_{j=1}^n |x_i|$ as follows by squaring both sides. Here $|x|$ is defined as in 3.20 and $|x_i|$ is equivalently defined in terms of this norm or simply as an absolute value. Thus:*

$$|T(y) - T(x) - T'(x)(y-x)| \leq \sum_{i=1}^{n} \left| t_i(y) - t_i(x) - \sum_{j=1}^{n} \partial_j t_i(x)(y_j - x_j) \right|.$$

Each t_i is a continuously differentiable function, so the mean value theorem obtains that for some $0 < \lambda_i < 1$:

$$t_i(y) - t_i(x) = \sum\nolimits_{j=1}^{n} \partial_j t_i(x + \lambda_i(y-x))(y_j - x_j).$$

Then:

$$\begin{aligned}&\left|T(y) - T(x) - T'(x)(y-x)\right| \\ \leq\ & \sum\nolimits_{i=1}^{n} \left|\sum\nolimits_{j=1}^{n} \left[\partial_j t_i(x + \lambda_i(y-x)) - \partial_j t_i(x)\right](y_j - x_j)\right| \\ \leq\ & |y-x| \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{n} \left|\partial_j t_i(x + \lambda_i(y-x)) - \partial_j t_i(x)\right|,\end{aligned}$$

since $|y_j - x_j| \leq |y-x|$ for all j. The result now follows because if $x, y \in A$, then $x + \lambda_i(y-x) \in A$. ■

Remark 3.30 *Note that while 3.21 implies that the error term $T(y) - T(x) - T'(x)(y-x)$ is only bounded by $O(|y-x|)$, meaning that*

$$\frac{O(|y-x|)}{|y-x|} \rightarrow C \text{ as } |y-x| \rightarrow 0,$$

it is in fact $o(|y-x|)$ as noted above since by continuity of $\partial_j t_i$ we have that $\beta \rightarrow 0$ as $|y-x| \rightarrow 0$.

2. ***Inverse Function Theorem:*** *If T is continuously differentiable at x_0 and $\det T'(x_0) \neq 0$, there are open sets U with $x_0 \in U$ and V with $T(x_0) \in V$ so that $T^{-1} : V \rightarrow U$ is continuously differentiable with*

$$\left(T^{-1}\right)'(v) = \left(T'\left[T^{-1}(v)\right]\right)^{-1}.$$

In words, the Jacobian matrix of T^{-1} at v is the inverse of the Jacobian matrix of T at $T^{-1}(v)$.

Remark 3.31 *This formula generalizes the one variable result whereby if $f'(x)$ is continuously differentiable and $f'(x_0) \neq 0$, then there are open intervals U with $x_0 \in U$ and V with $f(x_0) \in V$ so that $f^{-1} : V \rightarrow U$ is continuously differentiable with*

$$\left(f^{-1}\right)'(v) = 1/f'\left[f^{-1}(v)\right].$$

Generalizing the approach in 3.15 where T was a linear transformation, we define a measure μ on $\mathcal{B}(\mathbb{R}^n)$ or $\mathcal{M}_L^n(\mathbb{R}^n)$ by:

$$\mu(B) = \int_B \left|\det(T'(x))\right| dm^n, \tag{3.22}$$

where m^n is Lebesgue measure and $\det(T'(x))$ is the Jacobian determinant of T. When T is continuously differentiable this integral is well-defined and finite for bounded B, since continuous $|\det(T'(x))|$ is bounded on $\bar{B}$ the closure of B. More generally:

Proposition 3.32 *If T is a continuously differentiable transformation on $\mathbb{R}^n$, then μ in 3.22 defines a measure on $\mathcal{B}(\mathbb{R}^n)$ or $\mathcal{M}_L^n(\mathbb{R}^n)$.*
Proof. *This is a corollary to proposition 3.3 and remark 3.4 since $|\det(T'(x))|$ is nonnegative, and the assumption of continuity assures Lebesgue measurability.* ∎

This definition of μ generalizes that in 3.15 for a linear transformation. If $x = (x_1, x_2, ..., x_n)$ and T is a linear transformation as in 3.14, then $T'(x)$ is constant matrix given by the matrix of coefficients:

$$T'(x) = A \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Hence for linear T 3.22 becomes:

$$\mu(B) = |\det(A)|\, m^n(B),$$

which is 3.15.

We first restate proposition 3.6 and the results of 3.2 and 3.3 in the current context, recalling remark 3.4.

Proposition 3.33 *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable, one-to-one, and satisfy $\det(T'(x)) \neq 0$ for all x. With μ as in 3.22, then for any nonnegative Lebesgue measurable function $g(x)$ defined on $\mathbb{R}^n$:*

$$\int_{\mathbb{R}^n} g(x) d\mu = \int_{\mathbb{R}^n} g(x) \left|\det(T'(x))\right| dm^n, \tag{3.23}$$

although both integrals may be infinite.

In addition, Lebesgue measurable $g(x)$ is μ-integrable if and only if $g(x)\,|\det(T'(x))|$ is Lebesgue integrable, and when integrable,

$$\int_A g(x)d\mu = \int_A g(x)\left|\det(T'(x))\right| dm^n, \tag{3.24}$$

for any $A \in \mathcal{M}_L^n(\mathbb{R}^n)$.

Proof. *This is a corollary to proposition 3.3 and remark 3.4, since $f(x) \equiv |\det(T'(x))|$ is nonnegative and Lebesgue measurable.* ■

The following result on change of variables in Lebesgue integrals is technically a corollary of the general results in proposition 3.14 using the above proposition 3.33. But to apply the earlier results we must demonstrate the following. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable, one-to-one transformation with $\det(T'(x)) \neq 0$, then:

$$\mu_T(A) = m^n(A)$$

for all $A \in \mathcal{M}_L^n(\mathbb{R}^n)$. Here μ_T denotes the measure on the range space induced by T, where μ is defined as in 3.22 on the domain space. The demonstration of this identity is the essence of the following rather long proof.

That the following result is restricted to Borel measurable g, see remark 3.4.

Proposition 3.34 *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable, one-to-one transformation with $\det(T'(x)) \neq 0$ for all x. Then for any nonnegative Borel measurable function g:*

$$\int_{\mathbb{R}^n} g(Tx)\left|\det(T'(x))\right| dm^n = \int_{\mathbb{R}^n} g(y)dm^n, \tag{3.25}$$

although both integrals may be infinite.

More generally, a Borel measurable function g is Lebesgue integrable if and only if $g(Tx)\,|\det(T'(x))|$ is Lebesgue integrable, and when integrable,

$$\int_{T^{-1}A} g(Tx)\left|\det(T'(x))\right| dm^n = \int_A g(y)dm^n \tag{3.26}$$

for all $A \in \mathcal{M}_L^n(\mathbb{R}^n)$.

Proof. *As noted above, the statement of the result in 3.25 is exactly 3.10 if $\mu_T = m^n$ once 3.23 is applied to $g(T)$. In detail, applying 3.23 to $g(Tx)$, then 3.10 yields:*

$$\int_{\mathbb{R}^n} g(Tx)\left|\det(T'(x))\right| dm^n = \int_{\mathbb{R}^n} g(Tx)d\mu = \int_{\mathbb{R}^n} g(y)d\mu_T,$$

which is 3.25 if $\mu_T = m^n$. Similarly, 3.26 is exactly 3.11 and 3.24:

$$\int_{T^{-1}A} g(Tx)\left|\det(T'(x))\right| dm^n = \int_{T^{-1}A} g(Tx)d\mu = \int_A g(y)d\mu_T.$$

Turning to the proof of $\mu_T = m^n$, continuity of T obtains that $T^{-1}(G)$ is open for all open G, while the inverse mapping theorem yields that $T(G)$ is open for all open G. This then implies that $\mathcal{B}(\mathbb{R}^n) = T\left[\mathcal{B}(\mathbb{R}^n)\right] = T^{-1}\left[\mathcal{B}(\mathbb{R}^n)\right].$ Thus the definition $\mu_T(A) \equiv \mu\left[T^{-1}(A)\right]$ for all $A \in \mathcal{B}(\mathbb{R}^n)$ is equivalent to $\mu_T\left[T(A)\right] \equiv \mu\left[A\right]$ so for all $A \in \mathcal{B}(\mathbb{R}^n)$ because T is one-to-one and so $\left(T^{-1}\right)^{-1} = T$. To prove $\mu_T = m^n$, it is therefore sufficient to prove that for all $A \in \mathcal{B}(\mathbb{R}^n)$:

$$m^n\left[T(A)\right] = \int_A |J(x)|\, dm^n, \tag{(*)}$$

where $J(x) \equiv \det(T'(x))$ to simplify notation.

For this it is enough to restrict A to the semi-algebra of right semi-closed rectangles, $\mathcal{A}' \equiv \left\{\prod_{j=1}^n (a_j, b_j]\right\}$, since by the extension process of chapter 6 of book 1, this equality then applies to all $A \in \mathcal{B}(\mathbb{R}^n)$. This extension process is applicable because the set functions:

$$\mu_1(A) \equiv m^n\left[T(A)\right], \qquad \mu_2(A) \equiv \int_A |J(x)|\, dm^n,$$

are each countably additive on $\mathcal{A}'$. In the first case T is one-to-one, while in the second, 2.26 applies. By proposition 6.13 of book 1 both set functions can be extended to measures on the algebra $\mathcal{A}$ defined as the collection of all finite disjoint unions of $\mathcal{A}'$-sets, and looking to the proof reveals that then $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{A}$. Proposition 6.14 of book 1 now applies since both measures are σ-finite, and thus $\mu_1(A) = \mu_2(A)$ for all A in the smallest sigma algebra that contains $\mathcal{A}$, and this is $\mathcal{B}(\mathbb{R}^n)$.

Finally, we can also restrict the proof of $(*)$ to such rectangles $A \in \mathcal{A}'$ that are bounded since every $A \in \mathcal{A}'$ is at most a countable union of disjoint bounded rectangles, and thus the bounded case implies the general case by countable additivity.

So assume $A = \prod_{j=1}^{n}(a_j, b_j]$ *is bounded and hence the closure* $\overline{A} = \prod_{j=1}^{n}[a_j, b_j]$ *is compact. As* $J(x)$ *is continuous on* $\overline{A}$ *it is uniformly continuous, so given* $\epsilon > 0$ *there is a* δ_1 *such that for* $x, y \in \overline{A}$ *with* $|x - y| < \delta_1$ *we have that* $|J(x) - J(y)| < \epsilon$. *Partition the intervals* $(a_j, b_j]$ *so that* $A = \bigcup_{k=1}^{N} R_k$ *is a finite disjoint union of rectangles with the property that if* $x, y \in R_k$ *then* $|x - y| < \delta_2 \leq \delta_1$, *such* δ_2 *now being arbitrary but will be specified below. Thus for* $x, y \in R_k$, $|J(x) - J(y)| < \epsilon$ *implies that* $||J(x)| - |J(y)|| < \epsilon$, *and so:*

$$|J(y)| - \epsilon < |J(x)| < |J(y)| + \epsilon.$$

Arbitrarily choose $x = x_k \in R_k$ *and integrate this inequality over* R_k *to obtain:*

$$\int_{R_k} |J(y)|\, dm^n - \epsilon m^n [R_k] < |J(x_k)|\, m^n [R_k] < \int_{R_k} |J(y)|\, dm^n + \epsilon m^n [R_k] .$$

Note for later that because $\{R_k\}$ *is a finite disjoint collection of rectangles that this same inequality is proved for* A *in place of* R_k :

$$\int_{A} |J(y)|\, dm^n - \epsilon m^n [A] < \sum_{k=1}^{N} |J(x_k)|\, m^n [R_k] < \int_{A} |J(y)|\, dm^n + \epsilon m^n [A] , \tag{(**)}$$

where $x_k \in R_k$ *are arbitrary, but will be fixed next.*

The last step is to show that the middle terms in these inequalities, each $|J(x_k)|\, m^n [R_k]$, *can be made arbitrarily close to* $m^n [T(R_k)]$. *To this end, fix* $x_k \in R_k$ *as the "midpoint" of any such rectangle, meaning that if* $R_k = \prod_{j=1}^{n}(c_j, d_j]$ *then* $x_{k_j} = (c_j + d_j)/2$. *For* $\epsilon > 0$ *above, and we can assume* $\epsilon < 1$, *then define* $R_k^{+\epsilon}$ *and* $R_k^{-\epsilon}$ *to be rectangles centered on* x_k *but with respective sides of length* $(1 + \epsilon)(d_j - c_j)$ *and* $(1 - \epsilon)(d_j - c_j)$. *We now show that given such* ϵ *there exists* $\delta_2 \leq \delta_1$ *so that:*

$$T(x_k) + T'(x_k) \left[R_k^{-\epsilon} - x_k\right] \subset T(R_k) \subset T(x_k) + T'(x_k) \left[R_k^{+\epsilon} - x_k\right] , \tag{***}$$

where $R_k^{\pm\epsilon} - x_k$ *denotes the respective rectangles translated to be centered at the origin. Since Lebesgue measure is translation invariant, these inclusions assure inequalities for* $m^n [T(R_k)]$ *and* $m^n [A]$ *as follows.*

Recalling 3.18 in corollary 3.25, and noting that $m^n \left[R_k^{\pm\epsilon}\right] = (1 \pm \epsilon)^n m^n [R_k]$, *the set inclusions in* $(***)$ *imply that:*

$$(1 - \epsilon)^n |J(x_k)|\, m^n [R_k] \leq m^n [T(R_k)] \leq (1 + \epsilon)^n |J(x_k)|\, m^n [R_k] ,$$

and since T is one-to-one this obtains by summation:

$(1-\epsilon)^n \sum_{k=1}^{N} |J(x_k)|\, m^n [R_k] \leq m^n [T(A)] \leq (1+\epsilon)^n \sum_{k=1}^{N} |J(x_k)|\, m^n [R_k]$.

Combining with the inequalities in $(**)$ produces:

$$(1-\epsilon)^n \left[\int_A |J(y)|\, dm^n - \epsilon m^n [A]\right] \leq m^n [T(A)] \leq (1+\epsilon)^n \left[\int_A |J(y)|\, dm^n + \epsilon m^n [A]\right],$$

and $(*)$ then follows because $\epsilon > 0$ is arbitrary.

To prove the set inclusions in $(***)$, note first that because $\partial_j t_i(x)$ is continuous for all i, j, and hence uniformly continuous on $\overline{A} = \prod_{k=1}^{n} [a_k, b_k]$, given ϵ above there exists $\delta_1' \leq \delta_1$ so that if $|y - x| < \delta_1'$ then for all i, j,

$$|\partial_j t_i(y) - \partial_j t_i(x)| \leq \epsilon / n^2.$$

By 3.21 we have that on every R_k defined as above relative to δ_1' that for $x, y \in R_k$:

$$\left|T(y) - T(x) - T'(x)(y - x)\right| \leq \epsilon\, |y - x|\,.$$

So letting $x = x_k$ constructed as above, we conclude that if $y \in R_k$, $T(y)$ is contained in a ball of radius $\epsilon\, |y - x_k|$ centered on $T(x_k) + T'(x_k)\, [y - x_k]$. Thus $T\,(y) \in T(x_k) + T'(x_k) \left[R_k^{+\epsilon} - x_k\right]$, proving the inclusion on the right.

To prove the inclusion on the left let $y \in R_k^{-\epsilon}$, and we show that for some $\delta_1'' \leq \delta_1$, the R_k-rectangles defined as above relative to δ_1'' have the property that $T(x_k) + T'(x_k)\, [y - x_k] \in T\,(R_k)$, and we do this by proving that $T^{-1}\, [T(x_k) + T'(x_k)\, [y - x_k]] \in R_k$. For notational convenience, denote $T^{-1}(u) = (s_1(u), s_2(u), \cdots, s_n(u))$. Since $\partial_j s_i(u)$ is continuous for all i, j by the inverse function theorem and hence uniformly continuous on the compact set $T\left[\overline{A}\right] \equiv T\left[\prod_{k=1}^{n} [a_k, b_k]\right]$. So given ϵ above there is a $\delta' \leq \delta_1$ such that if $|u - v| < \delta'$ then for all i, j :

$$|\partial_j s_i(u) - \partial_j s_i(v)| \leq \epsilon / M n^2,$$

where M is defined by:

$$M = \max\left[1, \max_{x \in \overline{A}, |y|=1} \left|T'(x) y\right|\right].$$

Note that M is finite because it equals the maximum of a continuous function on a compact set. By 3.21 we then have that if $|u - v| < \delta'$:

$$\left|T^{-1}(v) - T^{-1}(u) - \left[T^{-1}\right]'(u)(v - u)\right| \leq \frac{\epsilon}{M}\, |v - u|\,.$$

Now let $v = T(x_k) + T'(x_k)\,[y - x_k]$ *and* $u = T(x_k)$*. Note that* $|u - v| = |T'(x_k)\,[y - x_k]| \leq M\,|y - x_k|\,,$ *and so* $|u - v| < \delta'$ *if* $|y - x_k| < \delta_1'' \equiv \delta'/M$*. Now construct the* R_k*-rectangles relative to* δ_1''*, and we have that* $T^{-1}(u) = x_k$*,* $\left[T^{-1}\right]'(u) = [T'(x_k)]^{-1}$ *by the inverse function theorem, and so* $T^{-1}(u) + \left[T^{-1}\right]'(u)(v - u) = y$ *and thus*

$$\left|T^{-1}(v) - y\right| \leq \frac{\epsilon}{M}\left|T'(x_k)\,[y - x_k]\right| \leq \epsilon\,|y - x_k|\,.$$

Hence, $T^{-1}(v)$ *is within* $\epsilon\,[y - x_k]$ *of* y*, and since* $y \in R_k^{-\epsilon}$ *this demonstrates that* $T^{-1}\,[v] \in R_k$ *as was to be shown.*

The proof is complete by setting $\delta_2 \equiv \min(\delta_1', \delta_1'')$ *in the fifth paragraph above, defining the* R_k*-sets.* ■

Remark 3.35 **1**. *As justified in the above proof,* $\mathcal{B}(\mathbb{R}^n) = T^{-1}\left[\mathcal{B}(\mathbb{R}^n)\right],$ *and so 3.26 can be restated in the alternative version of 3.12, that for all* $A \in \mathcal{B}(\mathbb{R}^n)$*:*

$$\int_A g(Tx)\left|\det(T'(x))\right| dm^n = \int_{T(A)} g(y) dm^n. \tag{3.27}$$

2. *The above proposition also applies to* $T : U \to V$ *with open* $U, V \subset \mathbb{R}^n$ *and* $T(U) = V$ *when* T *is a continuously differentiable, one-to-one transformation with* $\det(T'(x)) \neq 0$ *for all* $x \in U$*. To prove this requires that the general results be applied to measure spaces as follows:*

$$(U, \mathcal{B}(U), \mu) \quad \xrightarrow{T} \quad (V, \mathcal{B}(V), m^n) \quad \xrightarrow{g} \quad (\mathbb{R}, \mathcal{B}(\mathbb{R}), m).$$

The essential observation that makes the proof of this local result work is that the inverse mapping theorem is fundamentally a local result.

Generalizing corollary 3.25, we have the following result,

Proposition 3.36 *Let* $T : \mathbb{R}^n \to \mathbb{R}^n$ *be a continuously differentiable, one-to-one transformation with* $\det(T'(x)) \neq 0$*. Then for any* $A \in \mathcal{B}(\mathbb{R}^n)$*:*

$$m^n[T(A)] = \int_A \left|\det(T'(x))\right| dm^n. \tag{3.28}$$

Proof. *The proof follows from 3.27 with* $g \equiv 1$*.* ■

Chapter 4

The L_p Spaces

In this chapter we introduce an important collection of function spaces called the L_p Spaces and sometimes denoted L^p Spaces. While of interest generally in the development of measure spaces with special properties, these spaces will be crucial in the later books on stochastic integration. The L^p spaces are examples of **Banach spaces,** named for **Stefan Banach** (1892 – 1945), who axiomatized this theory in the early 1920s. Other notable early contributors to this theory are **Frigyes Riesz** (1880 – 1956), **Ernst Fischer** (1875 – 1954), **Hans Hahn** (1879 – 1934), and **Norbert Wiener** (1894 – 1964).

See section 4.1.4 of book 3 for a different example of a Banach space, and other results on bounded linear functionals.

4.1 Introduction to Banach Spaces

We begin with a collection of definitions. While the notion of a vector space initially appears very abstract, it is important to keep in mind that this definition states abstractly all the operations one takes for granted in the familiar vector space of $\mathbb{R}^n$, which is formally a vector space over the real field $\mathbb{R}$. It is useful to formally translate this abstract definition to the familiar context of $\mathbb{R}^n$ to provide concreteness. Beyond the notion of vector space are the notions of norm, convergence of a sequence of points, and completeness, collectively providing a definition of Banach space. Again, $\mathbb{R}^n$ with Euclidean norm $\|x\| = \sqrt{\sum_{j=1}^n x_j^2}$ for $x = (x_1, ..., x_n)$ is the quintessential example of a Banach space.

Note that the field $\mathcal{F}$ in this definition is virtually always $\mathbb{R}$ or $\mathbb{C}$.

Definition 4.1 *A space V is called a* ***vector space over a field*** *$\mathcal{F}$ or a* ***linear space over a field*** *$\mathcal{F}$, if it is a collection of* ***vectors*** *or* ***points*** *on which is defined* ***vector addition*** *and* ***scalar multiplication*** *which satisfy:*

1. *(Vector addition) For all $x, y, z \in V$:*

(**a**) ***Closure:*** $x + y \in V$

(**b**) ***Commutativity***: $x + y = y + x$

(**c**) ***Associativity:*** $(x + y) + z = y + (x + z)$

(**d**) ***Zero Vector:*** *There is an element $\theta \in V$ so that $x + \theta = x$ for all $x \in V$. This element is often denoted* 0.

2. *(Scalar multiplication) For all $x, y \in V$, $\alpha, \beta \in \mathcal{F}$:*

(**a**) $\alpha x \in V$

(**b**) $\alpha (x + y) = \alpha x + \alpha y$

(**c**) $(\alpha + \beta) x = \alpha x + \beta x$

(**d**) $\alpha (\beta x) = (\alpha\beta) x$

3. *For all $x \in V$:*

(**a**) $0x = \theta$, *where "0" denotes the additive unit in $\mathcal{F}$, meaning* $0+\alpha = \alpha$ *for all $\alpha \in \mathcal{F}$.*

(**b**) $1x = x$, *where "1" denotes the multiplicative unit in $\mathcal{F}$, meaning* $1\alpha = \alpha$ *for all $\alpha \in \mathcal{F}$.*

Remark 4.2 *Note that θ is unique by* 1.*b and* 1.*d. Given θ and θ' :*

$$\theta' = \theta' + \theta = \theta + \theta' = \theta.$$

Also, vector spaces contain additive inverses, $-x \equiv -1x$, where -1 is the additive inverse of 1 *in $\mathcal{F}$. That $x + (-x) = \theta$ follows from* 2.*c and* 3.*a.*

Definition 4.3 *A vector space is called a* ***normed vector space*** *or* ***normed linear space over a field*** *$\mathcal{F}$ if there exists a functional denoted $\|\| : V \to \mathbb{R}^+$, so that:*

4. $\|x\| = 0$ *if and only if* $x = 0$, *5. $\|\alpha x\| = |\alpha|\,\|x\|$ for all $x \in V$ and $\alpha \in \mathcal{F}$,*

*6. **Triangle inequality:** $\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in V$.*

Definition 4.4 *In a normed linear space, a sequence $\{s_n\}_{n=1}^{\infty} \subset V$ is a **Cauchy sequence** if for any $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that $\|s_n - s_m\| < \epsilon$ for all $n, m \ge N$.*

*A normed linear space is **complete** if given any Cauchy sequence $\{s_n\}_{n=1}^{\infty} \subset V$, there is an $s \in V$ so that $\|s_n - s\| \to 0$. In other words, for any $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that $\|s_n - s\| < \epsilon$ for all $n \ge N$. This is then denoted as $s_n \to s$ as well as $s = \lim_{n\to\infty} s_n$.*

Definition 4.5 *A **Banach space** is a complete normed linear space. A **real Banach space** is a complete normed linear space over $\mathcal{F} = \mathbb{R}$, and analogously for a **complex Banach space.***

Because addition is well defined in a linear space, there are also notions of convergence of a given infinite series.

Definition 4.6 *A series $\{x_n\}_{n=1}^{\infty} \subset V$ in a normed linear space is said to be **summable** if there exists $x \in V$ such that $\left\|\sum_{j=1}^{n} x_j - x\right\| \to 0$ as $n \to \infty$. The series $\{x_n\}_{n=1}^{\infty} \subset V$ is said to be **absolutely summable** if $\sum_{j=1}^{\infty} \|x_j\| < \infty$.*

Perhaps not surprisingly, the notion of completeness in a normed linear space can also be characterized in terms of the convergence properties of series.

Proposition 4.7 *A normed linear space V is complete if and only if every absolutely summable series is summable.*

Proof. *If V is complete and $\{x_n\}_{n=1}^{\infty} \subset V$ an absolutely summable series, then for any $\epsilon > 0$ there is an N so that $\sum_{j=N}^{\infty} \|x_j\| < \epsilon$. Let $s_n = \sum_{j=1}^{n} x_j$ be the partial summation sequence. Then for $n \ge m \ge N$ the triangle inequality obtains that*

$$\|s_n - s_m\| \le \sum\nolimits_{j=m+1}^{n} \|x_j\| < \epsilon.$$

Thus $\{s_n\}_{n=1}^{\infty} \subset V$ is a Cauchy sequence and hence converges to some point $x \in V$ by completeness.

Assume that absolutely summable series in V are summable, and let $\{s_n\}_{n=1}^{\infty} \subset V$ be a Cauchy sequence. By definition, for any k there is an N_k

so that $\|s_n - s_m\| \leq 1/2^k$ *for all* $n, m \geq n_k$, *and we can by construction assure that* $\{n_k\}_{k=1}^{\infty}$ *is an increasing sequence. Then* $\{s_{n_k}\}_{k=1}^{\infty}$ *is a subsequence of* $\{s_n\}_{n=1}^{\infty}$. *Define* $x_1 = s_{n_1}$ *and* $x_k = s_{n_k} - s_{n_{k-1}}$, *then* $\{x_k\}_{k=1}^{\infty}$ *is a series with partial sums given by* $\sum_{j=1}^{k} x_j = s_{n_k}$. *Also, since* $\|x_j\| \leq 1/2^{j-1}$ *for* $j \geq 2$, *it follows that* $\sum_{j=1}^{k} \|x_j\| \leq \|s_{n_1}\| + 1$ *and so* $\{x_k\}_{k=1}^{\infty}$*is an absolutely summable series which by assumption is summable to some* $x \in V$. *Hence, the subsequence* $\{s_{n_k}\}_{k=1}^{\infty}$ *satisfies* $\|s_{n_k} - x\| \to 0$. *But since* $\{s_n\}_{n=1}^{\infty}$ *is a Cauchy sequence, for any* $\epsilon > 0$ *there is an* N *so that* $\|s_n - s_m\| < \epsilon$ *for all* $n, m \geq N$. *For this same* ϵ *there is a* K *so that* $\|s_{n_k} - x\| < \epsilon$ *for* $k \geq K$. *Hence if* $n \geq \max\{N, n_K\}$*:*

$$\|s_n - x\| \leq \|s_n - s_{n_k}\| + \|s_{n_k} - x\| < 2\epsilon,$$

and so $s_n \to x$ *and* V *is complete.* ■

Example 4.8 *The classical example of a real Banach space is* ***Euclidean space*** $\mathbb{R}^n$ *under the* ***standard Euclidean norm*** *defined on* $x = (x_1, x_2, \ldots, x_n)$ *by:*

$$\|x\| \equiv \sqrt{\sum_{j=1}^{n} x_j^2}.$$

This same structure works in ***complex Euclidean space*** $\mathbb{C}^n$ *with standard norm defined on* $x = (x_1, x_2, \ldots, x_n)$ *by:*

$$\|x\| \equiv \sqrt{\sum_{j=1}^{n} |x_j|^2},$$

where $|x_j| \equiv \sqrt{x_j \bar{x}_j}$. *Recall that* $\bar{x}$ *denotes the* ***complex conjugate*** *of* $x = a + bi$ *, defined as* $\bar{x} = a - bi$. *In both cases, vector addition and scalar multiplication are defined component-wise:*

$$x + y \equiv (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$

$$\alpha x \equiv (\alpha x_1, \alpha x_2, \ldots, \alpha x_n).$$

The Euclidean spaces $\mathbb{R}^n$ *and* $\mathbb{C}^n$ *also have* ***inner products*** *which are compatible with the above norms. See definition 4.23 for properties of inner products. In these cases. if* $x = (x_1, x_2, \ldots, x_n)$ *and* $y = (y_1, y_2, \ldots, y_n)$, *then the inner product* (x, y) *is defined:*

$$(x, y) \equiv \sum_{j=1}^{n} x_j y_j, \qquad x, y \in \mathbb{R}^n,$$

$$(x, y) \equiv \sum_{j=1}^{n} x_j \overline{y}_j, \qquad x, y \in \mathbb{C}^n.$$

In both cases,

$$\|x\| = \sqrt{(x,x)}.$$

As it turns out, most Banach spaces do not have such compatible inner products, though there is a special class of Banach spaces that do. See the section below, The Special Case of $p = 2$*, Hilbert Space.*

These classical examples of $\mathbb{R}^n$ *and* $\mathbb{C}^n$ *can be generalized in two directions separately or together:*

1. ***Change to an*** l_p***-Norm***

   There are infinitely many norms that can be defined on $\mathbb{R}^n$ *and* $\mathbb{C}^n$*, and a popular collection is referred to as the* l_p*-norms which are parametrized by* $1 \leq p \leq \infty$ *and defined on* $y = (y_1, y_2, \ldots, y_n)$*, whether real or complex, by:*

$$\|y\|_p \equiv \begin{cases} \left(\sum_{j=1}^n |y_j|^p\right)^{1/p}, & 1 \leq p < \infty, \\ \sup_j |y_j|, & p = \infty. \end{cases} \tag{4.1}$$

   In a finite dimensional space one can equally well define $\|y\|_\infty = \max_j |y_j|$*, but we use supremum to make this definition applicable below where* $n = \infty$*. Also note that within this collection of norms that the standard norm is the* l_2*-norm, and so*

$$\|y\| = \|y\|_2 .$$

   In this context, one sometimes see the notation $\mathbb{R}_p^n$ *and* $\mathbb{C}_p^n$ *to denote that the* l_p*-norms are used in these linear spaces rather than the standard Euclidean norms.*

2. ***Infinite Dimension with Standard Norm***

   There is an infinite dimensional variation of $\mathbb{R}^n$ *and* $\mathbb{C}^n$*, which we denote by* $\mathbb{R}^{\mathbb{N}}$ *and* $\mathbb{C}^{\mathbb{N}}$*, defined as the collection of infinite real or complex sequences:*

$$y = (y_1, y_2, \ldots),$$

   which have finite norm:

$$\|y\| \equiv \sqrt{\sum_{j=1}^{\infty} |y_j|^2} < \infty.$$

3. ***Infinite Dimension with l_p-Norm***

These spaces are defined identically to $\mathbb{R}^{\mathbb{N}}$ and $\mathbb{C}^{\mathbb{N}}$ but with respect to the l_p-norms. Denoted l_p, or $l_p(\mathbb{R})$ and $l_p(\mathbb{C})$ if needed for clarity, these spaces include all sequences $y = (y_1, y_2, \ldots)$ with finite l_p-norm, $\|y\|_p < \infty$, where generalizing 4.1:

$$\|y\|_p \equiv \begin{cases} \left(\sum_{j=1}^{\infty} |y_j|^p\right)^{1/p}, & 1 \leq p < \infty, \\ \sup_j |y_j|, & p = \infty. \end{cases}$$

Remark 4.9 *Note that as a collection of points,*

$$\mathbb{R}_p^n = \mathbb{R}^n, \qquad \mathbb{C}_p^n = \mathbb{C}^n,$$

for any p with $1 \leq p \leq \infty$. In other words, while the norms of points change, the collection of points does not.

However, for the infinite dimensional spaces the value of p has an effect both on the norm of a sequence, as well as on the collection of points in the given space. For example, $y = (1, 1/2, 1/3, \ldots)$ is an element of l_p for $p > 1$ but not a point in l_1. In general,

$$l_{p'} \subsetneq l_p \qquad \text{for } p' < p.$$

4.2 The $L_p(X)$-Spaces

The l_p-spaces of sequences are generalized in this section to the L_p function spaces. We first introduce the necessary definitions, then develop some of the properties of these spaces. We restrict our definitions to real valued functions, since that is what we need in later books, but nothing significant changes for complex valued functions other than the definition of $|f(x)|$, as noted in the example above, and the use $\mathbb{C}$ instead of $\mathbb{R}$ as the underlying field $\mathcal{F}$.

Definition 4.10 *Given a measure space $(X, \sigma(X), \mu)$, we define* ***$L_p(X)$-space,*** *and sometimes* ***$L_p(X, \mu)$-space*** *when needed for clarity, as the space of all extended real-valued μ-measurable functions f on X with finite* ***$L_p(X)$-norm,*** *$\|f\|_p < \infty$. When $1 \leq p < \infty$, this norm is defined:*

$$\|f\|_p = \left(\int_X |f|^p \, d\mu\right)^{1/p}. \tag{4.2}$$

*For $p = \infty$, the $L_\infty(X)$**-norm** $\|f\|_\infty$ denotes the **essential supremum of** f, denoted* $\operatorname{ess\,sup} f$, *and defined by:*

$$\|f\|_\infty = \operatorname{ess\,sup} f \equiv \inf\{\alpha|\mu\left[\{x|\,|f(x)| > \alpha\}\right] = 0\}. \tag{4.3}$$

*So $L_\infty(X)$ is the space of μ-**measurable functions** f on X that are bounded outside a set of μ-**measure** 0. Such functions are sometimes referred to as **essentially bounded functions.***

*Given index p with $1 < p < \infty$, the index q is called the **conjugate index to** p if $\frac{1}{p} + \frac{1}{q} = 1$. In this case $1 < q < \infty$ since $q = \frac{p}{p-1}$. The pair (p, q) are called **conjugate indexes,** and this definition is sometimes applied to $(1, \infty)$, notationally defining $\frac{1}{0} = \infty$.*

Remark 4.11 *Note that $\|f\|_\infty$ is well-defined since for any α, $\{x|\,|f(x)| > \alpha\} \in \sigma(X)$ by the μ-measurability of f. Also, the notion of conjugacy for indexes is symmetric, so if (p, q) then (q, p).*

Notation 4.12 *When $(X, \sigma(X), \mu) = (\mathbb{R}^n, \mathcal{M}_L^n, m^n)$, n-dimensional Lebesgue measure space, the $L_p(X)$-spaces are often called the **classical** $L_p(\mathbb{R}^n)$**-spaces.***

Our goal is to prove that the $L_p(X)$-spaces are Banach spaces, but we have an apparent problem. As stated, $\|f\|_p$ is not a norm because it violates item 4 in the above definition, that $\|x\| = 0$ if and only if $x = 0$. By proposition 2.26, if $1 \leq p < \infty$ and $\|f\|_p = 0$ we can only assert that $f = 0$ μ-a.e., and the same conclusion follows from $\|f\|_\infty = 0$ by definition. The most common approach to circumventing this "problem" is to identify a function f of an $L_p(\mu)$-space with an **equivalence class** of functions, so that

$$f \equiv \{f_0 | f_0 = f \ \mu\text{-a.e.}\}.$$

In other words, we will not distinguish between different representatives of this equivalence class, and any statement about f applies equally well to any function in the equivalence class of f. Now with this convention, $\|f\|_p = 0$ if and only if $f = 0$ since the latter statement simply means that $f_0 \equiv 0 \in f$.

Remark 4.13 *Regarding this convention, the language that is used can take a little getting used to. Looking at the last sentence, the notion that $\|f\|_p = 0$ is well defined whether f denotes a function or an equivalence class. If this is true for a given f of the class, it is true for all f in that class. On the other hand, the statement $f = 0$ has a different meaning in these contexts.*

Interpreted as a statement about a function, this implies $f = f_0$ above. However, when understood as a statement about the equivalence class, this only implies that $f = 0$, μ-a.e. Sometimes the "μ-a.e." will be emphasized in a given conclusion if it is to be applied to a specific function, while this qualifier is unnecessary if applied to f when this denotes the equivalence class of f.

As many authors will now say: "The reader will rarely find this ambiguous or confusing."

With this convention, we now prove that for any measure space $(X, \sigma(X),\mu)$, the corresponding $L_p(X)$-space is a Banach space. This proof will be split into two parts, the first establishing that $L_p(X)$ is a normed linear space over $\mathbb{R}$, the second step proving completeness as a normed space.

Proposition 4.14 *Given a measure space $(X, \sigma(X),\mu)$, the space $L_p(X)$ is a normed linear space over $\mathbb{R}$ for $1 \leq p \leq \infty$.*
Proof. *It is apparent that these spaces satisfy most of the vector space and norm requirements, except that we need to verify 1.a., that $L_p(X)$ is closed under addition, and 6, that the $L_p(\mu)$-norm satisfies the triangle inequality. Both results are proved by the following result.* ■

Minkowski's inequality or **the Minkowski inequality** of the next result, is named for **Hermann Minkowski** (1864 – 1909).

Proposition 4.15 (Minkowski's Inequality) *If $1 \leq p \leq \infty$ and $f, g \in L_p(X)$, then $f + g \in L_p(X)$ and*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p\,. \tag{4.4}$$

Proof. *As $f + g$ is always measurable, that $f + g \in L_p(X)$ follows from a proof of 4.4. If $p = 1$ this norm inequality follows from the triangle inequality, that $|f + g| \leq |f| + |g|$, while for $p = \infty$ this result follows from the triangle inequality and the observation that:*

$$\begin{aligned} \{x|\,|f(x) + g(x)| > \alpha + \beta\} &\subset \{x|\,|f(x)| + |g(x)| > \alpha + \beta\} \\ &\subset \{x|\,|f(x)| > \alpha\} \cup \{x|\,|g(x)| > \beta\}. \end{aligned}$$

So assume that $1 < p < \infty$. The next step requires Hölder's inequality of proposition 3.46 of book 4, and this in turn requires knowledge that $f + g \in L_p(X)$. But this follows from the inequality: $|(f + g)/2|^p \leq (|f|^p + |g|^p)/2$, which in turn follows from the convexity (definition 3.36 of book 4) of $h(x) =$

x^p on $[0,\infty)$ for $p>1$. While this upper bound assures finiteness of $\|f+g\|_p$, it falls short of the triangle inequality. For the last step we will assume that $\|f+g\|_p \neq 0$, since 4.4 is certainly valid in this case.

To better estimate the norm of $f+g$, note that by the triangle inequality $|f+g| \leq |f|+|g|$ and so:

$$\int_X |f+g|^p \, d\mu \leq \int_X |f|\,|f+g|^{p-1} \, d\mu + \int_X |g|\,|f+g|^{p-1} \, d\mu.$$

Since $f, g, f+g \in L_p(X)$, Hölder's inequality is applicable with q conjugate to p because $|f+g|^{p-1} \in L_q(X)$. This follows because $(p-1)q = p$, and this obtains:

$$\left\| |f+g|^{p-1} \right\|_q = \left(\|f+g\|_p \right)^{p/q} < \infty.$$

Thus:

$$\begin{aligned} \int_X |f|\,|f+g|^{p-1} \, d\mu \ &\leq \ \|f\|_p \left\| |f+g|^{p-1} \right\|_q , \\ \int_X |g|\,|f+g|^{p-1} \, d\mu \ &\leq \ \|g\|_p \left\| |f+g|^{p-1} \right\|_q . \end{aligned}$$

Combining results,

$$\|f+g\|_p^p \leq \left(\|f\|_p + \|g\|_p \right) \left(\|f+g\|_p \right)^{p/q},$$

and 4.4 follows by division since $p - p/q = 1$. ■

The final step in proving that $L_p(X)$ is a Banach space is to prove that this space is complete. This result is called the **Riesz-Fischer theorem,** named for **Frigyes Riesz** (1880 – 1956) and **Ernst Fischer** (1875 – 1954), who independently derived this conclusion for the space L_2 in 1907.

