Chapter 2

A modicum of measure theory

SECTION 1 defines measures and sigma-fields.
SECTION 2 defines measurable functions.
SECTION 3 defines the integral with respect to a measure as a linear functional on a cone of measurable functions. The definition sidesteps the details of the construction of integrals from measures.
*SECTION *4 constructs integrals of nonnegative measurable functions with respect to a countably additive measure.*
SECTION 5 establishes the Dominated Convergence theorem, the Swiss Army knife of measure theoretic probability.
SECTION 6 collects together a number of simple facts related to sets of measure zero.
*SECTION *7 presents a few facts about spaces of functions with integrable pth powers, with emphasis on the case p=2, which defines a Hilbert space.*
SECTION 8 defines uniform integrability, a condition slightly weaker than domination. Convergence in $\mathcal{L}^1$ is characterized as convergence in probability plus uniform integrability.
SECTION 9 defines the image measure, which includes the concept of the distribution of a random variable as a special case.
SECTION 10 explains how generating class arguments, for classes of sets, make measure theory easy.
*SECTION *11 extends generating class arguments to classes of functions.*

1. Measures and sigma-fields

As promised in Chapter 1, we begin with measures as set functions, then work quickly towards the interpretation of integrals as linear functionals. Once we are past the purely set-theoretic preliminaries, I will start using the de Finetti notation (Section 1.4) in earnest, writing the same symbol for a set and its indicator function.

Our starting point is a ***measure space***: a triple $(\mathcal{X}, \mathcal{A}, \mu)$, with $\mathcal{X}$ a set, $\mathcal{A}$ a class of subsets of $\mathcal{X}$, and μ a function that attaches a nonnegative number (possibly $+\infty$) to each set in $\mathcal{A}$. The class $\mathcal{A}$ and the set function μ are required to have properties that facilitate calculations involving limits along sequences.

<1> **Definition.** *Call a class $\mathcal{A}$ a sigma-field of subsets of $\mathcal{X}$ if:*

(i) the empty set $\emptyset$ and the whole space $\mathcal{X}$ both belong to $\mathcal{A}$;

(ii) if A belongs to $\mathcal{A}$ then so does its complement A^c;

(iii) if $A_1, A_2, \ldots$ is a countable collection of sets in $\mathcal{A}$ then both the union $\cup_i A_i$ and the intersection $\cap_i A_i$ are also in $\mathcal{A}$.

Some of the requirements are redundant as stated. For example, once we have $\emptyset \in \mathcal{A}$ then (ii) implies $\mathcal{X} \in \mathcal{A}$. When we come to establish properties about sigma-fields it will be convenient to have the list of defining properties pared down to a minimum, to reduce the amount of mechanical checking. The theorems will be as sparing as possible in the amount the work they require for establishing the sigma-field properties, but for now redundancy does not hurt.

The collection $\mathcal{A}$ need not contain every subset of $\mathcal{X}$, a fact forced upon us in general if we want μ to have the properties of a countably additive measure.

<2> **Definition.** *A function μ defined on the sigma-field $\mathcal{A}$ is called a (countably additive, nonnegative) measure if:*

(i) $0 \le \mu A \le \infty$ for each A in $\mathcal{A}$;

(ii) $\mu\emptyset = 0$;

(iii) if $A_1, A_2, \ldots$ is a countable collection of pairwise disjoint sets in $\mathcal{A}$ then $\mu\left(\cup_i A_i\right) = \sum_i \mu A_i$.

A measure μ for which $\mu\mathcal{X} = 1$ is called a ***probability measure***, and the corresponding $(\mathcal{X}, \mathcal{A}, \mu)$ is called a ***probability space***. For this special case it is traditional to use a symbol like $\mathbb{P}$ for the measure, a symbol like Ω for the set, and a symbol like $\mathcal{F}$ for the sigma-field. A triple $(\Omega, \mathcal{F}, \mathbb{P})$ will always denote a probability space.

Usually the qualifications "countably additive, nonnegative" are omitted, on the grounds that these properties are the most commonly assumed—the most common cases deserve the shortest names. Only when there is some doubt about whether the measures are assumed to have all the properties of Definition <2> should the qualifiers be attached. For example, one speaks of "finitely additive measures" when an analog of property (iii) is assumed only for finite disjoint collections, or "signed measures" when the value of μA is not necessarily nonnegative. When finitely additive or signed measures are under discussion it makes sense to mention explicitly when a particular measure is nonnegative or countably additive, but, in general, you should go with the shorter name.

Where do measures come from? The most basic constructions start from set functions μ defined on small collections of subsets $\mathcal{E}$, such as the collection of all subintervals of the real line. One checks that μ has properties consistent with the requirements of Definition <2>. One seeks to extend the domain of definition while preserving the countable additivity properties of the set function. As you saw in Chapter 1, Theorems guaranteeing existence of such extensions were the culmination of a long sequence of refinements in the concept of integration (Hawkins 1979). They represent one of the great achievements of modern mathematics, even though those theorems now occupy only a handful of pages in most measure theory texts.

Finite additivity has several appealing interpretations (such as the fair-prices of Section 1.5) that have given it ready acceptance as an axiom for a model of real-world uncertainty. Countable additivity is sometimes regarded with suspicion, or justified as a matter of mathematical convenience. (However, see Problem [6] for an equivalent form of countable additivity, which has some claim to intuitive appeal.) It is difficult to develop a simple probability theory without countable additivity, which gives one the licence (for only a small fee) to integrate series term-by-term, differentiate under integrals, and interchange other limiting operations.

The classical constructions are significant for my exposition mostly because they ensure existence of the measures needed to express the basic results of probability theory. I will relegate the details to the Problems and to Appendix A. If you crave a more systematic treatment you might consult one of the many excellent texts on measure theory, such as Royden (1968).