Proposition 4.16 (Riesz-Fischer theorem) *Given a measure space $(X, \sigma(X), \mu)$, the space $L_p(X)$ is a complete normed linear space over $\mathbb{R}$ for $1 \leq p \leq \infty$, and hence a Banach space.*

Proof. *First addressing $p=\infty$, let $\{f_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $L_\infty(X)$ and choose an increasing sequence $\{n_k\}_{k=1}^{\infty}$ so that $\|f_n - f_m\|_\infty < 2^{-k}$ for $n, m \geq n_k$. Then $\left\|f_{n_{k+1}} - f_{n_k}\right\|_\infty < 2^{-k}$ and so (recall remark 4.13) there is $A_k \in \sigma(X)$ with $\mu(A_k) = 0$ and $\left|f_{n_{k+1}}(x) - f_{n_k}(x)\right| < 2^{-k}$ for $x \notin A_k$. In other words, such $\{A_k\}$ exists for any function sequence from these equivalence classes. With $A \equiv \cup_k A_k$ it follows that $\mu(A) = 0$ and for $x \notin A$,*

$\left|f_{n_{k+1}}(x) - f_{n_k}(x)\right| < 2^{-k}$ *for all* k*. Hence for each* $x \notin A$, $\{f_{n_k}(x)\}_{k=1}^{\infty}$ *is a Cauchy sequence in* $\mathbb{R}$*, which is complete, so define* $f(x) = \lim_{k\to\infty} f_{n_k}(x)$ *for* $x \notin A$*, and* $f(x) = 0$ *for* $x \in A$*. That* $f(x)$ *so defined is measurable can be seen as follows. Define* $\widetilde{f}_{n_k}(x) \equiv f_{n_k}(x)$ *for* $x \notin A$ *and* $\widetilde{f}_{n_k}(x) = 0$ *for* $x \in A$*. Then each* $\widetilde{f}_{n_k}(x)$ *is* μ*-measurable:*

$$\widetilde{f}_{n_k}^{-1}((a,\infty)) = \begin{array}{ll} f_{n_k}^{-1}((a,\infty)) \cap \widetilde{A}, & a > 0, \\ \left[f_{n_k}^{-1}((a,\infty)) \cap \widetilde{A}\right] \cup A, & a \leq 0, \end{array}$$

and thus $\widetilde{f}_{n_k}^{-1}((a,\infty)) \in \sigma(X)$ *since* A *is measurable. Since* $f(x) = \lim_{k\to\infty} \widetilde{f}_{n_k}(x)$ *for all* x*,* $f(x)$ *is then* μ*-measurable as a pointwise limit of measurable functions. Now if* $n \geq n_k$*, then for* $x \notin A$*,*

$$\begin{aligned} |f_n(x) - f(x)| &\leq |f_n(x) - f_{n_k}(x)| + |f_{n_k}(x) - f(x)| \\ &\leq 2^{-k} + 2^{-k}, \end{aligned}$$

and so $\|f_n - f\|_\infty \to 0$*. The last detail (which we will not repeat again) is to prove that given another function sequence* $\{\tilde{f}_n(x)\}_{n=1}^{\infty}$ *from the associated equivalence classes, and corresponding* $\{\tilde{A}_k\}$*-collection, the resulting function* $\tilde{f}$ *will satisfy* $f = \tilde{f}$ *,* μ*-a.e.. This assures that* $\tilde{f}$ *is in the same equivalence class as the above constructed* f*. This follows because* $f_n = \tilde{f}_n$ μ*-a.e. assures that* $\left\|\tilde{f}_n - f\right\|_\infty \to 0$*, and this with* $\left\|\tilde{f}_n - \tilde{f}\right\|_\infty \to 0$ *obtains the result.*

Next, assume $1 \leq p < \infty$ *and let* $\{f_n\}_{n=1}^{\infty}$ *be an absolutely summable sequence in* $L_p(X)$*, meaning* $\sum_{n=1}^{\infty} \|f_n\|_p = M < \infty$*. By proposition 4.7 completeness follows by showing that* $\{f_n\}_{n=1}^{\infty}$ *is summable, meaning there exists* $s \in L_p(X)$ *with* $\sum_{j=1}^{n} f_j \to s$ *in* $L_p(X)$*. To this end, define the partial summation sequence* $g_n(x) = \sum_{j=1}^{n} |f_j(x)|$ *. By the Minkowski inequality,* $g_n \in L_p(X)$ *and* $\|g_n\|_p \leq M$*, and hence:*

$$\int_X g_n^p d\mu \leq M^p. \tag{(*)}$$

Now for each x*,* $\{g_n(x)\}_{n=1}^{\infty}$ *is an increasing sequence of extended real numbers in* $\overline{\mathbb{R}}$*, so define* $g(x) = \lim_{n\to\infty} g_n(x)$*. As a limit of measurable functions* $g(x)$ *is measurable, and since* $g_n(x) \geq 0$ *we conclude by 2.9 of Fatou's lemma and* $(*)$ *that* $g \in L_p(X)$ *and* $\|g\|_p \leq M$*. This then implies that* $g(x) = \sum_{j=1}^{\infty} |f_j(x)| < \infty$ μ*-a.e., and thus there exists* $A \in \sigma(X)$ *with* $\mu(A) = 0$ *and for* $x \notin A$*,* $\sum_{j=1}^{\infty} f_j(x)$ *is absolutely convergent to a real number and hence is convergent. For* $x \notin A$*, define* $s(x) = \lim_{n\to\infty} s_n(x)$ *where*

$s_n(x) = \sum_{j=1}^{n} f_j(x)$, and $s(x) = 0$ on A . As above, defining $\widetilde{f}_j(x) = f_j(x)$ for $x \notin A$ and $\widetilde{f}_j(x) = 0$ on A, $\widetilde{f}_j(x)$ is μ-measurable and $s(x) = \sum_{j=1}^{\infty} \widetilde{f}_j(x)$ for all x. Hence $s(x)$ is measurable and $s \in L_p(X)$ since $|s| \leq g \in L_p(X)$. Also, $|s_n| \leq g$ and so:

$$|s(x) - s_n(x)| \leq 2\,|g(x)|\,,$$

and because $g \in L_p(\mu)$, $|s(x) - s_n(x)|^p$ is dominated by the integrable function $|g(x)|^p$. By Lebesgue's dominated convergence theorem, the almost everywhere pointwise convergence $|s(x) - s_n(x)|^p \to 0$ obtains

$$\int_X |s(x) - s_n(x)|^p\, d\mu \to 0,$$

and so $\|s - s_n\|_p \to 0$ and $\{f_n\}_{n=1}^{\infty}$ is summable to $s \in L_p(X)$. ■

There is an important corollary to the Riesz-Fischer theorem that connects the notion of L_p-convergence to convergence μ-a.e.

Corollary 4.17 (Riesz-Fischer theorem) *If $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence with limit f in $L_p(X)$ for $1 \leq p < \infty$, then there is a subsequence $\{f_{n_j}\}_{j=1}^{\infty}$ so that $f_{n_j} \to f$, μ-a.e. If $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence with limit f in $L_\infty(X)$, then $f_n \to f$ uniformly μ-a.e.*

Proof. *First assume that $p < \infty$. Generalizing the construction in the prior proof, let $\{f_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $L_p(X)$ and choose an increasing sequence $\{n_k\}_{k=1}^{\infty}$ so that $\|f_n - f_m\|_p < 2^{-k}$ for $n, m \geq n_k$. Thus $\left\|f_{n_{k+1}} - f_{n_k}\right\|_p < 2^{-k}$. Define $g_j(x) = \sum_{k=1}^{j} \left|f_{n_{k+1}}(x) - f_{n_k}(x)\right|$ and $g(x) = \sum_{k=1}^{\infty} \left|f_{n_{k+1}}(x) - f_{n_k}(x)\right|$. By construction $\|g_j\|_p \leq 1$ for all j, and since $p < \infty$ we can apply Fatou's lemma to $\{g_j^p\}_{j=1}^{\infty}$ to conclude that $\|g\|_p \leq 1$. Hence, $g(x) < \infty$ μ-a.e. and so the series:*

$$f_{n_1}(x) + \sum_{k=1}^{\infty} \left(f_{n_{k+1}}(x) - f_{n_k}(x)\right)$$

converges absolutely μ-a.e. Define f to equal this summation when convergent, and equal to 0 on the non-convergence set of μ-measure 0. Then since

$$f_{n_j}(x) = f_{n_1}(x) + \sum_{k=1}^{j-1} \left(f_{n_{k+1}}(x) - f_{n_k}(x)\right),$$

it follows that $f_{n_j} \to f$ μ-a.e.

If $p = \infty$, then $\|f_n - f\|_\infty \to 0$ implies that for each n there exists $A_n \subset X$ with $\mu(A_n) = 0$, $\sup|f_n(x) - f(x)| \leq c_n$ for $x \notin A_n$, and $c_n \to 0$. Let $A = \cup A_n$, then $\mu(A) = 0$ and $\sup|f_n(x) - f(x)| \to 0$ for $x \notin A$. That is, $f_n \to f$ uniformly on $\widetilde{A}$. ■

4.3 Approximating Functions in $L_p(X)$-Spaces

The following result is one of many and shows that when $1 \leq p \leq \infty$, every $f \in L_p(X)$ is the $L_p(X)$-limit of a sequence of simple functions, and when X is σ-finite and $p < \infty$ each such function can be defined to be zero outside X sets of finite μ-measure. For the proof we will recall some of the results from the above section, Approximating μ-Measurable Functions. Example 4.19 demonstrates that the second part of this result does not hold in $L_\infty(X)$.

When X has a topology and $\sigma(X)$ contains the open sets and is hence a Borel sigma algebra, such $f \in L_p(X)$ are also the $L_p(X)$-limit of a sequence of continuous functions on X, which again can be defined to be zero outside sets of finite μ-measure when X is σ-finite. We do not prove these results, but see for example Doob (1994) or Rudin (1974) in the references.

Proposition 4.18 *Given a measure space $(X, \sigma(X), \mu)$, let $\mathcal{S}$ denote the space of simple functions on X. Then $\mathcal{S}$ is dense in $L_p(X)$ for $1 \leq p \leq \infty$. In other words, for any $f \in L_p(X)$ there is a sequence $\{\varphi_n\}_{n=1}^\infty \subset \mathcal{S}$ so that*

$$\|f - \varphi_n\|_p \to 0.$$

If $\mathcal{S}_0 \subset \mathcal{S}$ is defined so that for all $\varphi \in \mathcal{S}_0$:

$$\mu\left[\{x \in X | \varphi(x) \neq 0\}\right] < \infty,$$

then $\mathcal{S}_0$ is dense in $L_p(X)$ for $1 \leq p < \infty$ if X is σ-finite.

Proof. *If $f \in L_p(X)$ is nonnegative, $f \geq 0$, define $\{\varphi_n\}_{n=1}^\infty$ as in proposition 1.18 so $\{\varphi_n\}_{n=1}^\infty \subset \mathcal{S}$ and $\varphi_n \to f$ pointwise. Since $0 \leq \varphi_n \leq f$ it follows that $\varphi_n \in L_p(X)$ for all n. By construction $|f - \varphi_n|^p \leq |f|^p$ for $p < \infty$ and so Lebesgue's dominated convergence theorem can be applied to conclude that as $n \to \infty$,*

$$\int_X |f(x) - \varphi_n(x)|^p \, d\mu \to 0,$$

and hence $\|f - \varphi_n\|_p \to 0$. When $p = \infty$ it follows by definition of $\|f\|_\infty < \infty$ that $|f| \leq M$ for $x \notin A \subset X$ with $\mu(A) = 0$. Then by the construction in proposition 1.18, if $n > M$ then $\sup |f - \varphi_n| \leq 2^{-n}$ for $x \notin A$, and thus $\|f - \varphi_n\|_\infty \to 0$.

For general f write $f(x) = f^+(x) - f^-(x)$ where $f^+(x)$ and $f^-(x)$ are nonnegative functions as in definition 2.36. Hence, there are sequences of simple functions $\{\varphi_n^+(x)\}$ and $\{\varphi_n^-(x)\}$ so that $\varphi_n^+(x) \to f^+(x)$ and $\varphi_n^-(x) \to$

$f^-(x)$ pointwise. Defining $\varphi_n(x) = \varphi_n^+(x) - \varphi_n^-(x)$, then $\varphi_n(x) \to f(x)$ pointwise. Also, since $f - \varphi_n = (f^+ - \varphi_n^+) - (f^- - \varphi_n^-)$, the triangle inequality obtains

$$\|f - \varphi_n\|_p \leq \left\|f^+ - \varphi_n^+\right\|_p + \left\|f^- - \varphi_n^-\right\|_p,$$

and the conclusion for nonnegative f then proves that $\|f - \varphi_n\|_p \to 0$.

If X is σ-finite and $p < \infty$ such simple functions can be chosen from $\mathcal{S}_0$ by corollary 1.22. ■

Example 4.19 *That the second statement in proposition 4.18 is not true in $L_\infty(X)$, let $X = \mathbb{R}$ with Lebesgue measure, which is a σ-finite measure space. If $f \equiv 1$ and g is any function that equals 0 outside a set of finite measure, then $\|f - g\|_\infty \geq 1$.*

The details of the proof of the following result are assigned as an exercise.

Proposition 4.20 *If $1 \leq p < \infty$, then $f_n(x) \to f(x)$ in $L_p(X)$ implies that:*

$$\|f_n(x)\|_{L_p} \to \|f(x)\|_{L_p}. \tag{4.5}$$

Proof. *Hint: For $p = 1$, note that*

$$||a| - |b|| \leq |a - b| \tag{(*)}$$

for all a and b. For $p > 1$, show using a Taylor series that for positive a, b:

$$|a^p - b^p| \leq p\,|a - b|\,(a + b)^{p-1},$$

and then apply the Hölder and Minkowski inequalities and $(*)$ to show that:

$$\left\||f|^p - |f_n|^p\right\|_{L_1} \leq p\,\|f - f_n\|_{L_p}\left(\|f - f_n\|_{L_p} + 2\,\|f\|_{L_p}\right)^{p/q}.$$

■

4.4 Further Properties of the $L_p(X)$-Spaces

In this section, we state some additional properties of the $L_p(X)$ spaces for completeness, but mostly without proof as these are not needed in what follows. See section 4.1.4 of book 3 for additional results on bounded linear functions.

4.4.1 Bounded Linear Functionals

A **linear functional** F on an arbitrary normed linear space $(V, \|\|)$ is defined as a mapping, $F : V \to \mathbb{R}$, so that for all $x, y \in V$, $\alpha, \beta \in \mathcal{F}$,

$$F(\alpha x + \beta y) = \alpha F(x) + \beta F(y).$$

By a **bounded linear functional** is meant a linear functional with the additional property that there is a constant M so that for all $x \in V$,

$$|F(x)| \leq M \|x\|.$$

Equivalently, it is an exercise to check that a linear functional is bounded if and only if $|F(x)| \leq M$ for all $x \in V$ with $\|x\| = 1$.

The **norm of a bounded linear functional,** denoted $\|F\|$, is then defined as the minimum value of all such M :

$$\begin{aligned} \|F\| &= \inf\{M \,|\, |F(x)| \leq M \|x\| \text{ all } x\} \\ &= \inf\{M \,|\, |F(x)| \leq M \text{ all } x \text{ with } \|x\| = 1\}. \end{aligned}$$

Example 4.21 *Given an $L_p(X)$ space with $1 \leq p \leq \infty$, let $g \in L_q(X)$, where q is conjugate to p, meaning $\frac{1}{p} + \frac{1}{q} = 1$ and where we invoke the convention that $\frac{1}{\infty} = 0$. Define F on $L_p(X)$ by:*

$$F(f) = \int_X fg d\mu.$$

If this integral is well defined, F is linear by proposition 2.40.

If $1 < p < \infty$, then by the triangle inequality and then Hölder's inequality of proposition 3.46 of book 4:

$$|F(f)| \leq \|fg\|_1 \leq \|g\|_q \|f\|_p.$$

So F is a bounded linear functional on $L_p(\mu)$ with $\|F\| \leq \|g\|_q$.

When $p = 1$ or $p = \infty$ the same inequality holds. If $p = 1$ and $g \in L_\infty(\mu)$, then the definition of $F(f)$ is unchanged by redefining g to be 0 on the set of μ-measure 0 where $|g| > \|g\|_\infty$. Hence by the triangle inequality,

$$|F(f)| \leq \|g\|_\infty \|f\|_1,$$

and so again $\|F\| \leq \|g\|_\infty$. The result for $p = \infty$ follows by symmetry relative to $p = 1$.

In all cases it is in fact true that

$$\|F\| = \|g\|_q\,,$$

though an extra assumption is needed when $p = 1$.

When $1 < p < \infty$, *let* $f = |g|^{q/p}\, sgn[g]$ *where* $sgn[g] = +1$ *if* $g \geq 0$, $sgn[g] = -1$ *if* $g < 0$. *Then* $f \in L_p(X)$ *and* $F(f) = \|g\|_q \|f\|_p$. *This implies* $\|F\| \geq \|g\|_q$ *and the result follows. When* $p = \infty$ *and* $g \in L_1(X)$, *then if* $f \equiv c \in L_\infty(X)$ *for* $c \neq 0$ *a constant yields* $\|F\| = \|g\|_1$. *Finally for* $p = 1$ *and* $g \in L_\infty(X)$, *let* $A = \{x|\, |g| > \|g\|_\infty - \epsilon\}$. *For arbitrary* $\epsilon > 0$, A *is* μ*-measurable with non-zero measure by definition of* $\|g\|_\infty$, *though* A *need not have finite measure. Assume that* μ *is* σ*-finite, meaning* $X = \cup X_i$ *with* $\mu(X_i) < \infty$, *and let* $f_n = \chi_{B_n}$ *where* $B_n = A \cap \cup_{i=1}^n X_i$ *with* n *large enough so that* $\mu[B_n] > 0$. *Then* $f \in L_1(X)$ *and* $F(f) \geq (\|g\|_\infty - \epsilon)\|f\|_1$ *and the result follows since* ϵ *is arbitrary.*

Hence given an $L_p(X)$ *space and* $1 < p \leq \infty$, *every function in the conjugate space* $L_q(X)$ *induces a bounded linear functional on* $L_p(X)$ *with norm equal to the* $L_q(X)$*-norm of that function. When* $p = 1$, *this result is again true but requires* σ*-finiteness of the measure space.*

An important question that arises with normed linear spaces is, can all bounded linear functionals be characterized in some accessible way? In the case of the $L_p(X)$ spaces, this question could be restated given the above discussion. Specifically, are **all** bounded linear functionals on $L_p(X)$ given by some $g \in L_q(X)$? The answer is in the affirmative for $1 \leq p < \infty$, but not so for $L_\infty(X)$. This famous result is the **Riesz Representation theorem** named for **Frigyes Riesz** (1880 – 1956), who also developed many results of this type for various normed linear spaces, all of which are referred to under this general name.

We state this theorem without proof.

Proposition 4.22 (Riesz Representation theorem) *Let* F *be a bounded linear functional on* $L_p(X)$ *for* $1 \leq p < \infty$. *Then there exists* $g \in L_q(X)$ *so that*

$$F(f) = \int_X fg\,d\mu. \tag{4.6}$$

Further, $\|F\| = \|g\|_q$.

Proof. *See for example Royden (1971) for* $X = \mathbb{R}$, *and Rudin (1974) for the general case.* ■

4.4.2 The Special Case of $p = 2$, Hilbert Space

That $p = 2$ is a special case within the $L_p(X)$ space hierarchy is initially appreciated with Hölder's inequality. Namely, $p = 2$ is the only case for which fg is integrable when both $f, g \in L_p(X)$. In this special case, Hölder's inequality is the **Cauchy-Schwarz inequality**:

$$\|fg\|_1 \leq \|f\|_2 \, \|g\|_2 \, ,$$

and this inequality alone implies that every $g \in L_2(X)$ induces a bounded linear functional on $L_2(X)$. That $p = 2$ is a special case within the $L_p(X)$ space hierarchy is also appreciated with the Riesz Representation theorem in that again, $L_2(X)$ is the only $L_p(X)$ space for which all bounded linear functionals are given by 4.6 with functions g within this same space.

In addition, $L_2(X)$ has more structure than the other $L_p(X)$ spaces, and as a result it is a special type of Banach space that is called a **Hilbert Space.** This space is named for **David Hilbert** (1862 – 1943) who investigated solutions to integral equations in this space in the early 1900s. In short, a Hilbert space has all the structure of the Euclidean space $\mathbb{R}^n$ or $\mathbb{C}^n$. Recalling example 4.8 above, this means that a Hilbert space is first and foremost a Banach space, but with the additional property that it possesses an **inner product** which is compatible with the given norm.

Definition 4.23 *A* ***real Hilbert space*** *H is a real Banach space on which is defined an inner product, denoted (x, y), which produces a mapping $H \times H \rightarrow \mathbb{R}$, with the properties that for $x_j, y \in H$ and $\alpha, \beta \in \mathbb{R}$:*

1. $(\alpha x_1 + \beta x_2, y) = \alpha(x_1, y) + \beta(x_2, y)$,

2. $(x, y) = (y, x)$,

3. $(x, x) = \|x\|^2$.

For ***complex Hilbert spaces****, the inner product is a mapping $H \times H \rightarrow \mathbb{C}$, and 2 is changed to:*

$2'$. $(x, y) = \overline{(y, x)}$

where $\overline{(y, x)}$ denotes the complex conjugate of (y, x). Thus if $(y, x) = a + bi$ then $\overline{(y, x)} = a - bi$.

Remark 4.24 *Note that by combining* 1 *and* 2 *we have that in a real Hilbert space:*

4. $(x, \alpha y_1 + \beta y_2) = \alpha(x, y_1) + \beta(x, y_2)$,

   while in a complex Hilbert space, combining 1 *and* $2'$ *produces for* $\alpha, \beta \in \mathbb{C}$:

4′. $(x, \alpha y_1 + \beta y_2) = \overline{\alpha}(x, y_1) + \overline{\beta}(x, y_2)$.

Example 4.25 *As noted above,* $\mathbb{R}^n$ *and* $\mathbb{C}^n$ *are real and complex Hilbert spaces respectively. For* $x, y \in \mathbb{R}^n$, *we have defined:*

$$(x, y) = \sum\nolimits_{i=1}^{n} x_i y_i,$$

while for $x, y \in \mathbb{C}^n$,

$$(x, y) = \sum\nolimits_{i=1}^{n} x_i \bar{y}_i.$$

It is a straightforward exercise to verify that these are indeed inner products on the respective spaces.

On general $L_2(X)$ *we define the inner product for* $f, g \in L_2(X)$ *by:*

$$(f, g) = \int_X f g d\mu,$$

for real-valued functions, or:

$$(f, g) = \int_X f \overline{g} d\mu.$$

for complex valued functions. Again, that (f, g) *so defined is an inner product on the respective space is readily verified.*

Proposition 4.26 *If* H *is a Hilbert space, then the inner product is a continuous function from* $H \times H$ *to* $\mathbb{R}$ *or* $\mathbb{C}$.

Proof. *Note that*

$$(x_1, y_1) - (x_2, y_2) = (x_1 - x_2, y_1) + (x_1 - x_2, y_1 - y_2) + (x_1, y_1 - y_2),$$

and so by the Cauchy-Schwarz inequality:

$$|(x_1, y_1) - (x_2, y_2)| \leq \|x_1 - x_2\|_2 \|y_1\|_2 + \|x_1 - x_2\|_2 \|y_1 - y_2\|_2 + \|x_1\|_2 \|y_1 - y_2\|_2 .$$

Hence, $|(x_1, y_1) - (x_2, y_2)| \to 0$ *as* $\|x_1 - x_2\|_2 \to 0$ *and* $\|y_1 - y_2\|_2 \to 0$. ∎

The significance of the additional structure in a Hilbert space is that an inner product allows one to introduce the notion of **orthogonality.**

Definition 4.27 *In a Hilbert space H, x is defined to be* ***orthogonal*** *to y, denoted $x \perp y$, if $(x, y) = 0$.*

This is precisely the notion of orthogonality in $\mathbb{R}^n$ or $\mathbb{C}^n$ with (x, y) defined above, and this notion can be used in the same way in this general setting.

Example 4.28 *Assume that $\{x_n\}_{n=1}^{\infty} \subset H$ are pairwise orthogonal, $(x_n, x_m) = 0$ for $n \neq m$, and have unit norm, $(x_n, x_n) = \|x_n\|^2 = 1$. Such a collection is called an* ***orthonormal system.*** *Assume further that for some constants $\{\alpha_n\}_{n=1}^{\infty}$ that $\sum_{n=1}^{\infty} \alpha_n x_n$ converges in H to x. That is, as $m \to \infty$:*

$$\left\| \sum\nolimits_{n=1}^{m} \alpha_n x_n - x \right\|_2 \to 0.$$

Then it is the case that $\alpha_n = (x, x_n)$ and this obtains the ***Fourier series expansion:***

$$x = \sum\nolimits_{n=1}^{\infty} (x, x_n) x_n,$$

where the coefficients $\{(x, x_n)\}_{n=1}^{\infty}$ are called the ***Fourier coefficients of*** *x. These are named for* ***Jean-Baptiste Joseph Fourier*** *(1768 – 1830) who developed these ideas in the context of what are called* ***Fourier series****, or trigonometric series expansions of integrable functions.*

For example, consider the complex Hilbert space $H \equiv L_2([-\pi, \pi])$, of complex-valued square integrable functions defined on the interval $[-\pi, \pi]$ with inner product defined by:

$$(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \overline{g(x)} dx.$$

Then $\left\{e^{inx}\right\}_{n=-\infty}^{\infty} \subset L_2([-\pi, \pi])$ is an orthonormal system, as can be verified as an exercise. Hence if $f \in L_2([-\pi, \pi])$ has a Fourier expansion $f = \sum_{n=-\infty}^{\infty} \alpha_n e^{inx}$, it must be the case that

$$\alpha_n = (f, e^{inx}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx.$$

The Fourier coefficients $\{\alpha_n\}_{n=-\infty}^{\infty}$ are often denoted $\left\{\widehat{f}(n)\right\}_{n=-\infty}^{\infty}$, and so:

$$f = \sum\nolimits_{n=-\infty}^{\infty} \widehat{f}(n) e^{inx}. \tag{4.7}$$

That this is a trigonometric series expansion of f follows from 6.12.

Remark 4.29 *We return to another topic in Fourier analysis below in the section entitled Fourier Transforms of a Finite Borel Measure, as well as in book 6 in the study of characteristic functions.*

Of course the first big question is, when does $f \in L_2([-\pi,\pi])$ have such a Fourier expansion? If $f \in L_1([-\pi,\pi]) \cap L_2([-\pi,\pi])$ then α_n is well defined with $|\alpha_n| \leq \|f\|_1$, but the general case is more subtle. It turns out, though we will not prove this, that every function $f \in L_2([-\pi,\pi])$ has this expansion in the sense that as $N \to \infty$,

$$\sum\nolimits_{n=-N}^{N} \widehat{f}(n)e^{inx} \to f.$$

Here convergence is defined in $L_2([-\pi,\pi])$:

$$\left\| \sum\nolimits_{n=-N}^{N} \widehat{f}(n)e^{inx} - f \right\|_2 \to 0.$$

By continuity of the inner product from proposition 4.26, we can therefore conclude that

$$\left(\sum\nolimits_{n=-N}^{N} \widehat{f}(n)e^{inx}, \sum\nolimits_{n=-N}^{N} \widehat{f}(n)e^{inx} \right) \to \|f\|_2^2 .$$

But by orthonormality, this inner product is readily evaluated:

$$\left(\sum\nolimits_{n=-N}^{N} \widehat{f}(n)e^{inx}, \sum\nolimits_{n=-N}^{N} \widehat{f}(n)e^{inx} \right) = \sum\nolimits_{n=-N}^{N} \left| \widehat{f}(n) \right|^2 ,$$

and so every $f \in L_2([-\pi,\pi])$ has a Fourier series expansion where as $N \to \infty$,

$$\sum\nolimits_{n=-N}^{N} \left| \widehat{f}(n) \right|^2 \to \|f\|_2^2 .$$

In other words, every $f \in L_2([-\pi,\pi])$ has a Fourier series expansion for which the Fourier coefficients are square summable, meaning a series in the Hilbert space $l_2(\mathbb{C})$, recalling example 4.28 but defined with respect to vectors: $x = (..., x_{-2}, x_{-1}, x_0, x_1, x_2, ...)$.

Another big question is, when is a given series $\{\alpha_n\}_{n=-\infty}^{\infty}$ equal to the Fourier coefficients of a function $f \in L_2([-\pi,\pi])$? It turns out that if $\{\alpha_n\}_{n=-\infty}^{\infty} \in l_2(\mathbb{C})$, so:

$$\sum\nolimits_{n=-\infty}^{\infty} |\alpha_n|^2 < \infty,$$

then there is an $f \in L_2([-\pi, \pi])$ with $\widehat{f}(n) = \alpha_n$. This is the original version of the **Riesz-Fischer theorem** of proposition 4.16, and the essential ingredient for its proof is to first prove that $L_2([-\pi, \pi])$ is complete.

Summarizing, these results assert that there is a one-to-one correspondence between $L_2([-\pi, \pi])$, the space of square integrable functions on $[-\pi, \pi]$, and the space $l_2(\mathbb{C})$, the space of square summable double-sided series. Moreover, this correspondence is norm preserving in that:

$$\|f\|_{L_2} = \left\| \left\{ \widehat{f}(n) \right\}_{n=-\infty}^{\infty} \right\|_{l_2}. \tag{4.8}$$

The identity in 4.8 is known as **Parseval's identity** and named for **Marc-Antoine Parseval** (1755 – 1836).

Chapter 5

Integrals in Product Spaces

The goal of this section is to develop **Fubini's theorem**, named for **Guido Fubini** (1879 – 1943), and **Tonelli's Theorem,** named for **Leonida Tonelli** (1885 – 1946). Both of these theorems provide results on the relationship between the value of a **product space integral**, by which is meant an integral defined relative to a product measure, and the value of the so-called **iterated integrals**, whereby the integrand is in essence integrated one space variable at a time. We begin this discussion with a review of the construction and essential features of the product space measure developed in book 1, and compare that to an alternative approach which is also popular. The reason for this discussion is that while the product space measures are quite similar, and indeed only differ on sets of measure 0, the statements of Fubini's and Tonelli's theorems reflect the approach taken and will be seen to differ in important ways which we attempt to elucidate.

5.1 Discussion on Product Space Sigma Algebras

The approach to product space measures taken in book 1 is consistent with that of the Royden (1971) text in the references, which also addresses the essential differences of that approach with the alternative approach. The Rudin (1974) text develops the alternative approach and then discusses the approach taken in this book. The alternative approach is also seen in the Billingsley (1995) and Doob (1994) texts.

Recalling chapter 7 of book 1, given σ-finite measure spaces $\{(X_i, \sigma(X_i), \mu_i)|i =$

$1, ..., n\}$, the **product space** $X = \Pi_{i=1}^{n} X_i$ is defined:

$$X = \{(x_1, x_2, ..., x_n) | x_i \in X_i\}.$$

A **measurable rectangle** in X is a set A so that with $A_i \in \sigma(X_i)$:

$$A = \prod\nolimits_{i=1}^{n} A_i = \{x \in X | x_i \in A_i\}.$$

Given a measurable rectangle $\prod_{i=1}^{n} A_i$, the **product set function** μ_0 is defined by:

$$\mu_0(A) = \prod\nolimits_{i=1}^{n} \mu_i(A_i). \tag{5.1}$$

This is the starting point for the approach taken in book 1, as well as for a common alternative approach.

Remark 5.1 (On σ-Finiteness) *We assume throughout this chapter that the component measure spaces $\{(X_i, \sigma(X_i), \mu_i) | i = 1, ..., n\}$ are σ-finite. This is not the most general development possible, but is more than adequate for the applications of these volumes, covering all probability spaces as well as all Borel measure spaces.*

*The product space existence result of proposition 7.20 of book 1 required this assumption. But as noted in remark 7.22 of book 1, the σ-finiteness assumption is not strictly necessary. Given the available tools of book 1, the proof of countable additivity of μ_0 as extended to the algebra $\mathcal{A}$ generated by the measurable triangles utilized a continuity from above argument, which required σ-finiteness. But as seen above in section 2.2.1, this extension of μ_0 to $\mathcal{A}$ is countably additive without this σ-finiteness assumption, and thus so too is the **existence result** of proposition 7.20.*

*However, the assumption of σ-finiteness is required for a reason other than existence. This assumption is required for the **uniqueness result** of the book 1 product measure as detailed by that book's proposition 6.14 and extended by proposition 6.24. Uniqueness is then needed in the current context to employ the result of proposition 7.24 of book 1, which proved that the product space of proposition 7.20 could be created in one step from the n component spaces, or in a sequence of steps, at each step creating product spaces of smaller dimensions. This result will be needed in the current chapter to generalize the Fubini and Tonelli theorems from the case of $n = 2$ which we explicitly prove, to general n, in the case where product spaces are constructed using the book 1 approach. As will be see below in proposition 5.17, to do this requires the identification of the product spaces using proposition 7.24 of book 1. On the other hand, when product spaces are constructed using the alternative method, this extension from $n = 2$ to general* *n does not require σ-finiteness, and this is seen in the first step in the proof of proposition 7.24.*

The essential distinction between the book 1 and alternative approaches is in the construction of the sigma algebra on which the definition of μ_0 can be extended to be a measure μ_X. In both constructions the resulting sigma algebras contain $\mathcal{A}'$, the semi-algebra of measurable rectangles in X, and hence also $\mathcal{A}$, the associated algebra of all finite unions of disjoint measurable rectangles. Also, because μ_0 is defined as above on $\mathcal{A}'$, the extension of μ_0 to $\mathcal{A}$ and denoted μ_A is well defined by finite additivity. Hence up to the algebra $\mathcal{A}$, the measurable sets and measures of these sets are identical between approaches.

The final step is the construction of the sigma algebras.

1. **Chapter 7, Book 1 Approach:** For the product measure space $(X, \sigma(X), \mu_X)$ of proposition 7.20, the sigma algebra $\sigma(X)$ is the collection of **Carathéodory measurable** sets defined in that book's 6.1. This collection is defined using the outer measure μ_A^* which is itself defined in terms of μ_A and the algebra $\mathcal{A}$ as in definition 5.16. By the Hahn–Kolmogorov extension theorem (proposition 6.4) and the first Carathéodory extension theorem (proposition 6.6.2), $\sigma(X)$ is a complete sigma algebra, and denoting by μ_X the restriction of μ_A^* on $\sigma(X)$, it follows that μ_X is a measure on $\sigma(X)$.

   A generalized version of **Littlewood's first principle** was then shown to apply in corollary 7.23, and stated that general measurable sets in $\sigma(X)$ could be approximated in various ways with simpler measurable sets. Specifically, if $B \in \sigma(X)$ and $\epsilon > 0$, then if either $\mu_X(B) < \infty$, or, μ_X is σ-finite:

   (**a**) There is a set $A \in \mathcal{A}_\sigma$, the collection of countable unions of sets in the algebra $\mathcal{A}$, so that $B \subset A$ and:
   $$\mu_X(A) \leq \mu_X(B) + \epsilon, \qquad \mu_X(A - B) < \epsilon.$$

   (**b**) There is a set $C \in \mathcal{A}_\delta$, the collection of countable intersections of sets in the algebra $\mathcal{A}$, so that $C \subset B$ and:
   $$\mu_X(B) \leq \mu_X(C) + \epsilon, \qquad \mu_X(B - C) < \epsilon.$$

   (**c**) There is a set $A' \in \mathcal{A}_{\sigma\delta}$, the collection of countable intersections of sets in $\mathcal{A}_\sigma$, and $C' \in \mathcal{A}_{\delta\sigma}$, the collection of countable unions of sets in $\mathcal{A}_\delta$, so that $C' \subset B \subset A'$ and
   $$\mu_X(A' - B) = \mu_X(B - C') = 0.$$

In other words, the key conclusion in part c is that every measurable set (i.e., every $B \in \sigma(X)$) is within measure 0 of subsets and supersets that can be constructed with countably many unions and intersections of sets from the algebra $\mathcal{A}$. Since the algebra $\mathcal{A}$ is the collection of finite unions of disjoint measurable rectangles in $\mathcal{A}'$, this conclusion can also be stated that every measurable set in $\sigma(X)$ is within measure 0 of subsets and supersets that can be constructed with countably many unions and intersections of sets from the semi-algebra $\mathcal{A}'$.

2. **Alternative Approach (See for example section 18 of Billingsley (1995)):** For the product measure space $(X, \sigma'(X), \mu'_X)$, the sigma algebra $\sigma'(X)$ is defined as the smallest sigma algebra which contains the algebra $\mathcal{A}$, so then by definition $\mathcal{A} \subset \sigma'(X) \subset \sigma(X)$. Thus the conclusion from 1.c. above is that every $B \in \sigma(X)$ is within μ_X-measure 0 of subsets and supersets from $\sigma'(X)$. But then, how is it proved that μ_A, the μ_0 set function extended to $\mathcal{A}$, can be further extended to a measure μ'_X on $\sigma'(X)$?

   The most common approach is to use results from integration theory. For example, assume that there are 2 component spaces to simplify notation, $(X_1, \sigma(X_1), \mu_1)$ and $(X_2, \sigma(X_2), \mu_2)$. For a set $A' \in \sigma'(X)$, define the **projection sets** or **cross-sections** of A' :

$$\begin{aligned} A'(x_1) &= \{x_2 | (x_1, x_2) \in A'\}, \\ A'(x_2) &= \{x_1 | (x_1, x_2) \in A'\}. \end{aligned}$$

   It is then shown that $A'(x_1) \in \sigma(X_2)$ **for all** x_1, and $A'(x_2) \in \sigma(X_1)$ **for all** x_2, and hence $\mu_2 [A'(x_1)]$ and $\mu_1 [A'(x_2)]$ are well defined.

   But more importantly, $\mu_2 [A'(x_1)]$ is a measurable function on X_1, and $\mu_1 [A'(x_2)]$ is a measurable function on X_2, and hence one can define:

$$\int \mu_2 \left[A'(x_1)\right] d\mu_1 \text{ and } \int \mu_1 \left[A'(x_2)\right] d\mu_2.$$

   The next step is to show that for a measurable rectangles, $A' = A'_1 \times A'_2$, these integrals have the desired common value $\mu_1 (A'_1) \mu_2 (A'_2) = \mu_0(A')$ as in 5.1, and further that these integrals agree for all $A' \in \sigma'(X)$. Finally, the common value of these integrals is then taken as the definition of $\mu'_X(A')$, and μ'_X is shown to be a measure using Lebesgue's monotone convergence theorem. The use of such powerful integration results to prove a result on measures was exemplified in the relatively simple proof that the set function μ_0 was countably additive on the semi-algebra $\mathcal{A}'$. See the above section, Product Space Measures Revisited.

Summary 5.2 *Since $\mu'_X(A') = \mu_X(A') = \mu_1(A'_1)\,\mu_2(A'_2)$ for all $A \in \mathcal{A}'$, if μ_X is σ-finite the uniqueness result of proposition 6.14 of book 1 applies. Namely, it follows that $\mu'_X(A') = \mu_X(A')$ for all $A \in \sigma'(X)$.*

Though not obvious, the sigma algebra $\sigma'(X)$ is not complete in general, and so:

$$\sigma'(X) \subsetneqq \sigma(X).$$

As the smallest sigma algebra that contains $\mathcal{A}$, $\sigma'(X)$ must contain any set which can be constructed based on countably many operations of union or intersection, and thus:

$$\mathcal{A}, \mathcal{A}_\sigma, \mathcal{A}_\delta, \mathcal{A}_{\sigma\delta}, \mathcal{A}_{\delta\sigma}..... \subset \sigma'(X) \subsetneqq \sigma(X).$$

So from 1.c. above, since $\sigma(X)$ is complete and is within measure 0 of $\mathcal{A}_{\sigma\delta}$-sets and $\mathcal{A}_{\delta\sigma}$-sets, we can conclude that:

Conclusion 5.3 **1**. *For any $A \in \sigma(X)$, there is $A'_1, A'_2 \in \sigma'(X)$ so that $A'_1 \subset A \subset A'_2$ and*

$$\mu_X(A'_1) = \mu_X(A) = \mu_X(A'_2).$$

2. *If the original spaces $\{(X_i, \sigma(X_i), \mu_i)\}$ are σ-finite then so too is $(X, \sigma(X), \mu_X)$ and so as noted above, $\mu'_X(A') = \mu_X(A')$ for all $A' \in \sigma'(X)$. Thus from Conclusion (1):*

$$\mu_X(A'_1) = \mu'_X(A'_1) = \mu_X(A) = \mu'_X(A'_2) = \mu_X(A'_2).$$

3. *When component measures spaces are σ-finite, $(X, \sigma(X), \mu_X)$ is the **completion** of $(X, \sigma'(X), \mu'_X)$ as constructed in proposition 6.20 of book 1.*

There are two important differences in the results below that are related to the approach taken in constructing the product measure space:

1. **Measurability of Cross-Sections:**

   (a) **Alternative Approach:** In the notation above, **every cross-section of a measurable set** $A' \in \sigma'(X)$ **is measurable** in the appropriate sigma algebra of component spaces.

(b) **Chapter 7, Book 1 Approach:** If $A \in \sigma(X)$, since sets of μ_X-measure 0 can be added or removed from A at will, we can only assert that **cross-sections of measurable sets are measurable except for cross-sections defined relative to sets of measure** 0.

Example 5.4 *In Lebesgue measure space on* $\mathbb{R}^2$, *the set* $A' \equiv \{(1, y)|0 \leq y \leq 1\}$ *is* $\sigma'(\mathbb{R}^2)$*-measurable as the intersection of nested measurable rectangles,* $A'_n \equiv \{(x, y)|1 - 1/n \leq x \leq 1,\ 0 \leq y \leq 1\}$. *As these rectangles have product measure* $\mu(A'_n) = 1/n$, *it follows from continuity from above that* $\mu(A') = 0$.

Now let $E \subset [0, 1]$ *be the nonmeasurable set constructed in proposition 2.31 of book 1, and define* $A'' \equiv \{(1, y)|y \in E\}$. *Then* $A'' \subset A'$ *and thus by completeness* $A'' \in \sigma(\mathbb{R}^2)$ *and* $\mu(A'') = 0$.

Now A'' *has a nonmeasurable cross-section when* $x = 1$, *and because of this it must be the case that* $A'' \notin \sigma'(\mathbb{R}^2)$ *though it is difficult to find a direct and rigorous demonstration of this.*

2. **Component Measurability of Measurable Functions:**

(a) **Alternative Approach:** To simplify notation, if $f(x, y)$ is a $\sigma'(\mathbb{R}^2)$-measurable function, meaning that $f^{-1}(B) \in \sigma'(\mathbb{R}^2)$ for any $B \in \mathcal{B}(\mathbb{R})$, then for every x the component function

$$f_x(y) \equiv f(x, y)$$

is a measurable function of y. Similarly, for every y,

$$f_y(x) \equiv f(x, y)$$

is a measurable function of x. In other words, the **component functions of a measurable function are measurable.**

(b) **Chapter 7, Book 1 Approach:** If $f(x, y)$ is a $\sigma(\mathbb{R}^2)$-measurable function, meaning that $f^{-1}(B) \in \sigma(\mathbb{R}^2)$ for any $B \in \mathcal{B}(\mathbb{R})$, then we can only prove that $f_x(y)$ is measurable for all x outside a set of measure 0, and similarly for $f_y(x)$. In other words, **component functions of a measurable function are measurable except for component functions defined relative to sets of measure** 0. This conclusion is a result of the fact that that $\sigma(\mathbb{R}^2)$-measurability is a weaker condition than $\sigma'(\mathbb{R}^2)$-measurability, because the value of f can be arbitrarily changed on sets in $\mathbb{R}^2$ of μ-measure 0 and remain measurable.

Example 5.5 *On Lebesgue measure space on $\mathbb{R}^2$ define $f(x,y)=1$. This is apparently a $\sigma'(\mathbb{R}^2)$-measurable function since $f^{-1}(B)$ is $\mathbb{R}^2$ or $\emptyset$ depending on whether $1\in B$ or not. Now redefine $f(x,y)$ on a set of measure 0 to:*

$$g(x,y)=\begin{cases}2, & (x,y)\in A'',\\ 1, & (x,y)\notin A'',\end{cases}$$

where A'' is defined in the above example. Then $g=f$ a.e., and thus g is a $\sigma(\mathbb{R}^2)$-measurable function by completeness. Directly, g is $\sigma(\mathbb{R}^2)$-measurable since $A''\in\sigma(\mathbb{R}^2)$. But note that g is not a $\sigma'(\mathbb{R}^2)$-measurable function since $A''\notin\sigma'(\mathbb{R}^2)$. Further, $g_1(y)$ is not measurable, since for example, if $B=[2,3]\in\mathcal{B}(\mathbb{R})$, then $g_1^{-1}(B)=E$.

Summary 5.6 *The chapter 7, book 1 approach to product spaces has the advantage that because the associated sigma algebras are always complete, the product space associated with n, 1-dimensional Lebesgue measure spaces is an n-dimensional Lebesgue measure space. In contrast, the alternative approach requires the product space to be completed to be equivalent to n-dimensional Lebesgue measure space.*

That said, completeness of $\sigma(X)$ imposes a somewhat weaker condition on a set to be measurable, or a function to be measurable, than does the alternative approach which derives $\sigma'(X)$. Corresponding to these somewhat weaker conditions, one obtains somewhat weaker conclusions in that many results are stated to be true "except on a set of measure 0" rather than the stronger statement which omits this qualifier.

More will be said on this below as the results emerge. But the primary focus here will be on the complete product measure space $(X,\sigma(X),\mu_X)$, since completeness is often desirable. We will also state results for the product measure space $(X,\sigma'(X),\mu'_X)$, usually without proof, where $\sigma'(X)$ is the smallest σ-algebra that contains the algebra $\mathcal{A}$ of finite unions of disjoint measurable rectangles. In this latter case, if it is assumed that the component spaces are σ-finite, then as noted above, $\mu'_X=\mu_X$ on $\sigma'(X)$ by proposition 6.14 of book 1.

Notation 5.7 *Reflecting the above discussion, given for example three σ-finite measure spaces $(X,\sigma(X),\mu)$, $(Y,\sigma(Y),\upsilon)$ and $(Z,\sigma(Z),\lambda)$, we denote by $\sigma(X\times Y\times Z)$ the complete sigma algebra of **Carathéodory measurable** sets defined in terms of the set function $\mu\times\nu\times\lambda$ defined on semi-algebra $\mathcal{A}'$ of rectangles in $\sigma(X)\times\sigma(Y)\times\sigma(Z)$ and the associated outer measure*

$(\mu \times \nu \times \lambda)^*$. *If needed for clarity this could be denoted* $\sigma(\sigma(X) \times \sigma(Y) \times \sigma(Z))$. *Similarly,* $\sigma'(X \times Y \times Z)$ *will denote the smallest sigma algebra that contains the semi-algebra* $\mathcal{A}'$ *of rectangles from* $\sigma(X) \times \sigma(Y) \times \sigma(Z)$*, denoted* $\sigma_0(X \times Y \times Z)$ *in book 1. If needed for clarity, this latter sigma algebra can also be denoted* $\sigma'(\sigma(X) \times \sigma(Y) \times \sigma(Z))$.

By the development of section 7.5 of book 1, with slightly different notation:

$$\sigma'(X \times Y \times Z) = \sigma'(\sigma'(X \times Y) \times \sigma(Z)),$$

and similarly by proposition 7.24 of book 1:

$$\sigma(X \times Y \times Z) = \sigma(\sigma(X \times Y) \times \sigma(Z)).$$

In each case the expression on the right is created in two steps, first creating $\sigma'(X \times Y)$ *or* $\sigma(X \times Y)$ *as defined, and then extending the constructions to the next step as defined.*

5.2 Introduction to the Fubini and Tonelli Theorems

We first develop results in a "2-dimensional" product space to simplify notation, and generalize below. But note that here, "dimension" is merely a notational device in that there is nothing in the following development that precludes either of these "1-dimensional spaces" from themselves being product spaces. With that noted, assume that the component spaces are denoted $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$, are σ-finite, and that the complete product space $(X \times Y, \sigma(X \times Y), \mu \times \upsilon)$ is constructed as in chapter 7 of book 1. As in that development, $\mathcal{A}'$ denotes the semi-algebra of measurable rectangles in $X \times Y$, $\mathcal{A}$ the associated algebra of finite disjoint unions, and $\mathcal{A}_\sigma, \mathcal{A}_\delta, \mathcal{A}_{\sigma\delta}$, and $\mathcal{A}_{\delta\sigma}$ are defined as above in the generalized statement of Littlewood's first principle.

Given a function $f(x, y)$ which is integrable on $X \times Y$, meaning

$$\int_{X \times Y} |f(x, y)|\, d(\mu \times \upsilon) < \infty,$$

Fubini's theorem states that

$$\int_X \left[\int_Y f(x, y) d\upsilon\right] d\mu = \int_{X \times Y} f(x, y) d(\mu \times \upsilon) = \int_Y \left[\int_X f(x, y) d\mu\right] d\upsilon.$$

Before addressing the equality of the values of the integrals, this theorem also has to justify that this statement makes sense, meaning that it must also address:

1. The integrability of the component functions, $f_x(y)$ and $f_y(x)$, so that $\int_Y f_x(y)d\upsilon$ and $\int_X f_y(x)d\mu$ are well defined.

2. The integrability of the **iterated integrals**, $\int_Y f_x(y)d\upsilon$ and $\int_X f_y(x)d\mu$, so that $\int_X \left[\int_Y f_x(y)d\upsilon\right] d\mu$ and $\int_Y \left[\int_X f_y(x)d\mu\right] d\upsilon$ are well defined.

Tonelli's theorem states the same conclusion on the equality of these integrals without assuming integrability of $f(x,y)$ on $X \times Y$, instead assuming only that $f(x,y)$ is non-negative and measurable, and then allows the possibility that all of these integrals are infinite. The point is, they will be infinite or finite together, and when finite, the values will agree.

The approach to the proof of either result is to first confirm that these results are true for characteristic functions of measurable sets of finite measure, and then to generalize using the integration tools developed in chapter 2. To address characteristic functions we begin with sets $A \in \mathcal{A}_{\sigma\delta} \subset \sigma'(X \times Y)$ defined above, and we will show that all cross-sections are measurable, and the measure of these cross-sections can be expressed as measurable functions of the defining variables. We then generalize to characteristic functions of sets $A \in \sigma(X \times Y)$, recalling the generalized Littlewood's first principle above that such sets are within $\mu \times \upsilon$-measure 0 of $\mathcal{A}_{\sigma\delta}$-sets. In this case we will see that cross-section measurability is now only true almost everywhere, and the measure of these cross-sections can be expressed as a measurable function that is defined almost everywhere.

5.3 Preliminary Results for Characteristic Functions

We begin by formalizing definitions and notations introduced above, and then show that Fubini's theorem holds for characteristic functions of sets $A \in \sigma(X \times Y)$, where $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$ are σ-finite measure spaces, and $(X \times Y, \sigma(X \times Y), \mu \times \upsilon)$ is the complete product space constructed as in chapter 7 of book 1.

Definition 5.8 *Given a set $A \subset X \times Y$, we define the* ***cross-sections of***

A by:

$$\begin{aligned} &x\text{-cross-section} : \quad A_x = \{y|(x,y) \in A\}, \\ &y\text{-cross-section} : \quad A_y = \{x|(x,y) \in A\}. \end{aligned} \tag{5.2}$$

By definition, $A_x \subset Y$ *for all* x*, and* $A_y \subset X$ *for all* y*.*

If $f(x,y)$ *is a function defined on* $X \times Y$*, we define the* ***component functions of*** $f(x,y)$ *by:*

$$\begin{aligned} &x\text{-component function} : \quad f_x(y) = f(x,y), \\ &y\text{-component function} : \quad f_y(x) = f(x,y). \end{aligned} \tag{5.3}$$

Recall that $\mathcal{A}' \subset \sigma(X \times Y)$ denotes the semi-algebra of measurable rectangles, $\mathcal{A}$ the associated algebra of finite disjoint unions, and related collections such as $\mathcal{A}_{\sigma\delta}$ are defined as before. The first result is that for sets $A \in \mathcal{A}_{\sigma\delta}$, all cross-sections are measurable and when $\mu \times \upsilon(A) < \infty$, the measures of these cross-sections are integrable functions of the defining variable.

Remark 5.9 *The next proposition is Fubini's theorem restricted to characteristic functions* $f(x,y) = \chi_A(x,y)$ *for* $A \in \mathcal{A}_{\sigma\delta}$ *with* $\mu \times \upsilon(A) < \infty$*. This follows because for all* x, y *:*

$$\chi_A(x,y) = \chi_{A_x}(y) = \chi_{A_y}(x),$$

and thus if appropriately measurable:

$$\upsilon(A_x) = \int \chi_A(x,y)d\upsilon(y), \qquad \mu(A_y) = \int \chi_A(x,y)d\mu(x).$$

Since

$$\int \chi_A(x,y)d(\mu \times \upsilon) \equiv \mu \times \upsilon(A),$$

the result in 5.4 can be expressed:

$$\int \left[\int \chi_A(x,y)d\upsilon\right] d\mu = \int \chi_A(x,y)d(\mu \times \upsilon) = \int \left[\int \chi_A(x,y)d\mu\right] d\upsilon.$$

Note that we do not assume σ*-finiteness of measure spaces for this next result, and hence the need for* $\mu \times \upsilon(A) < \infty$ *in part 2 to utilize the generalized version of Littlewood's first principle noted above.*

Proposition 5.10 *Let $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$ be complete measure spaces and $A \in \mathcal{A}_{\sigma\delta}$. Then:*

1. $A_x \in \sigma(Y)$ *for all* x, *and* $A_y \in \sigma(X)$ *for all* y.
2. *If* $\mu \times \upsilon(A) < \infty$, *then* $g(x) \equiv \upsilon(A_x)$ *is* μ*-integrable on* X, $h(y) \equiv \upsilon(A_y)$ *is* ν*-integrable on* Y, *and:*
$$\int g(x) d\mu = \mu \times \upsilon(A) = \int h(y) d\upsilon. \tag{5.4}$$

Proof.

1. *By symmetry, only the statement regarding* A_x *needs proof. If* $A \in \mathcal{A}'$, *a measurable rectangle, then the result follows by definition. Assume next that* $A \in \mathcal{A}_\sigma$, *say* $A = \cup_{j=1}^{\infty} A_j$ *with* $A_j \in \mathcal{A}$. *By definition, each such* A_j *is a finite union of disjoint* $\mathcal{A}'$*-sets, and hence this characterization can be modified by relabelling to assume that* $A_j \in \mathcal{A}'$. *Then* $\chi_A(x, y) = \sup_j \chi_{A_j}(x, y)$, *and so*
$$\chi_{A_x}(y) = \sup\nolimits_j \chi_{(A_j)_x}(y).$$
Each $(A_j)_x \in \sigma(Y)$ *because* $A_j \in \mathcal{A}'$ *is a measurable rectangle, and thus* $\chi_{(A_j)_x}(y)$ *is a* ν*-measurable function. As the supremum of* ν*-measurable functions, it follows that* $\chi_{A_x}(y)$ *is a* ν*-measurable function and so* A_x *is measurable. That is,* $A_x \in \sigma(Y)$.

   Finally, for $A \in \mathcal{A}_{\sigma\delta}$, *say* $A = \cap_{j=1}^{\infty} A_j$ *with* $A_j \in \mathcal{A}_\sigma$, *the same steps lead to*
$$\chi_{A_x}(y) = \inf\nolimits_j \chi_{(A_j)_x}(y),$$
and hence $\chi_{A_x}(y)$ *is a* ν*-measurable function and* A_x *is* ν*-measurable.*

2. *Again by symmetry we only prove the result for* $g(x)$, *and use the same sequential approach as part 1. If* $A \in \mathcal{A}'$ *then the results follow as an exercise and is left for the reader.*

   (**a**) *Assume* $A \in \mathcal{A}_\sigma$, *say* $A = \cup_{j=1}^{\infty} A_j$ *with* $A_j \in \mathcal{A}'$ *as noted in part 1. Define* $A^{(n)} = \cup_{j=1}^{n} A_j$, *then* $\{A^{(n)}\}$ *is a nested increasing sequence of measurable sets,* $A^{(n)} \subset A^{(n+1)}$, *and by continuity from below,*
$$\mu \times \upsilon(A) = \lim_{n \to \infty} \mu \times \upsilon\left(A^{(n)}\right).$$

Using the approach of proposition 7.13 in book 1, $A^{(n)}$ *can be expressed as a finite union of disjoint measurable rectangles. Specifically, if each* $A_j = B_j \times C_j$ *then* $\bigcup_{j=1}^{n} B_j$ *and* $\bigcup_{j=1}^{n} C_j$ *can each be expressed as unions of no more that* $M = 2^n - 1$ *disjoint sets, many of which could be empty. If* $\{B'_j\}_{j=1}^{M}$ *and* $\{C'_j\}_{j=1}^{M}$ *are the resulting collections then* $\{B'_j \times C'_k\}_{j,k=1}^{M}$ *are* M^2 *disjoint measurable rectangles. For every* j, k, *either* $B'_j \times C'_k \subset A^{(n)}$ *or* $B'_j \times C'_k \cap A^{(n)} = \emptyset$, *and we choose the subset rectangles and note by construction that they are disjoint and with union* $A^{(n)}$. *Relabel these so* $A^{(n)} = \cup_{j=1}^{N(n)} A_j^{(n)}$ *with* $\{A_j^{(n)}\}_{j=1}^{N(n)}$ *disjoint. If* $g_j^{(n)}(x) \equiv \upsilon\left[\left(A_j^{(n)}\right)_x\right]$, *then* $g_j^{(n)}(x)$ *is nonnegative and measurable since* $A_j^{(n)} \in \mathcal{A}'$. *Defining* $g^{(n)}(x) = \sum_{j=1}^{N(n)} g_j^{(n)}(x)$, *then* $g^{(n)}(x)$ *is again measurable and increasing in* n. *By disjointedness of* $\left\{\left(A_j^{(n)}\right)_x\right\}_{j=1}^{N(n)}$ *and finite additivity of* ν,

$$g^{(n)}(x) \equiv \sum_{j=1}^{N(n)} \upsilon\left[\left(A_j^{(n)}\right)_x\right] = \nu\left[A_x^{(n)}\right].$$

But then by continuity from below:

$$g(x) \equiv \nu\left[A_x\right] = \lim_{n\to\infty} g^{(n)}(x),$$

and so $g(x)$ *is measurable.*
Linearity of the integral and finite additivity of $\mu \times \upsilon$ *then yield:*

$$\begin{aligned} \int g^{(n)}(x)d\mu &= \textstyle\sum_{j=1}^{N(n)} \displaystyle\int g_j^{(n)}(x)d\mu \\ &= \textstyle\sum_{j=1}^{N(n)} \mu \times \upsilon\left(A_j^{(n)}\right) \\ &= \mu \times \upsilon(A^{(n)}). \end{aligned}$$

Finally, by Lebesgue's monotone convergence theorem and continuity from below:

$$\begin{aligned} \int g(x)d\mu &= \lim_{n\to\infty} \int g^{(n)}(x)d\mu \\ &= \lim_{n\to\infty} \mu \times \upsilon(A^{(n)}) \\ &= \mu \times \upsilon\left(A\right). \end{aligned}$$

Hence, part 2 of the proposition is proved for $A \in \mathcal{A}_\sigma$.

(**b**) *Assume $A \in \mathcal{A}_{\sigma\delta}$ with $\mu \times \upsilon(A) < \infty$. Then $A = \cap_{j=1}^{\infty} A_j$ with $A_j \in \mathcal{A}_{\sigma}$. We can assume that $\{A_j\}$ are nested with $A_{j+1} \subset A_j$ since given a general representation $A = \cap_{i=1}^{\infty} B_i$ with $B_i \in \mathcal{A}_{\sigma}$, we define $A_j = \cup_{i=j}^{\infty} B_i$, noting that $A_j \in \mathcal{A}_{\sigma}$ by definition. The assumption that $\mu \times \upsilon(A) < \infty$ implies by the approximation results noted above that A has a superset $D \in \mathcal{A}_{\sigma}$, so $A \subset D$, and the measure of D is no more than $\mu \times \upsilon(A) + \epsilon$. By redefining the A_j-set by $A_j \cap D \in \mathcal{A}_{\sigma}$ we can assume that $\mu \times \upsilon(A_1) < \infty$. If $g_j(x) \equiv \upsilon\left[(A_j)_x\right]$, then by part 2.a. $g_j(x)$ is nonnegative and measurable, and*

$$\int g_1(x)d\mu = \mu \times \upsilon(A_1) < \infty,$$

and so $g_1(x) < \infty$ μ-a.e. Hence for any x with $g_1(x) < \infty$, $\{(A_j)_x\}$ is a nested decreasing collection of sets of finite measure with intersection A_x. By continuity from above of υ we conclude that $g(x) = \lim_j g_j(x)$ μ-a.e. and hence $g(x)$ is measurable by completeness. Finally, because $0 \leq g_j(x) \leq g_1(x)$ and $g_1(x)$ is integrable, we can apply Lebesgue's dominated convergence theorem and continuity from above to conclude that:

$$\begin{aligned}\int g(x)d\mu &= \lim_{j\to\infty} \int g_j(x)d\mu \\ &= \lim_{j\to\infty} \mu \times \upsilon(A_j) \\ &= \mu \times \upsilon(A),\end{aligned}$$

noting that continuity from above of $\mu \times \upsilon$ is justified by $\mu \times \upsilon(A_1) < \infty$.

■

Remark 5.11 *One conclusion that can be drawn from the above proposition is that if $A \in \mathcal{A}_{\sigma\delta}$, then $\mu \times \upsilon(A) = 0$ if and only if both $\upsilon(A_x) = 0$ for almost all x on X, and, $\mu(A_y) = 0$ for almost all y in Y. This follows because $\mu \times \upsilon(A) = 0$ if and only if:*

$$\int g(x)d\mu = \int h(y)d\upsilon = 0,$$

and since nonnegative, this is true if and only if $g(x) = 0$ μ-a.e. and $h(y) = 0$ ν-a.e. Put another way, if $\mu \times \upsilon(A) < \infty$ and $\mu \times \upsilon(A) \neq 0$, then both $\upsilon(A_x)$ and $\mu(A_y)$ must be non-zero on sets of positive measure.

*An important corollary to this result is that for any set $A \in \sigma(X \times Y)$ with $\mu \times \upsilon(A) = 0$, **almost all** cross-sections have measure 0 relative to the component measures. This transition from $\mathcal{A}_{\sigma\delta}$ to $\sigma(X \times Y)$ is based on another application of the approximation results of corollary 7.23 of book 1 noted above.*

Corollary 5.12 *Let $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$ be complete measure spaces. If $A \in \sigma(X \times Y)$ and $\mu \times \upsilon(A) = 0$, then $\upsilon(A_x) = 0$ for μ-almost-all x, and $\mu(A_y) = 0$ for ν-almost-all y.*

Proof. *By corollary 7.23 of book 1 there is a set $A' \in \mathcal{A}_{\sigma\delta}$ so that $A \subset A'$ and $\mu \times \upsilon(A') = 0$. Hence $\upsilon(A'_x) = 0$ for almost all x and $\mu(A'_y) = 0$ for almost all y by remark 5.11. Then from $A_x \subset A'_x$ and $A_y \subset A'_y$ and the completeness of the measure spaces, the result follows.* ■

The next result generalizes proposition 5.10 to $A \in \sigma(X \times Y)$. As above, we do not assume σ-finiteness. But note that for measurable sets outside $\mathcal{A}_{\sigma\delta}$, the conclusions in general switch from "all x" to "almost all x", and similarly for the y-statements. However, this proposition is one step closer to Fubini's theorem in that by 5.5, remark 5.9 now applies to characteristic functions $f(x, y) = \chi_A(x, y)$ for general $A \in \sigma(X \times Y)$ with $\mu \times \upsilon(A) < \infty$.

Proposition 5.13 *Let $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$ be complete measure spaces. If $A \in \sigma(X \times Y)$ with $\mu \times \upsilon(A) < \infty$, then:*

1. *$A_x \in \sigma(Y)$ for μ-almost-all x, and $A_y \in \sigma(X)$ for ν-almost-all y.*

2. *$g(x) \equiv \upsilon(A_x)$ is defined for μ-almost-all x and is μ-integrable, $h(y) \equiv \upsilon(A_y)$ is defined for ν-almost-all y and is υ-integrable, and,*

$$\int g(x) d\mu = \mu \times \upsilon(A) = \int h(y) d\upsilon. \tag{5.5}$$

Proof. *As in the proof of the above corollary, given such A there is a set $A' \in \mathcal{A}_{\sigma\delta}$ so that $A \subset A'$ and $\mu \times \upsilon(A') = \mu \times \upsilon(A)$. If $B \equiv A' - A$, then $B \in \sigma(X \times Y)$ and finite additivity applied to $A' = A \cup B$ obtains $\mu \times \upsilon(B) = 0$. By the above corollary $\upsilon(B_x) = 0$ for μ-almost-all x, and $\mu(B_y) = 0$ for ν-almost-all y. Hence $g(x) \equiv \upsilon(A_x) = \upsilon(A'_x)$ μ-a.e., and $h(y) \equiv \upsilon(A_y) = \upsilon\left(A'_y\right)$ υ-a.e. But $\upsilon(A'_x)$ is μ-measurable by proposition 5.10 and thus by completeness of $(X, \sigma(X), \mu)$, $g(x)$ is equal to a μ-measurable function μ-a.e and thus by completeness is μ-measurable. The same argument applies to the ν-measurability of $h(y)$.*

By 5.4:

$$\int g(x)d\mu = \int \upsilon\left(A'_x\right)d\mu = \mu \times \upsilon\left(A'\right) = \mu \times \upsilon\left(A\right),$$

and similarly for $\int h(y)d\upsilon$, *which proves 5.5.* ■

For completeness, we present a parallel result for the sigma algebra $\sigma'(X \times Y)$.

Proposition 5.14 *Let* $(X, \sigma(X), \mu)$ *and* $(Y, \sigma(Y), \upsilon)$ *be measure spaces and* $A \in \sigma'(X \times Y)$, *the smallest* σ*-algebra that contains the algebra* $\mathcal{A}$. *Then:*

1. $A_x \in \sigma(Y)$ *for all* x, *and* $A_y \in \sigma(X)$ *for all* y.
2. $g(x) \equiv \upsilon(A_x)$ *is defined for all* x *and is* μ*-integrable,* $h(y) \equiv \upsilon(A_y)$ *is defined for all* y *is a* ν*-integrable, and 5.5 is satisfied, though all integrals may be infinite.*

Proof. *We prove part 1 because it is needed in book 7 in the section, Stochastic Processes and Their Measurability. For other details see Billingsley (1995) for example.*

By symmetry only the first statement of part 1 requires proof. Given $x \in X$, *define a mapping* $f_x : Y \to X \times Y$ *by* $f_x(y) = (x, y)$. *If* $A \in \mathcal{A}' \subset \sigma'(X \times Y)$ *is a measurable rectangle, say* $A = E \times F$, *then* $f_x^{-1}(A) \in \sigma(Y)$ *since* $f_x^{-1}(A) = F$ *if* $x \in E$ *and* $f_x^{-1}(A) = \emptyset$ *if* $x \notin E$. *Similarly, if* $A \in \mathcal{A}$ *is a finite union of measurable rectangles,* $A = \bigcup_{j=1}^{n} E_j \times F_j$, *then* $f_x^{-1}(A) \in \sigma(Y)$ *as a finite union of the* F_j*-sets for which* $x \in E_j$. *For such* $A \in \mathcal{A}$ *it follows that* $f_x^{-1}(\widetilde{A}) = Y - f_x^{-1}(A) \in \sigma(Y)$, *while if* $\{A_j\}_{j=1}^{\infty} \subset \mathcal{A}$, *then* $f_x^{-1}(\bigcup_{j=1}^{\infty} A_j) = \bigcup_{j=1}^{\infty} f_x^{-1}(A_j) \in \sigma(Y)$. *Thus since* $\mathcal{A}$ *generates* $\sigma'(X \times Y)$, *it follows that* $f_x^{-1}[\sigma'(X \times Y)] \subset \sigma(Y)$. *As* $f_x^{-1}(A) = A_x$, *this proves that* $A_x \in \sigma(Y)$ *for all* x. ■

5.4 Fubini's Theorem

As noted above, **Fubini's theorem** is named for **Guido Fubini** (1879 – 1943). By remark 5.9 and proposition 5.13, this result is already proved for $f(x, y) = \chi_A(x, y)$ for $A \in \sigma(X \times Y)$ with $\mu \times \upsilon(A) < \infty$. Such functions are automatically $\mu \times \upsilon$-integrable. The final result below will state that for any such integrable function, that the properties proved above for special characteristic functions generalize. The transition from characteristic functions to general measurable functions will require the earlier results of corollary 1.22 on approximating measurable functions with simple functions.

Proposition 5.15 (Fubini's theorem) *Let $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$ be complete, σ-finite measure spaces, and $f(x, y)$ an integrable function on $(X \times Y, \sigma\left(X \times Y\right), \mu \times \upsilon)$:*

$$\int_{X \times Y} |f(x, y)|\, d\left(\mu \times \upsilon\right) < \infty. \tag{5.6}$$

Then:

1. *For almost all x, $f_x(y)$ is υ-integrable on Y, and,*

1'. *For almost all y, $f_y(x)$ is μ-integrable on X.*

2. *$\int_Y f_x(y) d\upsilon \equiv \int_Y f(x, y) d\upsilon$ is defined for μ-almost-all x and is μ-integrable on X, and,*

2'. *$\int_X f_y(x) d\mu \equiv \int_X f(x, y) d\mu$ is defined for ν-almost-all y and is ν-integrable on Y.*

3. *The $\mu \times \upsilon$-integral of $f(x, y)$ can be evaluated as an iterated integral:*

$$\int_X \left[\int_Y f(x, y) d\upsilon\right] d\mu = \int_{X \times Y} f(x, y) d\left(\mu \times \upsilon\right) = \int_Y \left[\int_X f(x, y) d\mu\right] d\upsilon. \tag{5.7}$$

Proof. *By symmetry, we will only prove half of the results, specifically, parts 1, 2 and the first half of 3. And for this we can assume that $f(x, y)$ is nonnegative, since by linearity and the assumption that $f(x, y)$ is integrable, this result can then be applied to general $f \equiv f^+ - f^-$.*

Recalling remark 5.9, proposition 5.13 proves the desired results for $f(x, y) = \chi_A(x, y)$ for $A \in \sigma\left(X \times Y\right)$ with $\mu \times \upsilon\left(A\right) < \infty$. Hence by linearity these results are true for simple functions which are zero outside sets of finite measure. By corollary 1.22, given nonnegative integrable $f(x, y)$ there is an increasing sequence of nonnegative simple functions $\{\varphi_n(x, y)\}$, each equal to zero outside a set of finite measure, and $\varphi_n(x, y) \to f(x, y)$. Hence $\left(\varphi_n\right)_x(y) \to f_x(y)$, and thus $f_x(y)$ is ν-measurable. By Lebesgue's monotone convergence theorem,

$$\int_Y f_x(y) d\upsilon = \lim_{n \to \infty} \int_Y \varphi_n(x, y) d\upsilon,$$

and so the μ-measurability of $\int_Y \varphi_n(x,y)dv$ proves the same for $\int_Y f_x(y)dv$. Finally, applying Lebesgue's theorem again,

$$\begin{aligned}\int_X \left[\int_Y f(x,y)dv\right] d\mu &= \lim_{n\to\infty} \int_X \left[\int_Y \varphi_n(x,y)dv\right] d\mu \\ &= \lim_{n\to\infty} \int_{X\times Y} \varphi_n(x,y)d\left(\mu \times v\right) \\ &= \int_{X\times Y} f(x,y)d\left(\mu \times v\right),\end{aligned}$$

which proves 5.7. ■

Remark 5.16 *It should be noted that there was only one step in the proof of Fubini's theorem where the assumption of integrability of $f(x,y)$ was needed, and that was in the step where we approximated $f(x,y)$ by a sequence of simple functions. The integrability of $f(x,y)$ allowed the assumption that these simple functions were equal to zero outside sets of finite measure. This conclusion was needed in order to apply the result in proposition 5.13, which confirmed the validity of Fubini's results for characteristic functions of sets of finite measure.*

As will be seen below, Tonelli's assumptions allow the same approximating sequence of simple functions.

See the section below, Fubini's Theorem in $\sigma'(X \times Y)$, for another version of Fubini's theorem applicable in the measure space $(X\times Y, \sigma'(X \times Y), \mu\times v)$ discussed in this chapter's introduction. In this case the "almost all " statements of 1, $1'$, 2, and $2'$ are changed to "all."

5.4.1 Generalizing Fubini's Theorem

The above theorem, as well as Tonelli's theorem below, are explicitly stated and proved in terms of a bivariate function $f(x,y)$. As noted in the introduction, while these results appear to apply only in a "2-dimensional" product space, "dimension" here is merely a notational device and there was nothing in the above development that precluded either of these "1-dimensional spaces" from themselves being product spaces. But more importantly, these results are more generally applicable to higher "dimensions" as is verified by an iterative application of the two dimensional result.

For example, assume that $(X, \sigma(X), \mu)$, $(Y, \sigma(Y), \upsilon)$ and $(Z, \sigma(Z), \lambda)$ are complete, σ-finite measure spaces and $f(x, y, z)$ an integrable function on $(X \times Y \times Z, \sigma(X \times Y \times Z), \mu \times \upsilon \times \lambda)$:

$$\int_{X \times Y \times Z} |f(x, y, z)|\, d(\mu \times \upsilon \times \lambda) < \infty.$$

The key step in iterating the above result is the identification of $(X \times Y \times Z, \sigma(X \times Y \times Z), \mu \times \upsilon \times \lambda)$ and $(X \times (Y \times Z), \sigma(X \times (Y \times Z)), \mu \times (\upsilon \times \lambda))$ as measure spaces. The notation implies that the triple product measure space is identical with the product measure space created with $(X, \sigma(X), \mu)$ and $(Y \times Z, \sigma(Y \times Z), \upsilon \times \lambda)$. This identification was proved in proposition 7.24 of book 1, for which the needed uniqueness result of that book's proposition 6.14 required that the measure spaces be σ-finite.

We can conclude from this result the following re-interpretation of proposition 5.15. We leave further generalizations to the reader.

Proposition 5.17 *Assume that $(X, \sigma(X), \mu)$, $(Y, \sigma(Y), \upsilon)$ and $(Z, \sigma(Z), \lambda)$ are complete, σ-finite measure spaces, and that $f(x, y, z)$ an integrable function on $(X \times Y \times Z, \sigma(X \times Y \times Z), \mu \times \upsilon \times \lambda)$. Then:*

1. *For μ-almost-all x, $f_x(y, z) = f(x, y, z)$ is $\upsilon \times \lambda$-integrable on $Y \times Z$, and,*

1'. *For $\upsilon \times \lambda$-almost-all (y, z), $f_{(y,z)}(x) = f(x, y, z)$ is μ-integrable on X.*

2. *$\int_{Y \times Z} f(x, y, z) d(\upsilon \times \lambda)$ is μ-integrable on X, and,*

2'. *$\int_X f(x, y, z) d\mu$ is $\nu \times \lambda$-integrable on $Y \times Z$.*

3. *The $\mu \times \upsilon \times \lambda$-integral of $f(x, y, z)$ can be evaluated as an iterated integral:*

$$\begin{aligned}\int_X \left[\int_{Y \times Z} f(x, y, z) d(\upsilon \times \lambda)\right] d\mu &= \int_{X \times Y \times Z} f(x, y, z) d(\mu \times \upsilon \times \lambda) \\ &= \int_{Y \times Z} \left[\int_X f(x, y, z) d\mu\right] d(\upsilon \times \lambda).\end{aligned}$$

Remark 5.18 *This is not the final statement of Fubini's theorem in three variables, but provides the iterative step. By 2', we can apply Fubini's theorem to the $\upsilon \times \lambda$-integral on the right in 3 to conclude that*

$$\begin{aligned}\int_Y \left[\int_Z \left[\int_X f(x, y, z) d\mu\right] d\lambda\right] d\nu &= \int_{Y \times Z} \left[\int_X f(x, y, z) d\mu\right] d(\upsilon \times \lambda) \\ &= \int_Z \left[\int_Y \left[\int_X f(x, y, z) d\mu\right] d\nu\right] d\lambda.\end{aligned}$$

Similarly, by 1 and 2 we can apply Fubini's theorem to the $\upsilon \times \lambda$-integral on the left of 3 to conclude that

$$\int_X \left[\int_Z \left[\int_Y f(x,y,z)d\nu\right] d\lambda\right] d\mu = \int_X \left[\int_{Y\times Z} f(x,y,z)d\left(\upsilon \times \lambda\right)\right] d\mu$$
$$= \int_X \left[\int_Y \left[\int_Z f(x,y,z)d\lambda\right] d\nu\right] d\mu.$$

Hence, under the assumption of integrability under the product measure, the value of the integral under this measure equals the value of the iterated integrals:

$$\begin{aligned} &\int_X \left[\int_Y \left[\int_Z f(x,y,z)d\lambda\right] d\nu\right] d\mu \\ = &\int_{X\times Y\times Z} f(x,y,z)d\left(\mu \times \upsilon \times \lambda\right) \qquad (5.8) \\ = &\int_Z \left[\int_Y \left[\int_X f(x,y,z)d\mu\right] d\nu\right] d\lambda. \end{aligned}$$

5.4.2 Fubini's Theorem in $\sigma'(X \times Y)$

In the above development of Fubini's theorem, the first step was the verification that this theorem was true for characteristic functions of measurable sets of finite measure. The first and easiest step in this verification was for characteristic functions of $A \in \mathcal{A}'$, the semi-algebra of measurable rectangles. Recalling the above section on the **functional monotone class theorem**, this step also provides one of the key properties to verify in order to conclude from that theorem that the same Fubini result was true for all bounded measurable functions. Here measurability is defined relative to $\sigma(\mathcal{A}')$, the smallest sigma algebra generated by $\mathcal{A}'$. It is thus natural to inquire as to the potential for this earlier result to prove a version of Fubini's theorem.

For this inquiry, we will therefore work with the sigma algebra $\sigma'(X \times Y)$, defined as the smallest sigma algebra which contains the algebra $\mathcal{A}$, which in turn is generated by the semi-algebra of measurable rectangles $\mathcal{A}'$. So in the notation of the functional monotone class theorem,

$$\sigma(\mathcal{A}') \equiv \sigma'(X \times Y) \subset \sigma(X \times Y).$$

To ensure the uniqueness of the product set function in 5.1 as an extension from $\mathcal{A}'$ to $\sigma(\mathcal{A}')$, we will require that $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$ be σ-finite measure spaces.

With this introduction, let $(X \times Y, \sigma'(X \times Y), \mu \times \upsilon)$ denote the product space referenced in this chapter's introduction under the **Alternative Approach,** whereby $\sigma'(X \times Y)$ is defined as above, and $\mu \times \upsilon$ is uniquely determined on $\sigma'(X \times Y)$ as an extension of the measure of measurable rectangles $A \in \mathcal{A}'$ as defined in 5.1. In the notation of the functional monotone class theorem, let $\mathcal{L}$ denote a class of functions $f(x, y)$ on $X \times Y$ which satisfy a given statement of Fubini's theorem. To demonstrate that $\mathcal{L}$ contains all bounded $\sigma'(X \times Y)$-measurable functions on $X \times Y$, it must be shown that:

1. $\chi_A \in \mathcal{L}$ for all $A \in \mathcal{A}'$ and $\chi_X \in \mathcal{L}$.

2. $\mathcal{L}$ is a vector space: If $f, g \in \mathcal{L}$ then $af + bg \in \mathcal{L}$ for all $a, b \in \mathbb{R}$.

3'. If $f : X \to \mathbb{R}^+$ is bounded and nonnegative and the pointwise limit of increasing $\{f_n\} \subset \mathcal{L}$, then $f \in \mathcal{L}$.

Note that condition 3 of proposition 1.32 is replaced by $3'$ as discussed in remark 1.34.

The next result is Fubini's theorem applied to $(X \times Y, \sigma'(X \times Y), \mu \times \upsilon)$ in the more limited case of finite component measure spaces, such as probability spaces, without the assumption on completeness. See Billingsley (1995) for the more general result.

Proposition 5.19 (Fubini's theorem) *Let $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$ be finite measure spaces, and $f(x, y)$ a bounded and measurable, and thus integrable, function on $(X \times Y, \sigma'(X \times Y), \mu \times \upsilon)$. Then:*

1. *For all x, $f_x(y)$ is υ-integrable on Y, and,*

1'. For all y, $f_y(x)$ is μ-integrable on X.

2. $\int_Y f_x(y) d\upsilon \equiv \int_Y f(x, y) d\upsilon$ is defined for all x and is μ-integrable on X, and,

2'. $\int_X f_y(x) d\mu \equiv \int_X f(x, y) d\mu$ is defined for all y and is ν-integrable on Y.