The constructions do not—indeed cannot, in general—lead to countably additive measures on the class of all subsets of a given $\mathcal{X}$. Typically, they extend a set function defined on a class of sets $\mathcal{E}$ to a measure defined on the ***sigma-field*** $\sigma(\mathcal{E})$ ***generated by*** $\mathcal{E}$, or to only slightly larger sigma-fields. By definition,

$$\begin{aligned}\sigma(\mathcal{E}) &:= \text{smallest sigma-field on } \mathcal{X} \text{ containing all sets from } \mathcal{E}\\ &= \{A \subseteq \mathcal{X} : A \in \mathcal{F} \text{ for every sigma-field } \mathcal{F} \text{ with } \mathcal{E} \subseteq \mathcal{F}\}.\end{aligned}$$

The representation given by the second line ensures existence of a smallest sigma-field containing $\mathcal{E}$. The method of definition is analogous to many definitions of "smallest . . . containing a fixed class" in mathematics—think of generated subgroups or linear subspaces spanned by a collection of vectors, for example. For the definition to work one needs to check that sigma-fields have two properties:

(i) If $\{\mathcal{F}_i : i \in \mathcal{I}\}$ is a nonempty collection of sigma-fields on $\mathcal{X}$ then $\cap_{i\in\mathcal{I}}\mathcal{F}_i$, the collection of all the subsets of $\mathcal{X}$ that belong to every $\mathcal{F}_i$, is also a sigma-field.

(ii) For each $\mathcal{E}$ there exists at least one sigma-field $\mathcal{F}$ containing all the sets in $\mathcal{E}$.

You should check property (i) as an exercise. Property (ii) is trivial, because the collection of all subsets of $\mathcal{X}$ is a sigma-field.

REMARK. Proofs of existence of nonmeasurable sets typically depend on some deep set-theoretic principle, such as the Axiom of Choice. Mathematicians who can live with different rules for set theory can have bigger sigma-fields. See Dudley (1989, Section 3.4) or Oxtoby (1971, Section 5) for details.

<3> **Exercise.** Suppose $\mathcal{X}$ consists of five points a, b, c, d, and e. Suppose $\mathcal{E}$ consists of two sets, $E_1 = \{a, b, c\}$ and $E_2 = \{c, d, e\}$. Find the sigma-field generated by $\mathcal{E}$.

SOLUTION: For this simple example we can proceed by mechanical application of the properties that a sigma-field $\sigma(\mathcal{E})$ must possess. In addition to the obvious $\emptyset$ and $\mathcal{X}$, it must contain each of the sets

$$\begin{aligned}F_1 &:= \{a, b\} = E_1 \cap E_2^c \qquad \text{and} \qquad F_2 := \{c\} = E_1 \cap E_2,\\ F_3 &:= \{d, e\} = E_1^c \cap E_2 \qquad \text{and} \qquad F_4 := \{a, b, d, e\} = F_1 \cup F_3.\end{aligned}$$

Further experimentation creates no new members of $\sigma(\mathcal{E})$; the sigma-field consists of the sets

$$\emptyset, \quad F_1, \quad F_2, \quad F_3, \quad F_1 \cup F_3, \quad F_1 \cup F_2 = E_1, \quad F_2 \cup F_3 = E_2, \quad \mathcal{X}.$$

The sets F_1, F_2, F_3 are the ***atoms*** of the sigma-field; every member of $\sigma(\mathcal{E})$ is a union of some collection (possibly empty) of F_i. The only measurable subsets of F_i are the empty set and F_i itself. There are no measurable protons or neutrons hiding inside these atoms. □

An unsystematic construction might work for finite sets, but it cannot generate all members of a sigma-field in general. Indeed, we cannot even hope to list all the members of an infinite sigma-field. Instead we must find a less explicit way to characterize its sets.

<4> **Example.** By definition, the Borel sigma-field on the real line, denoted by $\mathcal{B}(\mathbb{R})$, is the sigma-field generated by the open subsets. We could also denote it by $\sigma(\mathcal{G})$ where $\mathcal{G}$ stands for the class of all open subsets of $\mathbb{R}$. There are several other generating classes for $\mathcal{B}(\mathbb{R})$. For example, as you will soon see, the class $\mathcal{E}$ of all intervals $(-\infty, t]$, with $t \in \mathbb{R}$, is a generating class.

It might appear a hopeless task to prove that $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$ if we cannot explicitly list the members of both sigma-fields, but actually the proof is quite routine. You should try to understand the style of argument because it is often used in probability theory.

The equality of sigma-fields is established by two inclusions, $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{G})$ and $\sigma(\mathcal{G}) \subseteq \sigma(\mathcal{E})$, both of which follow from more easily established results. First we must prove that $\mathcal{E} \subseteq \sigma(\mathcal{G})$, showing that $\sigma(\mathcal{G})$ is one of the sigma-fields $\mathcal{F}$ that enter into the intersection defining $\sigma(\mathcal{E})$, and hence $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{G})$. The other inclusion follows similarly if we show that $\mathcal{G} \subseteq \sigma(\mathcal{E})$.

Each interval $(-\infty, t]$ in $\mathcal{E}$ has a representation $\bigcap_{n=1}^{\infty}(-\infty, t+n^{-1})$, a countable intersection of open sets. The sigma-field $\sigma(\mathcal{G})$ contains all open sets, and it is stable under countable intersections. It therefore contains each $(-\infty, t]$. That is, $\mathcal{E} \subseteq \sigma(\mathcal{G})$.