3. The $\mu \times \upsilon$-integral of $f(x, y)$ can be evaluated as an iterated integral:

$$\int_X \left[\int_Y f(x, y) d\upsilon\right] d\mu = \int_{X \times Y} f(x, y) d(\mu \times \upsilon) = \int_Y \left[\int_X f(x, y) d\mu\right] d\upsilon. \tag{5.9}$$

Proof. *Let $\mathcal{L}$ denote the class of measurable functions which satisfy all five statements above. Checking the criteria for the functional monotone class theorem, the above proposition 5.13 proved that $\chi_A(x,y) \in \mathcal{L}$ for $A \in \mathcal{A}'$ with $\mu \times \upsilon(A) < \infty$, while the finiteness of the measure space makes this true for all $A \in \mathcal{A}'$. Also, $\mathcal{L}$ is a vector space over $\mathbb{R}$ by earlier proved properties of measurable functions and linearity of integrals.*

Now if $f : X \to \mathbb{R}^+$ is bounded and nonnegative and the pointwise limit of increasing $\{f_n\} \subset \mathcal{L}$, then 1, 1′, 2, 2′ are satisfied by the monotone convergence theorem. For 3, by definition of $\mathcal{L}$:

$$\int_X \left[\int_Y f_n(x,y)d\upsilon\right] d\mu = \int_{X\times Y} f_n(x,y)d\left(\mu \times \upsilon\right) = \int_Y \left[\int_X f_n(x,y)d\mu\right] d\upsilon$$

for all n. Lebesgue's monotone convergence theorem then obtains:

$$\int_{X\times Y} f_n(x,y)d\left(\mu \times \upsilon\right) \to \int_{X\times Y} f(x,y)d\left(\mu \times \upsilon\right).$$

Similarly, for all x :

$$\int_Y f_n(x,y)d\upsilon \to \int_Y f(x,y)d\upsilon,$$

while another application of monotone convergence yields the final result. \

Thus by the functional monotone class theorem, $\mathcal{L}$ contains all bounded measurable functions. ■

Corollary 5.20 *Proposition 5.19 is valid for all integrable functions on $(X \times Y, \sigma'(X \times Y), \mu \times \upsilon)$.*

Proof. *To prove that the class $\mathcal{L}$ of proposition 5.19 contains all integrable functions, given nonnegative integrable $f(x,y)$, define $f_n(x,y) = \max\left[f(x,y), n\right]$. Then $f_n(x,y) \to f(x,y)$ pointwise, and $f_n(x,y) \in \mathcal{L}$ since bounded. Thus 5.9 obtains:*

$$\int_X \left[\int_Y f_n(x,y)d\upsilon\right] d\mu = \int_{X\times Y} f_n(x,y)d\left(\mu \times \upsilon\right) = \int_Y \left[\int_X f_n(x,y)d\mu\right] d\upsilon.$$

Since $\{f_n(x,y)\}$ is increasing, Lebesgue's monotone convergence theorem yields $\int_{X\times Y} f_n(x,y)d\left(\mu \times \upsilon\right) \to \int_{X\times Y} f(x,y)d\left(\mu \times \upsilon\right)$, and focusing on the first equality, $\int_Y f_n(x,y)d\upsilon \to \int_Y f(x,y)d\upsilon$ for all x. Letting $g_n(x) \equiv \int_Y f_n(x,y)d\upsilon$, then $g_n(x)$ is measurable by proposition 5.19 and increasing. Again by Lebesgue's monotone convergence theorem, $\int_X \left[\int_Y f_n(x,y)d\upsilon\right] d\mu \to \int_X \left[\int_Y f(x,y)d\upsilon\right] d\mu$ where this limit function is finite by 5.9. Thus:

$$\int_X \left[\int_Y f(x,y)d\upsilon\right] d\mu = \int_{X\times Y} f(x,y)d\left(\mu \times \upsilon\right).$$

A similar argument applies to the second equality in 5.9.

The result now applies to general integrable f, using the decomposition into nonnegative functions in 2.17: $f = f^+ - f^-$, and linearity of integrals.

■

Remark 5.21 *Note that if $f(x,y)$ is integrable, then 5.9 only implies that $\int_Y f(x,y)d\upsilon$ is finite μ-a.e, and similarly for $\int_X f(x,y)d\mu$. In the above proof, while $\int_Y f_n(x,y)d\upsilon \to \int_Y f(x,y)d\upsilon$ for all x, this limit need not be finite for all x. But as $\int_X \left[\int_Y f(x,y)d\upsilon\right] d\mu$ is finite, this assures that this limit is finite for almost all x.*

5.5 Tonelli's Theorem

As noted above, **Tonelli's Theorem** is named for **Leonida Tonelli** (1885 – 1946) and addresses the same question as does Fubini's theorem. Specifically, it addresses the relationship between product space and iterated integrals, but circumvents Fubini's sometimes difficult to establish assumption on the integrability of $f(x,y)$. Tonelli requires only that the function be nonnegative and measurable.

Because integrability of the function $f(x,y)$ is not assumed, all the statements in Fubini's result related to integrability are now changed to statements of measurability. In addition, the identity in 5.7 repeated below in 5.10 must now also allow for the case where all integrals are infinite. In other words, measurability of $f(x,y)$ and σ-finiteness of the component spaces does not assure the integrability of $f(x,y)$. But these conditions do assure that when integrable, the product space and iterated integrals agree, and when not integrable, all integrals are infinite.

The proof is nearly identical with that of Fubini's result, and is included for completeness. As noted in the above section on Generalizing Fubini's Theorem, this next result applies more generally to integrals in product spaces of more than two factors.

Proposition 5.22 (Tonelli's Theorem) *Let $(X, \sigma(X), \mu)$ and $(Y, \sigma(Y), \upsilon)$ be complete, σ-finite measure spaces, and $f(x,y)$ a nonnegative measurable function on $(X \times Y, \sigma(X \times Y), \mu \times \upsilon)$. Then:*

1. *$f_x(y) = f(x,y)$ is υ-measurable on Y, and,*

1'. *$f_y(x) = f(x,y)$ is μ-measurable on X.*

2. $\int_Y f(x,y)d\upsilon$ *is* μ*-measurable on* X*, and,*

2'. $\int_X f(x,y)d\mu$ *is* ν*-measurable on* Y*.*

3. The $\mu\times\upsilon$*-integral of* $f(x,y)$ *can be evaluated as an iterated integral:*

$$\int_X\left[\int_Y f(x,y)d\upsilon\right]d\mu=\int_{X\times Y}f(x,y)d\left(\mu\times\upsilon\right)=\int_Y\left[\int_X f(x,y)d\mu\right]d\upsilon. \tag{5.10}$$

Proof. *Again by symmetry, we will only prove half of the results, specifically, parts 1, 2 and the first half of 3.*

The proof of proposition 5.13 proves the desired measurability results and 5.10 for $f(x,y)=\chi_A(x,y)$ *for* $A\in\sigma\left(X\times Y\right)$ *with* $\mu\times\upsilon\left(A\right)<\infty$*, and hence these results are true for simple functions which are zero outside sets of finite measure. By corollary 1.22, given nonnegative* $f(x,y)$ *on a* σ*-finite measure space, there is an increasing sequence of nonnegative simple functions,* $\{\varphi_n(x,y)\}$*, each equal to zero outside a set of finite measure, and with* $\varphi_n(x,y)\to f(x,y)$ *pointwise. Hence* $\left(\varphi_n\right)_x(y)\to f_x(y)$ *and so* $f_x(y)$ *is measurable.*

By Lebesgue's monotone convergence theorem,

$$\int_Y f(x,y)d\upsilon=\lim_{n\to\infty}\int_Y\varphi_n(x,y)d\upsilon,$$

and the μ*-measurability of* $\int_Y\varphi_n(x,y)d\upsilon$ *proves the same for* $\int_Y f(x,y)d\upsilon$*. Applying Lebesgue's theorem again:*

$$\begin{aligned}\int_X\left[\int_Y f(x,y)d\upsilon\right]d\mu &= \lim_{n\to\infty}\int_X\left[\int_Y\varphi_n(x,y)d\upsilon\right]d\mu\\ &= \lim_{n\to\infty}\int_{X\times Y}\varphi_n(x,y)d\left(\mu\times\upsilon\right)\\ &= \int_{X\times Y}f(x,y)d\left(\mu\times\upsilon\right),\end{aligned}$$

which proves 5.10. ■

Remark 5.23 *From a theoretical point of view, Fubini's theorem appears more satisfying because it assures that under the assumption of the integrability of* $f(x,y)$*, that the iterated integrals also exist and provide the same numerical value. Thus integrability of* $f(x,y)$ *assures that the component functions are integrable almost everywhere, and that the integrals of these* *component functions are also integrable. So the step-by-step evaluation of the iterated integral makes sense at each step.*

On the other hand, from a practical point of view Tonelli's theorem is more useful than is Fubini's because we do not need to ascertain integrability of nonnegative $f(x,y)$ in advance. The Tonelli results state that we can in fact establish integrability of nonnegative measurable $f(x,y)$ by demonstrating the finiteness of the iterated integrals, and this can sometimes be easier than establishing the finiteness of a joint integral. In addition, for a general function $f(x,y)$ Tonelli's theorem can be applied to $|f(x,y)|$ to determine if $f(x,y)$ is integrable by verifying the finiteness of the iterated integral of $|f(x,y)|$. If integrable, we can conclude by Fubini's theorem that $\int_{X\times Y} f(x,y)d\,(\mu\times \upsilon)$ can be evaluated as an iterated integral.

Example 5.24 *An example of the application of Fubini's or Tonelli's theorem is, perhaps surprisingly, to the evaluation of the one-variable integral of the* ***normal probability density function*** *introduced in book 2. Recall that this density function is non-negative, defined on $(-\infty,\infty)$, depends on a location parameter $\mu\in\mathbb{R}$ and a scale parameter $\sigma^2>0$, and is defined by:*

$$f_N(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\left(x-\mu\right)^2/\left(2\sigma^2\right)\right). \tag{5.11}$$

Because:

$$\exp\left(-x^2/2\right)<|x|^{-N},$$

for any N as $|x|\to\infty$, it follows that $\int_{-\infty}^{\infty} f_N(x)dx<\infty$ as an improper Riemann integral. By proposition 2.64 of book 3, this integral therefore equals the corresponding Lebesgue integral. By either a traditional change of variables in the Riemann integral, or the corresponding transformation of the Lebesgue integral under the measurable transformation $T: x\to (x-\mu)/\sigma$, we obtain that

$$\int_{-\infty}^{\infty} f_N(x)dx=\int_{-\infty}^{\infty}\phi(x)dx,$$

where $\phi(x)$ denotes the ***unit normal probability density:***

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-x^2/2\right). \tag{5.12}$$

Example 5.25 *To use this section's results we consider the square of the* *integral of $\phi(x)$:*

$$\begin{aligned}\left[\int_{-\infty}^{\infty}\phi(x)dx\right]^2 &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\left(x^2+y^2\right)/2\right]dxdy \\ &= \frac{1}{2\pi}\iint_{\mathbb{R}^2}\exp\left[-\left(x^2+y^2\right)/2\right]dm^2 \\ &= \frac{2}{\pi}\iint_{Q}\exp\left[-\left(x^2+y^2\right)/2\right]dm^2.\end{aligned}$$

While the first step is notational, the transition from iterated integrals to product space integral is justified by Tonelli's theorem since $\exp\left[-\left(x^2+y^2\right)/2\right]$ *is continuous and hence measurable, or by Fubini's theorem, noting that this function is integrable because as noted above,* $\exp\left[-\left(x^2+y^2\right)/2\right] < \left[x^2+y^2\right]^{-N}$ *for all N as $x^2+y^2\to\infty$. Then by symmetry this integral over $\mathbb{R}^2$ equals four times the integral over the first quadrant, $Q=\{(x,y)|x>0,y>0\}$.*

The final step is to use ***polar coordinates*** *to evaluate this joint integral, using the measurable transformation $T:(0,\infty)\times(0,\pi/2)\to Q$ defined by:*

$$\begin{gathered}T:(r,\theta)\to(x,y),\\ x=r\cos\theta,\qquad y=r\sin\theta.\end{gathered}$$

This transformation is continuously differentiable with Jacobian matrix:

$$T'(r,\theta)\equiv\begin{pmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{pmatrix},$$

and Jacobian determinant $\det\left(T'(r,\theta)\right)=r$. *The inverse transformation is given:*

$$\begin{gathered}T^{-1}:(x,y)\to(r,\theta),\\ r=\sqrt{x^2+y^2},\qquad \theta=\arctan(y/x).\end{gathered}$$

From 3.25 of proposition 3.34 and 2 *of remark 3.35 it follows that with $T:(0,\infty)\times(0,\pi/2)\to Q$ and $g(x,y)=\exp\left[-\left(x^2+y^2\right)/2\right]$,*

$$\begin{aligned}\frac{2}{\pi}\int_Q\exp\left[-\left(x^2+y^2\right)/2\right]dm^2 &= \frac{2}{\pi}\int_{T^{-1}Q}g(T(r,\theta))\left|\det(T'(r,\theta))\right|dm^2 \\ &= \frac{2}{\pi}\int_{T^{-1}Q}r\exp\left(-r^2/2\right)dm^2.\end{aligned}$$

Another application of Fubini's theorem to the last integral, justified since $r\exp\left(-r^2/2\right)$ *is integrable over* $T^{-1}Q = (0,\infty)\times(0,\pi/2)$, *produces the iterated integral:*

$$\frac{2}{\pi}\int_Q \exp\left[-\left(x^2+y^2\right)/2\right]dm^2 = \frac{2}{\pi}\int_0^{\pi/2}\int_0^{\infty} r\exp\left(-r^2/2\right)drd\theta = 1.$$

Recalling that the integral on the left equals $\left[\int_{-\infty}^{\infty}\phi(y)dy\right]^2$ *completes the proof that the Normal density integrates to* 1.

Example 5.26 *The following result is another example of the application of Fubini's theorem to the evaluation of a one-variable integral, and one which is needed in the section below on Fourier transforms. The trick is to embed the one variable function that we seek to integrate into a* 2*-variable function for which the multiple integral is well defined, and where the integral relative to the second variable is easy. Then by Fubini's theorem we can reverse the orders of integration to put the easy integral first.*

Proposition 5.27 *The function* $f(x) = \frac{\sin x}{x}$ *is Riemann and Lebesgue integrable on* $[0,\infty)$ *and:*

$$\lim_{t\to\infty}\int_0^t \frac{\sin x}{x}dx = \frac{\pi}{2}. \tag{5.13}$$

Proof. *Note that for* $x > 0$,

$$f(x) = \int_0^{\infty} e^{-ux}\sin x du$$

as a Riemann integral, but also as a Lebesgue integral by proposition 2.64 of book 3 since $|e^{-ux}\sin x| \leq e^{-ux}$ *is Riemann integrable. Further, the integrand* $e^{-ux}\sin x$ *is continuous and hence measurable on* $A_t \equiv \{(x,u)|0 < x \leq t,\ 0 < u < \infty\}$, *and is in fact integrable over* A_t *since by Tonelli's theorem:*

$$\begin{aligned}\iint_{A_t}\left|e^{-ux}\sin x\right|dm^2 &= \int_0^t\left[\int_0^{\infty}\left|e^{-ux}\sin x\right|du\right]dx\\ &= \int_0^t \frac{|\sin x|}{x}dx\\ &< t.\end{aligned}$$

The last inequality follows because $|\sin x| \leq x$ *for* $x \geq 0$.

Fubini's theorem now applies and allows the interchange in the order of integration as follows:

$$\begin{aligned}
\int_0^t \frac{\sin x}{x} dx &= \int_0^\infty \int_0^t e^{-ux} \sin x dx du \\
&= \int_0^\infty \left[\frac{1}{u^2+1} \left(1 - e^{-ut} \left[u \sin t + \cos t\right]\right)\right] du \\
&= \int_0^\infty \frac{1}{u^2+1} du - \int_0^\infty \frac{e^{-ut} \left[u \sin t + \cos t\right]}{u^2+1} du,
\end{aligned}$$

where the middle equation follows by two applications of integration by parts. The final step is to note that by a substitution of $u = \tan x$, *the first integral equals* $\pi/2$, *and by the substitution* $y = ut$ *the second converges to* 0 *as* $t \to \infty$. *In detail:*

$$\begin{aligned}
\left| \int_0^\infty \frac{e^{-ut} \left[u \sin t + \cos t\right]}{u^2+1} du \right| &\leq \int_0^\infty \frac{e^{-y}}{y^2+t^2} \left| y \sin t + t \cos t \right| dy \\
&\leq \int_0^\infty \frac{(y+t)}{y^2+t^2} e^{-y} dy \\
&\leq \frac{c}{t}.
\end{aligned}$$

The last inequality follows from a standard calculus maximization of $f(y) = (y+t) / \left(y^2 + t^2\right)$ *for* $y \geq 0$ *to show that* $f(y) \leq c/t$ *for* $t > 0$. ■

Chapter 6

Applications of Fubini/Tonelli

In this chapter, the results of the Fubini/Tonelli theorems will be applied in three very different contexts:

1. **Integration by Parts for Lebesgue-Stieltjes Integrals**

   The first section will develop results which extend the corresponding Lebesgue result of section 3.4, and Riemann-Stieltjes result of section 4.1.2, both of book 3, and which in turn extended the familiar result for Riemann integrals.

2. **Integrability of the Convolution of Integrable Functions**

   Convolutions of functions were informally introduced in section 1.4.1 of book 4. This notion will be formally defined in the second section below and the all-important question of integrability of convolutions studied. As introduced in book 4, the convolution of two density functions produces the density function of the sum of the associated random variables, with similar results for distribution functions, and this study will be continued in book 6.

3. **Fourier Transforms of Finite Borel Measures**

   The final section will focus on the development of Fourier transform theory, the application of which will be seen in book 6 when the **characteristic function** of a distribution function will be introduced and applied. Characteristic functions provide a far more robust tool for studying properties of distribution functions than do moment generating functions, since as will be seen, they always exist. With this tool we derive a more general central limit theorem than that developed in book 4, as well as important results on sums of random variables.

6.1 Lebesgue-Stieltjes Integration by Parts

In this section we present an extension of the Lebesgue and Riemann-Stieltjes integration by parts results developed in book 3, which in turn extended the familiar Riemann integration by parts formula. Recall that given a real-valued function $f(x)$ on an interval $[a,b]$, we say that $f(x)$ **is of bounded variation on** $[a,b]$**,** and abbreviated $f \in B.V.$ or $f \in B.V.[a,b]$, if the **total variation** $T_{[a,b]} < \infty$, with:

$$T_{[a,b]} = \sup_{\Pi} \sum\nolimits_{i=1}^{n} |f(x_i) - f(x_{i-1})| \,.$$

The supremum here is taken over all partitions $\Pi = \{x_0, x_1, ..., x_n\}$ with

$$a = x_0 < x_1 ... < x_n = b.$$

In proposition 3.27 of book 3 was proved that a function $f(x)$ defined on $[a,b]$ is of bounded variation if and only if

$$f(x) = I_1(x) - I_2(x),$$

where $I_1(x)$ and $I_2(x)$ are monotonically increasing real valued functions. As a difference of increasing functions, each of which is measurable and by proposition 3.15 of book 3 differentiable m-a.e., meaning almost everywhere relative to Lebesgue measure, it follows that bounded variation functions are measurable and differentiable m-a.e.

For the current application we would like to generalize chapter 2 and define integrals relative to certain "signed" measures which will be defined relative to functions of bounded variation. To this end, we first generalize the bounded variation definition from $[a,b]$ to $\mathbb{R}$, and then identify a useful subset of this class of functions in the following definition. The notion of signed measures will be studied in more detail in the section below, A Digression into Signed Measures.

Definition 6.1 *A real-valued function $f(x)$ defined on $\mathbb{R}$* ***is of bounded variation on*** $\mathbb{R}$***,*** *abbreviated $f \in B.V.[\mathbb{R}]$, if the* ***total variation*** $T < \infty$

where $T = \sup_{[a,b]} T_{[a,b]}$. *We say that* $f(x)$ ***is of normalized bounded variation on*** $\mathbb{R}$***,*** *abbreviated* $f \in N.B.V.[\mathbb{R}]$, *if* $f \in B.V.[\mathbb{R}]$, f *is right continuous, and* $f(-\infty) = 0$.

Remark 6.2 *We make a few observations, leaving some of the details as exercises.*

1. *If* $f \in B.V.[\mathbb{R}]$ *then* $f \in B.V.[a,b]$ *for all* $[a,b]$, *and hence* $f(x) = I_1^{[a,b]}(x) - I_2^{[a,b]}(x)$ *on* $[a,b]$ *for monotonically increasing real valued functions,* $I_1^{[a,b]}(x)$ *and* $I_2^{[a,b]}(x)$. *This implies that* $f(x)$ *has this same representation as defined on* $\mathbb{R}$, *since if* $[c,d] \subset [a,b]$, $I_j^{[a,b]}(x)$ *can be normalized so that* $I_j^{[a,b]}(x) = I_j^{[c,d]}(x)$ *for* $x \in [c,d]$.

2. *If* $f \in B.V.[\mathbb{R}]$ *then* f *is differentiable* m*-a.e. by 1 and proposition 3.15 of book 3.*

3. *If* $f \in B.V.[\mathbb{R}]$ *then* f *is bounded, so* $|f(x)| \leq M$ *for some* M. *This does not imply that* $I_1(x)$ *and* $I_2(x)$ *are bounded, but since*

$$I_2(x) - M \leq I_1(x) \leq I_2(x) + M,$$

either both of $I_1(x)$ *and* $I_2(x)$ *are bounded, or neither is bounded.*

4. *If* $f \in B.V.[\mathbb{R}]$ *then* $\lim_{x\to\pm\infty} f(x)$ *exist. For example, if*

$$\liminf_{x\to\infty} f(x) = l < L = \limsup_{x\to\infty} f(x),$$

then for any $\epsilon > 0$ *there exist* x_n, $x'_n \to \infty$ *so that* $l \leq f(x_n) \leq l + \epsilon$ *and* $L - \epsilon \leq f(x'_n) \leq L$. *Hence if* $\epsilon < (L-l)/2$ *then any collection of disjoint intervals of the form* $[x_n, x'_m]$ *or* $[x'_n, x_m]$ *would provide a contradiction to the assumption that* $f \in B.V.[\mathbb{R}]$.

5. *By 2, if* $f \in B.V.[\mathbb{R}]$ *then* f *is continuous* m*-a.e. Hence,* $g(x) \equiv f(x^+)$ *is right continuous,* $g \in B.V.[\mathbb{R}]$, *and* $g(x) = f(x)$ m*-a.e. Recall that* $f(x^+) \equiv \lim_{y\to x^+} f(y)$. *See section 3.2.1 for more on this notion of one-sided limits.*

6. *By 4 and 5, if* $f \in B.V.[\mathbb{R}]$ *then* $g(x) \equiv f(x^+) - f(-\infty)$ *is right continuous,* $g \in B.V.[\mathbb{R}]$ *and* $g(-\infty) = 0$. *Hence* $g \in N.B.V.[\mathbb{R}]$.

7. *If* $f \in B.V.[\mathbb{R}]$ *then with* $I_1(x)$ *and* $I_2(x)$ *as defined in 1, and with* $g(x) \equiv f(x^+)$ *as in 5, we can take:*

$$g(x) = I_1(x^+) - I_2(x^+).$$

In other words, that $g \in B.V.[\mathbb{R}]$ implies the existence of increasing functions so that $g(x) = J_1(x) - J_2(x)$, but the right continuity of g does not assure the right continuity of $J_1(x)$ and $J_2(x)$. However, such g can be defined in terms of the right continuous versions of the I_j increasing functions because by right continuity:

$$g(x) \equiv \lim_{y \to x^+} I_1(y) - \lim_{y \to x^+} I_2(y),$$

and hence defining $J_1(x)$ and $J_2(x)$ to equal these right limits does not change the value of $g(x)$.

8. *If $f \in N.B.V.[\mathbb{R}]$ is defined by increasing $I_1(x)$ and $I_2(x)$, then:*

$$\lim_{x\to-\infty} I_1(x) = \lim_{x\to-\infty} I_2(x), \qquad \lim_{x\to\infty} I_1(x) = \lim_{x\to\infty} I_2(x)+C.$$

This follows since $f(-\infty) = 0$, meaning $\lim_{x\to-\infty} f(x) = 0$, and $f(\infty) = \lim_{x\to\infty} f(x) = C < \infty$. But note that from 3, these individual limits need not be finite.

Example 6.3 *If μ is a finite Borel measure then $F_\mu(x) \equiv \mu(-\infty, x] \in N.B.V.[\mathbb{R}]$, since it is increasing, right continuous, $F(-\infty) = 0$ and $F(\infty) < \infty$ by proposition 5.7 of book 1. By the same result, if μ is a general Borel measure, then $F_\mu(y) \in B.V.[a, b]$ for all $[a, b]$, where:*

$$F_\mu(y) = \begin{cases} \mu((0, y]), & y > 0, \\ 0, & y = 0, \\ -\mu((y, 0]), & y < 0. \end{cases}$$

Such $F_\mu(y)$ is again right continuous reflecting continuity from above of μ.

Conversely, if $F(x)$ is right continuous and increasing then $F \in B.V.[a, b]$ for all $[a, b]$, and by proposition 5.23 of book 1, F induces a measure μ_F on the Borel sigma algebra, $\mathcal{B}(\mathbb{R})$. If F is bounded then $F \in B.V.[\mathbb{R}]$ and μ_F is a finite measure. In this case this measure is identical with the measure μ_G where $G(x) \equiv F(x) - F(-\infty) \in N.B.V.[\mathbb{R}]$.

This identification between right continuous increasing functions and Borel measures is generalized in the next result which identifies right continuous $B.V.[\mathbb{R}]$ with countably additive set functions definable on bounded $A \in \mathcal{B}(\mathbb{R})$. This set function is "nearly" a measure as is discussed below, and also "nearly" a signed measure as discussed in chapter 8.

Proposition 6.4 *If $f \in B.V.[\mathbb{R}]$ is right continuous, define the set function $\mu_f[(a,b]] = f(b) - f(a)$ for all bounded $(a,b] \in \mathcal{A}'$, with $\mathcal{A}'$ the semi-algebra of right semi-closed intervals. Then there exist Borel measures μ_1 and μ_2 so that $\mu_f = \mu_1 - \mu_2$ on $\mathcal{A}'$, and μ_f so given is well defined for bounded $A \in \mathcal{B}(\mathbb{R})$.*

Proof. *By 7 of remark 6.2 above, $f(x) = I_1(x) - I_2(x)$ where $I_1(x)$ and $I_2(x)$ are right continuous, increasing functions. Consequently by the development of chapter 5 of book 1, there exist Borel measures μ_1 and μ_2 so that for all right semi-closed intervals $(a,b]$:*

$$\mu_j[(a,b]] = I_j(b) - I_j(a),$$

where $j = 1,2$. Thus for all such bounded intervals, $\mu_f[(a,b]] = \mu_1[(a,b]] - \mu_2[(a,b]]$, so $\mu_f = \mu_1 - \mu_2$ on $\mathcal{A}'$. The set function μ_f :

$$\mu_f[A] \equiv \mu_1[A] - \mu_2[A], \tag{6.1}$$

is therefore defined for bounded Borel sets $A \in \mathcal{B}(\mathbb{R})$.

To see that μ_f is well-defined, assume that also $f(x) = I_1'(x) - I_2'(x)$ where $I_1'(x)$ and $I_2'(x)$ are right continuous, increasing functions. Then since $\mu_f[(a,b]] = f(b) - f(a)$, it follows by the same argument that for all bounded $(a,b]$:

$$\mu_1[(a,b]] - \mu_2[(a,b]] = \mu_1'[(a,b]] - \mu_2'[(a,b]].$$

As each of μ_j and μ_j' can be extended to measures on $\mathcal{B}(\mathbb{R})$, it then follows that for all bounded $A \in \mathcal{B}(\mathbb{R})$,

$$\mu_1[A] - \mu_2[A] = \mu_1'[A] - \mu_2'[A].$$

Hence μ_f is well defined by 6.1 for bounded $A \in \mathcal{B}(\mathbb{R})$. ■

Remark 6.5 *Note that μ_f is not in general a measure since by construction, it need not be the case that $\mu_f[A] \geq 0$. Further, this proposition falls short of declaring that μ_f extends to a countably additive set function on the Borel sigma algebra $\mathcal{B}(\mathbb{R})$. The problem is that as noted in 3 of remark 6.2 above, that either both of $I_1(x)$ and $I_2(x)$ are bounded, or neither is bounded. In the later case, if A is an unbounded interval, then formally, $\mu_f[A] = \infty - \infty$ is not well defined. Similarly, such μ_f is not countably additive on disjoint $\{A_i\}$ with $\mu_j[\cup A_j] = \infty$. However, μ_f is perfectly well-defined on bounded Borel measurable sets, and countably additive when both $\mu_j[\cup A_j] < \infty$, which is enough for the development below.*

Exercise 6.6 *Prove that if $f(x) = I_1(x) - I_2(x)$ where $I_1(x)$ and $I_2(x)$ are right continuous, increasing and bounded functions, then μ_f is a countably additive set function on $\mathcal{B}(\mathbb{R})$. This is an example of a* ***signed measure*** *of chapter 8.*

Proposition 6.7 *If $f \in B.V.[\mathbb{R}]$ is right continuous, then for all bounded Borel measurable functions g and bounded $A \in \mathcal{B}(\mathbb{R})$, the* ***Lebesgue-Stieltjes integral,***

$$\int_A g d\mu_f,$$

is well defined by

$$\int_A g d\mu_f = \int_A g d\mu_1 - \int_A g d\mu_2, \tag{6.2}$$

where μ_1 and μ_2 are the measures defined in 6.1.

Proof. *Given a decomposition $f(x) = I_1(x) - I_2(x)$ where $I_1(x)$ and $I_2(x)$ are right continuous, increasing functions, the integrals on the right in 6.2 are well defined by chapter 2. If also $f(x) = I_1'(x) - I_2'(x)$ where $I_1'(x)$ and $I_2'(x)$ are right continuous, increasing functions, then for all bounded $A \in \mathcal{B}(\mathbb{R})$, $\mu_1 - \mu_2 = \mu_1' - \mu_2'$ as noted above. Hence for all simple functions φ with bounded domain:*

$$\int \varphi d\mu_1 - \int \varphi d\mu_2 = \int \varphi d\mu_1' - \int \varphi d\mu_2'.$$

If g is nonnegative, bounded and measurable and A bounded, then by corollary 1.22 there is an increasing sequence $\{\varphi_n\}$ so that $\varphi_n \to g$ pointwise on A. By Lebesgue's monotone convergence theorem,

$$\int_A g d\mu = \lim_n \int_A \varphi_n d\mu$$

where μ denotes any of the above four measures, and each integral is finite by the assumptions on g and A. Hence $\int_A g d\mu_f$ is well defined by 6.2. For g bounded and measurable, this nonnegative result applies to g^+ and g^- of definition 2.36 and the proof is complete. ■

With the aid of Fubini's theorem we are now ready to state and prove an integration by parts result for Lebesgue-Stieltjes integrals. The first result compares most closely with the corresponding result on Riemann-Stieltjes integrals of proposition 4.15 of book 3.

Remark 6.8 *To simplify the statements, the following proposition and corollaries assume that $f, g \in B.V.[\mathbb{R}]$ and then present results that are true for any bounded interval $(a,b]$. But given fixed $(a,b]$, it seems logical that a local condition on f, g, such as $f, g \in B.V.[[a-\epsilon, b]]$, would be enough to assure the same conclusions on this interval. This is indeed the case since such functions can be extended to $B.V.[\mathbb{R}]$ functions by defining, for example, $f(x) = f(a-\epsilon)$ for $x < a-\epsilon$ and $f(x) = f(b)$ for $x > b$. The details are left as an exercise.*

Proposition 6.9 (Lebesgue-Stieltjes Integration by Parts) *If $f, g \in B.V.[\mathbb{R}]$ are right continuous and at least one continuous, then for any bounded interval $(a,b]$:*

$$\int_{(a,b]} g d\mu_f = f(b)g(b) - f(a)g(a) - \int_{(a,b]} f d\mu_g. \tag{6.3}$$

Proof. *First note that both f and g are bounded and measurable since $f, g \in B.V.[\mathbb{R}]$, and so by proposition 6.7 both integrals are well defined. Define $E \subset \mathbb{R}^2$ by:*

$$E = \{(x,y) | a < x \leq y \leq b\}.$$

Then $E \in \mathcal{B}(\mathbb{R}^2)$ as the intersection of open and closed sets:

$$E = \{x \leq y\} \cap \{y \leq b\} \cap \{x > a\}.$$

Also, since $E \subset A$ with $A = (a,b] \times (a,b]$, and by definition $\left(\mu_f \times \mu_g\right)[A] = (f(b) - f(a))\,(g(b) - g(a)) < \infty$, it follows that:

$$\left(\mu_f \times \mu_g\right)(E) = \int_A \chi_E(x,y) d\left(\mu_f \times \mu_g\right) < \infty.$$

Thus Fubini's theorem applies and this integral can be evaluated as an iterated integral. Since either measure can be assigned to the x-variable, these iterated integrals will have the following form, with h and k denoting f and g in some order (see below):

$$\int_{(a,b]} \left[\int_{(a,y]} d\mu_h(x)\right] d\mu_k(y) = \int_{(a,b]} \left[\int_{[x,b]} d\mu_k(y)\right] d\mu_h(x).$$

Now for any Borel measure, $\int_{(a,y]} d\mu_h(x) = h(y) - h(a)$, but the integral $\int_{[x,b]} d\mu_k(y)$ is more subtle. However, if the $B.V$ function k underlying μ_k is continuous, then $\mu_k\,[[x,b]] = \mu_k\,[(x,b]]$ by exercise 5.22 of book 1, and so *again it follows that* $\int_{[x,b]} d\mu_k(y) = k(b) - k(x)$. *So whichever of* f *and* g *is continuous is assigned the role of* k, *notationally. To be specific, assume that* g *is the continuous function. Then:*

$$\begin{aligned}
\left(\mu_f \times \mu_g\right)(E) &= \int_{(a,b]} \left[\int_{(a,y]} d\mu_f(x)\right] d\mu_g(y) \\
&= \int_{(a,b]} \left[f(y) - f(a)\right] d\mu_g(y) \\
&= \int_{(a,b]} f d\mu_g - f(a)\left[g(b) - g(a)\right].
\end{aligned}$$

Similarly,

$$\begin{aligned}
\left(\mu_f \times \mu_g\right)(E) &= \int_{(a,b]} \left[\int_{[x,b]} d\mu_g(y)\right] d\mu_f(x) \\
&= \int_{(a,b]} \left[g(b) - g(x)\right] d\mu_f(x) \\
&= g(b)[f(b) - f(a)] - \int_{(a,b]} g d\mu_f.
\end{aligned}$$

The proof is complete by simplifying the resulting expressions. ■

Corollary 6.10 *If* $f, g \in B.V.[\mathbb{R}]$ *are right continuous and at least one continuous, then for any bounded interval,* $[a,b]$:

$$\int_{[a,b]} g d\mu_f = f(b)g(b) - f(a^-)g(a^-) - \int_{[a,b]} f d\mu_g, \tag{6.4}$$

where $f(a^-)$ *and* $g(a^-)$ *denote the lefts limits at* a.
Proof. *Define* $E' = \{(x,y) | a \leq x \leq y \leq b\}$, *then* $\left(\mu_f \times \mu_g\right)(E') < \infty$ *and again Fubini's theorem applies. The details are left as an exercise.* ■

Of course by the continuity requirement of this corollary, at least one of $f(a^-)$ or $g(a^-)$ equals the value of the respective function at a. When both f and g are continuous, $f(a^-)g(a^-) = f(a)g(a)$, and we have the following:

Corollary 6.11 *If* $f, g \in B.V.[\mathbb{R}]$ *are continuous, then for any bounded interval,* $[a,b]$:

$$\int_{[a,b]} g d\mu_f = f(b)g(b) - f(a)g(a) - \int_{[a,b]} f d\mu_g. \tag{6.5}$$

The next corollary is an application of the above general result to Lebesgue integration by parts, and is thus a restatement of proposition 3.63 of book 3. Recall from definition 3.54 of book 3 that a function f is defined to be **absolutely continuous** if for any $\epsilon > 0$ there is a δ so that

$$\sum\nolimits_{i=1}^{n} \left| f(x_i) - f(x_i') \right| < \epsilon$$

for any finite collection of disjoint subintervals $\{(x_i', x_i)\}_{i=1}^n$ with

$$\sum\nolimits_{i=1}^{n} \left| x_i - x_i' \right| < \delta.$$

An absolutely continuous function is apparently continuous, but also by proposition 3.58 of book 3, if $f(x)$ is absolutely continuous on $[a,b]$ then $f(x) \in B.V.[a,b]$ and thus $f'(x)$ exists m-a.e.

Corollary 6.12 (Lebesgue Integration by Parts) *If f, g are absolutely continuous functions, then for any bounded interval $[a,b]$:*

$$\int_{[a,b]} g(x)f'(x)dm = f(b)g(b) - f(a)g(a) - \int_{[a,b]} f(x)g'(x)dm, \tag{6.6}$$

where dm denotes Lebesgue measure.

Proof. *Because f is absolutely continuous, $f'(x)$ exists m-a.e. and by proposition 3.61 of book 3:*

$$\int_{[a,b]} f'(x)dm = f(b) - f(a).$$

Also by definition and continuity of f (see exercise 5.22 of book 1):

$$\int_{[a,b]} d\mu_f \equiv \mu_f\left[[a,b]\right] = f(b) - f(a).$$

Now $\int_{[a,b]} d\mu_f$ and $\int_{[a,b]} f'(x)dm$ are countably additive on the semi-algebra $\mathcal{A}'$ of right semi-closed intervals by 2.20 and proposition 2.60 of book 3 respectively. Thus it follows by the uniqueness theorem of proposition 6.14 of book 1 that for all $A \in \mathcal{B}(\mathbb{R})$:

$$\int_A d\mu_f = \int_A f'(x)dm.$$

Proposition 3.6 then states in 3.3, with a slight notational change, that for all measurable functions g :

$$\int_A g d\mu_f = \int_A g(x)f'(x)dm.$$

A similar derivation obtains,

$$\int_A f d\mu_g = \int_A f(x)g'(x)dm.$$

The result in 6.6 now follows by continuity of f and g, and 6.5. ■

6.2 Integrability of the Convolution of Functions

It is not a surprise after a little thought that if $f(x)$ and $g(x)$ are μ-integrable functions, then the product $f(x)g(x)$ need not be μ-integrable. For example on $[0,1]$, a function $f(x) = x^a$ is Lebesgue integrable if and only if $a > -1$, so the product of such functions, say both with $a = -1/2$, need not be integrable. However, there is a special type of product, called a **convolution product** or just a **convolution**, for which integrability is preserved. Introduced in section 1.4 of book 4 and to be further developed in book 6, convolutions play an important role in probability theory.

Notation 6.13 *To simplify notation we will denote Lebesgue integrals by $\int f(x)dm$ rather than $(\mathcal{L})\int f(x)dx$ as above. In integrals involving more than one variable, the notation such as $\int f(x,y)dm(y)$ will be used for clarity.*

Definition 6.14 *Let $f(x)$ and $g(x)$ be integrable functions on the complete Lebesgue measure space $(\mathbb{R}, \mathcal{M}_L, m)$. Then the **convolution of** f **and** g, denoted $f * g$, is defined by:*

$$f * g(x) = \int f(x-y)g(y)dm(y) \tag{6.7}$$

when this integral exists.

Remark 6.15 *Informally it is easy to gain an insight into the probability theory application of convolutions by considering for x fixed, the variates on which f and g are valued for this integral. As can be seen, f is valued on $x-y$ and g is valued on y, and so for all y these variates add to x. In other words, this convolution formula reflects all pairs of variates that add to x. While introduced in book 4, convolutions will be further investigated and seen to play a key role in the development of the distribution function of the sum of "independent" random variables in book 6.*

It is an exercise using results on change of variables to show that for any x for which the integral in 6.7 exists, that $f*g(x) = g*f(x)$. In other words,

$$\int f(x-y)g(y)dm(y) = \int g(x-y)f(y)dm(y). \tag{6.8}$$

The main result of this section is that if f and g are integrable, then $f(x-y)g(y)$ is integrable for almost all x, and $f*g(x)$ is in fact integrable. Moreover, this last result will also yield an upper bound for the value of the integral $\int f*g(x)dx$.

Lebesgue product spaces are σ-finite, as is proved by considering the collection of compact rectangles defined by $A_m = \{(x_1, x_2, ..., x_n) | -m \leq x_j \leq m \text{ for all } j\}$, and noting that A_m has Lebesgue measure $(2m)^n$. Thus to justify the use of Tonelli's theorem we must prove "only" that $f(x-y)g(y)$ is a Lebesgue measurable function on the complete product space $(\mathbb{R}^2, \mathcal{M}_L^2, m^2)$ constructed with two copies of the original space $(\mathbb{R}, \mathcal{M}_L, m)$ as in chapter 7 of book 1.

Regarding Lebesgue measurability of $h(x,y) \equiv f(x-y)g(y)$, meaning

$$h : (\mathbb{R}^2, \mathcal{M}_L^2, m^2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$$

is measurable, it is enough to establish measurability of the component functions by proposition 1.5. While the measurability of $g(y)$ is readily established, perhaps surprisingly the Lebesgue measurability of $f(x-y)$ on $\mathbb{R}^2$ is not obvious and in fact somewhat challenging to establish. One reference that explicitly addresses this question is Hewitt and Stromberg (1965), whose approach we follow. Given Lebesgue measurable $f : \mathbb{R} \to \mathbb{R}$, the idea is to identify a property of $\varphi : \mathbb{R}^2 \to \mathbb{R}$ which assures the Lebesgue measurability of $k(x,y) \equiv f(\varphi(x,y))$, and then to prove that $\varphi(x,y) \equiv x-y$ has this property.

Proposition 6.16 *Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be Borel measurable and with the additional property that if $N \in \mathcal{M}_L$ with $m(N) = 0$, then $\varphi^{-1}(N) \in \mathcal{M}_L^2$ and $m^2\left[\varphi^{-1}(N)\right] = 0$. Then $f(\varphi) : \mathbb{R}^2 \to \mathbb{R}$ is Lebesgue measurable for every Lebesgue measurable function f.*

Proof. *First let $f(x) \equiv \chi_A(x)$, the characteristic function of a set $A \in \mathcal{M}_L$. By proposition 2.42 of book 1 there exists disjoint $G \in \mathcal{B}(\mathbb{R})$ and $N \in \mathcal{M}_L$ so that $A = G \cup N$ and $m(N) = 0$. Then $\chi_A(\varphi)(x,y) \equiv \chi_{\varphi^{-1}(A)}(x,y)$ and:*

$$\chi_{\varphi^{-1}(A)}(x,y) = \chi_{\varphi^{-1}(G)}(x,y) + \chi_{\varphi^{-1}(N)}(x,y).$$

Thus $\chi_A(\varphi)$ is Lebesgue measurable since $\varphi^{-1}(G) \in \mathcal{B}(\mathbb{R}^2)$ by Borel measurability of φ, and $\varphi^{-1}(N) \in \mathcal{M}_L^2$ by hypothesis. It then follows from *proposition 3.30 of book 1 that* $f(\varphi)$ *is Lebesgue measurable for all simple functions* $f = \sum_{j=1}^{n} a_j \chi_{A_j}$, *since then* $f(\varphi) = \sum_{j=1}^{n} a_j \chi_{A_j}(\varphi)$.

Given a Lebesgue measurable function f, *part 2 of the proof of proposition 3.49 of book 1 (see also proposition 1.18 and remark 1.19 above) assure the existence of a sequence of simple functions* f_k *with* $f_k \to f$. *Thus* $f_k(\varphi) \to f(\varphi)$ *and measurability of* $f(\varphi)$ *follows from proposition 3.47 of book 1.* ■

Example 6.17 *The function* $\varphi(x-y) = x-y$ *satisfies the requirements of proposition 6.16. First,* φ *is Borel measurable since by continuity,* $\varphi^{-1}(A)$ *is open and hence* $\varphi^{-1}(A) \in \mathcal{B}(\mathbb{R})$ *for open* $A \subset \mathbb{R}^2$. *Thus* $\varphi^{-1}\left[\mathcal{B}(\mathbb{R}^2)\right] \subset \mathcal{B}(\mathbb{R})$ *since such open sets generate* $\mathcal{B}(\mathbb{R}^2)$. *Next, if* $N \in \mathcal{M}_L$ *with* $m(N) = 0$, *then* $\varphi^{-1}(N) = \bigcup_{n=1}^{\infty} P_n$ *with*

$$P_n \equiv \{(x,y) | x - y \in N \text{ and } |y| \leq n\}.$$

The desired result that $m^2\left[\varphi^{-1}(N)\right] = 0$ *will follow from* $m^2[P_n] = 0$ *for all* n *and countable subadditivity. By proposition 2.43 of book 1,* m *is outer regular and there exists open sets* $\{G'_k\}$ *so that* $N \subset G'_k$ *and* $\inf m[G'_k] = m(N) = 0$. *Further by proposition 2.42 of book 1, given* $\epsilon > 0$ *such sets can be chosen to have* $m[G'_k] < \epsilon$ *for all* k. *Defining* $G_k = \bigcap_{j \leq k} G'_j$ *it follows that* $\{G_k\}$ *are open and nested,* $G_{k+1} \subset G_k$, $m[G_1] < \epsilon$, *and* $\lim_{k\to\infty} m[G_k] = m(N) = 0$. *For given* n *define the sets* $\{H_k\}$ *by* $H_k = \{(x,y)|x-y \in G_k\} \cap \{|y| \leq n\}$. *Note that* $\{(x,y)|x-y \in G_k\} = \varphi^{-1}(G_k)$ *is open as the pre-image of an open set under a continuous function, and so each* $H_k \in \mathcal{B}(\mathbb{R}^2)$ *as the intersection of an open and closed set,* $H_{k+1} \subset H_k$, *and* $P_n \subset \bigcap H_k$. *To prove that* $m^2[P_n] = 0$ *we prove that* $m^2[\bigcap H_k] = 0$.

To this end, since $\{H_k\}$ *are nested and* $m[H_1] < \epsilon$ *by construction, it follows by continuity from above of proposition 2.44 of book 1 that:*

$$m^2[\bigcap H_k] = \lim_{k\to\infty} m[H_k].$$

Now with χ_A *denoting the characteristic function of a set* A, *it follows by either Fubini's or Tonelli's theorem that:*

$$\begin{aligned}
m[H_k] &\equiv \int \chi_{H_k}(x,y)dm^2 \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} \chi_{H_k}(x,y)dm(x)\right]dm(y) \\
&= \int_{-n}^{n}\left[\int_{-\infty}^{\infty} \chi_{G_k}(x-y)dm(x)\right]dm(y).
\end{aligned}$$

By a change of variable,

$$\int_{-\infty}^{\infty} \chi_{G_k}(x-y)dm(x) = \int_{-\infty}^{\infty} \chi_{G_k}(x)dm(x) = m(G_k),$$

and thus:

$$m\left[H_k\right] = 2nm(G_k).$$

So $m^2\left[\bigcap H_k\right] = 0$ *since* $\lim_{k\to\infty} m\left[G_k\right] = m(N) = 0$.

Proposition 6.18 *Let* $f(x)$ *and* $g(x)$ *be integrable functions on the Lebesgue measure space* $(\mathbb{R}, \mathcal{M}_L, m)$. *Then:*

1. *The function* $f(x-y)g(y)$ *is* $m(y)$*-integrable for almost all* x. *That is,*

$$\int |f(x-y)g(y)|\, dm(y) < \infty, \ m\text{-a.e.}$$

2. *Defining* $f * g(x)$ *as in 6.7, then* $f * g(x)$ *is* m*-integrable with:*

$$\int |f * g(x)|\, dm(x) \leq \int |f(x)|\, dm \int |g(y)|\, dm, \tag{6.9}$$

and

$$\int f * g(x)dm(x) = \int f(x)dm \int g(y)dm. \tag{6.10}$$

Proof. *To apply Tonelli's theorem, it must be shown that* $F(x,y) \equiv f(x-y)g(y)$ *is Lebesgue measurable and this follows from proposition 6.16 and example 6.17. To prove integrability of* $F(x,y)$, *an application of Tonelli's theorem obtains:*

$$\begin{aligned}
\iint |f(x-y)g(y)|\, dm^2(x,y) &= \int \int |f(x-y)g(y)|\, dm(x)dm(y) \\
&= \int |f(x)|\, dm(x) \int |g(y)|\, dm(y).
\end{aligned}$$

The second step uses a change of variable in the x*-integral:*

$$\int |f(x-y)|\, dm(x) = \int |f(x)|\, dm(x) \text{ for all } y.$$

Consequently, the integrability of $f(x)$ *and* $g(x)$ *assure the integrability of* $f(x-y)g(y)$ *in* $(\mathbb{R}^2, \mathcal{M}_L^2, m^2)$. *Because* $f(x-y)g(y)$ *is integrable, Fubini's* *theorem assures that this function is y-integrable for almost all x. In other words, $\int |f(x-y)g(y)|\, dm(y) < \infty$ m-a.e., which is part 1.*

Moreover, Fubini's theorem and the triangle inequality assure the integrability of $f * g(x) = \int f(x-y)g(y)dm(y)$:

$$\begin{aligned} \int |f * g(x)|\, dm(x) &\leq \int \int |f(x-y)g(y)|\, dm(x)dm(y) \\ &= \int |f(x)|\, dm(x) \int |g(y)|\, dm(y), \end{aligned}$$

which is 6.9.

Applying Fubini's theorem to $f(x-y)g(y)$:

$$\begin{aligned} \textstyle\iint f(x-y)g(y)dm^2(x,y) &= \int \int f(x-y)g(y)dm(x)dm(y) \\ &= \int f(x)dm(x) \int g(y)dm(y), \end{aligned}$$

as above. But note that the m^2-integral can also be expressed as iterated integrals in reversed order, and:

$$\int \int f(x-y)g(y)dm(y)dm(x) = \int f * g(x)dm(x).$$

Combining obtains 6.10. ■

Example 6.19 *Let $f(x)$ and $g(y)$ be the density functions for the unit normal random variable as in 5.12, for example:*

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right).$$

Then

$$\begin{aligned} f * g(x) &= \frac{1}{2\pi} \int \exp\left[-(x-y)^2/2\right] \exp\left(-y^2/2\right) dy \\ &= \frac{1}{2\pi} \exp\left(-x^2/4\right) \int \exp\left[-(y-x/2)^2\right] dy. \end{aligned}$$

This second integral equals $\sqrt{\pi}$ by substitution, and hence

$$f * g(x) = \frac{1}{\sqrt{4\pi}} \exp\left(-x^2/4\right).$$

Comparing this with 5.11, it can be concluded that $f * g(x)$ is the density of a normal variate with $\mu = 0$ and $\sigma^2 = 2$.

With more algebra this conclusion generalizes.

Exercise 6.20 *Prove that if $f(x)$ and $g(y)$ are density functions for normal variates with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), then $f * g(x)$ is the density of a normal variate with $\mu = \mu_1 + \mu_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Generalize this to n density functions by induction.*

6.3 Fourier Transforms of Finite Borel Measures

As noted in the introduction to this chapter, Fourier transform theory will be applied in book 6 where the **characteristic function** of a distribution function will be introduced and its properties developed. Characteristic functions provide a far more robust tool for studying properties of distribution functions than do the moment generating functions of book 4, since they will be seen to always exist. Consequently, a more general central limit theorem is possible with this tool, as are important results on sums of random variables.

This section will focus on the 1-dimensional theory of Fourier transforms, which will be sufficient for both a 1-dimensional and n-dimensional theory of characteristic functions in book 6.

6.3.1 Integration of Complex Valued Functions

The definitions of Fourier transform and Fourier-Stieltjes transform require an integration theory applicable to **complex valued functions** of a real variable. That is, functions $f(x)$ with $f : \mathbb{R} \to \mathbb{C}$. As it turns out, the above theory easily extends as we discuss here. But note that a good deal more work and justification would be needed for an integration theory for complex valued functions of a complex variable, $f : \mathbb{C} \to \mathbb{C}$, which fortunately is a theory not needed for our purposes.

Given a function $f(x)$ with $f : \mathbb{R} \to \mathbb{C}$, it follows that $f(x) = u(x) + iv(x)$ where $u, v : \mathbb{R} \to \mathbb{R}$, and i is the standard notation for the complex or "imaginary" unit, $i \equiv \sqrt{-1}$.

Definition 6.21 *Given a real measure space $(\mathbb{R}, \sigma(\mathbb{R}), \mu)$, define $f(x) : \mathbb{R} \to \mathbb{C}$ to be **measurable** if $f(x) = u(x) + iv(x)$ and both $u(x)$ and $v(x)$ are measurable. That is, $f(x)$ is measurable if $u^{-1}(A) \in \sigma(\mathbb{R})$ and $v^{-1}(A) \in \sigma(\mathbb{R})$ for all Borel sets $A \in \mathcal{B}(\mathbb{R})$. A function $f : X \to \mathbb{C}$ is similarly defined to be measurable relative to $(X, \sigma(X), \mu)$ based on the analogous measurability of $u(x)$ and $v(x)$.*

*If $u(x)$ and $v(x)$ are μ-integrable, we define **μ-integral of** $f(x)$ by:*

$$\int f(x)d\mu \equiv \int u(x)d\mu + i\int v(x)d\mu,$$

and similarly for $\int_A f(x)d\mu$.

Remark 6.22 *Since $|f(x)| = \sqrt{u^2(x)+v^2(x)}$, we have that:*

$$\max[|u(x)|, |v(x)|] \leq |f(x)| \leq |u(x)| + |v(x)|,$$

and hence $f(x)$ is μ-integrable, meaning $\int |f(x)|\, d\mu < \infty$, if and only if $u(x)$ and $v(x)$ are μ-integrable.

With this definition, all of the properties of the integral summarized in proposition 2.40 are readily seen to be satisfied, with perhaps the exception of the triangle inequality. To justify a triangle inequality requires a limiting argument, because the naive approach using the triangle inequality for real functions does not produce the desired result. Specifically, by definition of $|z|^2$ for complex z:

$$\begin{aligned}\left|\int f(x)d\mu\right|^2 &\equiv \left|\int u(x)d\mu\right|^2 + \left|\int v(x)d\mu\right|^2 \\ &\leq \left[\int |u(x)|\, d\mu\right]^2 + \left[\int |v(x)|\, d\mu\right]^2 \\ &\leq \left[\int |u(x)|\, d\mu + \int |v(x)|\, d\mu\right]^2.\end{aligned}$$

Hence:

$$\left|\int f(x)d\mu\right| \leq \int [|u(x)| + |v(x)|]\, d\mu.$$

This upper bound exceeds the desired result of $\int |f(x)|\, d\mu = \int \sqrt{u^2(x)+v^2(x)}d\mu$.

Exercise 6.23 (Triangle inequality) *If $f(x) = u(x) + iv(x)$ where $u(x)$ and $v(x)$ are μ-integrable, prove that with:*

$$\int f(x)d\mu \equiv \int u(x)d\mu + i\int v(x)d\mu,$$

that the triangle inequality holds:

$$\left|\int f(x)d\mu\right| \leq \int |f(x)|\, d\mu.$$

Hint: Start with simple functions $u(x)$ and $v(x)$ defined with the same partition. Then develop approximating sequences of simple functions for $u(x)$ and $v(x)$, justifying the use of the same partitions using Lebesgue's dominated convergence theorem and the integrability of $f(x)$.

Finally, **Lebesgue's dominated convergence theorem** and corollaries again apply in this context. To see this, assume that $\{f_n(x)\} = \{u_n(x) + iv_n(x)\}$ is a sequence of μ-measurable functions on μ-measurable set E, with $\lim_{n\to\infty} f_n(x) = f(x)$ where $f(x) = u(x) + iv(x)$. Assume also that there is a μ-measurable function $g(x)$, integrable on E, so that

$$|f_n(x)| \leq g(x), \text{ all } n.$$

This inequality then applies to both $|u_n(x)|$ and $|v_n(x)|$ and so both $u(x)$ and $v(x)$ are μ-integrable on E and by Lebesgue's dominated convergence theorem:

$$\int_E u(x)d\mu = \lim_{n\to\infty} \int_E u_n(x)d\mu, \quad \int_E v(x)d\mu = \lim_{n\to\infty} \int_E v_n(x)d\mu.$$

Hence $f(x)$ is μ-integrable on E and by linearity,

$$\int_E f(x)d\mu = \lim_{n\to\infty} \int_E f_n(x)d\mu.$$

Further,

$$\begin{aligned} \int_E |f_n(x) - f(x)|\, d\mu \ &\leq \ \int_E |u_n(x) - u(x)|\, d\mu + \int_E |v_n(x) - v(x)|\, d\mu \\ &\to \ 0, \text{ as } n \to \infty. \end{aligned}$$

6.3.2 Fourier Transforms

Fourier transforms are named for **Jean Baptiste Joseph Fourier** (1768 – 1830) who introduced the underlying mathematical tools in the context of **Fourier series** which he developed to study the solutions of certain partial differential equations. An example of a Fourier series was seen in 4.7. As we will see in book 6, Fourier transforms also have important applications to probability theory, where this transform is known as the **characteristic function.**

With i the standard notation for the complex unit $i \equiv \sqrt{-1}$, recall that as is verified by Taylor series analysis, e^{ix} is well defined by the power series:

$$e^{ix} \equiv \sum\nolimits_{j=0}^{\infty} \frac{(ix)^j}{j!}. \tag{6.11}$$

This series is absolutely convergent for all $x \in \mathbb{R}$, and so the summation can be split into even and odd terms:

$$\sum\nolimits_{j=0}^{\infty} \frac{(ix)^j}{j!} = \sum\nolimits_{j=0}^{\infty} \frac{x^{2j}}{(2j)!} + i \sum\nolimits_{j=0}^{\infty} \frac{x^{2j+1}}{(2j+1)j!}.$$

This obtains **Euler's formula:**

$$e^{ix} = \cos x + i \sin x, \tag{6.12}$$

named for **Leonhard Euler** (1707 – 1783).

A simple consequence of Euler's formula is that

$$\left|e^{ix}\right| = 1, \text{ all } x \in \mathbb{R}, \tag{6.13}$$

since $\cos^2 x + \sin^2 x = 1$ for all such x. This formula also contains the beautiful result known as **Euler's identity:**

$$e^{\pi i} = 1. \tag{6.14}$$

Finally, considering $e^{\pm ix}$ yields:

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}, \qquad \sin x = \frac{e^{ix} - e^{-ix}}{2i}. \tag{6.15}$$

Remark 6.24 *The formula in 6.11 generalizes to $z = a + bi$, $a, b \in \mathbb{R}$ using $e^z = e^a e^{ib}$ and a reordering of the series that is justified by absolute convergence:*

$$e^z \equiv \sum\nolimits_{k=0}^{\infty} \frac{z^k}{k!} = e^a \left(\cos b + i \sin b\right). \tag{6.16}$$

The error in the partial summations in 6.11 is given by the following proposition, and will be useful in the development that follows.

Proposition 6.25 *For every n :*

$$\left|e^{ix} - \sum\nolimits_{j=0}^{n} \frac{(ix)^j}{j!}\right| \leq \min\left[\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right]. \tag{6.17}$$

Proof. *First, $e^{ix} = 1 + i\int_0^x e^{it}dt$, and integration by parts yields:*

$$\int_0^x (x-t)^n e^{it} dt = \frac{x^{n+1}}{n+1} + \frac{i}{n+1} \int_0^x (x-t)^{n+1} e^{it} dt. \tag{(*)}$$

Starting with $n = 0$ and applying this formula iteratively obtains:

$$e^{ix} = \sum_{j=0}^{n} \frac{(ix)^j}{j!} + \frac{i^{n+1}}{n!} \int_0^x (x-t)^n e^{it} dt. \qquad ((**))$$

Hence by 6.13 and the triangle inequality, considering $x < 0$ and $x > 0$ separately produces:

$$\left| e^{ix} - \sum_{j=0}^{n} \frac{(ix)^j}{j!} \right| \leq \frac{|x|^{n+1}}{(n+1)!}.$$

Rewriting the equality in $(*)$:

$$\begin{aligned} \frac{i}{n} \int_0^x (x-t)^n e^{it} dt &= \int_0^x (x-t)^{n-1} e^{it} dt - \frac{x^n}{n} \\ &= \int_0^x (x-t)^{n-1} \left(e^{it} - 1\right) dt, \end{aligned}$$

and so

$$\frac{i^{n+1}}{n!} \int_0^x (x-t)^n e^{it} dt = \frac{i^n}{(n-1)!} \int_0^x (x-t)^{n-1} \left(e^{it} - 1\right) dt.$$

Since $\left|e^{it} - 1\right| \leq 2$, we estimate again the error term from the above expression for e^{ix} in $(**)$ to get:

$$\left| e^{ix} - \sum_{j=0}^{n} \frac{(ix)^j}{j!} \right| \leq \frac{2\,|x|^n}{n!}.$$

■

With the above preliminary results, we now turn to the main topic of this section.

Definition 6.26 *If $f(x)$ is a Lebesgue integrable function, $f : \mathbb{R} \to \mathbb{C}$, the* ***Fourier transform of*** *$f(x)$, denoted $\widehat{f}(t)$, is defined by:*

$$\widehat{f}(t) = \int_{-\infty}^{\infty} f(x) e^{itx} dm, \qquad (6.18)$$

where dm denotes Lebesgue measure. If μ is a finite Borel measure on $\mathbb{R}$, the ***Fourier-Stieltjes transform of the finite measure*** *μ, denoted $\phi_\mu(t)$, is defined by:*

$$\phi_\mu(t) = \int_{-\infty}^{\infty} e^{itx} d\mu. \qquad (6.19)$$

Notation 6.27 *As derived in chapter 5 of book 1, every finite Borel measure μ is induced by an increasing, bounded and right continuous function $F(x)$ in the sense that:*

$$\mu\left[(a,b]\right] \equiv F(b) - F(a).$$

Thus it is common to denote $\phi_\mu(t)$ in 6.19 as $\phi_F(t)$, and then it is also common to express this integral as a Lebesgue-Stieltjes integral:

$$\phi_F(t) = \int_{-\infty}^{\infty} e^{itx} dF.$$

Remark 6.28 **1**. *See remark 6.52 below on other conventions for the definition of $\widehat{f}(t)$. The particular formulation here was chosen to be consistent with the notion of a characteristic function studied in book 6.*

2. *Note that by the discussion above on integration of complex valued functions, both $\widehat{f}(t)$ and $\phi_\mu(t)$ are well defined for all t for the integrable functions and finite measures considered. By 6.13 and the triangle inequality,*

$$\left|\widehat{f}(t)\right| \leq \int_{-\infty}^{\infty} |f(x)|\, dm, \qquad \left|\phi_\mu(t)\right| \leq \int_{-\infty}^{\infty} d\mu,$$

and by definition,

$$\widehat{f}(0) = \int_{-\infty}^{\infty} f(x) dm, \qquad \phi_\mu(0) = \int_{-\infty}^{\infty} d\mu.$$

3. *When the "distribution" function F induced by μ is in turn defined by a nonnegative integrable "density" function $f(x)$ in the sense that:*

$$F(x) = \int_{-\infty}^{x} f(y) dm,$$

then by an application of the change of variables formula 3.2:

$$\phi_F(t) = \widehat{f}(t),$$

and hence these definitions agree. That is, the Fourier-Stieltjes transform of the finite measure μ equals the Fourier transform of the underlying "density" function f when this function exists. Consequently, results below for $\phi_F(t)$ are automatically true for $\widehat{f}(t)$ when $F(x)$ is defined by an integrable function $f(x)$.

Conversely, given a nonnegative integrable function $f(x)$ we can always create the associated finite Borel measure μ by the above steps, and thus $\widehat{f}(t) = \phi_F(t)$ again. Consequently, results below for $\widehat{f}(t)$ are automatically true for $\phi_F(t)$ for F defined by f.

4. *By 6.12,*

$$\widehat{f}(t) = \int_{-\infty}^{\infty} f(x)\cos xdm + i\int_{-\infty}^{\infty} f(x)\sin xdm, \tag{6.20}$$

so:

(**a**) *If $f(x)$ is an **even function**, meaning $f(x) = f(-x)$, then $\int_{-\infty}^{\infty} f(x)\sin xdm = 0$ and $\widehat{f}(t)$ is real for all t,*

(**b**) *If $f(x)$ is an **odd function**, meaning $f(x) = -f(-x)$, then $\int_{-\infty}^{\infty} f(x)\cos xdm = 0$ and $\widehat{f}(t)$ is purely imaginary for all t.*

Properties of Fourier Transforms

Some of the important properties of $\widehat{f}(t)$ and $\phi_F(t)$ are developed in the following propositions. See exercise 6.30 below for other results. The first proposition states that the smoothness of the Fourier-Stieltjes transform $\phi_F(t)$, is determined by the rate of decay of the μ_F-measure of sets as these sets move toward $\pm\infty$. In a probability context, this decay rate is captured in the notion of **fat tails** for probability measures which decay slowly, and **thin** or **skinny tails** for probability measures that decay rapidly.

More formally, part 1 of the next proposition states that for a finite measure μ_F, meaning:

$$\int_M^N d\mu_F < C_0 < \infty, \text{ all } N, M,$$

$\phi_F(t)$ will be uniformly continuous. However, if

$$\int_M^N |x|\, d\mu_F < C_1 < \infty, \text{ all } N, M,$$

then $\phi_F(t)$ will be once continuously differentiable, while if

$$\int_M^N |x|^n\, d\mu_F < C_n < \infty, \text{ all } N, M,$$

$\phi_F(t)$ will be n-times continuously differentiable. Infinite differentiability is also characterized.

In the case where μ_F is given by a distribution function $F(x)$ with associated density function $f(x)$, this general result can be restated with 3.2 as follows. If

$$\int_M^N |x|^n f(x)dm < C_n < \infty, \text{ all } N, M,$$

then $\widehat{f}(t)$ will be n-times continuously differentiable. Thus by this formula we see that the decay rate of $f(x)$ at $\pm\infty$ implies the degree of smoothness of $\widehat{f}(t)$. Specifically, if $|f(x)| \leq C|x|^{-n-1-\epsilon}$ as $|x| \to \infty$, then $\widehat{f}(t)$ will be n-times continuously differentiable.

Remark 6.29 *Note that when μ_F is a* ***probability measure****, the finiteness of these integrals can be stated in terms of the existence of moments. See chapter 3 of book 4, but more will be made of this connection in book 6.*

Exercise 6.30 *Derive the following properties of the Fourier transform, assuming that $f(x)$ and $g(x)$ are integrable. Justify the integrability of the other functions.*

1. *If $h(x) = f(x)e^{iax}$ for $a \in \mathbb{R}$, then $\widehat{h}(x) = \widehat{f}(x-a)$.*
2. *If $h(x) = f(x-a)$ for $a \in \mathbb{R}$, then $\widehat{h}(x) = \widehat{f}(x)e^{-iax}$.*
3. *If $h(x) = f(x/a)$ for $a > 0$, then $\widehat{h}(x) = a\widehat{f}(ax)$.*
4. *If $h(x) = af(x) + bg(x)$ for $a, b \in \mathbb{R}$, then $\widehat{h}(x) = a\widehat{f}(x) + b\widehat{g}(x)$.*
5. *If $h(x) = f(ax+b)$ for $a, b \in \mathbb{R}$, then $\widehat{h}(x) = e^{ibt}\widehat{f}(at)$.*

Proposition 6.31 *Given a finite measure μ_F:*

1. *$\phi_F(t)$ is uniformly continuous on $\mathbb{R}$.*
2. *If $\int_{-\infty}^{\infty} |x|^n d\mu_F < \infty$ for positive integer n, then $\phi_F(t)$ is differentiable up to order n,*

$$\phi_F^{(k)}(t) = \int_{-\infty}^{\infty} (ix)^k e^{itx} d\mu_F, \qquad \text{for } 1 \leq k \leq n, \tag{6.21}$$

and $\phi_F^{(k)}(t)$ is uniformly continuous for $1 \leq k \leq n$.

3. *If* $\int_{-\infty}^{\infty} e^{|sx|} d\mu_F < \infty$ *for any* $s \neq 0$, *then* $\phi_F(t)$ *is infinitely differentiable and for* $|t| \leq |s|$:

$$\phi_F(t) = \sum_{j=0}^{\infty} \frac{(it)^j}{j!} \int_{-\infty}^{\infty} x^k d\mu_F. \tag{6.22}$$

Proof. *Taking these statements in turn:*

1. *By 6.13,* $\left|e^{i(t+h)x} - e^{itx}\right| = \left|e^{ihx} - 1\right|$ *is independent of* t, *and so:*

$$\begin{aligned} |\phi_F(t+h) - \phi_F(t)| &\leq \int_{-\infty}^{\infty} \left|e^{ihx} - 1\right| d\mu_F \\ &\leq 2\int_{-\infty}^{\infty} d\mu_F. \end{aligned}$$

By continuity $e^{ihx} - 1 \to 0$ *for all* x *as* $h \to 0$, *and thus by Lebesgue's dominated convergence theorem,*

$$|\phi_F(t+h) - \phi_F(t)| \to 0$$

as $h \to 0$, *and this convergence is independent of* t.

2. *We prove the result for* $n = 1$, *and illustrate the general case by induction. Note that by 6.17:*

$$\begin{aligned} \left|\frac{\phi_F(t+h) - \phi_F(t)}{h}\right| &\leq \int_{-\infty}^{\infty} \left|\frac{e^{ihx} - 1}{h}\right| d\mu_F \\ &\leq \int_{-\infty}^{\infty} |x| \, d\mu_F, \end{aligned}$$

so by Lebesgue's dominated convergence it follows that:

$$\phi_F'(t) \equiv \lim_{h\to 0} \int_{-\infty}^{\infty} \left[\frac{e^{i(t+h)x} - e^{itx}}{h}\right] d\mu_F = \int_{-\infty}^{\infty} [ix] \, e^{itx} d\mu_F.$$

Moreover, by the argument in part 1, $\phi_F'(t)$ *is again seen to be uniformly continuous.*

The induction step then works as follows, illustrating $n = 2$ *for simplicity. First:*

$$\begin{aligned} \left|\frac{\phi_F'(t+h) - \phi_F'(t)}{h}\right| &\leq \int_{-\infty}^{\infty} |x| \left|\frac{e^{ihx} - 1}{h}\right| d\mu_F \\ &\leq \int_{-\infty}^{\infty} |x|^2 \, d\mu_F, \end{aligned}$$

and by Lebesgue's dominated convergence:

$$\phi_F''(t) \equiv \lim_{h\to 0}\int_{-\infty}^{\infty}[ix]\left[\frac{e^{i(t+h)x}-e^{itx}}{h}\right]d\mu_F = \int_{-\infty}^{\infty}[ix]^2 e^{itx}d\mu_F.$$

Uniform continuity of $\phi_F''(t)$ *follows as above.*

3. *Infinite differentiability follows since for any* n, $|x|^n \leq e^{|sx|}$ *for* $|x|/\ln|x| \geq n/|s|$ *and so the requirement of part 2 is satisfied for all* n. *For this conclusion note that this inequality is equivalent to* $|x| \geq c_n$ *since* $|x|/\ln|x|$ *is increasing and unbounded. For the expansion in 6.25, 6.17 provides:*

$$\left|\phi_F(t) - \sum\nolimits_{j=0}^{n}\frac{(it)^j}{j!}\int_{-\infty}^{\infty}x^k d\mu_F\right| \leq \frac{|t|^{n+1}}{(n+1)!}\int_{-\infty}^{\infty}|x|^{n+1}d\mu_F,$$

and the proof will be completed by showing that this upper bound can be made small uniformly in n *if* $|t| \leq |s|$.

To this end, the μ_F*-integrability of* $e^{|sx|}$ *justifies the application of corollary 2.48 to Lebesgue's dominated convergence theorem to conclude that*

$$\int_{-\infty}^{\infty}e^{|sx|}d\mu_F = \sum\nolimits_{j=0}^{\infty}\frac{(|s|)^j}{j!}\int_{-\infty}^{\infty}|x|^j d\mu_F.$$

This implies that for any $\epsilon > 0$ *there is an* N *so that for* $n \geq N$:

$$\sum\nolimits_{j=n}^{\infty}\frac{(|s|)^j}{j!}\int_{-\infty}^{\infty}|x|^j d\mu_F < \epsilon.$$

Hence if $|t| \leq |s|$, *this can be applied in the* $\phi_F(t)$ *estimate above to obtain that for any* $\epsilon > 0$,

$$\left|\phi_F(t) - \sum_{j=0}^{n}\frac{(it)^j}{j!}\int_{-\infty}^{\infty}x^k d\mu_F\right| < \epsilon$$

for $n \geq N(\epsilon)$, *which is 6.22.*

■

As noted above, the prior proposition and 3.2 imply that the degree of the decay rate of $f(x)$ at $\pm\infty$ determines the degree of smoothness of

$\widehat{f}(t)$ as defined in terms of the existence of continuous derivatives. Specifically, if $|x|^n f(x)$ is integrable then $\widehat{f}(t)$ will be n-times continuously differentiable. Also noted above, this integrability condition is assured if $|f(x)| \leq C|x|^{-n-1-\epsilon}$ as $|x| \to \infty$. Hence the decay rate of $f(x)$ implies smoothness of $\widehat{f}(t)$.

The next proposition provides another connection between decay rate and smoothness between $f(x)$ and $\widehat{f}(t)$. Here it is the smoothness of $f(x)$, defined in terms of the integrability of $f(x)$ and its derivatives, that determines the decay rate of $\widehat{f}(t)$ as $t \to \pm\infty$. Part 1 of this result is the **Riemann-Lebesgue Lemma,** named for **Bernhard Riemann** (1826 – 1866) and **Henri Lebesgue** (1875 – 1941). It states that if $f(x)$ is integrable then $\left|\widehat{f}(t)\right| \to 0$ as $t \to \pm\infty$. Part 2 generalizes this in that if $f'(x)$ exists and is integrable, then $\left|\widehat{f}(t)\right| = o(t^{-1})$ as $t \to \pm\infty$, and if all derivatives up to the nth derivative $f^{(n)}(x)$ exist and are integrable, then $\left|\widehat{f}(t)\right| = o(t^{-n})$ as $t \to \pm\infty$.

Recall the following "big-O" and "little-o" terminology.

Definition 6.32 (big O, little o) *Let $f(x)$ and $g(x) > 0$ be two functions defined on $\mathbb{R}$. The expression*

$$f(x) = O(g(x)) \text{ as } x \to \infty,$$

*or in words, "$f(x)$ **is big-O of** $g(x)$ **as** $x \to \infty$," means that there is an x_0 and positive constant M so that:*

$$|f(x)| \leq Mg(x) \text{ for all } x \geq x_0.$$

The expression

$$f(x) = o(g(x)) \text{ as } x \to \infty,$$

*or in words, "$f(x)$ **is little-o of** $g(x)$ **as** $x \to \infty$," if:*

$$\frac{f(x)}{g(x)} \to 0 \text{ as } x \to \infty.$$

Remark 6.33 *To highlight the connection between 1 and 2 of this proposition, 6.23 is sometimes written as $\left|\widehat{f}(t)\right| = o(1)$ as $t \to \pm\infty$.*

Proposition 6.34 *Given a Lebesgue measurable function $f(x)$:*

1. *(**Riemann-Lebesgue lemma**) If $f(x)$ is integrable then:*

$$\left|\widehat{f}(t)\right| \to 0 \text{ as } t \to \pm\infty \ . \tag{6.23}$$

2. *If $f^{(k)}(x)$ exists and is an integrable function for $k \leq n$, then:*

$$\left|\widehat{f}(t)\right| = o(|t|^{-n}) \text{ as } t \to \pm\infty \ . \tag{6.24}$$

Proof. *For the Riemann-Lebesgue lemma, assume $f(x)$ is nonnegative and integrable as the general case follows from applying this result to $f^{+}(x)$ and $f^{-}(x)$ of definition 2.36. By proposition 1.24 there is an increasing sequence of step functions $\{\varphi_n(x)\}_{j=1}^{\infty}$, each defined on a disjoint finite collection of right semi-closed intervals, each is 0 outside a set of finite measure, and $\varphi_n(x) \to f(x)$ for all x. Hence by Lebesgue's dominated convergence theorem (above, or proposition 2.61 of book 3),*

$$\int_{-\infty}^{\infty} \varphi_n(x)dm \to \int_{-\infty}^{\infty} f(x)dm,$$

and for any $\epsilon > 0$ there is an M so that for $n \geq M$:

$$\int_{-\infty}^{\infty} |f(x) - \varphi_n(x)|\, dm < \epsilon/2.$$

Since $||a| - |b|| \leq |a - b|$ for $a, b \in \mathbb{R}$, it follows from this, 6.13, and the triangle inequality that for $n \geq M$ and all t :

$$\begin{aligned} & \left|\left|\int_{-\infty}^{\infty} f(x)e^{itx}dm\right| - \left|\int_{-\infty}^{\infty} \varphi_n(x)e^{itx}dm\right|\right| \\ \leq & \left|\int_{-\infty}^{\infty} f(x)e^{itx}dm - \int_{-\infty}^{\infty} \varphi_n(x)e^{itx}dm\right| \\ < & \ \epsilon/2. \end{aligned}$$

If for such n, $\varphi_n(x) = \sum_{j=1}^{N_n} c_j \chi_{(a_j,b_j]}(x)$, then:

$$\int_{-\infty}^{\infty} \varphi_n(x)e^{itx}dm = \sum\nolimits_{j=1}^{N_n} c_j \left[e^{ita_j} - e^{itb_j}\right]/it,$$

and by 6.13:

$$\left|\int_{-\infty}^{\infty} \varphi_n(x)e^{itx}dm\right| \leq 2\sum\nolimits_{j=1}^{N_n} |c_j|\,/t.$$

Hence as a function of t, $\left|\int_{-\infty}^{\infty} f(x)e^{itx}dm\right|$ *can be made arbitrarily close to* $\left|\int_{-\infty}^{\infty} \varphi_n(x)e^{itx}dm\right|$, *which in turn converges to* 0 *as* $t \to \infty$, *and this obtains the Riemann-Lebesgue lemma.*

To prove 6.24, recall Lebesgue integration by parts in 6.6. To justify this application recall 4 of remark 3.55 of book 3 to conclude that the assumption on the existence of $f^{(k)}(x)$ *for* $k \leq n$ *implies that* $f^{(k)}(x)$ *is absolutely continuous for* $0 \leq k \leq n-1$. *Similarly, as* e^{itx} *is infinitely differentiable as a function of t, it is also absolutely continuous and so by 6.6:*

$$\begin{aligned} \widehat{f}(t) &= \int_{-\infty}^{\infty} f(x)e^{itx}dm \\ &= \frac{-1}{it}\int_{-\infty}^{\infty} f'(x)e^{itx}dm \\ &\vdots \\ &= \frac{1}{(-it)^n}\int_{-\infty}^{\infty} f^{(n)}(x)e^{itx}dm. \qquad ((*)) \end{aligned}$$

For this derivation, the Riemann-Lebesgue lemma and the integrability of $f^{(k)}(x)$ *for* $0 \leq k \leq n-1$ *imply that the limit terms at* $\pm\infty$ *equal* 0. *The integrability of* $f^{(n)}(x)$ *then implies that*

$$\left|\widehat{f}(t)\right| = |t|^{-n}\left|\int_{-\infty}^{\infty} f^{(n)}(x)e^{itx}dm\right|,$$

which is a big-O estimate: $\left|\widehat{f}(t)\right| = O(|t|^{-n})$ *by integrability of* $f^{(n)}(x)$. *But then one last application of the Riemann-Lebesgue lemma to integrable* $f^{(n)}(x)$ *improves this result to* $\left|\widehat{f}(t)\right| = o(|t|^{-n})$. ■

Corollary 6.35 *If* $f^{(k)}(x)$ *is an integrable function for* $k \leq n$, *then for any such* $k:$

$$\widehat{f^{(k)}}(x) = (-it)^k\,\widehat{f}(t). \qquad (6.25)$$

Proof. *Immediate from* $(*)$ *in the above proof.* ■

The final result of this section states that Fourier transforms work naturally with convolutions. Note that it makes sense to contemplate $\widehat{f*g}(t)$, the Fourier transform of the convolution of integrable f and g, since $f*g(x)$ is then integrable by proposition 6.18, and thus $\widehat{f*g}(t)$ well defined.

Proposition 6.36 *For integrable functions $f(x)$ and $g(x)$:*

$$\widehat{f * g}(t) = \widehat{f}(t)\widehat{g}(t). \tag{6.26}$$

Proof. *Not surprisingly, this result requires an application of Fubini's theorem. First by 6.12:*

$$\begin{aligned}
\widehat{f * g}(t) &= \int_{-\infty}^{\infty} f * g(x) e^{itx} dm \\
&\equiv \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(y) g(x-y) dm(y) \right] e^{itx} dm(x). \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(y) g(x-y) dm(y) \right] \cos tx dm(x) \\
&\quad + i \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(y) g(x-y) dm(y) \right] \sin tx dm(x).
\end{aligned}$$

To apply Fubini's result, it must be verified that these integrands are integrable on $(\mathbb{R}^2, \mathcal{M}_L^2, m^2)$. Proposition 6.18 proved the m^2-integrability of $f(y)g(x-y)$, and since $\cos tx$ and $\sin tx$ are absolutely bounded by 1 the conclusion follows. An application of Fubini's theorem is hence justified and the order of these iterated integrals can be reversed without changing the values of the integrals:

$$\begin{aligned}
\widehat{f * g}(t) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x-y) \cos tx dm(x) \right] f(y) dm(y) \\
&\quad + i \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x-y) \sin tx dm(x) \right] f(y) dm(y) \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x-y) e^{it(x-y)} dm(x) \right] f(y) e^{ity} dm(y).
\end{aligned}$$

The final step is to use 3.10 on the x-integral. Specifically, for any y define $T : (\mathbb{R}, \mathcal{M}_L, m) \to (\mathbb{R}, \mathcal{M}_L, m_T)$ by $T(x) = x - y$. Then $m_T = m$ by the translation invariance of Lebesgue measure, and by 3.10,

$$\int_{-\infty}^{\infty} g(x-y) e^{it(x-y)} dm(x) = \int_{-\infty}^{\infty} g(z) e^{itz} dm_T(z) = \widehat{g}(t),$$

and the result follows. ■

6.3.3 Fourier Inversion: From $\phi_F(t)$ to μ_F

The previous section demonstrated that when considered in pairs: $(\mu_F, \phi_F(t))$ or $(f(x), \widehat{f}(t))$, certain properties of μ_F or $f(x)$ would predict properties of the associated $\phi_F(t)$ or $\widehat{f}(t)$. For example, proposition 6.31 states that the smoothness of $\phi_F(t)$ is determined by the rate of decay of the measure of sets near $\pm\infty$, a notion made precise by the finiteness of $\int_{-\infty}^{\infty} |x^n|\, d\mu_F$. When this integral is finite, we are assured that $\phi_F(t)$ is n-times continuously differentiable. Applied to a statement about $f(x)$, this result states that if $\int_{-\infty}^{\infty} |x|^n f(x)dm$ is finite then $\widehat{f}(t)$ is n-times continuously differentiable. In addition, proposition 6.34 states that when $f^{(k)}(x)$ exists and is Lebesgue integrable for $k \leq n$, the rate of decay of $\widehat{f}(t)$ at $\pm\infty$ is predictable, and in particular, $\widehat{f}(t) = o(|t|^{-n})$ as $t \to \pm\infty$.