The argument for $\mathcal{G} \subseteq \sigma(\mathcal{E})$ is only slightly harder. It depends on the fact that an open subset of the real line can be written as a countable union of open intervals. Such an interval has a representation $(a, b) = (-\infty, b) \cap (-\infty, a]^c$, and $(-\infty, b) = \bigcup_{n=1}^{\infty}(-\infty, b - n^{-1}]$. That is, every open set can be built up from sets in $\mathcal{E}$ using operations that are guaranteed not to take us outside the sigma-field $\sigma(\mathcal{E})$.

My explanation has been moderately detailed. In a published paper the reasoning would probably be abbreviated to something like "a generating class argument shows that . . . ," with the routine details left to the reader. □

REMARK. The generating class argument often reduces to an assertion like: $\mathcal{A}$ is a sigma-field and $\mathcal{A} \supseteq \mathcal{E}$, therefore $\mathcal{A} = \sigma(\mathcal{A}) \supseteq \sigma(\mathcal{E})$.

<5> **Example.** A class $\mathcal{E}$ of subsets of a set $\mathcal{X}$ is called a ***field*** if it contains the empty set and is stable under complements, finite unions, and finite intersections. For a field $\mathcal{E}$, write $\mathcal{E}_\delta$ for the class of all possible intersections of countable subclasses of $\mathcal{E}$, and $\mathcal{E}_\sigma$ for the class of all possible unions of countable subclasses of $\mathcal{E}$.

Of course if $\mathcal{E}$ is a sigma-field then $\mathcal{E} = \mathcal{E}_\delta = \mathcal{E}_\sigma$, but, in general, the inclusions $\sigma(\mathcal{E}) \supseteq \mathcal{E}_\delta \supseteq \mathcal{E}$ and $\sigma(\mathcal{E}) \supseteq \mathcal{E}_\sigma \supseteq \mathcal{E}$ will be proper. For example, if $\mathcal{X} = \mathbb{R}$ and $\mathcal{E}$ consists of all finite unions of half open intervals $(a, b]$, with possibly $a = -\infty$ or $b = +\infty$, then the set of rationals does not belong to $\mathcal{E}_\sigma$ and the complement of the same set does not belong to $\mathcal{E}_\delta$.

Let μ be a finite measure on $\sigma(\mathcal{E})$. Even though $\sigma(\mathcal{E})$ might be much larger than either $\mathcal{E}_\sigma$ or $\mathcal{E}_\delta$, a generating class argument will show that all sets in $\sigma(\mathcal{E})$ can be ***inner approximated by*** $\mathcal{E}_\delta$, in the sense that,

$$\mu A = \sup\{\mu F : A \supseteq F \in \mathcal{E}_\delta\} \qquad \text{for each } A \text{ in } \sigma(\mathcal{E}),$$

and ***outer approximated by*** $\mathcal{E}_\sigma$, in the sense that,

$$\mu A = \inf\{\mu G : A \subseteq G \in \mathcal{E}_\sigma\} \qquad \text{for each } A \text{ in } \sigma(\mathcal{E}).$$

REMARK. Incidentally, I chose the letters G and F to remind myself of open and closed sets, which have similar approximation properties for Borel measures on metric spaces—see Problem [12].

It helps to work on both approximation properties at the same time. Denote by $\mathcal{B}_0$ the class of all sets in $\sigma(\mathcal{E})$ that can be both innner and outer approximated. A set B belongs to $\mathcal{B}_0$ if and only if, to each $\epsilon > 0$ there exist $F \in \mathcal{E}_\delta$ and $G \in \mathcal{E}_\sigma$ such that $F \subseteq B \subseteq G$ and $\mu(G \backslash F) < \epsilon$. I'll call the sets F and G an ϵ-sandwich for B.

Trivially $\mathcal{B}_0 \supseteq \mathcal{E}$, because each member of $\mathcal{E}$ belongs to both $\mathcal{E}_\sigma$ and $\mathcal{E}_\delta$. The approximation result will follow if we show that $\mathcal{B}_0$ is a sigma-field, for then we will have $\mathcal{B}_0 = \sigma(\mathcal{B}_0) \supseteq \sigma(\mathcal{E})$.

Symmetry of the definition ensures that $\mathcal{B}_0$ is stable under complements: if $F \subseteq B \subseteq G$ is an ϵ-sandwich for B, then $G^c \subseteq B^c \subseteq F^c$ is an ϵ-sandwich for B^c. To show that $\mathcal{B}_0$ is stable under countable unions, consider a countable collection $\{B_n : n \in \mathbb{N}\}$ of sets from $\mathcal{B}_0$. We need to slice the bread thinner as n gets larger: choose $\epsilon/2^n$-sandwiches $F_n \subseteq B_n \subseteq G_n$ for each n. The union $\cup_n B_n$ is sandwiched between the sets $G := \cup_n G_n$ and $H = \cup_n F_n$; and the sets are close in μ measure because

$$\mu\left(\cup_n G_n \backslash \cup_n F_n\right) \le \sum_n \mu(G_n \backslash F_n) < \sum_n \epsilon/2^n = \epsilon.$$

REMARK. Can you prove this inequality? Do you see why $\cup_n G_n \backslash \cup_n F_n \subseteq \cup_n (G_n \backslash F_n)$ and why countable additivity implies that the measure of a countable union of (not necessarily disjoint) sets is smaller than the sum of their measures? If not, just wait until Section 3, after which you can argue that $\cup_n G_n \backslash \cup_n F_n \le \sum_n (G_n \backslash F_n)$, as an inequality between indicator functions, and $\mu\left(\sum_n (G_n \backslash F_n)\right) = \sum_n \mu(G_n \backslash F_n)$ by Monotone Convergence.