The question addressed in this section relates to the invertibility of the Fourier transform, or in general, the Fourier-Stieltjes transform. In other words, if we are given $\phi_F(t)$ or $\widehat{f}(t)$, can the defining μ_F or $f(x)$ be recovered? If so, this would imply that the pairings $(\mu_F, \phi_F(t))$ or $(f(x), \widehat{f}(t))$ are unique, which has important probability theory applications in book 6, but also that given $\phi_F(t)$ or $\widehat{f}(t)$, there is a constructive method for finding the associated μ_F or $f(x)$.

To prepare for the statement and proof of the main result, we require a small modification of the result in 5.13 from proposition 5.27 that with $S(t) \equiv \int_0^t \frac{\sin x}{x} dx$:

$$\lim_{t\to\infty} S(t) = \frac{\pi}{2}.$$

Defining $S_\theta(t) \equiv \int_0^t \frac{\sin \theta x}{x} dx$ for $\theta \in \mathbb{R}$, it follows by a simple substitution that $S_\theta(t) = S(t\theta)$. To eliminate the ambiguity as to whether θ is positive or negative, write:

$$S_\theta(t) = sgn(\theta) S(t\,|\theta|),$$

where the **sign function** or **signum function** $sgn(\theta)$ is defined by:

$$sgn(\theta) = \begin{cases} 1, & \theta > 0, \\ 0, & \theta = 0, \\ -1, & \theta < 0. \end{cases}$$

We then can conclude from 5.13 that

$$\lim_{t\to\infty} S_\theta(t) = \frac{\pi}{2} sgn(\theta). \tag{6.27}$$

Proposition 6.37 (Inverse Fourier-Stieltjes transform) *If μ_F is a finite Borel measure on $\mathbb{R}$ and $\phi_F(t)$ its Fourier-Stieltjes transform, then for $b > a$,*

$$\lim_{T\to\infty}(2\pi)^{-1}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\phi_F(t)dm = \mu_F[(a,b)] + \frac{1}{2}\mu_F[\{a,b\}], \quad (6.28)$$

where $\mu_F[\{a,b\}]$ denotes the Borel measure of these points: $\mu_F[a]+\mu_F[b]$.
Proof. *Note that the integrand in 6.28 does not have a singularity at $t=0$ since*

$$\frac{e^{-iat}-e^{-ibt}}{it} = \int_a^b e^{-itx}dx,$$

and this is bounded in absolute value by $b-a$ by 6.13. By definition of $\phi_F(t)$,

$$\begin{aligned}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\phi_F(t)dm &= \int_{-T}^{T}\int_{-\infty}^{\infty}\frac{e^{-iat}-e^{-ibt}}{it}e^{itx}d\mu_F(x)dm(t)\\ &= \int_{-T}^{T}\int_{-\infty}^{\infty}\frac{e^{i(x-a)t}-e^{i(x-b)t}}{it}d\mu_F(x)dm(t).\end{aligned}$$

This integrand is continuous on $[-T,T]\times\mathbb{R}$ and bounded by $b-a$ as noted above, and hence this function is integrable on $([-T,T]\times\mathbb{R}, \sigma([-T,T]\times\mathbb{R}), m\times\mu_F)$ since μ_F is a finite Borel measure. By Fubini's theorem the order of the iterated integrals can be reversed, and then by Euler's formula in 6.12:

$$\begin{aligned}\int_{-T}^{T}\frac{e^{i(x-a)t}-e^{i(x-b)t}}{it}dm(t) &= \int_{-T}^{T}\frac{\cos[(x-a)t]-\cos[(x-b)t]}{it}dm(t)\\ &\quad+\int_{-T}^{T}\frac{\sin[(x-a)t]-\sin[(x-b)t]}{t}dm(t).\end{aligned}$$

The first integral equals 0 because the integrand is an odd function of t, meaning $g(-t)=-g(t)$ and so $\int_{-T}^{0}g(t)dt = -\int_0^T g(t)dt$. In addition, the integrand in the second integral is an even function, meaning $h(-t)=h(t)$, so $\int_{-T}^{0}h(t)dt = \int_0^T h(t)dt$. Hence

$$(2\pi)^{-1}\int_{-T}^{T}\frac{e^{-iat}-e^{-ibt}}{it}\phi_F(t)dm = \frac{1}{\pi}\int_{-\infty}^{\infty}\left[S_{(x-a)}(T)-S_{(x-b)}(T)\right]d\mu_F(x).$$

By 6.27 this integrand is bounded and converges for all x as $T\to\infty$ to the

μ_F-integrable function $\lambda_{a,b}(x)$ defined by:

$$\lambda_{a,b}(x) = \begin{cases} 0, & x < a, \\ \frac{1}{2}, & x = a, \\ 1, & a < x < b, \\ \frac{1}{2}, & x = b, \\ 0, & x > b. \end{cases}$$

Because μ_F is a finite measure and the integrand is bounded, Lebesgue's dominated convergence theorem can be applied to derive:

$$\begin{aligned} \lim_{T\to\infty} (2\pi)^{-1} \int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} \phi_F(t) dm &= \int_{-\infty}^{\infty} \lambda_{a,b}(x) d\mu_F(x) \\ &= \mu_F[(a,b)] + \frac{1}{2}\mu_F[\{a,b\}]. \end{aligned}$$

■

Notation 6.38 *The limit in 6.28 is called the* ***inverse Fourier-Stieltjes transform of*** *$\phi_F(t)$ since this limit reproduces the measure on which the Fourier-Stieltjes transform is defined.*

Remark 6.39 *Note that for any T, the integral in 6.28 is also defined as a Riemann integral since the integrand is continuous by proposition 6.31, and for such Riemann integrable functions, proposition 2.31 of book 3 applies.*

Also, the use of a limit in 6.28 and of the symmetrical integral over $[-T, T]$ is essential in general. This is because the integrand need not be integrable over $\mathbb{R}$, and hence the limit need not exist if more generally defined. This approach to the evaluation of an improper integral $\int_{-\infty}^{\infty} f dm$ is called the ***Cauchy principal value of the improper integral****, named for* ***Augustin-Louis Cauchy*** *(1789 – 1857). Cauchy introduced this and related evaluations of improper integrals which were not formally definable by traditional means. Of course, this limit produces the correct value when the integral over $\mathbb{R}$ exists.*

It is not difficult to develop examples of functions for which the improper integrals exist in the sense of the principal value but not formally as improper integrals. For example $f(x) = \sin x$ has limit 0 since $\int_{-T}^{T} f dm = 0$ for all T, but this integral does not exist more generally. The same is true of many odd functions. The next example illustrates that the Cauchy principal value in 6.28 is also necessary, since the integrand there need not be integrable.

Example 6.40 *Let $F(x) = 0$ for $x < 0$ and $F(x) = 1$ otherwise, so that μ_F is the finite measure with $\mu_F(A) = 1$ if $0 \in A$ and $\mu_F(A) = 0$ otherwise. Then,*

$$\phi_F(t) = \int_{-\infty}^{\infty} e^{itx} d\mu_F(x) = 1,$$

which is clearly not Lebesgue integrable. However:

$$\int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} dm(t) = \int_{-T}^{T} \int_{a}^{b} e^{-itx} dx dm(t),$$

and this integrand is continuous and bounded and hence integrable on the bounded domain $[-T, T] \times [a, b]$. By Fubini's theorem,

$$\begin{aligned} \int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} dm(t) &= \int_{a}^{b} \int_{-T}^{T} e^{-itx} dm(t) dx \\ &= \int_{a}^{b} \frac{e^{iTx} - e^{-iTx}}{ix} dx \\ &= 2 \int_{a}^{b} \frac{\sin Tx}{x} dx \\ &= 2 \int_{Ta}^{Tb} \frac{\sin y}{y} dy. \end{aligned}$$

Thus, notationally inserting $\phi_F(t) = 1$ into this expression:

$$(2\pi)^{-1} \int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} \phi_F(t) dm = \frac{1}{\pi} \int_{Ta}^{Tb} \frac{\sin y}{y} dy.$$

If $a, b > 0$ or $a, b < 0$, this integral converges to 0 by 5.13, which agrees with the value of $\mu_F[(a, b)]$. However, if $a < 0 < b$, then

$$\frac{1}{\pi} \int_{Ta}^{Tb} \frac{\sin y}{y} dy = \frac{1}{\pi} \left[\int_{0}^{Tb} \frac{\sin y}{y} dy + \int_{0}^{-Ta} \frac{\sin y}{y} dy \right],$$

which converges to 1, again the value of $\mu_F[(a, b)]$. In all such cases, $\mu_F[a] + \mu_F[b] = 0$ and thus these results agree with 6.28. The reader is invited to investigate the cases where $a = 0$ or $b = 0$ to again confirm 6.28, which in this case is a limit of $1/2$.

Exercise 6.41 *Let $F(x) = 0$ for $x < 0$, and $F(x) = 1$ otherwise, so that μ_F is the finite measure with $\mu_F(A) = 1$ if $0 \in A$ and $\mu_F(A) = 0$ otherwise. Show that*

$$\lim_{S,T \to \infty} (2\pi)^{-1} \int_{-S}^{T} \frac{e^{-iat} - e^{-ibt}}{it} \phi_F(t) dm$$

does not exist in general, meaning that this limit is not independent of how S *and* T *approach* ∞.

Corollary 6.42 (Uniqueness of the Fourier-Stieltjes transform) *If* μ_F *and* μ_G *are finite Borel measures on* $\mathbb{R}$ *and their Fourier-Stieltjes transforms satisfy* $\phi_F(t) = \phi_G(t)$ *for all* t*, then* $\mu_F(A) = \mu_G(A)$ *for all Borel sets,* $A \in \mathcal{B}(\mathbb{R})$.

Proof. *Recalling proposition 5.7 of book 1, there is no loss of generality to assume that the right continuous increasing functions underlying these measures have been normalized to satisfy* $F(-\infty) = G(-\infty) = 0$*, since changing such functions by a constant does not change the resulting measures. On* $(a, b] \in \mathcal{A}'$*, the semi-algebra of right semi-closed intervals, recall that:*

$$\mu_F\left[(a, b]\right] = F(b) - F(a), \qquad \mu_G\left[(a, b]\right] = G(b) - G(a).$$

If it can be shown that $\mu_F(A) = \mu_G(A)$ *for all* $A \in \mathcal{A}$*, the algebra generated by* $\mathcal{A}'$*, then the uniqueness theorem of proposition 6.14 of book 1 assures that* $\mu_F(A) = \mu_G(A)$ *for all* $A \in \sigma(\mathcal{A})$*. Here* $\sigma(\mathcal{A})$ *denotes the smallest sigma algebra that contains* $\mathcal{A}$*, and thus* $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ *and the proof will be complete.*

To show that $\mu_F(A) = \mu_G(A)$ *for all* $A \in \mathcal{A}$ *it is enough to show this for* $A \in \mathcal{A}'$ *since* $\mathcal{A}$ *is the collection of finite disjoint unions of* $\mathcal{A}'$*-sets. To this end, if* $A = (a, b]$ *is a right semi-closed interval for which both* a *and* b *are continuity points of both* $F(x)$ *and* $G(x)$*, then* $\mu_F[\{a, b\}] = 0$ *and thus* $\mu_F[(a, b]] = \mu_F[(a, b)]$*. As the same is true for* μ_G*, the assumption that* $\phi_F(t) = \phi_G(t)$ *implies from 6.28 that* $\mu_F[(a, b]] = \mu_G[(a, b]]$.

Since F *and* G *can have at most countably many discontinuity points, for any set* $A = (a, b] \in \mathcal{A}'$*, define decreasing sequences of continuity points of* F *and* G*,* $a_j \to a$ *and* $b_j \to b$ *where* $a_j > a$ *and* $b_j > b$*. Then*

$$(a, b] = \bigcup\nolimits_{j=1}^{\infty} (a_{j+1}, a_j] \cup (a_1, b],$$

and by countable additivity,

$$\mu_F\left[(a, b]\right] = \sum\nolimits_{j=1}^{\infty} \mu_F\left[(a_{j+1}, a_j]\right] + \mu_F\left[(a_1, b]\right],$$

and an identical statement holds for $\mu_G\left[(a, b]\right]$*. The summations contain only* $\mathcal{A}'$*-sets with endpoints on which both* $F(x)$ *and* $G(x)$ *are continuous, and so*

$$\sum\nolimits_{j=1}^{\infty} \mu_F\left[(a_{j+1}, a_j]\right] = \sum\nolimits_{j=1}^{\infty} \mu_G\left[(a_{j+1}, a_j]\right].$$

By right continuity,

$$\mu_F\left[(a_1,b]\right] = \inf_j \mu_F\left[(a_1,b_j]\right],$$

with an identical statement for $\mu_G\left[(a_1,b]\right]$. *Since all endpoints are continuity points of* F *and* G:

$$\inf_j \mu_F\left[(a_1,b_j]\right] = \inf_j \mu_G\left[(a_1,b_j]\right],$$

and thus $\mu_F\left[(a,b]\right] = \mu_G\left[(a,b]\right]$ *and the proof is complete.* ■

Corollary 6.43 *If* μ_F *is a finite Borel measure on* $\mathbb{R}$ *and* $\phi_F(t) = 0$, *then* $\mu_F(A) = 0$ *for all* $A \in \mathcal{B}(\mathbb{R})$.
Proof. *Define* $G(x) \equiv 0$ *and use the prior proposition.* ■

6.3.4 The Special Case of Integrable $\phi_F(t)$

As demonstrated in the example of the previous section, $\phi_F(t)$ need not be integrable for a finite measure μ_F. Yet even in the general case the integral in 6.28, which is also definable as a Riemann integral due to the continuity of the integrand, converges to a limit as $T \to \infty$. This limit then provides the value of $\mu_F\left[(a,b]\right]$ for all such intervals where a and b are continuity points of $F(x)$, and more generally provides the value of $\mu_F[(a,b)] + \frac{1}{2}\mu_F[\{a,b\}]$.

In this section we investigate the implications for μ_F in the special case when $\phi_F(t)$ is in fact integrable, meaning:

$$\int_{-\infty}^{\infty} |\phi_F(t)|\, dm < \infty.$$

Extending the connections between properties of a measure and those of its Fourier transform, the following proposition then states that integrability of $\phi_F(t)$ assures that the increasing function $F(x)$ underlying the measure μ_F has a continuous density function $f(x)$ for which $F(x) = \int_{-\infty}^{x} f(y)dy$.

This next result is in fact related to the Riemann-Lebesgue lemma and generalization given in proposition 6.34. The Riemann-Lebesgue lemma states that if the finite measure μ_F is given by a density function $f(x)$, there presumed only Lebesgue integrable, then $|\phi_F(t)| \to 0$ as $t \to \pm\infty$. The generalization states that if this density function $f(x)$ has an integrable first derivative then $|\phi_F(t)| = o(|t|^{-1})$ as $t \to \pm\infty$, and if n integrable derivatives then $|\phi_F(t)| = o(|t|^{-n})$ as $t \to \pm\infty$. For $n \geq 2$, continuity of $\phi_F(t)$ and this estimate therefore assure the Riemann integrability of $\phi_F(t)$.

The main result of this section reverses this implication to state that if $\phi_F(t) \to 0$ as $t \to \pm\infty$ fast enough and/or appropriately enough to ensure integrability of $\phi_F(t)$, then the associated density function $f(x)$ must be continuous. Before developing this result, the next section investigates relationships between integrability and growth at infinity.

Integrability Versus Growth at $\pm\infty$

First, integrability of a function $h(x)$ does not ensure $h(x) \to 0$ as $x \to \pm\infty$, nor even that $h(x)$ is bounded.

Example 6.44 *Define $h(x) = n$ on $[n - 1/(n2^{n+1}), n + 1/(n2^{n+1})]$, for $n = 1, 2, 3...$, and $h(x) = 0$ otherwise. Then $h(x)$ is Lebesgue (and indeed, Riemann) integrable, with $\int h(x)dm = 1$, but $h(x) \nrightarrow 0$ as $x \to \pm\infty$, and $h(x)$ is unbounded.*

However, since $\phi_F(t)$ is always uniformly continuous and the above function $h(x)$ is not, perhaps integrability of a uniformly continuous function $h(x)$ ensures $h(x) \to 0$ as $x \to \pm\infty$. The next result is often called **Barbalat's Lemma**, though a historical reference is unavailable.

Proposition 6.45 (Barbalat's Lemma) *If $h(x)$ is uniformly continuous and integrable on $\mathbb{R}$, then $|h(x)| \to 0$ as $x \to \pm\infty$.*
Proof. *Assume not, then for some $\epsilon > 0$ there is an unbounded increasing sequence $\{x_n\}$ so that $|h(x_n)| \geq \epsilon$ for all n. By uniform continuity, there is a δ so that*

$$\left||h(x)| - \left|h(x')\right|\right| \leq \left|h(x) - h(x')\right| < \epsilon/2$$

if $|x - x'| < \delta$, and hence $|h(x)| \geq \epsilon/2$ on $[x_n - \delta, x_n + \delta]$ for all n, contradicting the integrability of $h(x)$. ■

Conclusion 6.46 (1.) *Integrability of uniformly continuous $h(x)$ implies that $|h(x)| \to 0$ as $t \to \pm\infty$.*

Next we investigate by a series of examples if there is a more quantitative implication for the degree of decay at $\pm\infty$ if a function is uniformly continuous and integrable.

Example 6.47 1. *If $h(x) = O(|x|^{-1-\epsilon})$ as $x \to \pm\infty$ for any $\epsilon > 0$ and is continuous $\Longrightarrow h(x)$ is integrable.*

By definition of $O(|x|^{-1-\epsilon})$ there is an x_0 so that for $|x| \geq |x_0|$, we have $|h(x)| \leq M\,|x|^{-1-\epsilon}$ and so by this and continuity $h(x)$ is integrable *on $|x_0| \geq |x|$. By continuity $h(x)$ is bounded and hence integrable on $|x| \leq |x_0|$.*

2. *If $h(x) = O(|x|^{-1})$ as $x \to \pm\infty$ and is uniformly continuous $\not\Rightarrow$ $h(x)$ is integrable.*

   Define $g(x) = 1$ for $|x| \leq 1$ and $g(x) = 1/|x|$ otherwise. Then $g(x)$ is uniformly continuous, $g(x) = O(|x|^{-1})$ as $x \to \pm\infty$, but $g(x)$ is not integrable.

3. *If $h(x) = o(|x|^{-1})$ as $x \to \pm\infty$ and is uniformly continuous $\not\Rightarrow$ $h(x)$ is integrable.*

   Define $h(x) = e/(|x| \ln |x|)$ on $e \leq |x| < \infty$ and $h(x) = 1$ otherwise. Then $h(x)$ is uniformly continuous and $h(x) = o(|x|^{-1})$ as $x \to \pm\infty$, but $h(x)$ is not integrable since by substitution:

$$\int_{-\infty}^{\infty} h(x)dm = 2e\left[1 + \int_1^{\infty} \frac{dy}{y}\right].$$

4. *Given any strictly decreasing function $g(x)$ on $[x_0, \infty)$ for $x_0 > 0$ with $g(x) \to 0$ as $x \to \infty$, there is an integrable and uniformly continuous function $h(x)$ with $h(x) = O(g(|x|))$ as $x \to \pm\infty$.*

   First let $g(x) = 1/\ln x$ on $[e, \infty)$, and define $h(x) = 1/\ln |x|$ on $\left[\bigcup_n [e^{n^2}, e^{n^2}+1]\right] \cup \left[\bigcup_n [-e^{n^2}-1, -e^{n^2}]\right]$, and $h(x) = 0$ otherwise. Then $h(x) = O(g(|x|))$ as $x \to \pm\infty$, and $h(x)$ is integrable since

$$\int_{-e^{n^2}-1}^{-e^{n^2}} h(x)dm = \int_{e^{n^2}}^{e^{n^2}+1} h(x)dm < e/n^2.$$

   In the general case, since $1/g(x)$ is strictly increasing and unbounded, replace e^{n^2} in the above example by a_n where $1/g(a_n) = n^2$.

   It is then an exercise to show that such $h(x)$ can be modified to be continuous without affecting integrability.

Conclusion 6.48 (2.) *Integrability of uniformly continuous $h(x)$ need not imply any decay rate at infinity stronger than $|h(x)| \to 0$ as $x \to \pm\infty$.*

Remark 6.49 *This second conclusion is initially surprising perhaps, because one often thinks of the examples $h(x) = |x|^{-1-\epsilon}$ as integrable on $|x| \geq 1$ for $\epsilon > 0$ and $h(x) = |x|^{-1+\epsilon}$ as not integrable for $\epsilon \geq 0$. But* *the key insight is that a function's order of magnitude at $\pm\infty$ is not changed by redefining the function to be zero on even infinitely many intervals, as long as the original definition is preserved for infinitely large x. On the other hand, the value of, and even existence of, the integral of a function can be materially modified by such redefinitions. The above construction shows that one can redefine the function enough to make it integrable without changing the original order of magnitude.*

Looked at a different way, Lebesgue measure is translation invariant, while the notion of order of magnitude is clearly not. For example, the integral of $h(x)$ defined above with $g(x) = 1/\ln x$ is the same as the integral of $k(x)$ defined as a translated version of $h(x)$. Specifically, on $[n, n{+}1]$ define $k(x)$ to have the same values as does $h(x)$ on $[e^{n^2}, e^{n^2}+1]$. In other words, $k(n) = 1/n^2$ and in general,

$$k(n+\lambda) = \left[n^2 + \ln(1+\lambda e^{-n^2})\right]^{-1}, \ 0 \leq \lambda \leq 1.$$

So while the integral of $k(x)$ equals that of $h(x)$, we now see that $k(x) = O\left(|x|^{-2}\right)$.

Lebesgue Integrable $\phi_F(t)$

The main result for Lebesgue integrable $\phi_F(t)$ is summarized in the next proposition. Namely, the associated measure μ_F is then induced by an increasing function F with continuous and necessarily nonnegative density function $f(x)$ with:

$$F(x) = \int_{-\infty}^{x} f(y)dm.$$

Since μ_F is finite, F is bounded and hence f is Lebesgue integrable. It then follows by continuity that f is also Riemann integrable. And extending 6.28, such f is recoverable from ϕ_F.

Proposition 6.50 (Inverse Fourier transform) *Let μ_F be a finite measure and $\phi_F(t)$ its Fourier-Stieltjes transform. If $\phi_F(t)$ is Lebesgue integrable, then there is a continuous function $f(x)$ so that:*

$$\mu_F[A] = \int_A f(x)dm, \ A \in \mathcal{B}(\mathbb{R}), \tag{6.29}$$

and thus $\phi_F(t) = \widehat{f}(t)$. In addition, $f(x)$ is given by:

$$f(x) = (2\pi)^{-1}\int_{-\infty}^{\infty} \phi_F(t)e^{-ixt}dm. \tag{6.30}$$

Proof. *By 6.17 and 6.13:*

$$\left|\frac{e^{-iat}-e^{-ibt}}{it}\right| = \frac{\left|e^{i(b-a)t}-1\right|}{|t|} \leq (b-a),$$

and thus because $\phi_F(t)$ is Lebesgue integrable 6.28 can be rewritten as:

$$(2\pi)^{-1}\int_{-\infty}^{\infty}\frac{e^{-iat}-e^{-ibt}}{it}\phi_F(t)dm = \mu_F[(a,b)] + \frac{1}{2}\mu_F[\{a,b\}].$$

The triangle inequality yields:

$$\left|\int_{-\infty}^{\infty}\frac{e^{-iat}-e^{-ibt}}{it}\phi_F(t)dm\right| \leq (b-a)\int_{-\infty}^{\infty}|\phi_F(t)|\,dm \leq c(b-a),$$

and hence μ_F has no point masses, $\mu_F[\{a\}]=0$ for all a. This follows by letting $b=a+1/n$ in 6.28:

$$\mu_F[(a,a+1/n)] + \frac{1}{2}\left(\mu_F[\{a\}]+\mu_F[\{a+1/n\}]\right) \leq c/(2\pi n) \to 0.$$

Defining the distribution function $F(x) \equiv \mu_F\left[(-\infty,x]\right]$ associated with finite μ_F, we now show that $F'(x)$ exists and is continuous. For $h \neq 0$,

$$\frac{F(x+h)-F(x)}{h} = (2\pi)^{-1}\int_{-\infty}^{\infty}\frac{e^{-ixt}-e^{-i(x+h)t}}{ith}\phi_F(t)dm,$$

since μ_F has no point masses. Because $\left|\frac{e^{-ixt}-e^{-i(x+h)t}}{ith}\right| \leq 1$ by 6.17, we apply Lebesgue's dominated convergence theorem with any sequence $h_n \to 0$ and conclude that $F(x)$ is differentiable:

$$\begin{aligned} F'(x) &\equiv \lim_{h\to 0}\frac{F(x+h)-F(x)}{h} \\ &= (2\pi)^{-1}\int_{-\infty}^{\infty}\lim_{h\to 0}\left[\frac{e^{-ixt}-e^{-i(x+h)t}}{ith}\right]\phi_F(t)dm \\ &= (2\pi)^{-1}\int_{-\infty}^{\infty}\phi_F(t)e^{-ixt}dm. \end{aligned}$$

Defining $f(x)=F'(x)$ obtains 6.30.

Then as in the proof of the continuity of $\phi_F(t)$:

$$\begin{aligned} |f(x+h)-f(x)| &\leq \int_{-\infty}^{\infty}\left|e^{-i(x+h)t}-e^{-ixt}\right||\phi_F(t)|\,dm \\ &\leq 2\int_{-\infty}^{\infty}|\phi_F(t)|\,dm. \end{aligned}$$

Since $e^{-i(x+h)t} - e^{-ixt} \to 0$ as $h \to 0$ for all x, t, Lebesgue's dominated convergence theorem assures that $|f(x+h) - f(x)| \to 0$ as $h \to 0$ and $F'(x)$ is thus continuous.

With $F'(x)$ continuous, the fundamental theorem of calculus yields:

$$(\mathcal{R})\int_a^b F'(x)dx = F(b) - F(a),$$

and because μ_F is a finite measure and $F(x) \equiv \mu_F[(-\infty, x]]$ this formula holds for all finite and infinite intervals. Since this integral must be non-negative for all intervals and the integrand is continuous, this assures that $F'(x) \geq 0$ for all x. By propositions 2.31 and 2.64 of book 3, if $F'(x)$ is bounded and Riemann integrable on $[a, b]$, or absolutely Riemann integrable on $\mathbb{R}$, it is Lebesgue integrable on these intervals with the same values. Hence for all such intervals:

$$\begin{aligned}\int_a^b F'(x)dm &= F(b) - F(a) \\ &\equiv \mu_F[(a, b]].\end{aligned}$$

In summary, with $f(x)$ defined by 6.30,

$$\mu_F[(a, b]] = \int_a^b f(x)dm,$$

for all right semi-closed intervals.

Defining $\widetilde{\mu}_F(A)$ by 6.29 for all $A \in \mathcal{B}(\mathbb{R})$, it follows that

$$\widetilde{\mu}_F(A) = \mu_F(A)$$

for all $A \in \mathcal{A}$, the algebra of finite disjoint unions of right semi-closed intervals. Since $\mathcal{A}$ generates $\mathcal{B}(\mathbb{R})$ and $\mathbb{R}$ is σ-finite, it follows by the uniqueness theorem of proposition 6.14 of book 1 that $\widetilde{\mu}_F(A) = \mu_F(A)$ for all $A \in \mathcal{B}(\mathbb{R})$.

As noted in remark 6.28, since μ_F is given by the density function $f(x)$, it follows that $\phi_F(t) = \widehat{f}(t)$. ■

Notation 6.51 *When $\phi_F(t)$ is integrable, the formula in 6.30 is called the **inverse Fourier transform of** $\phi_F(t)$. Because $\phi_F(t) = \widehat{f}(t)$ in this case, $f(x)$ is the **inverse Fourier transform of** $\widehat{f}(t)$ since this integral reproduces the function f on which the Fourier transform was defined.*

Remark 6.52 *When both $f(x)$ and $\widehat{f}(t)$ and integrable, there is a great deal of symmetry between the formula for the Fourier transform, and that for the inverse Fourier transform given in 6.18 and 6.30, respectively:*

$$\widehat{f}(t) = \int_{-\infty}^{\infty} f(x)e^{itx}dm, \qquad f(x) = (2\pi)^{-1}\int_{-\infty}^{\infty} \phi_F(t)e^{-ixt}dm.$$

The symmetry observed motivates a variety of notational conventions in general Fourier analysis.

While it is always the case that the exponential switches from positive to negative in these formulas, in some texts the Fourier transform is defined with the negative exponential, and then the inverse transform defined with the positive. Also, because there is an $(2\pi)^{-1}$ factor in one formula and not the other, it is common in general Fourier analysis to see the factor of $(2\pi)^{-1/2}$ in both formulas, adding to the apparent symmetry. It is also possible to eliminate this coefficient entirely by defining the Fourier transform with a $\pm 2\pi itx$ in the exponential, and then the inverse transform will have $\mp 2\pi itx$ in the exponential and with no coefficient in either case.

The approach taken in this text is consistent with the ultimate application we have in mind, and that is to use Fourier methods in the context of probability theory. In this case, the Fourier transform is called the ***characteristic function*** *and the notational convention for this latter function is nearly universally consistent with that used here, whereby the transform reflects the positive exponent itx and with no numerical coefficient.*

But in general, it is always a good idea to verify the notational conventions for Fourier analysis used by a given reference.

6.3.5 Continuity Theorem for Fourier Transforms

The last topic investigated in this section is the "continuity property" of the Fourier-Stieltjes transform. Namely, if $\{\mu_{F_n}\}$ is a sequence of finite measures for which $\mu_{F_n} \to \mu_F$ in some sense, must $\phi_{F_n}(t) \to \phi_F(t)$, and if so, in what sense? Conversely, if $\phi_{F_n}(t) \to \phi(t)$, what if anything does this tell us about the convergence of the associated measure sequence, $\{\mu_{F_n}\}$. In particular, must $\mu_{F_n} \to \mu_F$ for some measure μ_F, and if so, must the associated $\phi_F(t) = \phi(t)$?

Introduced in book 2 and studied further in book 4, the notion of **weak convergence of probability measures** is a natural one to investigate for this purpose. Specifically, we investigate the implications of $\mu_{F_n} \Rightarrow \mu_F$, meaning $\{\mu_{F_n}\}$ converges weakly to μ_F, for the convergence $\phi_{F_n}(t) \to \phi_F(t)$.

We also investigate the reverse implication. Recall that by definition 8.2 of book 2:

Definition 6.53 (Weak convergence) 1. *A sequence of distribution functions on* $\mathbb{R}$, $\{F_n(x)\}$, *will be said to* ***converge weakly*** *to a distribution function* $F(x)$, *denoted* $F_n \Rightarrow F$, *if* $F_n(x) \to F(x)$ *for every continuity point of* $F(x)$.

2. *A sequence of probability measures on* $\mathbb{R}$, $\{\mu_n\}$, *will be said to* ***converge weakly*** *to a probability measure* μ, *denoted* $\mu_n \Rightarrow \mu$, *if* $\mu_n((-\infty, x])$ *converges to* $\mu((-\infty, x])$ *for all* x *for which* $\mu\{x\} = 0$.

By exercise 8.4 of that book these notions are equivalent in the sense that if F_n is defined by the measure μ_n, meaning that $F_n(x) \equiv \mu_n((-\infty, x])$, and similarly that F is defined by μ, then $F_n \Rightarrow F$ if and only if $\mu_n \Rightarrow \mu$.

That weak convergence of measures has implications for convergence of certain associated integrals was already seen in proposition 3.66 of book 4 in the context of convergence of moments. The more general results on convergence of integrals will be seen in book 6 in the so-called **portmanteau theorem,** which was proved in 1940 by **Aleksandr Aleksandrov** (1912 – 1999) and in some references identified with his name.

The following continuity theorem provides a very strong result on the above question but is not self-contained, requiring one result from Aleksandrov's theorem from the book 6 chapter, General Results on Weak Convergence of Measures.

Proposition 6.54 (Fourier Transform Continuity theorem) *Let* $\{\mu_{F_n}, \mu_F\}$ *be a collection of probability measures and* $\{\phi_{F_n}, \phi_F\}$ *the associated Fourier transforms. Then* $\mu_{F_n} \Rightarrow \mu_F$ *if and only if* $\phi_{F_n}(t) \to \phi_F(t)$ *for all* t.

Proof. *Assume that* $\mu_{F_n} \Rightarrow \mu_F$. *Then by the portmaneau theorem of book 6,* $\int g(x)d\mu_n \to \int g(x)d\mu$ *for every bounded, continuous real-valued function,* $g(x)$. *Since* $\cos tx$ *and* $\sin tx$ *are bounded and continuous for any* t, *we have by 6.12,*

$$\begin{aligned}
\phi_{F_n}(t) &= \int_{-\infty}^{\infty} e^{itx} d\mu_{F_n} \\
&= \int_{-\infty}^{\infty} \cos tx d\mu_{F_n} + i\int_{-\infty}^{\infty} \sin tx d\mu_{F_n} \\
&\to \int_{-\infty}^{\infty} \cos tx d\mu_F + i\int_{-\infty}^{\infty} \sin tx d\mu_F \\
&= \phi_F(t).
\end{aligned}$$

Thus $\phi_{F_n}(t) \rightarrow \phi_F(t)$ *for all* t.

Conversely, assume that $\phi_{F_n}(t) \rightarrow \phi_F(t)$ *for all* t. *We first show that* $\{\mu_{F_n}\}$ *is a tight sequence of measures. Recalling definition 8.16 of book 2, this means that for any* $\epsilon > 0$ *there is a finite interval* $(a, b]$ *so that* $\mu_{F_n}((a, b]) > 1-\epsilon$ *for all* n. *Equivalently, there is an* a *so that* $\mu_{F_n}(\{|x| \geq a\}) \leq \epsilon$ *for all* n. *To this end, first note that since* $|x| \geq 2/|u|$ *implies that* $1 \leq 2(1 - 1/|ux|)$:

$$\begin{aligned}
\mu_{F_n}(\{|x| \geq 2/|u|\}) &\leq 2\int_{|x|\geq 2/|u|} (1 - 1/|ux|)\, d\mu_{F_n} \\
&\leq 2\int_{|x|\geq 2/|u|} \left(1 - \frac{\sin ux}{ux}\right) d\mu_{F_n} \\
&\leq 2\int_{-\infty}^{\infty} \left(1 - \frac{\sin ux}{ux}\right) d\mu_{F_n}.
\end{aligned}$$

Note that the second step follows since if $|xu| \geq c > 0$, *the inequality* $1 - 1/|ux| \leq 1 - \frac{\sin ux}{ux}$ *is equivalent to* $\sin ux \leq 1$. *Now:*

$$2\left(1 - \frac{\sin ux}{ux}\right) = \frac{1}{u}\int_{-u}^{u}(1 - e^{itx})dt,$$

and an application of Fubini's theorem is justified since μ_{F_n} *is a probability measure and thus* $1 - e^{itx}$ *is* $m \times \mu_{F_n}$*-integrable on* $[-u, u] \times \mathbb{R}$. *It then follows that*

$$\begin{aligned}
\mu_{F_n}(\{|x| \geq 2/|u|\}) &\leq \frac{1}{u}\int_{-u}^{u}\int_{-\infty}^{\infty}(1 - e^{itx})d\mu_{F_n}dt \\
&\leq \frac{1}{u}\int_{-u}^{u}\left(1 - \phi_{F_n}(t)\right)dt.
\end{aligned}$$

Since $\phi_F(t)$ *is continuous by proposition 6.31 and* $\phi_F(0) = 1$, *it follows that*

$$\left|\frac{1}{u}\int_{-u}^{u}(1 - \phi_F(t))\, dt\right| \leq 2\sup\nolimits_{[-u,u]}|1 - \phi_F(t)|,$$

and this upper bound converges to 0 *as* $u \rightarrow 0$. *So given* $\epsilon > 0$ *there is a* u *so that*

$$\frac{1}{u}\int_{-u}^{u}(1 - \phi_F(t))\, dt < \epsilon.$$

But $\phi_{F_n}(t) \rightarrow \phi_F(t)$ *for all* t, *and the bounded convergence theorem obtains:*

$$\frac{1}{u}\int_{-u}^{u}\left(1 - \phi_{F_n}(t)\right)dt \rightarrow \frac{1}{u}\int_{-u}^{u}(1 - \phi_F(t))\, dt$$

for any u. Thus for the above given $\epsilon > 0$ and u there is an N so that for $n \geq N$,

$$\frac{1}{u}\int_{-u}^{u}\left(1-\phi_{F_n}(t)\right)dt < 2\epsilon.$$

Since all $\phi_{F_n}(t)$ are continuous and $\phi_{F_n}(0) = 1$, by reducing u this inequality will also remain valid for all $n < N$. Hence for this u and all n :

$$\mu_{F_n}\left(\{|x| \geq 2/|u|\}\right) < 2\epsilon.$$

Letting $a \equiv \frac{2}{|u|}$ this proves that $\{\mu_{F_n}\}$ is tight.

Now recall corollary 8.22 of book 2. If it can be shown that any subsequence of $\{\mu_{F_n}\}$ that converges weakly in fact converges weakly to μ_F, then we can conclude that $\mu_{F_n} \Rightarrow \mu_F$. Let $\{\mu_{F_{n_k}}\}$ be such a subsequence, and assume that $\mu_{F_{n_k}} \Rightarrow \upsilon$. Then by the first part of this theorem, $\phi_{F_{n_k}}(t) \to \phi(t)$ where $\phi(t)$ is the Fourier transform of υ. But by assumption, $\phi_{F_n}(t) \to \phi_F(t)$ and hence $\phi(t) = \phi_F(t)$, and then by the Fourier inversion formula, $\upsilon = \mu_F$.
■

The next corollary almost certainly provides the most important application of the continuity theorem.