We have an ϵ-sandwich, but the bread might not be of the right type. It is certainly true that $G \in \mathcal{E}_\sigma$ (a countable union of countable unions is a countable union), but the set H need not belong to $\mathcal{E}_\delta$. However, the sets $H_N := \cup_{n \le N} F_n$ do belong to $\mathcal{E}_\delta$, and countable additivity implies that $\mu H_N \uparrow \mu H$.

REMARK. Do you see why? If not, wait for Monotone Convergence again.

□ If we choose a large enough N we have a 2ϵ-sandwich $H_N \subseteq \cup_n B_n \subseteq G$.

The measure $\mathfrak{m}$ on $\mathcal{B}(\mathbb{R})$ for which $\mathfrak{m}(a,b] = b - a$ is called ***Lebesgue measure***. Another sort of generating class argument (see Section 10) can be used to show that the values $\mathfrak{m}(B)$ for B in $\mathcal{B}(\mathbb{R})$ are uniquely determined by the values given to intervals; there can exist at most one measure on $\mathcal{B}(\mathbb{R})$ with the stated property. It is harder to show that at least one such measure exists. Despite any intuitions you might have about length, the construction of Lebesgue measure is not trivial—see Appendix A. Indeed, Henri Lebesgue became famous for proving existence of the measure and showing how much could be done with the new integration theory.

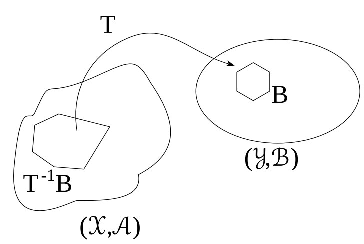
The name Lebesgue measure is also given to an extension of $\mathfrak{m}$ to a measure on a sigma-field, sometimes called the Lebesgue sigma-field, which is slightly larger than $\mathcal{B}(\mathbb{R})$. I will have more to say about the extension in Section 6.

Borel sigma-fields are defined in similar fashion for any topological space $\mathcal{X}$. That is, $\mathcal{B}(\mathcal{X})$ denotes the sigma-field generated by the open subsets of $\mathcal{X}$.

Sets in a sigma-field $\mathcal{A}$ are said to be ***measurable*** or $\mathcal{A}$-measurable. In probability theory they are also called ***events***. Good functions will also be given the title measurable. Try not to get confused when you really need to know whether an object is a set or a function.

2. Measurable functions

Let $\mathcal{X}$ be a set equipped with a sigma-field $\mathcal{A}$, and $\mathcal{Y}$ be a set equipped with a sigma-field $\mathcal{B}$, and T be a function (also called a map) from $\mathcal{X}$ to $\mathcal{Y}$. We say that T ***is*** $\mathcal{A}\backslash\mathcal{B}$***-measurable*** if the inverse image $\{x \in \mathcal{X} : Tx \in B\}$ belongs to $\mathcal{A}$ for each B in $\mathcal{B}$. Sometimes the inverse image is denoted by $\{T \in B\}$ or $T^{-1}B$. Don't be fooled by the T^{-1} notation into treating T^{-1} as a function from $\mathcal{Y}$ into $\mathcal{X}$: it's not, unless T is one-to-one (and onto, if you want to have domain $\mathcal{Y}$). Sometimes an $\mathcal{A}\backslash\mathcal{B}$-measurable map is referred to in abbreviated form as just $\mathcal{A}$-measurable, or just $\mathcal{B}$-measurable, or just measurable, if there is no ambiguity about the unspecified sigma-fields.



For example, if $\mathcal{Y} = \mathbb{R}$ and $\mathcal{B}$ equals the Borel sigma-field $\mathcal{B}(\mathbb{R})$, it is common to drop the $\mathcal{B}(\mathbb{R})$ specification and refer to the map as being $\mathcal{A}$-measurable, or as being Borel measurable if $\mathcal{A}$ is understood and there is any doubt about which sigma-field to use for the real line. *In this book, you may assume that any sigma-field on* $\mathbb{R}$ *is its Borel sigma-field, unless explicitly specified otherwise.* It can get confusing if you misinterpret where the unspecified sigma-fields live. My advice would be that you imagine a picture showing the two spaces involved, with any missing sigma-field labels filled in.

Sometimes the functions come first, and the sigma-fields are chosen specifically to make those functions measurable.

<6> **Definition.** *Let $\mathcal{H}$ be a class of functions on a set $\mathcal{X}$. Suppose the typical h in $\mathcal{H}$ maps $\mathcal{X}$ into a space $\mathcal{Y}_h$ equipped with a sigma-field $\mathcal{B}_h$. Then the sigma-field $\sigma(\mathcal{H})$ generated by $\mathcal{H}$ is defined as $\sigma\{h^{-1}(B) : B \in \mathcal{B}_h,\ h \in \mathcal{H}\}$. It is the smallest sigma-field $\mathcal{A}_0$ on $\mathcal{X}$ for which each h in $\mathcal{H}$ is $\mathcal{A}_0\backslash\mathcal{B}_h$-measurable.*

<7> **Example.** If $\mathcal{B} = \sigma(\mathcal{E})$ for some class $\mathcal{E}$ of subsets of $\mathcal{Y}$ then a map T is $\mathcal{A}\backslash\sigma(\mathcal{E})$-measurable if and only if $T^{-1}E \in \mathcal{A}$ for every E in $\mathcal{E}$. You should prove this assertion by checking that $\{B \in \mathcal{B} : T^{-1}B \in \mathcal{A}\}$ is a sigma-field, and then arguing from the definition of a generating class.

In particular, to establish $\mathcal{A}\backslash\mathcal{B}(\mathbb{R})$-measurability of a map into the real line it is enough to check the inverse images of intervals of the form (t, ∞), with t ranging over $\mathbb{R}$. (In fact, we could restrict t to a countable dense subset of $\mathbb{R}$, such as the set of rationals: How would you build an interval (t, ∞) from intervals (t_i, ∞) with rational t_i?) That is, a real-valued function f is Borel-measurable if $\{x \in \mathcal{X} : f(x) > t\} \in \mathcal{A}$ for each real t. There are many similar assertions obtained by using other generating classes for $\mathcal{B}(\mathbb{R})$. Some authors use particular generating classes for the definition of measurability, and then derive facts about inverse images of Borel sets as theorems. □

It will be convenient to consider not just real-valued functions on a set $\mathcal{X}$, but also functions from $\mathcal{X}$ into the extended real line $\overline{\mathbb{R}} := [-\infty, \infty]$. The Borel sigma-field $\mathcal{B}(\overline{\mathbb{R}})$ is generated by the class of open sets, or, more explicitly, by all sets in $\mathcal{B}(\mathbb{R})$ together with the two singletons $\{-\infty\}$ and $\{\infty\}$. It is an easy exercise to show that $\mathcal{B}(\overline{\mathbb{R}})$ is generated by the class of all sets of the form $(t, \infty]$, for t in $\mathbb{R}$, and by the class of all sets of the form $[-\infty, t)$, for t in $\mathbb{R}$. We could even restrict t to any countable dense subset of $\mathbb{R}$.

<8> **Example.** Let a set $\mathcal{X}$ be equipped with a sigma-field $\mathcal{A}$. Let $\{f_n : n \in \mathbb{N}\}$ be a sequence of $\mathcal{A}\backslash\mathcal{B}(\mathbb{R})$-measurable functions from $\mathcal{X}$ into $\mathbb{R}$. Define functions f and g by taking pointwise suprema and infima: $f(x) := \sup_n f_n(x)$ and $g(x) := \inf_n f_n(x)$. Notice that f might take the value $+\infty$, and g might take the value $-\infty$, at some points of $\mathcal{X}$. We may consider both as maps from $\mathcal{X}$ into $\overline{\mathbb{R}}$. (In fact, the whole argument is unchanged if the f_n functions themselves are also allowed to take infinite values.)

The function f is $\mathcal{A}\backslash\mathcal{B}(\overline{\mathbb{R}})$-measurable because

$$\{x : f(x) > t\} = \cup_n \{x : f_n(x) > t\} \in \mathcal{A} \qquad \text{for each real } t:$$

for each fixed x, the supremum of the real numbers $f_n(x)$ is strictly greater than t if and only if $f_n(x) > t$ for at least one n. Example <7> shows why we have only to check inverse images for such intervals.

The same generating class is not as convenient for proving measurability of g. It is not true that an infimum of a sequence of real numbers is strictly greater than t if and only if all of the numbers are strictly greater than t: think of the sequence $\{n^{-1} : n = 1, 2, 3, \ldots\}$, whose infimum is zero. Instead you should argue via the identity $\{x : g(x) < t\} = \cup_n \{x : f_n(x) < t\} \in \mathcal{A}$ for each real t. □

From Example <8> and the representations $\limsup f_n(x) = \inf_{n\in\mathbb{N}} \sup_{m\ge n} f_m(x)$ and $\liminf f_n(x) = \sup_{n\in\mathbb{N}} \inf_{m\ge n} f_m(x)$, it follows that the lim sup or lim inf of a sequence of measurable (real- or extended real-valued) functions is also measurable. In particular, if the limit exists it is measurable.

Measurability is also preserved by the usual algebraic operations—sums, differences, products, and so on—provided we take care to avoid illegal pointwise calculations such as $\infty - \infty$ or $0/0$. There are several ways to establish these stability properties. One of the more direct methods depends on the fact that $\mathbb{R}$ has a countable dense subset, as illustrated by the following argument for sums.

<9> **Example.** Let f and g be $\mathcal{B}(\mathbb{R})$-measurable functions, with pointwise sum $h(x) = f(x) + g(x)$. (I exclude infinite values because I don't want to get caught up with inconclusive discussions of how we might proceed at points x where $f(x) = +\infty$ and $g(x) = -\infty$, or $f(x) = -\infty$ and $g(x) = +\infty$.) How can we prove that h is also a $\mathcal{B}(\mathbb{R})$-measurable function?

It is true that

$$\{x : h(x) > t\} = \cup_{s\in\mathbb{R}} \big(\{x : f(x) = s\} \cap \{x : g(x) > t - s\}\big),$$

and it is true that the set $\{x : f(x) = s\} \cap \{x : g(x) > t - s\}$ is measurable for each s and t, but sigma-fields are not required to have any particular stability properties for uncountable unions. Instead we should argue that at each x for which $f(x)+g(x) > t$ there exists a rational number r such that $f(x) > r > t - g(x)$. Conversely if there is an r lying strictly between $f(x)$ and $t - g(x)$ then $f(x) + g(x) > t$. Thus

$$\{x : h(x) > t\} = \cup_{r\in\mathbb{Q}} (\{x : f(x) > r\} \cap \{x : g(x) > t - r\}),$$

where $\mathbb{Q}$ denotes the countable set of rational numbers. A countable union of intersections of pairs of measurable sets is measurable. The sum is a measurable function. □

As a little exercise you might try to extend the argument from the last Example to the case where f and g are allowed to take the value $+\infty$ (but not the value $-\infty$). If you want practice at playing with rationals, try to prove measurability of products (be careful with inequalities if dividing by negative numbers) or try Problem [4], which shows why a direct attack on the lim sup requires careful handling of inequalities in the limit.

The real significance of measurability becomes apparent when one works through the construction of integrals with respect to measures, as in Section 4. For the moment it is important only that you understand that the family of all measurable functions is stable under most of the familiar operations of analysis.