Corollary 6.55 *Let $\{\mu_{F_n}\}$ be a collection of probability measures and $\{\phi_{F_n}\}$ the associated Fourier transforms. Then if $\phi_{F_n}(t) \to \phi_0(t)$ for all t where $\phi_0(t)$ is continuous at $t = 0$, then there exists a probability measure μ so that $\mu_{F_n} \Rightarrow \mu$ and $\phi_0(t)$ is the Fourier transform of μ.*

Proof. *Following the proof of the above proposition, the continuity of $\phi(t)$ at $t = 0$ is enough to prove that $\{\mu_{F_n}\}$ is tight, since $\phi(0) = 1$ follows from $\phi_{F_n}(0) = 1$ for all n. By Helly's selection theorem of proposition 8.14 and proposition 8.20, both of book 2, there exists a subsequence $\{\mu_{F_{n_k}}\}$ and a probability measure μ with $\mu_{F_{n_k}} \Rightarrow \mu$. The above theorem yields that $\phi_{F_{n_k}}(t) \to \phi(t)$ where $\phi(t)$ is the Fourier transform of μ. But by assumption, $\phi_{F_n}(t) \to \phi_0(t)$ and hence $\phi(t) = \phi_0(t)$. By Fourier inversion we conclude that all subsequences which converge weakly must then converge to the same μ, and so $\mu_{F_n} \Rightarrow \mu$ by corollary 8.22 of book 2, and $\phi_0(t)$ is the Fourier transform of μ.* ■

Remark 6.56 *Jumping ahead to book 6, the* ***characteristic function of a distribution function*** *F will be defined as the Fourier-Stieltjes transform of the associated finite measure μ_F. As will be seen, the continuity* *results above will provide a powerful tool for investigations into weak convergence of distribution functions, $F_n \Rightarrow F$, or equivalently, the convergence in distribution of the underlying random variable sequence, $X_n \Rightarrow X$.*

*And indeed this will be a much more powerful tool for such results than the **moment generating function** developed in book 4 because:*

1. *Most importantly, characteristic functions exist for every probability measure and hence the above results are potentially applicable to any distribution function sequence. By propositions 3.22, 3.24 and 3.55 of book 4, moment generating functions only exist for distribution functions with infinitely many moments $\{\mu'_n\}$, and for which $\sum_{n=0}^{\infty} \mu'_n t^n/n!$ converges absolutely on $(-t_0, t_0)$ for some $t_0 > 0$.*

2. *Even given existence, the section 3.2.8 of book 4 **method of moments** results for moment generating functions assure similar but somewhat weaker conclusions:*

   (**a**) *(Corollary 3.72, book 4) If $F_n \Rightarrow F$ where the associated $M_n(t)$ and $M(t)$ exist on a common interval $(-t_0, t_0)$ with $t_0 > 0$, and if $\{M_n(t)\}$ is bounded on this interval for each t, then $M_n(t) \rightarrow M(t)$ all $t \in (-t_0, t_0)$.*

   (**b**) *(Corollary 3.74, book 4) Given $\{F_n\}$ and F with associated $M_n(t)$ and $M(t)$ which exist on a common interval $(-t_0, t_0)$ with $t_0 > 0$. If $M_n(t) \rightarrow M(t)$ for all $t \in (-t_0, t_0)$, then $F_n \Rightarrow F$.*

Example 6.57 *In proposition 5.5 of book 4 was presented a moment generating function proof of the **Poisson limit theorem**, that the binomial probability measure, properly calibrated, converged to the Poisson probability measure. Recall that this binomial measure has parameters $n \in \mathbb{N}$, and p with $0 < p < 1$, and is defined on $[0, 1, ..., n]$ by:*

$$\mu_{B_n}(j) = \binom{n}{j} p^j (1-p)^{n-j}, \qquad j = 0, 1, .., n, .$$

*The **binomial coefficient** $\binom{n}{j}$ is defined for $0 \leq j \leq n$ by $\binom{n}{j} = n!/[j!(n-j)!]$. The Poisson distribution, named for **Siméon-Denis Poisson** (1781 – 1840), is characterized by a single parameter $\lambda > 0$, and is defined on the nonnegative integers by:*

$$\mu_P(j) = e^{-\lambda} \frac{\lambda^j}{j!}, \qquad j = 0, 1, 2, ...$$

The Poisson limit theorem then states that for $\lambda = np$ *fixed, that for all* j :

$$\binom{n}{j} p^j (1-p)^{n-j} \longrightarrow e^{-\lambda}\frac{\lambda^j}{j!} \text{ as } n \to \infty.$$

This implies that $F_{B_n}(x) \to F_P(x)$ *for every* ***continuity point*** *of* F_P, *notationally for all* $x \in CP(F^P)$, *where:*

$$CP(F^P) = (-\infty, 0) \cup \bigcup\nolimits_{n=0}^{\infty} (n, n+1),$$

since for both F_{B_n} *and* F_P :

$$F(x) = \sum\nolimits_{j \leq x} F(j).$$

Thus by definition $\mu_{B_n} \Rightarrow \mu_P$.

We now prove this result again using the above Fourier transform tools. In book 6 these tools will applied in situations where moment generating functions do not exist, and thus will produce much greater conclusions.

Proposition 6.58 (Poisson limit theorem) *For* $\lambda = np$ *fixed, let* μ_{B_n} *denote the binomial measure defined on* $j = 0, 1, .., n$, *and* μ_P *the Poisson measure. Then as* $n \to \infty$,

$$\mu_{B_n} \Rightarrow \mu_P. \tag{6.31}$$

Proof. *By the Fourier transform continuity theorem, 6.31 follows by proving that for all* t, $\phi_{F_n}(t) \to \phi_{F_P}(t)$, *where to simplify notation* $F_n \equiv F_{B_n}$. *By definition,*

$$\phi_{F_n}(t) = \int_{-\infty}^{\infty} e^{itx} d\mu_{B_n}.$$

With μ_{B_n} *defined above and supported on integers* j *with* $0 \leq j \leq n$, *it follows that with* $p = \lambda/n$,

$$\phi_{F_n}(t) = \sum\nolimits_{j=0}^{n} \binom{n}{j} (\lambda/n)^j (1-\lambda/n)^{n-j} e^{itj}.$$

With $\phi_{F_1}(t) = (1-\lambda/n) + (\lambda/n) e^{it}$, *a calculation show that:*

$$\phi_{F_{n+1}}(t) = \phi_{F_n}(t)\phi_{F_1}(t),$$

and hence by iteration,

$$\begin{aligned} \phi_{F_n}(t) &= \left[(1-\lambda/n) + (\lambda/n) e^{it}\right]^n \\ &= \left[1 + \lambda\left(e^{it}-1\right)/n\right]^n. \end{aligned}$$

As $n \to \infty$,

$$\left(1+\frac{y}{n}\right)^n \to e^y \tag{6.32}$$

for all y, *so it follows that for all* t,

$$\phi_{F_n}(t) \to \exp\left[\lambda\left(e^{it}-1\right)\right], \text{ as } n \to \infty.$$

The final step is to evaluate $\phi_{F_P}(t)$ *for which we recall 6.11 as well as corollary 2.48:*

$$\begin{aligned}
\phi_{F_P}(t) &= \int_{-\infty}^{\infty} e^{itx} d\mu^P \\
&= \sum\nolimits_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!}e^{itj} \\
&= e^{-\lambda}\sum\nolimits_{j=0}^{\infty} \frac{\left(\lambda e^{it}\right)^j}{j!} \\
&= \exp\left[\lambda\left(e^{it}-1\right)\right].
\end{aligned}$$

Hence since $\phi_{F_n}(t) \to \phi_{F_P}(t)$ *for all* t, *6.31 follows.* ■

Chapter 7

General Measure Relationships

In this final chapter we investigate ways in which one measure can be decomposed relative to another measure. In the basic set-up we are given a measure space $(X, \sigma(X), \mu)$, as well as a second measure on this space υ, assumed to be defined on the same sigma algebra $\sigma(X)$. The question is, can υ be decomposed relative to μ in the sense that:

$$\upsilon = \upsilon_1 + \upsilon_2,$$

where in some sense υ_1 is closely related to μ, while in another sense υ_2 is completely unrelated to μ. Once the notions of "closely related" and "completely unrelated" have been formalized, the **Lebesgue Decomposition theorem** will address this. Named for **Henri Lebesgue** (1875 – 1941), it will state that if μ and υ are σ-finite measures, then this decomposition exists and is unique.

In addition it will turn out that υ_1, the measure closely related to μ, can in fact be defined in terms of the μ-integral of a μ-measurable function f. That is, for all $A \in \sigma(X)$,

$$\upsilon_1(A) = \int_A f d\mu.$$

This last result is called the **Radon-Nikodým theorem,** named for **Johann Radon** (1887 – 1956) who proved this result when $X = \mathbb{R}^n$, and **Otto Nikodým** (1887 – 1974) who generalized Radon's result to σ-finite measure spaces.

As we have already developed much of this theory in the special case of Borel measures defined on the Lebesgue measure space, so $(X, \sigma(X), \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ and $\nu = \mu_F$, the next section summarizes these results and sets the stage for the more general development that follows.

7.1 Decomposition of Borel Measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$

Let μ be a Borel measure on $\mathbb{R}$ and define $F_\mu(x)$ as in 5.1 of book 1, or equivalently when μ is finite, simply define $F_\mu(x) \equiv \mu\left[(-\infty, x]\right]$ as in 5.3. Then by construction, μ and F_μ are related in the sense that for any right semi-closed interval $(a, b]$:

$$\mu\left[(a, b]\right] = F_\mu(b) - F_\mu(a).$$

Given an increasing, right continuous function F one can also define a set function μ as above on the semi-algebra $\mathcal{A}'$ of right semi-closed intervals $(a, b]$, and extend this set function to a measure defined on the Borel sigma algebra $\mathcal{B}(\mathbb{R})$ using the methods of chapter 5 of book 1. In brief, μ is extended from $\mathcal{A}'$ to the complete sigma algebra of sets which are **Carathéodory measurable with respect to** $\mu^*_{\mathcal{A}}$, the outer measure induced by μ. On this complete sigma algebra which is denoted $\mathcal{M}_{\mu_F}(\mathbb{R})$, as well as the Borel sigma algebra, $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}_{\mu_F}(\mathbb{R})$, μ is defined to equal $\mu^*_{\mathcal{A}}$. As $\mathcal{B}(\mathbb{R})$ is the smallest sigma algebra that contains $\mathcal{A}'$, the extension to $\mathcal{B}(\mathbb{R})$ is unique by proposition 6.14 of book 1, meaning that if there is a measure μ' with $\mu' = \mu$ on $\mathcal{A}'$, then also $\mu' = \mu$ on $\mathcal{B}(\mathbb{R})$.

The function $F_\mu(x)$ defined above from given μ is an increasing function, $F_\mu(x) \leq F_\mu(x')$ if $x < x'$, since μ is nonnegative and additive, and hence F_μ is a Lebesgue measurable function. Further by proposition 3.15 of book 3, $F_\mu(x)$ is differentiable m-a.e. and by that book's proposition 3.16, $F'_\mu(x)$ is Lebesgue measurable. Hence by proposition 2.29 of book 3, $F'_\mu(x)$ is Lebesgue integrable on bounded measurable sets so define $F_1(x)$ for $x \geq 0$ by:

$$F_1(x) = (\mathcal{L})\int_0^x F'_\mu(y)dm.$$

Analogously, for $x < 0$ define $F_1(x)$ in terms of $-\int_x^0$. By monotonicity, $F'_\mu(x) \geq 0$ m-a.e., so $F_1(x)$ is increasing and by proposition 3.19 of book 3 it follows that for every interval $[a, b]$:

$$F_1(b) - F_1(a) = (\mathcal{L})\int_a^b F'_\mu(y)dm \leq F_\mu(b) - F_\mu(a).$$

Next define $F_2(x) = F_\mu(x) - F_1(x)$. Then $F_2(x)$ is measurable, and increasing since if $x < x'$:

$$\begin{aligned} F_2(x') - F_2(x) &\equiv \left[F_\mu(x') - F_1(x')\right] - \left[F_\mu(x) - F_1(x)\right] \\ &= \left[F_\mu(x') - F_\mu(x)\right] - \left[F_1(x') - F_1(x)\right] \\ &\geq 0. \end{aligned}$$

Hence $F_2(x)$ is again differentiable m-a.e. By proposition 3.37 of book 3, $F_1'(x) = F_\mu'(x)$ m-a.e. and thus:

$$\begin{aligned} F_2'(x) &= F_\mu'(x) - F_1'(x) \\ &= 0, \qquad m\text{-a.e.} \end{aligned}$$

The induced function F_μ can thus be decomposed as a sum of increasing functions,

$$F_\mu(x) = F_2(x) + F_1(x).$$

Now define set functions on right semi-closed intervals $(a,b] \in \mathcal{A}'$:

$$\begin{aligned} \upsilon_{F_1}\left[(a,b]\right] &\equiv F_1(b) - F_1(a), \\ \upsilon_{F_2}\left[(a,b]\right] &\equiv F_2(b) - F_2(a), \end{aligned}$$

and let υ_1 and υ_2 be defined as the extensions of υ_{F_1} and υ_{F_2} to $\mathcal{B}(\mathbb{R})$ as in chapter 5 of book 1 and summarized above. Then by definition, $\mu = \upsilon_{F_1} + \upsilon_{F_2}$ on $\mathcal{A}'$, as well as the algebra $\mathcal{A}$ of finite disjoint unions of such intervals.

If $A \subset \mathbb{R}$, then by the definition of outer measure in definition 5.16 of book 1 (see also remark 5.17):

$$\begin{aligned} \mu^*(A) &= \inf\left\{\sum\nolimits_n \mu(A_n) \mid A \subset \bigcup A_n,\ A_n \in \mathcal{A}' \text{ disjoint}\right\} \\ &= \inf\left\{\sum\nolimits_n \left[\upsilon_{F_1}(A_n) + \upsilon_{F_2}(A_n)\right] \mid A \subset \bigcup A_n,\ A_n \in \mathcal{A}' \text{ disjoint}\right\} \\ &\geq \inf\left\{\sum\nolimits_n \upsilon_{F_1}(A_n) \mid A \subset \bigcup A_n\right\} + \inf\left\{\sum\nolimits_n \upsilon_{F_2}(A_n') \mid A \subset \bigcup A_n',\right\} \\ &= \upsilon_{F_1}^*(A) + \upsilon_{F_2}^*(A). \end{aligned}$$

On the other hand if $A \subset \bigcup A_n$ and $A \subset \bigcup A_n'$, then $A \subset (\bigcup A_n) \cap (\bigcup A_n')$ and this induces another disjoint countable union of right semi-closed intervals with $A \subset \bigcup A_n''$ where $A_n'' = A_{n_1} \cap A_{n_2}'$. Thus as υ_{F_1} and υ_{F_2} are monotonic:

$$\begin{aligned}
\upsilon_{F_1}^*(A)+\upsilon_{F_2}^*(A) &= \inf\left\{\sum\nolimits_n \upsilon_{F_1}(A_n) \mid A\subset\cup A_n\right\}+\inf\left\{\sum\nolimits_n \upsilon_{F_2}(A_n') \mid A\subset\cup A_n',\right\} \\
&\geq \inf\left\{\sum\nolimits_n \upsilon_{F_1}(A_n'') \mid A\subset\cup A_n''\right\}+\inf\left\{\sum\nolimits_n \upsilon_{F_2}(A_n'') \mid A\subset\cup A_n''\right\} \\
&= \inf\left\{\sum\nolimits_n \left[\upsilon_{F_1}(A_n)+\upsilon_{F_2}(A_n)\right] \mid A\subset\cup A_n'',\ A_n''\in\mathcal{A}' \text{ disjoint}\right\} \\
&= \mu^*(A).
\end{aligned}$$

Therefore,

$$\mu^*=\upsilon_{F_1}^*+\upsilon_{F_2}^*$$

and since $\mu=\mu^*$ on $\mathcal{B}(\mathbb{R})$ and similarly for υ_1 and υ_2, this analysis produces the **Lebesgue decomposition** of a Borel measure μ on $(\mathbb{R},\mathcal{B}(\mathbb{R}),m)$:

$$\mu=\upsilon_1+\upsilon_2. \tag{7.1}$$

The first measure υ_1 is very closely related to m in the following sense. By definition, $\upsilon_1\left[(a,b]\right]=(\mathcal{L})\int_a^b F_\mu'(y)dm$ and it is an exercise to show that by extension:

$$\upsilon_1[A]=(\mathcal{L})\int_A F_\mu'(y)dm$$

for all $A\in\mathcal{B}(\mathbb{R})$. This implies that if $m(A)=0$ then $\upsilon_1[A]=0$, and so $\upsilon_1[A]>0$ only on sets for which $m(A)>0$. In this sense, υ_1 puts all of its non-zero measure on sets with non-zero Lebesgue measure.

The second measure υ_2 is completely unrelated to m in the following sense. Define D as the set on which F_2' exists, and on which as noted above $F_2'(x)=0$. Again by proposition 3.15 of book 3, the set on which F_2' does not exist satisfies $m(\widetilde{D})=0$. But while D and $\widetilde{D}$ are Lebesgue measurable, they need not be Borel sets and this is required for the next step. So let $\widetilde{E}$ be a Borel set with $\widetilde{D}\subset\widetilde{E}$ and $m(\widetilde{E}-\widetilde{D})=0$ as given in proposition 5.28 of book 1 with the Borel measure there, μ_F, replaced by m. Specifically, $\widetilde{E}\in\mathcal{A}_{\sigma\delta}$, the collection of countable intersections of sets in $\mathcal{A}_\sigma$, which in turn is the collection of countable unions of sets in the algebra $\mathcal{A}$. By finite additivity $m(\widetilde{E})=0$, and defining E as the complement of $\widetilde{E}$, it follows again by finite additivity that for any $A\in\mathcal{B}(\mathbb{R})$,

$$\mu[A]=\mu[A\cap E]+\mu\left[A\cap\widetilde{E}\right].$$

Note that this step required E, $\widetilde{E}\in\mathcal{B}(\mathbb{R})$ and would not have been valid with Lebesgue measurable D and $\widetilde{D}$.

Now $m(\widetilde{E}) = 0$ implies $m\left[A \cap \widetilde{E}\right] = 0$, and this in turn assures that

$$\upsilon_1\left[A \cap \widetilde{E}\right] = 0$$

as noted above. Thus:

$$\mu\left[A \cap \widetilde{E}\right] = \upsilon_2\left[A \cap \widetilde{E}\right]. \tag{(*)}$$

In addition, for any Borel set A:

$$\upsilon_2\left[A \cap E\right] = 0$$

by proposition 5.30 of book 1 since $A \cap E \subset D$, and the above analysis shows that $D = \{F_2'(x) = 0\}$. Thus υ_2 puts all of its non-zero measure on subsets of $\widetilde{E}$, which in turn have Lebesgue measure 0. Thus:

$$\mu\left[A \cap E\right] = \upsilon_1\left[A \cap E\right]. \tag{(*)}$$

Putting the pieces together, it can be concluded that for any $A \in \mathcal{B}(\mathbb{R})$,

$$\mu\left[A\right] = \upsilon_1\left[A \cap E\right] + \upsilon_2\left[A \cap \widetilde{E}\right].$$

That is, the μ-measure of any Borel set A naturally splits into two components, the μ-measures of $A \cap E$ and $A \cap \widetilde{E}$. On the first set, $\mu = \upsilon_1$, and on the second set, $\mu = \upsilon_2$.

Summary 7.1 *In terms of general results that are forthcoming, the above discussion has demonstrated that for any Borel measure μ defined on the Lebesgue measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$:*

1. ***Lebesgue decomposition theorem:*** *There exist Borel measures υ_1 and υ_2, with:*

   (a) *$\mu = \upsilon_1 + \upsilon_2$, where:*

   - **i.** *υ_1 is **absolutely continuous with respect to** m, denoted $\upsilon_1 \ll m$, which means that $\upsilon_1(A) = 0$ for all $A \in \mathcal{B}(\mathbb{R})$ with $m(A) = 0$.*
   - **ii.** *υ_2 and m are **mutually singular,** denoted $\upsilon_2 \perp m$, which by definition means that there is a set $E \in \mathcal{B}(\mathbb{R})$, for which $m(\widetilde{E}) = \upsilon_2(E) = 0$.*

(**b**) *For all $A \in \mathcal{B}(\mathbb{R})$, and the set E identified in 1.a.ii:*

$$\begin{aligned} \upsilon_1[A] &= \upsilon_1[A \cap E], \\ \upsilon_2[A] &= \upsilon_2\left[A \cap \widetilde{E}\right], \end{aligned}$$

and hence

$$\mu[A] = \upsilon_1[A \cap E] + \upsilon_2\left[A \cap \widetilde{E}\right].$$

2. ***Radon-Nikodým theorem:*** *Given a Borel measure υ which is **absolutely continuous with respect to** m, there is a nonnegative m-measurable function f so that*

$$\upsilon[A] = \int_A f dm.$$

Remark 7.2 *Note that while it may appear that this result has only been proved for such υ_1 constructed in part 1 for which $f(x) \equiv F'_\mu(x)$, the more general statement is obtained as follows. The part 1 decomposition can be applied again but to such υ to derive that*

$$\upsilon = \omega_1 + \omega_2,$$

where ω_1 is absolutely continuous with respect to υ, ω_2 and υ are mutually singular, and $\omega_1[A]$ is given by the above integral representation with $f(x) \equiv F'_\upsilon(x)$. Further by 1.a.ii there is a set $E \in \mathcal{B}(\mathbb{R})$ for which $m(\widetilde{E}) = \omega_2(E) = 0$. But the absolute continuity of υ with respect to m assures that $\omega_2(\widetilde{E}) = 0$, and this along with $\omega_2(E) = 0$ proves that $\omega_2 \equiv 0$ and thus $\upsilon \equiv \omega_1$.

3. ***Lebesgue decomposition refinement:*** *In the special case of Borel measures, we can in fact refine this decomposition by splitting the **singular component** υ_2 into*

$$\upsilon_2 = \upsilon_2^D + \upsilon_2^C.$$

Here υ_2^D is a discrete, pure point Borel measure which allocates nonzero measure to an at most countable set of points, and υ_2^C is a Borel measure induced by a continuous function. In essence, this is accomplished by splitting F_2, the monotonic function which induces υ_2, into the sum of a "singular" function and a "saltus" function as in chapter 1 of book 4. See section 1.1, A Characterization of Distribution Functions on $\mathbb{R}$, for details on this construction.

In the next section we will generalize the results in 1 and 2 to apply to a σ-finite measure ν on a σ-finite measure space $(X, \sigma(X), \mu)$.

7.2 Decomposition of σ-Finite Measures

In this section we seek to generalize the **Lebesgue decomposition theorem** of the above section to the decomposition of a σ-finite measure ν defined on a σ-finite measure space $(X, \sigma(X), \mu)$, and also generalize the **Radon-Nikodým theorem** in the same setting but where such ν is also absolutely continuous with respect to μ. We begin by formalizing the definitions introduced above, and then pursue a necessary digression into the notion of a **signed measure** which was briefly introduced in the above section, Integration by Parts for Lebesgue-Stieltjes Integrals.

Definition 7.3 *Given a measure space $(X, \sigma(X), \mu)$ and measures ν_1 and ν_2 defined on $\sigma(X)$:*

1. *ν_1 is **absolutely continuous with respect to** μ, denoted*

$$\nu_1 \ll \mu,$$

if $\nu_1(A) = 0$ for all $A \in \sigma(X)$ with $\mu(A) = 0$.

2. *ν_2 and μ are **mutually singular,** denoted*

$$\nu_2 \perp \mu,$$

if there is a measurable set $E_\mu \in \sigma(X)$ for which

$$\mu(\widetilde{E}_\mu) = \nu_2(E_\mu) = 0.$$

Remark 7.4 *1. The notion of **absolute continuity** was introduced in chapter 3 of book 3 as a property of a function. In the section below on the Radon-Nikodým theorem, this earlier definition and that related to measures above will be reconciled in proposition 7.18.*

*2. The notion of a **singular function** was introduced in chapter 1 of book 4. In the section below on the Lebesgue Decomposition Theorem, this earlier definition and that related to measures above will be reconciled in proposition 7.28.*

*3. Given a measure μ defined on $(X, \sigma(X), \mu)$, a measurable set E_μ is called **a support of** μ if $\mu(\widetilde{E}_\mu) = 0$. In this case, it is also said that μ is **concentrated on** E_μ. In 2 of definition 7.3, E_μ is a support of μ*

and $E_{\nu_2} \equiv \widetilde{E}_\mu$ is a support of ν_2. Consequently, in this definition, the defining relationship can be stated as:

$$\nu_2 \perp \mu \iff \nu_2 \text{ and } \mu \text{ have disjoint supports.}$$

The support of a measure is not unique. If $E_\mu \subset E'_\mu$ with E'_μ measurable, then E'_μ is also a support of μ since $\widetilde{E}'_\mu \subset \widetilde{E}_\mu$. The significance of a support set E_μ is that for any $A \in \sigma(X)$,

$$\mu(A) = \mu(A \cap E_\mu),$$

because $\mu(A \cap \widetilde{E}_\mu) = 0$.

7.2.1 A Digression into Signed Measures

In the development of this section we will have the need to contemplate properties of the difference of two measures, $\nu \equiv \mu_1 - \mu_2$ say, defined on the sigma algebra $\sigma(X)$. To avoid ever having the expression $\infty - \infty$, it will always be assumed that at least one of these measures is finite. In general, the set function ν will not be nonnegative, and hence will not be a measure on $\sigma(X)$ unless $\mu_1(A) > \mu_2(A)$ for all $A \in \sigma(X)$. But this set function has the important property of countable additivity, as well as $\nu(\emptyset) = 0$, and so it is called a **signed measure.** More formally:

Definition 7.5 *A set function ν defined on a sigma algebra $\sigma(X)$ is a **signed measure** if the range of ν includes at most one of $\pm\infty$, and if $\{A_i\} \subset \sigma(X)$ is a finite or countable collection of disjoint sets, then*

$$\nu\left[\bigcup\nolimits_i A_i\right] = \sum\nolimits_i \nu\left[A_i\right].$$

Here it is assumed that the summation is absolutely convergent if $\nu\left[\bigcup_i A_i\right]$ is finite, and divergent to one of $\pm\infty$ otherwise.

Remark 7.6 *While initially defined as a more general notion than that of a difference of two measures of exercise 6.6, it will be seen below that every signed measure can in fact be expressed as such a difference of measures. See the **Hahn decomposition** and **Jordan decomposition** theorems below.*

The essential question we need to answer for the development of the next section is as follows. Given a signed measure ν, can X be decomposed into measurable sets:

$$X = A^+ \cup A^-,$$

so that:

1. $A^+ \cap A^- = \emptyset$,

2. $\nu(A) \geq 0$ for all measurable $A \subset A^+$, and,

3. $\nu(A) \leq 0$ for all measurable $A \subset A^-$.

Such a decomposition is not unique in general, since if A^0 is a measurable set for which $\nu(A) = 0$ for all $A \subset A^0$ then A^0 can be part of either set, or measurably split and divided between the A^+ and A^- sets.

Momentarily ignoring the requirement that $X = A^+ \cup A^-$, we introduce some terminology.

Definition 7.7 *Given $\sigma(X)$ and a signed measure ν, a measurable set A^+ is called a* ***positive set for*** *ν if*

$$\nu(A) \geq 0 \text{ for all measurable } A \subset A^+.$$

Similarly, a measurable set A^- is called a ***negative set for*** *ν if*

$$\nu(A) \leq 0 \text{ for all measurable } A \subset A^-,$$

and a measurable set A^0 is called a ***null set for*** *ν if*

$$\nu(A) = 0 \text{ for all measurable } A \subset A^0.$$

Example 7.8 *In the case where $\nu \equiv \mu_1 - \mu_2$, a difference of measures as in exercise 6.6, this splitting can be understood as:*

1. A^+ : the set on which $\mu_1(A) \geq \mu_2(A)$ for all measurable $A \subset A^+$,

2. A^- : the set on which $\mu_1(A) \leq \mu_2(A)$ for all measurable $A \subset A^-$,

3. A^0 : the set on which $\mu_1(A) = \mu_2(A)$ for all measurable $A \subset A^0$.

It is apparent by definition that a subset of a positive set is positive, and the intersection of a collection of positive sets is a positive set. Less obvious and assigned as an exercise below is the proof that the union of a collection of positive sets is a positive set. Analogously, the same statements are true of negative sets and null sets. Unfortunately, the terminology for these sets can be misleading in that it is natural to assume that if A^+ is a positive set, then $\widetilde{A^+}$ is a negative set, and conversely for negative sets. But this need not be true, and is also assigned as an exercise.

Example 7.9 *Let f be a Lebesgue measurable function and $f = f^+ - f^-$ its decomposition into positive and negative parts as in definition 2.36. Assume that at least one of these components is Lebesgue integrable, and define*

$$\nu(A) = \int_A f dm.$$

Then ν is countably additive by corollary 2.49 if both component functions are Lebesgue integrable over the unioned set $A \equiv \cup A_i$. In the general case, if one of the component functions is not integrable, $\int_{\cup A_i} f dm = \infty$ say, then it follows that $\int_{\cup A_i} f^+ dm = \infty$ and thus $\sum_i \int_{A_i} f^+ dm = \infty$ by Lebesgue's monotone convergence theorem. This then implies$\sum_i \int_{A_i} f dm = \infty$ since $\int_{\cup A_i} f^- dm < \infty$ by assumption that at least one component function is integrable.

In this case it is not difficult to identify a positive set and negative set for ν. For example, $A^+ \equiv \{x|f \geq 0\}$ is a positive set and $A^- \equiv \{x|f < 0\}$ a negative set. But any measurable subsets of these are also valid candidates, as are definitions which allocate $A^0 = \{x|f = 0\}$ differently, in whole or measurably in part.

If we seek to also have $A^+ \cup A^- = \mathbb{R}$, then the definitions posed are best possible subject to the ambiguity on the placement of A^0.

Exercise 7.10 *If ν is a signed measure defined on a sigma algebra $\sigma(X)$, prove that if $\{A_i^+\}$ is a countable collection of positive sets then $\cup A_i^+$ is a positive set. (Hint: If $A \subset \cup A_i^+$, consider $A_n \equiv A \cap A_n^+ \cap \widetilde{A}_{n-1}^+ \cap ... \cap \widetilde{A}_1^+$. Note that this result is also true for negative and null sets by the same proof.*

Show by examples that if A^+ is a positive set, then $\widetilde{A^+}$ may be a negative set, but need not be a negative set.

The following result, that one can always determine a positive set and negative set for a signed measure, is known as the **Hahn decomposition theorem** and named for **Hans Hahn** (1879 – 1934). This decomposition theorem assures that except for the ambiguity related to the location of null sets denoted A^0 above, that this positive set and negative set are "maximal" in that $A^+ \cup A^- = X$ and $A^+ \cap A^- = \emptyset$. For its proof we require a result that at first seems apparent, that every measurable set of positive measure contains a positive subset of positive measure. And yet its demonstration is subtle because to be deemed a positive set requires a verification of a property of all measurable subsets.

The result needed is presented next.

Proposition 7.11 *Given a signed measure υ on a sigma algebra $\sigma(X)$, if $A \in \sigma(X)$ satisfies $0 < \nu(A) < \infty$, then there exists a positive set $A^+ \subset A$ with $\nu(A^+) > 0$.*

Proof. *Given any such A, our goal is to define A^+ as the set that is left over after removing all measurable subsets of A with negative measure. If there is no such subset of negative measure then A is a positive set and we are done. Otherwise, let k_1 be the smallest positive integer so that there is a set $A_1 \subset A$ with $\nu(A_1) < -1/k_1$. Inductively, once $\{A_i\}_{i=1}^{n-1}$ are chosen let k_n be the smallest positive integer so that there is a measurable set $A_n \subset A - \cup_{i=1}^{n-1} A_i$ with $\nu(A_n) < -1/k_n$. By construction, $k_{n+1} \geq k_n$ for all n. Finally, define $A^+ = A - \cup_{i=1}^{\infty} A_i$, noting that this union could be finite if there are no sets of negative measure after the Nth step.*

Since $A = A^+ \cup (\cup_{i=1}^{\infty} A_i)$ and this union is disjoint, it follows by definition that

$$\nu(A) = \nu(A^+) + \sum\nolimits_{i=1}^{\infty} \nu(A_i).$$

If this summation contains a finite number of terms then by assumption $0 < \nu(A)$ and by construction $\nu(A_n) < 0$ for all n assures that $\nu(A^+) > 0$ and we are done. If this is an infinite sum, because $\nu(A) < \infty$ this series converges absolutely by definition and hence $k_n \to \infty$. Also, $0 < \nu(A)$ and absolute convergence assure that $\nu(A^+) > 0$ as in the finite sum case.

The final step is to show that A^+ is a positive set, and this is done by showing that given $\epsilon > 0$, A^+ can contain no measurable set with υ-measure less than $-\epsilon$. To this end, given ϵ there is an n so that $1/(k_n - 1) < \epsilon$ since $k_n \to \infty$. Now $A^+ \subset A - \cup_{i=1}^{n} A_i$ and by construction, $A - \cup_{i=1}^{n} A_i$ and hence A^+ can contain no set with measure less that $-1/(k_n - 1)$. This follows because at step n we defined k_n as the smallest integer with $\nu(A_n) < -1/k_n$. But $-\epsilon < -1/(k_n - 1)$ and hence A^+can contain no set with measure less that $-\epsilon$. Since ϵ was arbitrary, it follows that A^+ contains no measurable set with negative measure, and thus A^+ is a positive set. ■

As is seen next, this technical result is the key ingredient to prove the main result of this section. For its statement, recall the definition of a **symmetric set difference**:

$$A \bigtriangleup B \equiv (A - B) \cup (B - A). \tag{7.2}$$

Note that by definition: $A \bigtriangleup B = B \bigtriangleup A$.

Proposition 7.12 (Hahn Decomposition theorem) *Let ν be a signed measure on a sigma algebra $\sigma(X)$. Then there is a positive set and a negative set for ν, respectively A^+ and A^-, for which $A^+ \cup A^- = X$ and $A^+ \cap A^- = \emptyset$.*

Further this decomposition is unique up to null sets. Specifically, if $B^+ \cup B^- = X$ *is another such decomposition, then* $A^+ \triangle B^+ = A^- \triangle B^-$ *is a null set.*

Proof. *For specificity, assume that* $-\infty \leq \nu(A) < \infty$ *for all* $A \in \sigma(X)$. *If* $-\infty < \nu(A) \leq \infty$, *this proved result applies to the signed measure* $-\nu$, *and we can then simply switch positive and negative sets to obtain the desired result for* ν.

Let $P = \{A \in \sigma(X) | A \text{ is positive}\}$, *noting that* P *is not empty since* $\emptyset \in P$, *and define* $\rho = \sup\{\nu(A)|A \in P\}$. *Then* $\rho \geq 0$ *since* $\nu(\emptyset) = 0$, *and let* $\{A_i\}_{i=1}^{\infty} \subset P$ *be chosen so that* $\rho = \lim_{i\to\infty} \nu(A_i)$. *This collection can be assumed to be nested and increasing since given arbitrary* $\{A_i'\}_{i=1}^{\infty} \subset P$ *define* $A_i = \cup_{j=1}^{i} A_j'$. *Now define* $A^+ = \cup_{i=1}^{\infty} A_i$, *and it is now left to show that this set satisfies the requirements of the proposition.*

Firstly, A^+ *is indeed a positive set since it is the union of positive sets and we can apply exercise 7.10 above. Since* $A^+ \in P$ *it follows by definition of* ρ *that* $\nu(A^+) \leq \rho$. *But for any* i, $A^+ - A_i \subset A^+$ *is a positive set and*

$$\nu(A^+) = \nu(A_i) + \nu(A^+ - A_i) \geq \nu(A_i).$$

This implies $\nu(A^+) \geq \rho$ *and thus* $\nu(A^+) = \rho$.

Now define $A^- = \widetilde{A^+}$, *and we show* A^- *is a negative set by contradiction. Assume that there is a set* $B \subset A^-$ *with* $\nu(B) > 0$. *Then by the above proposition there is a positive set* $B' \subset B$ *with* $\nu(B') > 0$, *and* B' *is disjoint from* A^+ *by construction. But then* $A^+ \cup B'$ *is a positive set, and by disjointedness* $\nu(A^+ \cup B') = \nu(A^+) + \nu(B') > \rho$, *contradicting the definition of* ρ. *Hence no such* B *can exist and* A^- *is therefore a negative set.*

For uniqueness, $\widetilde{B^+} = B^-$ *and so* $A^+ - B^+ = A^+ \cap B^-$ *is a positive set and a negative set and thus a null set. Similarly for* $B^+ - A^+$ *and then* $A^+ \triangle B^+$. ■

Remark 7.13 *Although we will not use it later, it is worth noting an important result known as the* ***Jordan decomposition theorem,*** *named for* ***Camille Jordan*** *(1838 – 1922). This result can be derived directly using the tools underlying the proof of the Hahn decomposition theorem, but now that this result is proved, we can obtain Jordan's result as an interesting corollary.*

Given a signed measure ν *and the measurable sets* A^+ *and* A^- *defined in the Hahn decomposition theorem, define for* $A \in \sigma(X)$:

$$\mu^+(A) = \nu(A \cap A^+), \qquad \mu^-(A) = -\nu(A \cap A^-).$$

Then μ^+ and $\mu^-$are measures, and:

$$\nu = \mu^+ - \mu^-.$$

Further, since $A^+ \cap A^- = \emptyset$, these measures have disjoint supports and so $\mu^+$and μ^- are mutually singular.

Of course, the decomposition of a signed measure ν into a difference of measures is far from unique, since for any measure λ it also follows that

$$\nu = \left(\mu^+ + \lambda\right) - \left(\mu^- + \lambda\right).$$

*However, the Jordan decomposition of ν as a difference of **mutually singular** measures is unique.*

Proposition 7.14 (Jordan Decomposition theorem) *Let ν be a signed measure on a sigma algebra $\sigma(X)$. Then there are mutually singular measures μ^+ and μ^- defined on $\sigma(X)$ with $\nu = \mu^+ - \mu^-$. Further, this decomposition is unique. If μ_0^+ and μ_0^- are mutually singular measures with $\nu = \mu_0^+ - \mu_0^-$, then $\mu^+ = \mu_0^+$ and $\mu^- = \mu_0^-$.*

Proof. *Existence is demonstrated above using the Hahn decomposition theorem..*

For uniqueness, let A^+ and A^- be the supports for μ^+ and μ^-, and A_0^+ and A_0^- the supports of μ_0^+ and μ_0^-. In other words, $A^+ \cup A^- = X$, $A^+ \cap A^- = \emptyset$, and $\mu^+(A^-) = \mu^-(A^+) = 0$, and similarly for A_0^+ and A_0^-. Then for any A, since $\mu^+(A) = \mu^+(A \cap A^+)$ and similarly for the other measures:

$$\begin{aligned}\nu(A) &= \mu^+(A \cap A^+) - \mu^-(A \cap A^-) \\ &= \mu_0^+(A \cap A_0^+) - \mu_0^-(A \cap A_0^-).\end{aligned}$$

Letting $A = A^+ \cap A_0^-$ obtains that $\mu^+(A^+ \cap A_0^-) = -\mu_0^-(A^+ \cap A_0^-)$ and thus since measures are nonnegative:

$$\mu^+(A^+ \cap A_0^-) = -\mu_0^-(A^+ \cap A_0^-) = 0.$$

Similarly, letting $A = A^- \cap A_0^+$ obtains:

$$-\mu^-(A^- \cap A_0^+) = \mu_0^+(A^- \cap A_0^+) = 0.$$

This then implies that both μ^+ and μ_0^+ are supported on $A^+ \cap A_0^+$, and both μ^- and μ_0^- are supported on $A^- \cap A_0^-$. In other words, as above:

$$\begin{aligned}\nu(A) &= \mu^+(A \cap A^+ \cap A_0^+) - \mu^-(A \cap A^- \cap A_0^-) \\ &= \mu_0^+(A \cap A^+ \cap A_0^+) - \mu_0^-(A \cap A^- \cap A_0^-).\end{aligned}$$

Using this, if $A \subset A^+ \cap A_0^+$ *then* $\mu^+(A) = \mu_0^+(A)$ *and thus* $\mu^+ = \mu_0^+$ *since for general* A :

$$\mu^+(A) = \mu^+(A \cap A^+ \cap A_0^+), \qquad \mu_0^+(A) = \mu_0^+(A \cap A^+ \cap A_0^+).$$

Similarly, $\mu^- = \mu_0^-$. ■

7.2.2 The Radon-Nikodým theorem

As noted in the introductory section, the **Radon-Nikodým theorem** is named for **Johann Radon** (1887 – 1956) who proved this result when $X = \mathbb{R}^n$, and **Otto Nikodým** (1887 – 1974) who generalized Radon's result to σ-finite measure spaces. To set the stage, let $(X, \sigma(X), \mu)$ be a σ-finite measure space and υ a σ-finite measure also defined on $\sigma(X)$, and which is absolutely continuous with respect to μ :

$$\upsilon \ll \mu.$$

Example 7.15 *If* $f : X \to \mathbb{R}$ *is a nonnegative* μ*-measurable function, then defining*

$$\upsilon(A) = \int_A f d\mu$$

produces a σ*-finite measure. This follows because* υ *is countably additive by corollary 2.49 to Lebesgue's monotone convergence theorem and example 7.9. Also, such* υ *is absolutely continuous with respect to* μ *since if* $\mu(A) = 0$ *then* $\upsilon(A) = 0$.

The goal of the Radon-Nikodým theorem is to reverse this conclusion to state that every σ-finite measure υ that is **absolutely continuous** with respect to μ arises in exactly this way. Further, the associated function f is called the **Radon-Nikodým derivative of** υ **with respect to** μ, and is unique μ-a.e. In other words, if also $\upsilon(A) = \int_A g d\mu$, then $f = g$ except on a set of μ-measure 0.

Notation 7.16 *The* ***Radon-Nikodým derivative of*** υ ***with respect to*** μ*, is often denoted* $\left[\frac{d\upsilon}{d\mu}\right]$ *or simply* $\frac{d\upsilon}{d\mu}$*, reflecting both by terminology and notation the notion of a derivative from calculus. This also allows the notationally compelling re-interpretation of 3.6, that a function* $g(x)$ *is* υ*-integrable if and only if* $g(x)\frac{d\upsilon}{d\mu}$ *is* μ*-integrable, and when integrable:*

$$\int_A g(x) d\upsilon = \int_A g(x) \frac{d\upsilon}{d\mu} d\mu. \tag{7.3}$$

for any $A \in \sigma(X)$.