<10> **Definition.** *The class $\mathcal{M}(\mathcal{X}, \mathcal{A})$, or $\mathcal{M}(\mathcal{X})$ or just $\mathcal{M}$ for short, consists of all $\mathcal{A}\backslash\mathcal{B}(\overline{\mathbb{R}})$-measurable functions from $\mathcal{X}$ into $\overline{\mathbb{R}}$. The class $\mathcal{M}^+(\mathcal{X}, \mathcal{A})$, or $\mathcal{M}^+(\mathcal{X})$ or just $\mathcal{M}^+$ for short, consists of the nonnegative functions in $\mathcal{M}(\mathcal{X}, \mathcal{A})$.*

If you desired exquisite precision you could write $\mathcal{M}(\mathcal{X}, \mathcal{A}, \overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, to eliminate all ambiguity about domain, range, and sigma-fields.

The collection $\mathcal{M}^+$ is a cone (stable under sums and multiplication of functions by positive constants). It is also stable under products, pointwise limits of sequences, and suprema or infima of countable collections of functions. It is not a vector space, because it is not stable under subtraction; but it does have the property that if f and g belong to $\mathcal{M}^+$ and g takes only real values, then the positive part $(f-g)^+$, defined by taking the pointwise maximum of $f(x)-g(x)$ with 0, also belongs to $\mathcal{M}^+$. You could adapt the argument from Example <9> to establish the last fact.

It proves convenient to work with $\mathcal{M}^+$ rather than with the whole of $\mathcal{M}$, thereby eliminating many problems with $\infty - \infty$. As you will soon learn, integrals have some convenient properties when restricted to nonnegative functions.

For our purposes, one of the most important facts about $\mathcal{M}^+$ will be the possibility of approximation by ***simple functions*** that is by measurable functions of the form $s := \sum_i \alpha_i A_i$, for finite collections of real numbers α_i and events A_i from $\mathcal{A}$. If the A_i are disjoint, $s(x)$ equals α_i when $x \in A_i$, for some i, and is zero otherwise. If the A_i are not disjoint, the nonzero values taken by s are sums of various subsets of the $\{\alpha_i\}$. Don't forget: the symbol A_i gets interpreted as an indicator function when we start doing algebra. I will write $\mathcal{M}^+_{\text{simple}}$ for the cone of all simple functions in $\mathcal{M}^+$.

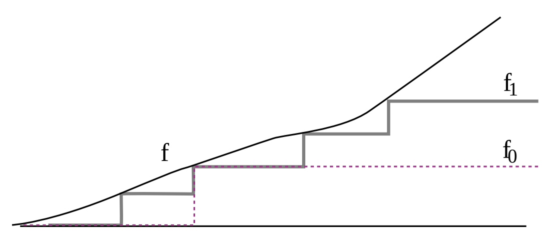
<11> **Lemma.** *For each f in $\mathcal{M}^+$ the sequence $\{f_n\} \subseteq \mathcal{M}^+_{\text{simple}}$, defined by*

$$f_n := 2^{-n} \sum_{i=1}^{4^n} \{ f \geq i/2^n \},$$

has the property $0 \leq f_1(x) \leq f_2(x) \leq \ldots \leq f_n(x) \uparrow f(x)$ *at every x.*

REMARK. The definition of f_n involves algebra, so you must interpret $\{f \geq i/2^n\}$ as the indicator function of the set of all points x for which $f(x) \geq i/2^n$.

Proof. At each x, count the number of nonzero indicator values. If $f(x) \geq 2^n$, all 4^n summands contribute a 1, giving $f_n(x) = 2^n$. If $k2^{-n} \leq f(x) < (k+1)2^{-n}$, for some integer k from $\{0, 1, 2, \ldots, 4^n - 1\}$, then exactly k of the summands contribute a 1, giving $f_n(x) = k2^{-n}$. (Check that the last assertion makes sense when k equals 0.) That is, for $0 \leq f(x) < 2^n$, the function f_n rounds down to an integer multiple of 2^{-n}, from which the convergence and monotone increasing properties follow.



If you do not find the monotonicity assertion convincing, you could argue, more formally, that

$$f_n = \frac{1}{2^{n+1}} \sum_{i=1}^{4^n} 2\left\{ f \geq \frac{2i}{2^{n+1}} \right\} \leq \frac{1}{2^{n+1}} \sum_{i=1}^{4\times 4^n} \left(\left\{ f \geq \frac{2i}{2^{n+1}} \right\} + \left\{ f \geq \frac{2i-1}{2^{n+1}} \right\} \right) = f_{n+1},$$

which reflects the effect of doubling the maximum value and halving the step size when going from the nth to the (n+1)st approximation. □

As an exercise you might prove that the product of functions in $\mathcal{M}^+$ also belongs to $\mathcal{M}^+$, by expressing the product as a pointwise limit of products of simple functions. Notice how the convention $0 \times \infty = 0$ is needed to ensure the correct limit behavior at points where one of the factors is zero.

3. Integrals

Just as $\int_a^b f(x)\,dx$ represents a sort of limiting sum of $f(x)$ values weighted by small lengths of intervals—the $\int$ sign is a long "S", for sum, and the dx is a sort of limiting increment—so can the general integral $\int f(x)\,\mu(dx)$ be defined as a limit of weighted sums but with weights provided by the measure μ. The formal definition involves limiting operations that depend on the assumed measurability of the function f. You can skip the details of the construction (Section 4) by taking the following result as an axiomatic property of the integral.

<12> **Theorem.** *For each measure μ on $(\mathcal{X}, \mathcal{A})$ there is a uniquely determined functional, a map $\tilde{\mu}$ from $\mathcal{M}^+(\mathcal{X}, \mathcal{A})$ into $[0, \infty]$, having the following properties:*

(i) $\tilde{\mu}(\mathbb{I}_A) = \mu A$ *for each* A *in* $\mathcal{A}$*;*

(ii) $\tilde{\mu}(0) = 0$*, where the first zero stands for the zero function;*

(iii) *for nonnegative real numbers* α*,* β *and functions* f*,* g *in* $\mathcal{M}^+$*,*

$$\tilde{\mu}(\alpha f + \beta g) = \alpha \tilde{\mu}(f) + \beta \tilde{\mu}(g);$$

(iv) *if* f*,* g *are in* $\mathcal{M}^+$ *and* $f \le g$ *everywhere then* $\tilde{\mu}(f) \le \tilde{\mu}(g)$*;*

(v) *if* $f_1, f_2, \ldots$ *is a sequence in* $\mathcal{M}^+$ *with* $0 \le f_1(x) \le f_2(x) \le \ldots \uparrow f(x)$ *for each* x *in* $\mathcal{X}$ *then* $\tilde{\mu}(f_n) \uparrow \tilde{\mu}(f)$*.*

I will refer to (iii) as ***linearity***, even though $\mathcal{M}^+$ is not a vector space. It will imply a linearity property when $\tilde{\mu}$ is extended to a vector subspace of $\mathcal{M}$. Property (iv) is redundant because it follows from (ii) and nonnegativity. Property (ii) is also redundant: put $A = \emptyset$ in (i); or, interpreting $0 \times \infty$ as 0, put $\alpha = \beta = 0$ and $f = g = 0$ in (iii). We need to make sure the bad case $\tilde{\mu} f = \infty$, for all f in $\mathcal{M}^+$, does not slip through if we start stripping away redundant requirements.

Notice that the limit function f in (v) automatically belongs to $\mathcal{M}^+$. The limit assertion itself is called the ***Monotone Convergence property***. It corresponds directly to countable additivity of the measure. Indeed, if $\{A_i : i \in \mathbb{N}\}$ is a countable collection of disjoint sets from $\mathcal{A}$ then the functions $f_n := A_1 + \ldots + A_n$ increase pointwise to the indicator function of $A = \cup_{i \in \mathbb{N}} A_i$, so that Monotone Convergence and linearity imply $\mu A = \sum_i \mu A_i$.

REMARK. You should ponder the role played by $+\infty$ in Theorem <12>. For example, what does $\alpha \tilde{\mu}(f)$ mean if $\alpha = 0$ and $\tilde{\mu}(f) = \infty$? The interpretation depends on the convention that $0 \times \infty = 0$.

In general you should be suspicious of any convention involving $\pm\infty$. Pay careful attention to cases where it operates. For example, how would the five assertions be affected if we adopted a new convention, whereby $0 \times \infty = 6$? Would the Theorem still hold? Where exactly would it fail? I feel uneasy if it is not clear how a convention is disposing of awkward cases. My advice: be very, very careful with any calculations involving infinity. Subtle errors are easy to miss when concealed within a convention.

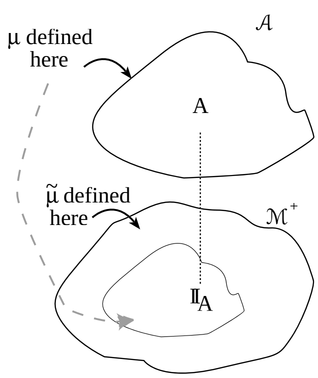
There is a companion to Theorem <12> that shows why it is largely a matter of taste whether one starts from measures or integrals as the more primitive measure theoretic concept.

<13> **Theorem.** *Let $\tilde{\mu}$ be a map from $\mathcal{M}^+$ to $[0, \infty]$ that satisfies properties (ii) through (v) of Theorem <12>. Then the set function defined on the sigma-field $\mathcal{A}$ by (i) is a (countably additive, nonnegative) measure, with $\tilde{\mu}$ the functional that it generates.*

Lemma <11> provides the link between the measure μ and the functional $\tilde{\mu}$. For a given f in $\mathcal{M}^+$, let $\{f_n\}$ be the sequence defined by the Lemma. Then

$$\tilde{\mu} f = \lim_{n\to\infty} \tilde{\mu} f_n = \lim_{n\to\infty} 2^{-n} \sum_{i=1}^{4^n} \mu\{f \geq i/2^n\},$$

the first equality by Monotone Convergence, the second by linearity. The value of $\tilde{\mu} f$ is uniquely determined by μ, as a set function on $\mathcal{A}$. It is even possible to use the equality, or something very similar, as the basis for a direct construction of the integral, from which properties (i) through (v) are then derived, as you will see from Section 4.



In summary: There is a one-to-one correspondence between measures on the sigma-field $\mathcal{A}$ and increasing linear functionals on $\mathcal{M}^+(\mathcal{A})$ with the Monotone Convergence property. To each measure μ there is a uniquely determined functional $\tilde{\mu}$ for which $\tilde{\mu}(\mathbb{I}_A) = \mu(A)$ for every A in $\mathcal{A}$. The functional $\tilde{\mu}$ is usually called an ***integral*** with respect to μ, and is variously denoted by $\int f\, d\mu$ or $\int f(x)\, \mu(dx)$ or $\int_{\mathcal{X}} f\, d\mu$ or $\int f(x)\, d\mu(x)$. With the de Finetti notation, where we identify a set A with its indicator function, the functional $\tilde{\mu}$ is just an extension of μ from a smaller domain (indicators of sets in $\mathcal{A}$) to a larger domain (all of $\mathcal{M}^+$). Accordingly, we should have no qualms about denoting it by the same symbol. I will write μf for the integral. With this notation, assertion (i) of Theorem <12> becomes: $\mu A = \mu A$ for all A in $\mathcal{A}$. You probably can't tell that the A on the left-hand side is an indicator function and the μ is an integral, but you don't need to be able to tell—that is precisely what (i) asserts.