Remark 7.17 *It is natural to wonder in what way the function $f \equiv \frac{d\upsilon}{d\mu}$ can be understood as a derivative. Assume that $(X, \sigma(X), \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$, the Borel measure space with Lebesgue measure, and let υ_F be a Borel measure induced by an increasing, right continuous function F. Equivalently, let υ be a Borel measure, then $\upsilon = \upsilon_F$ for such F given in 5.1 or 5.3 of book 1. We consider two perspectives.*

1. *From proposition 3.15 of book 3, such F is differentiable m-a.e., and in the special case where F is absolutely continuous by definition 3.54 of book 3, we have by that book's proposition 3.61 that F can be recovered by the Lebesgue integral of F'. In particular it follows that on any interval $[a, b]$ on which F is absolutely continuous,*

$$F(x) = F(a) + \int_a^x F'(y)dm.$$

Hence since υ_F is defined on the semi-algebra $\mathcal{A}'$ by $\upsilon_F[(a, b]] = F(b) - F(a)$, it follows that

$$\upsilon_F[(a, b]] = \int_a^b F'(y)dm,$$

and this generalizes to any $A \in \mathcal{B}(\mathbb{R})$:

$$\upsilon_F[A] = \int_A F'(y)dm.$$

This generalization follows from proposition 6.14 of book 1. However, this representation is not unique in that if $f = F'$ m-a.e., then f works equally well in this representation of υ_F.

In summary, in the special case where υ_F is a Borel measure induced by an increasing, absolutely continuous function F, the function f identified in the Radon-Nikodým theorem satisfies $f = F'$ m-a.e. In other words, in this special case:

$$\frac{d\upsilon_F}{dm} = \frac{dF}{dx}, \ m\text{-a.e},$$

where $\frac{d\upsilon_F}{dm}$ is the Radon-Nikodým derivative and equals m-a.e. the integrand $\frac{dF}{dx}$ that defines υ_F.

2. *Another insight to the notation $\frac{dv}{dm}$ as a calculus derivative is a corollary to the above remark. If $F'(x)$ is continuous on (a,b), then for any $x \in (a,b)$ and $\epsilon_n, \epsilon'_n \to 0$,*

$$\frac{\upsilon_F\left[(x-\epsilon_n, x+\epsilon'_n]\right]}{m\left[(x-\epsilon_n, x+\epsilon'_n]\right]} \to F'(x).$$

The limit on the left is reasonably interpreted as the derivative $\frac{d\upsilon_F}{dm}$, since for example for finite measures:

$$\upsilon_F\left[(x-\epsilon_n, x+\epsilon'_n]\right] = \upsilon_F\left[(-\infty, x+\epsilon'_n]\right] - \upsilon_F\left[(-\infty, x-\epsilon'_n]\right],$$

by 5.3 of book 1, while $m\left[(x-\epsilon_n, x+\epsilon'_n]\right] = \epsilon_n + \epsilon'_n$. For ν in the general case, a similar representation is possible using 5.3 of book 1 but depends on the sign of x. The limiting result is then proved as an exercise directly from remark 1, or as a corollary to proposition 5.30 of book 3 using the continuity of $F'(x)$.

The next result addresses the perhaps surprising double use of the terminology "absolutely continuous," first in the context of a function as introduced in chapter 3 of book 3, and now again in the context of a measure. The final conclusion below is that if F is an increasing right continuous function, then F is absolutely continuous by definition 3.54 of book 3 if and only if $\upsilon_F \ll m$.

Recall that a function F is absolutely continuous if given $\epsilon > 0$ there is a δ so that

$$\sum\nolimits_{i=1}^{n} \left|F(x_i) - F(x'_i)\right| < \epsilon$$

for any finite collection of disjoint intervals $\{(x'_i, x_i)\}$, with

$$\sum\nolimits_{i=1}^{n} \left|x_i - x'_i\right| < \delta.$$

But note that if F is increasing and we use right semi-closed intervals in this definition, then F is absolutely continuous if given $\epsilon > 0$ there is a δ so that with υ_F denoting the induced Borel measure,

$$\sum\nolimits_{i=1}^{n} \upsilon_F\left[(x'_i, x_i]\right] < \epsilon$$

for any finite collection of disjoint intervals $\{(x'_i, x_i]\}$, with

$$\sum\nolimits_{i=1}^{n} m\left[(x'_i, x_i]\right] < \delta.$$

So indeed, absolute continuity of an increasing function reflects a property between two measures, and begins to resemble the measure context, that $\upsilon_F \ll m$ when $m(A) = 0$ implies that $\upsilon_F(A) = 0$.

Proposition 7.18 (Absolute Continuity Definitions) *Let υ_F be a Borel measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ induced by an increasing, right continuous function F. Then F is absolutely continuous by definition 3.54 of book 3 if and only if $\upsilon_F \ll m$.*

Proof. *If F is absolutely continuous then as in the above remark it follows that for any for any $A \in \mathcal{B}(\mathbb{R})$:*

$$\upsilon_F[A] = \int_A F'(y)dm,$$

and so $\upsilon_F \ll m$.

If $\upsilon_F \ll m$, the Radon-Nikodým theorem below will identify a Lebesgue measurable function f so that $\upsilon_F[A] = \int_A f(y)dm$ for all Borel sets A. Letting $A = (a, x]$ it follows that:

$$\upsilon_F[A] = F(x) - F(a) = \int_a^x f(y)dm.$$

Hence $F(x)$ is defined as the indefinite Lebesgue integral of a measurable function, and is thus absolutely continuous by proposition 3.57 of book 3. ■

The Radon-Nikodým theorem requires that $(X, \sigma(X), \mu)$ be a σ-finite measure space, and υ a σ-finite measure on $\sigma(X)$ which is absolutely continuous with respect to μ. In the next example we illustrate the need for σ-finiteness of the underlying space.

Example 7.19 (Need for Measure Space σ-Finiteness) *Let $(X, \sigma(X), \mu) = ((0,1), \mathcal{B}((0,1), \mu)$ and define μ as the counting measure where $\mu(A)$ is defined as the number of points in A if finite, and $\mu(A) = \infty$ otherwise. Let $\upsilon = m$, Lebesgue measure. Then m is finite and hence σ-finite, but μ is not a σ-finite measure since $(0,1)$ has uncountably many points and it can therefore not be represented as a countable union of sets of finite measure. However $m \ll \mu$ since $\mu(A) = 0$ implies $A = \emptyset$ and hence $m(A) = 0$.*

Now if $m(A) = \int_A f d\mu$ for some μ-measurable and thus Borel measurable function f, express $f = f^+ - f^-$ as in definition 2.36 and apply the following argument to each nonnegative function. Denoting either by f for simplicity, it follows from 2.6 that:

$$\int_A f(x)d\mu = \sup_{\varphi \leq f} \int_A \varphi(x)d\mu,$$

where each $\varphi(x) = \sum_{i=1}^{n} a_i \chi_{A_i}(x)$ *is a simple function, and* $\{A_i\}$ *a disjoint collection of* μ*-measurable sets. Because*

$$\int_A \varphi(x) d\mu = \sum_{i=1}^{n} a_i \mu(A_i),$$

for the supremum of such integrals to be finite it must be the case that $f = 0$ *except on a finite collection of points. But if* $f(a) \neq 0$ *for some* a*, then letting* $A = \{a\}$:

$$\int_A f(x) d\mu = f(a) \neq 0 = m(A),$$

in contradiction to $m \ll \mu$.

Hence, even though $m \ll \mu$*, there is no* μ*-measurable function* f *for which* $m(A) = \int_A f d\mu$.

The proof of the Radon-Nikodým theorem is constructive, and for this proof a technical result will be needed on the existence of measurable functions with specified level sets defined as $A_\alpha \equiv \{x | f(x) \leq \alpha\}$. If f is measurable on $(X, \sigma(X), \mu)$, then by definition $A_\alpha \in \sigma(X)$ for all α, $A_\alpha \subset A_\beta$ if $\alpha < \beta$, and $f > \alpha$ on $X - A_\alpha$. The next results go the other way. Given $\{A_\alpha\} \subset \sigma(X)$, when does an associated measurable f exist, and what are its properties?

The first lemma provides the existence result for such a measurable function if $A_\alpha \subset A_\beta$ when $\alpha < \beta$, and yields the conclusion that $f \geq \alpha$ on $X - A_\alpha$. The second then generalizes this to a μ-a.e. result in the case where the assumption on sets for $\alpha < \beta$ is weakened from $A_\alpha - A_\beta = \emptyset$ to $\mu(A_\alpha - A_\beta) = 0$. In other words, if $\alpha < \beta$ then $A_\alpha \subset A_\beta$ except for a subset of A_α of μ-measure 0.

Lemma 7.20 *If* $\{A_\alpha\} \subset \sigma(X)$ *is a countable collection of sets with* $A_\alpha \subset A_\beta$ *for* $\alpha < \beta$*, then there is a measurable extended real-valued function* f *on* X *with* $f \leq \alpha$ *on* A_α *and* $f \geq \alpha$ *on* $X - A_\alpha$.

Proof. *Given* $x \in X$ *define* $f(x) = \inf\{\alpha' | x \in A_{\alpha'}\}$*, defining* $\inf \emptyset \equiv \infty$. *Hence if* $x \in A_\alpha$ *then* $f(x) \equiv \inf\{\alpha' | x \in A_{\alpha'}\} \leq \alpha$*, while if* $x \in X - A_\alpha$*, then* $\{\alpha' | x \in A_{\alpha'}\} \subset (\alpha, \infty)$ *and so* $f(x) = \inf\{\alpha' | x \in A_{\alpha'}\} \geq \alpha$*. To show that* f *is measurable, note that if* $f(x) < \alpha$ *then* $x \in A_\beta$ *for some* $\beta < \alpha$*, while if* $x \in A_\beta$ *for some* $\beta < \alpha$*, then* $f(x) \leq \beta < \alpha$*. Hence* $\{x | f(x) < \alpha\} = \bigcup_{\beta < \alpha} A_\beta \in \sigma(X)$*, since this is a countable union, and* f *is measurable.* ■

Lemma 7.21 *If* $\{A_\alpha\} \subset \sigma(X)$ *is a countable collection of sets with* $\mu(A_\alpha - A_\beta) = 0$ *for* $\alpha < \beta$*, then there is a measurable extended real-valued function* f *on* X *with* $f \leq \alpha$ μ*-a.e. on* A_α*, and* $f \geq \alpha$ μ*-a.e. on* $X - A_\alpha$.

Proof. *Define $C \equiv \bigcup_{\alpha<\beta} (A_\alpha - A_\beta)$, a countable union. As a countable union of sets of measure zero, $\mu(C) = 0$. Define $A'_\alpha = A_\alpha \cup C$. Then $A'_\alpha \subset A'_\beta$ if $\alpha < \beta$ since:*

$$\begin{aligned} A'_\alpha - A'_\beta &= (A_\alpha \cup C) \cap \left(\widetilde{A}_\beta \cap \widetilde{C}\right) \\ &= \left(A_\alpha \cap \widetilde{A}_\beta \cap \widetilde{C}\right) \cup \left(C \cap \widetilde{A}_\beta \cap \widetilde{C}\right) \\ &= (A_\alpha - A_\beta) - C. \end{aligned}$$

This last set is empty since α, β are among the countable collection of pairs that defined C.

Hence there is a measurable function f on X with $f \leq \alpha$ on A'_α and $f \geq \alpha$ on $X - A'_\alpha$. Thus except possibly on C, a set of measure 0, $f \leq \alpha$ on A_α and $f \geq \alpha$ on $X - A_\alpha$. ■

With these technical results in hand, we are now ready for the statement and proof of the Radon-Nikodým theorem. It should be noted that the ideas underlying this theorem and the Lebesgue decomposition theorem which follows are intimately related, as are the tools used in their proofs. In fact it is largely a matter of taste whether one proves Radon-Nikodým first and derives the Lebesgue decomposition as a corollary, or uses the opposite logic. In fact, some texts combine the results as the Lebesgue-Radon-Nikodým theorem.

Proposition 7.22 (Radon-Nikodým theorem) *Let $(X, \sigma(X), \mu)$ be a σ-finite measure space, and υ a σ-finite measure on $\sigma(X)$ which is absolutely continuous with respect to μ, so $\upsilon \ll \mu$. Then there exists a nonnegative measurable function $f : X \to \mathbb{R}$, also denoted $f \equiv \frac{\partial \upsilon}{\partial \mu}$, so that for all $A \in \sigma(X)$:*

$$\upsilon(A) = \int_A f d\mu. \tag{7.4}$$

Further, f is unique μ-a.e., meaning if g is a measurable function so that 7.4 is true with g, then $g = f$, μ-a.e.

Proof.

1. ***Reduction to Finite Measures:*** *Both μ and υ are σ-finite measures, and thus there is a disjoint collect $\{A_n\} \subset \sigma(X)$ so that $\bigcup_n A_n = X$, and both $\mu(A_n) < \infty$ and $\upsilon(A_n) < \infty$ for all n. This follows because such a collection exists for each measure separately, so define the A_n-sets in terms of intersections of these sets. For $A \in \sigma(X)$,*

define finite measures $\mu_n(A) \equiv \mu(A \cap A_n)$ *and* $\upsilon_n(A) \equiv \nu(A \cap A_n)$, *and note that then* $\upsilon_n \ll \mu_n$. *Now assume that the existence statement of the theorem has been proven for finite measures. Then for each* n *there is a nonnegative* $\sigma(X)$*-measurable function,* $f_n : X \to \mathbb{R}$, *so that* $\upsilon_n(A) = \int_A f_n d\mu_n$ *for* $A \in \sigma(X)$. *By definition,* $\int_A f_n d\mu_n = \int_A f_n \chi_{A_n} d\mu$, *and so by countable additivity and disjointedness of* $\{A_n\}$:

$$\begin{aligned} \upsilon(A) &= \sum \upsilon_n(A) \\ &= \sum \int_A f_n \chi_{A_n} d\mu \\ &= \int_A \sum f_n \chi_{A_n} d\mu. \end{aligned}$$

Hence the existence statement of the theorem is true in the σ*-finite case with* $f = \sum f_n \chi_{A_n}$.

2. ***Existence Proof for Finite Measures:***

   (**a**) ***Construction of*** f : *For rational nonnegative* r_k, *let* $\{A^+_{r_k}, A^-_{r_k}\}$ *be the Hahn decomposition of the signed measure* $\upsilon - r_k\mu$, *so that* $A^+_{r_k} \cup A^-_{r_k} = X$ *and* $A^+_{r_k} \cap A^-_{r_k} = \emptyset$. *Define* $A^+_0 = X$ *and* $A^-_0 = \emptyset$. *For any* r_j, r_k, *note that* $A^-_{r_k} - A^-_{r_j} = A^-_{r_k} \cap A^+_{r_j}$ *and hence this is a positive set for* $\upsilon - r_j\mu$ *and a negative set for* $\upsilon - r_k\mu$. *But if* $r_k < r_j$ *then for any* A :

   $$(\upsilon - r_j\mu)[A] = (\upsilon - r_k\mu)[A] - (r_j - r_k)\,\mu[A] \leq (\upsilon - r_k\mu)[A],$$

   and so if A *is a positive set for* $\upsilon - r_j\mu$ *and a negative set for* $\upsilon - r_k\mu$ *then* $\mu[A] = 0$. *Letting* $A = A^-_{r_k} - A^-_{r_j}$, *it follows that if* $r_k < r_j$ *then* $\mu[A^-_{r_k} - A^-_{r_j}] = 0$. *By lemma 7.21, let* f *be the measurable function on* X *so that* $f \leq r_k$ μ*-a.e. on* $A^-_{r_k}$, *and* $f \geq r_k$ μ*-a.e. on* $X - A^-_{r_k} = A^+_{r_k}$. *Since* $A^-_0 = \emptyset$, *it follows that* $f \geq 0$ μ*-a.e. on* X.

   b. ***μ-Integral of*** f : *Choose integer* N *and the subset of rationals with* $r_k = k/N$, $k \geq 0$. *Let* $A \in \sigma(X)$ *be given and define:*

   $$A_k = A \cap \left(A^-_{r_{k+1}} - A^-_{r_k}\right), \qquad A_\infty = A - \bigcup\nolimits_{k=0}^{\infty} A^-_{r_k}.$$

   Then $A = A_\infty \cup \bigcup_{k=0}^\infty A_k$ *is a disjoint union and so* $\mu[A] = \mu[A_\infty] + \sum_{k=0}^\infty \mu[A_k]$. *Now since*

   $$A_k \subset A^-_{r_{k+1}} - A^-_{r_k} = A^-_{r_{k+1}} \cap A^+_{r_k},$$

it follows that $k/N \leq f \leq (k+1)/N$ *on* A_k *and so*

$$\frac{k}{N}\mu[A_k] \leq \int_{A_k} f d\mu \leq \frac{k+1}{N}\mu[A_k].$$

Also, A_k *is by definition a positive set for* $\upsilon - r_j\mu$ *and a negative set for* $\upsilon - r_k\mu$ *and thus* $(\upsilon - r_k\mu)[A_k] \geq 0$ *and* $(\upsilon - r_{k+1}\mu)[A_k] \leq 0$. *This obtains that* $r_k\mu[A_k] \leq \upsilon[A_k] \leq r_{k+1}\mu[A_k]$, *and so for all* $k < \infty$:

$$\upsilon[A_k] - \frac{1}{N}\mu[A_k] \leq \int_{A_k} f d\mu \leq \upsilon[A_k] + \frac{1}{N}\mu[A_k].$$

Now by construction $f = \infty$ μ*-a.e. on* A_∞. *If* $\mu[A_\infty] > 0$ *then* $\upsilon[A_\infty] = \infty$ *since* $(\upsilon - r_k\mu)[A_\infty] \geq 0$ *for all* k, *contradicting the assumption that* υ *is a finite measure. So it must be the case that* $\mu[A_\infty] = 0$ *and hence* $\upsilon[A_\infty] = 0$ *since* $\upsilon \ll \mu$, *and thus:*

$$\upsilon[A_\infty] = \int_{A_\infty} f d\mu.$$

By the disjointedness of the union, $A = A_\infty \cup \bigcup_{k=0}^{\infty} A_k$, *these estimates can be added together by corollary 2.49 to obtain:*

$$\upsilon[A] - \frac{1}{N}\mu[A] \leq \int_A f d\mu \leq \upsilon[A] + \frac{1}{N}\mu[A].$$

This is now true for every subset of rationals of the form $r_k = k/N$, $k \geq 0$, *and hence* $\int_A f d\mu = \upsilon[A]$ *for all* $A \in \sigma(X)$.

3. ***Uniqueness of*** f : *Let* g *be another measurable function which satisfies 7.4. Then with* $h = f - g$,

$$\int_A h d\mu = 0$$

for all $A \in \sigma(X)$. *For any rational* $a, b > 0$, *let* $A_{a,b} = \{x | a \leq h \leq b\}$, *then*

$$a\mu[A_{a,b}] \leq \int_A h d\mu \leq b\mu[A_{a,b}],$$

and hence $\mu[A_{a,b}] = 0$. *A similar conclusion is found for any rational* $a, b < 0$. *So* $\{x | h \neq 0\}$ *is a countable union of sets of* μ*-measure zero.*

■

Remark 7.23 *It is worth a moment to better understand this result and proof. By definition,*

$$\begin{aligned} \upsilon(A) &= \int_A \chi_A d\upsilon \\ &= \sup_{\varphi \leq \chi_A} \int_A \varphi(x) d\upsilon \\ &= \sup_{\varphi \leq \chi_A} \sum_{i=1}^{n} \nu\left[A_i\right]. \end{aligned}$$

Here $\varphi(x) = \sum_{i=1}^{n} a_i \chi_{A_i}(x)$, *but since* $\varphi \leq \chi_A$ *there is no loss of generality by assuming* $\cup A_i = A$ *is a disjoint union, and all* $a_i = 1$.

Similarly, approximate nonnegative measurable f *with simple functions as in proposition 1.18, where for any* n *let* $N \equiv n2^n+1$ *and define measurable sets* $\{A_j^{(n)}\}_{j=1}^{N}$, *and a simple function* $\varphi_n(x)$, *by:*

$$A_j^{(n)} = \begin{cases} \{x \in A | (j-1)2^{-n} \leq f(x) < j2^{-n}\}, & 1 \leq j \leq N-1, \\ \{x \in A | n \leq f(x)\}, & j = N, \end{cases}$$

$$\varphi_n(x) = \left\{ a_j^{(n)} \equiv (j-1)2^{-n}, \ x \in A_j^{(n)}, \ 1 \leq j \leq N. \right.$$

Then since $\varphi_n(x) \to f(x)$ *if follows by Lebesgue's monotone convergence theorem:*

$$\int_A f d\mu = \sup_n \sum\nolimits_{i=1}^{N} a_j^{(n)} \mu\left[A_j^{(n)}\right].$$

So the essence of linking $\upsilon(A)$ *and* $\int_A f d\mu$ *is in finding sets for which* $\nu\left[A_i\right]$ *and* $a_j^{(n)} \mu\left[A_j^{(n)}\right]$ *can be linked, and defining* f *accordingly in terms of* $\{a_j^{(n)}\}$.

To this end, the above proof identified a countable collection of "almost nested" sets, $\{A_{r_k}^{-}\}$ *for rational* r_k, *so that if* $r_k < r_j$ *then* $\mu[A_{r_k}^{-} - A_{r_j}^{-}] = 0$. *A measurable function* f *was then defined on* X *so that* $f \leq r_k$ μ*-a.e. on* $A_{r_k}^{-}$, *and* $f \geq r_k$ μ*-a.e. on* $X - A_{r_k}^{-} = A_{r_k}^{+}$. *With this construction, it followed that for any* N *and* $r_k = k/N$, *if* $A_k \subset A_{r_{k+1}}^{-} - A_{r_k}^{-}$, *then*

$$r_k \mu[A_k] \leq \upsilon[A_k] \leq r_{k+1} \mu[A_k].$$

In addition, on any such A_k *we have by construction that* $r_k \leq f \leq r_{k+1}$ *and so*

$$r_k \mu[A_k] \leq \int_{A_k} f d\mu \leq r_{k+1} \mu[A_k].$$

These results linked the μ and υ measures for any set $A_k \subset A^-_{r_{k+1}} - A^-_{r_k}$. Since any $A \in \sigma(X)$ can be expressed as a disjoint union: $A = A_\infty \cup \bigcup_{k=0}^\infty A_k$, where $A_k \subset A^-_{r_{k+1}} - A^-_{r_k}$ if $k < \infty$ and $A_\infty \subset X_\infty \equiv X - \bigcup_{k=0}^\infty A^-_{r_k}$, this analysis then led to estimates for A. But as it cannot be assumed that $X = \cup_k \left(A^-_{r_{k+1}} - A^-_{r_k} \right)$, this construction leaves an "open question" regarding the υ-measure of the set $X_\infty = X - \bigcup_{k=0}^\infty A^-_{r_k}$.

While it was relatively straightforward to show that $\mu[A_\infty] = 0$ for any A, the value of $\nu[A_\infty]$ is in general not constrained by this construction. But in the special case of this theorem where $\upsilon \ll \mu$, this ensured that $\nu[A_\infty] = 0$. So then for every component of the disjoint union for A, the υ measures could be estimated based on the μ integral of f.

In summary, by studying the signed measures $\upsilon - r_k\mu$ and applying Hahn's decomposition theorem, we identified sets $\{A^-_{r_{k+1}} - A^-_{r_k}\}$ for which υ-measures could be bounded by multiples of μ-measures, and absolute continuity was the property needed to take care of the loose end of the ν-measure of X_∞. The function f was then defined to have these sets as level sets, and since the μ-integral of measurable f can then always be bounded by the μ-measures of the supporting sets, we could transform the integral $\int_A \chi_A d\upsilon$ into $\int_A f d\mu$.

Corollary 7.24 (Radon-Nikodým theorem) *Let $(X, \sigma(X), \mu)$ be a σ-finite measure space, and υ a σ-finite measure on $\sigma(X)$ which is absolutely continuous with respect to μ, so $\upsilon \ll \mu$. Then there exists a nonnegative measurable function $f : X \to \mathbb{R}$ so that for all $A \in \sigma(X)$, and all measurable functions $g : X \to \mathbb{R}$:*

$$\int_A g d\nu = \int_A g f d\mu, \tag{7.5}$$

although both integrals may be infinite. However, g is ν-integrable if and only if gf is μ-integrable.

Proof. *Since the measure ν is given by 7.4, this result is an immediate application of proposition 3.6 as noted in 7.3.* ■

Corollary 7.25 (Radon-Nikodým theorem) *Let $(X, \sigma(X), \mu)$ be a σ-finite measure space, and υ a σ-finite measure on $\sigma(X)$ for which both $\upsilon \ll \mu$ and $\mu \ll \nu$. Then:*

$$\frac{\partial \upsilon}{\partial \mu} = \left[\frac{\partial \mu}{\partial \upsilon} \right]^{-1}, \text{ a.e.} \tag{7.6}$$

Remark 7.26 *Note that the "a.e." in 7.6 is unambiguous, since by assumption $\mu(A) = 0$ if and only if $\nu(A) = 0$. Thus a.e. means both μ-a.e. and υ-a.e.*

Proof. *By the above proposition there exists nonnegative measurable functions* $f_\upsilon \equiv \frac{\partial \upsilon}{\partial \mu}$ *and* $f_\mu \equiv \frac{\partial \mu}{\partial \upsilon}$ *defined on* $X \to \mathbb{R}$ *so that for all* $A \in \sigma(X)$:

$$\upsilon(A) = \int_A f_\nu d\mu, \qquad \mu(A) = \int_A f_\mu d\upsilon.$$

Letting $g = f_\mu$ *in 7.5 obtains that for all* $A \in \sigma(X)$*:*

$$\mu(A) = \int_A f_\mu d\upsilon = \int_A f_\mu f_\nu d\mu.$$

Since both functions are nonnegative, $f_\mu f_\nu = 1$ *a.e.* ■

7.2.3 The Lebesgue Decomposition Theorem

The Lebesgue decomposition theorem is named for **Henri Lebesgue** (1875 – 1941), and can be proved based on a small adaptation of the constructive proof of the Radon-Nikodým theorem.

Proposition 7.27 (Lebesgue Decomposition Theorem) *Let* $(X, \sigma(X), \mu)$ *be a* σ*-finite measure space, and* υ *a* σ*-finite measure defined on* $\sigma(X)$*. Then there are measures* υ_{ac} *and* υ_s *defined on* $\sigma(X)$ *with* $\upsilon_{ac} \ll \mu$ *and* $\upsilon_s \perp \mu$*, so that for all* $A \in \sigma(X)$:

$$\upsilon(A) = \upsilon_{ac}(A) + \upsilon_s(A). \tag{7.7}$$

Further, this decomposition is unique in that if $\upsilon(A) = \upsilon'_{ac}(A) + \upsilon'_s(A)$ *with* $\upsilon'_{ac} \ll \mu$ *and* $\upsilon'_s \perp \mu$*, then* $\upsilon'_{ac} = \upsilon_{ac}$ *and* $\upsilon'_s = \upsilon_s$*.*
Proof.

1. ***Existence:*** *Recalling the construction in the proof of the Radon-Nikodým theorem, for any* N *let* $r_k = k/N$, $k \geq 0$, *and define:*

   $$E_{ac} = \bigcup\nolimits_{k=0}^{\infty} A^-_{r_k}, \qquad E_s = X - \bigcup\nolimits_{k=0}^{\infty} A^-_{r_k}.$$

   For any $A \in \sigma(X)$, *let:*

   $$\upsilon_{ac}(A) = \upsilon(A \cap E_{ac}), \qquad \upsilon_s(A) = \upsilon(A \cap E_s).$$

   Then 7.7 is satisfied since $X = E_{ac} \cup E_s$ *is a disjoint union.*

   To prove that $\upsilon_{ac} \ll \mu$*, the above construction obtains that for any* N :

   $$\upsilon[A \cap E_{ac}] - \frac{1}{N}\mu[A \cap E_{ac}] \leq \int_{A \cap E_{ac}} f d\mu \leq \upsilon[A \cap E_{ac}] + \frac{1}{N}\mu[A \cap E_{ac}].$$

As N is arbitrary it follows that for all $A \in \sigma(X)$:

$$\upsilon_{ac}(A) = \int_{A \cap E_{ac}} f d\mu,$$

and consequently $\upsilon_{ac}(A) = 0$ if $\mu(A) = 0$. For the proof that $\upsilon_s \perp \mu$, the above construction proved that $\mu(E_s) = 0$, and thus $\upsilon_{ac}(E_s) = 0$ by absolute continuity. But then since $E_{ac} \cap E_s = \emptyset$, it follows by definition above that $\upsilon_s(\tilde{E}_s) = \upsilon_s(E_{ac}) = 0$.

2. ***Uniqueness:*** *If also $\upsilon(A) = \upsilon'_{ac}(A) + \upsilon'_s(A)$, then for all $A \in \sigma(X)$ we have $\upsilon_{ac}(A) + \upsilon_s(A) = \upsilon'_s(A) + \upsilon'_{ac}(A)$, which we would like to express as:*

$$\upsilon_{ac}(A) - \upsilon'_{ac}(A) = \upsilon'_s(A) - \upsilon_s(A). \qquad ((*))$$

To avoid A-sets that produce $\infty - \infty$, we need to reduce this proof to the case of finite measures. As in the proof of the Radon-Nikodým theorem, the σ-finiteness of μ and ν allows the identification of disjoint $\{X_i\}$ with $\cup X_i = X$, and where each X_i has finite μ- and ν-measure. For either measure denoted λ, define the finite measure λ_i on $\sigma(X)$ by

$$\lambda_i(A) = \lambda(A \cap X_i).$$

Then with this notation we have proved from part 1 that for all i, $\upsilon_i(A) = \upsilon_{ac_i}(A) + \upsilon_{s_i}(A)$ with $\upsilon_{ac_i} \ll \mu_i$ and $\upsilon_{s_i} \perp \mu_i$. Now if also $\upsilon(A) = \upsilon'_{ac}(A) + \upsilon'_s(A)$, then this is also true with subscripts i, and once again $\upsilon'_{ac_i} \ll \mu_i$ and $\upsilon'_{s_i} \perp \mu_i$. Thus if we can prove that $\upsilon'_{ac_i} = \upsilon_{ac_i}$ and $\upsilon'_{s_i} = \upsilon_{s_i}$ the proof will be complete. Said differently, we can now drop the notationally burdensome i-subscripts and complete the proof assuming all measures are finite, and thus the subtractions in $(*)$ are well defined.

Now $(\upsilon_{ac} - \upsilon'_{ac}) \ll \mu$ by definition. Since $\upsilon_s \perp \mu$ and $\upsilon'_s \perp \mu$, let E_s be defined as above with $\mu(E_s) = \upsilon_s(\tilde{E}_s) = 0$, and let E'_s be the analogously defined set for υ'_s, so that $\mu(E'_s) = \upsilon_s(\tilde{E}'_s) = 0$. Defining $E''_s = E_s \cup E'_s$ obtains $\mu(E''_s) = 0$. Thus $\mu(A) = 0$ for all measurable $A \subset E''_s$, and since $(\upsilon_{ac} - \upsilon'_{ac}) \ll \mu$ and this implies that $(\upsilon_{ac} - \upsilon'_{ac})(A) = 0$ for all such A, and by $(*)$ this in turn obtains $(\upsilon'_s - \upsilon_s)(A) = 0$. Thus $\upsilon_{ac} = \upsilon'_{ac}$ and $\upsilon'_s = \upsilon_s$ on all measurable subsets of E''_s. But $\upsilon_s(\tilde{E}''_s) = \upsilon'_s(\tilde{E}''_s) = 0$ since $\tilde{E}''_s = \tilde{E}_s \cap \tilde{E}'_s$, and so $\upsilon_s(A) = \upsilon'_s(A) = 0$ for all measurable $A \subset \tilde{E}''_s$, and by $(*)$ the same is true that $\upsilon_{ac}(A) = \upsilon'_{ac}(A) = 0$. Combining obtains $\upsilon_{ac} = \upsilon'_{ac}$ and $\upsilon'_s = \upsilon_s$ on $\sigma(X)$ and the proof of uniqueness is complete.

■

Proposition 7.18 above investigated the notion of being **absolutely continuous**, both as this terms applies to a given function and as a property between two measures. The next result addresses the notion of being **singular,** again as this term applies to a function as in definition 3.47 of book 3 and as a property between two measures. The conclusion below is that if F is an increasing right continuous function, then F is singular by definition 3.47 if and only if $\upsilon_F \perp m$.

This result may seem surprising initially. A function F is singular if $F' = 0$ except on a set of measure 0. When applied to two measures this term reflects a comparative concept, and the terminology **mutually singular** highlights this fact. By definition, $\upsilon_F \perp m$ means that there is a set A so that $m(A) = \upsilon_F(\widetilde{A}) = 0$. In other words, m and υ_F have disjoint supports with union equal to X, or in the special case here, $\mathbb{R}$.

But note that for a function F, it could have been stated that F is **Lebesgue singular** if $m[A] = 0$ where $A = \{x|F'(x) \neq 0\}$. So indeed, singularity of a function reflects a comparative measure concept if it can be shown that for such A, $\upsilon_F(\widetilde{A}) = 0$ where υ_F is the Borel measure induced by F.

Proposition 7.28 (Singular Definitions) *Let υ_F be a Borel measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ induced by an increasing, right continuous function F. Then F is a singular function by definition 3.47 of book 3 if and only if $\upsilon_F \perp m$.*

Proof. *If F is a singular function then $F'(x) = 0$ m-a.e. Let $\widetilde{A} = \{x|F'(x) = 0\}$ and $A = \{x|F'(x) \neq 0\}$. Then $m[A] = 0$ and so $\upsilon_F \perp m$ will follow if we show that $\upsilon_F\left[\widetilde{A}\right] = 0$, and we do this by showing that $\upsilon_F\left[\widetilde{A} \cap [-a, a]\right] = 0$ for any interval $[-a, a]$. By proposition 5.30 of book 1, if $F'(x) \leq \epsilon$ on a Borel set B then $\upsilon_F(B) \leq \epsilon m(B)$. But $F'(x) = 0$ on $\widetilde{A}$ implies that for every $\epsilon > 0$ that*

$$\upsilon_F\left[\widetilde{A} \cap [-a, a]\right] \leq 2a\epsilon,$$

and so $\upsilon_F\left[\widetilde{A} \cap [-a, a]\right] = 0$ for all a. Using continuity from below obtains that $\upsilon_F\left[\widetilde{A}\right] = 0$, and hence, $\upsilon_F \perp m$.

Conversely, if $\upsilon_F \perp m$ then for some Borel set A, $m(A) = \upsilon_F(\widetilde{A}) = 0$. Again by proposition 5.30 of book 3, if $F'(x) \geq \epsilon$ on a Borel set B then

$\epsilon m(B) \leq \upsilon_F(B)$. Since $m(A) = 0$, this implies that:

$$\begin{aligned} \epsilon m\left[\{x|F'(x) \geq \epsilon\}\right] &= \epsilon m\left[\{x \in \widetilde{A}|F'(x) \geq \epsilon\}\right] \\ &\leq \upsilon_F(\widetilde{A}) \\ &= 0. \end{aligned}$$

Hence $m\left[\{x|F'(x) \geq \epsilon\}\right] = 0$ for all ϵ and so $m\left[\{x|F'(x) \neq 0\}\right] = 0$ and F is a singular function. ■

References

I have listed below a number of textbook references for the mathematics and finance presented in this series of books. All provide both theoretical and applied materials in their respective areas that are beyond those developed here and are worth pursuing by those interested in gaining a greater depth or breadth of knowledge. This list is by no means complete and is intended only as a guide to further study. In addition, these references include various published research papers if they have been identified in this book's chapters.

The reader will no doubt observe that the mathematics references are somewhat older than the finance references and upon web searching will find that several of the older texts in each category have been updated to newer editions, sometimes with additional authors. Since I own and use the editions below, I decided to present these editions rather than reference the newer editions which I have not reviewed. As many of these older texts are considered "classics", they are also likely to be found in university and other libraries.

That said, there are undoubtedly many very good new texts by both new and established authors with similar titles that are also worth investigating. One that I will at the risk of immodesty recommend for more introductory materials on mathematics, probability theory and finance is:

0. Reitano, Robert, R. *Introduction to Quantitative Finance: A Math Tool Kit.* Cambridge, MA: The MIT Press, 2010.

Topology, Measure, and Integration

1. Dugundji, James. *Topology.* Boston, MA: Allyn and Bacon, 1970.

2. Doob, J. L. *Measure Theory.* New York, NY: Springer-Verlag, 1994.

3. Edwards, Jr., C. H. *Advanced Calculus of Several Variables.* New York, NY: Academic Press, 1973.

4. Gemignani, M. C. *Elementary Topology.* Reading, MA: Addison-Wesley Publishing, 1967.

5. Halmos, Paul R. *Measure Theory.* New York, NY: D. Van Nostrand, 1950.

6. Hewitt, Edwin, and Karl Stromberg. *Real and Abstract Analysis.* New York, NY: Springer-Verlag, 1965.

7. Royden, H. L. *Real Analysis,* 2nd Edition. New York, NY: The MacMillan Company, 1971.

8. Rudin, Walter. *Principals of Mathematical Analysis,* 3rd Edition. New York, NY: McGraw-Hill, 1976.

9. Rudin, Walter. *Real and Complex Analysis,* 2nd Edition. New York, NY: McGraw-Hill, 1974.

10. Shilov, G. E., and B. L. Gurevich. Integral, Measure & Derivative: A Unified Approach. New York, NY: Dover Publications, 1977.

Probability Theory & Stochastic Processes

11. Billingsley, Patrick. *Probability and Measure,* 3rd Edition. New York, NY: John Wiley & Sons, 1995.

12. Chung, K. L., and R. J. Williams. *Introduction to Stochastic Integration.* Boston, MA: Birkhäuser, 1983.

13. Davidson, James. *Stochastic Limit Theory.* New York, NY: Oxford University Press, 1997.

14. de Haan, Laurens, and Ana Ferreira. *Extreme Value Theory, An Introduction.* New York, NY: Springer Science, 2006.

15. Durrett, Richard. *Probability: Theory and Examples,* 2nd Edition. Belmont, CA: Wadsworth Publishing, 1996.

16. Durrett, Richard. *Stochastic Calculus, A Practical Intriduction.* Boca Raton, FL: CRC Press, 1996.

17. Feller, William. *An Introduction to Probability Theory and Its Applications,* Volume I. New York, NY: John Wiley & Sons, 1968.

18. Feller, William. *An Introduction to Probability Theory and Its Applications,* Volume II, 2nd Edition. New York, NY: John Wiley & Sons, 1971.

19. Friedman, Avner. *Stochastic Differential Equations and Application, Volume 1 and 2.* New York, NY: Academic Press, 1975.

20. Ikeda, Nobuyuki, and Shinzo Watanabe. Stochastic Differential Equations and Diffusion Processes. Tokyo, Japan: Kodansha Scientific, 1981.

21. Karatzas, Ioannis, and Steven E. Shreve. *Brownian Motion and Stochastic Calculus.* New York, NY: Springer-Verlag, 1988.

22. Kloeden, Peter E., and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations.* New York, NY: Springer-Verlag, 1992.

23. Nelson, Roger B. *An Introduction to Copulas,* 2nd Edition. New York, NY: Springer Science, 2006.

24. Øksendal, Bernt. *Stochastic Differential Equations, An Introduction with Applications,* 5th Edition. New York, NY: Springer-Verlag, 1998.

25. Protter, Phillip. *Stochastic Integration and Differential Equations, A New Approach.* New York, NY: Springer-Verlag, 1992.

26. Revuz, Daniel, and Marc Yor. *Continuous Martingales and Brownian Motion,* 3rd Edition. New York, NY: Springer-Verlag, 1999.

27. Rogers, L. C. G., and D. Williams. *Diffusions, Markov Processes and Martingales*, Volume 1, Foundations, 2nd Edition. Cambridge, UK: Cambridge University Press, 2000.

28. Rogers, L. C. G., and D. Williams. *Diffusions, Markov Processes and Martingales*, Volume 2, Itô Calculus 2nd Edition. Cambridge, UK: Cambridge University Press, 2000.

29. Schilling, René L. and Lothar Partzsch. *Brownian Motion: An Introduction to Stochastic Processes,* 2nd Edition. Berlin/Boston: Walter de Gruyter GmbH, 2014.

30. Schuss, Zeev, *Theory and Applications of Stochastic Differential Equations.* New York, NY: John Wiley and Sons, 1980.

Finance Applications

31. Etheridge, Alison. *A Course in Financial Calculus.* Cambridge, UK: Cambridge University Press, 2002.

32. Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance.* New York, NY: Springer-Verlag, 1997.

33. Hunt, P. J., and J. E. Kennedy. *Financial Derivatives in Theory and Practice,* Revised Edition. Chichester, UK: John Wiley & Sons, 2004.

34. McLeish, Don L. *Monte Carlo Simulation and Finance.* New York, NY: John Wiley, 2005.

35. McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton, NJ.:Princeton University Press, 2005.

Index