REMARK. In elementary algebra we rely on parentheses, or precedence, to make our meaning clear. For example, both $(ax) + b$ and $ax + b$ have the same meaning, because multiplication has higher precedence than addition. With traditional notation, the $\int$ and the $d\mu$ act like parentheses, enclosing the integrand and separating it from following terms. With linear functional notation, we sometimes need explicit parentheses to make the meaning unambiguous. As a way of eliminating some parentheses, I often work with the convention that integration has lower precedence than exponentiation, multiplication, and division, but higher precedence than addition or subtraction. Thus I intend you to read $\mu fg + 6$ as $(\mu(fg)) + 6$. I would write $\mu(fg + 6)$ if the 6 were part of the integrand.

Some of the traditional notations also remove ambiguity when functions of several variables appear in the integrand. For example, in $\int f(x, y)\,\mu(dx)$ the y variable is held fixed while the μ operates on the first argument of the function. When a similar ambiguity might arise with linear functional notation, I will append a superscript, as in $\mu^x f(x, y)$, to make clear which variable is involved in the integration.

<14> **Example.** Suppose μ is a finite measure (that is, $\mu\mathcal{X} < \infty$) and f is a function in $\mathcal{M}^+$. Then $\mu f < \infty$ if and only if $\sum_{n=1}^{\infty} \mu\{f \ge n\} < \infty$.

The assertion is just a pointwise inequality in disguise. By considering separately values for which $k \le f(x) < k+1$, for $k = 0, 1, 2, \ldots$, you can verify the pointwise inequality between functions,

$$\textstyle\sum_{n=1}^{\infty}\{f \ge n\} \le f \le 1 + \sum_{n=1}^{\infty}\{f \ge n\}.$$

In fact, the sum on the left-hand side defines $\lfloor f(x) \rfloor$, the largest integer $\le f(x)$, and the right-hand side denotes the smallest integer $> f(x)$. From the leftmost inequality,

$$\begin{aligned} \mu f &\ge \mu\big(\textstyle\sum_{n=1}^{\infty}\{f \ge n\}\big) && \text{increasing} \\ &= \lim_{N\to\infty} \mu\big(\textstyle\sum_{n=1}^{N}\{f \ge n\}\big) && \text{Monotone Convergence} \\ &= \lim_{N\to\infty} \textstyle\sum_{n=1}^{N} \mu\{f \ge n\} && \text{linearity} \\ &= \textstyle\sum_{n=1}^{\infty} \mu\{f \ge n\}. \end{aligned}$$

A similar argument gives a companion upper bound. Thus the pointwise inequality integrates out to $\sum_{n=1}^{\infty} \mu\{f \ge n\} \le \mu f \le \mu\mathcal{X} + \sum_{n=1}^{\infty} \mu\{f \ge n\}$, from which the asserted equivalence follows. □

Extension of the integral to a larger class of functions

Every function f in $\mathcal{M}$ can be decomposed into a difference $f = f^+ - f^-$ of two functions in $\mathcal{M}^+$, where $f^+(x) := \max\big(f(x), 0\big)$ and $f^-(x) := \max\big(-f(x), 0\big)$. To extend μ from $\mathcal{M}^+$ to a linear functional on $\mathcal{M}$ we should define $\mu f := \mu f^+ - \mu f^-$. This definition works if at least one of μf^+ and μf^- is finite; otherwise we get the dreaded $\infty - \infty$. If both $\mu f^+ < \infty$ and $\mu f^- < \infty$ (or equivalently, f is measurable and $\mu|f| < \infty$) the function f is said to be ***integrable*** or μ-integrable. The linearity property (iii) of Theorem <12> carries over partially to $\mathcal{M}$ if $\infty - \infty$ problems are excluded, although it becomes tedious to handle all the awkward cases involving $\pm\infty$. The constants α and β need no longer be nonnegative. Also if both f and g are integrable and if $f \le g$ then $\mu f \le \mu g$, with obvious extensions to certain cases involving ∞.

<15> **Definition.** *The set of all real-valued, μ-integrable functions in $\mathcal{M}$ is denoted by $\mathcal{L}^1(\mu)$, or $\mathcal{L}^1(\mathcal{X}, \mathcal{A}, \mu)$.*

The set $\mathcal{L}^1(\mu)$ is a vector space (stable under pointwise addition and multiplication by real numbers). The integral μ defines an increasing linear functional on $\mathcal{L}^1(\mu)$, in the sense that $\mu f \ge \mu g$ if $f \ge g$ pointwise. The Monotone Convergence property implies other powerful limit results for functions in $\mathcal{L}^1(\mu)$, as described in Section 5. By restricting μ to $\mathcal{L}^1(\mu)$, we eliminate problems with $\infty - \infty$.

For each f in $\mathcal{L}^1(\mu)$, its $\mathcal{L}^1$ ***norm*** is defined as $\|f\|_1 := \mu|f|$. Strictly speaking, $\|\cdot\|_1$ is only a seminorm, because $\|f\|_1 = 0$ need not imply that f is the zero function—as you will see in Section 6, it implies only that $\mu\{f \neq 0\} = 0$. It is common practice to ignore the small distinction and refer to $\|\cdot\|_1$ as a norm on $\mathcal{L}^1(\mu)$.

<16> **Example.** Let Ψ be a convex, real-valued function on $\mathbb{R}$. The function Ψ is measurable (because $\{\Psi \le t\}$ is an interval for each real t), and for each x_0 in $\mathbb{R}$ there is a constant α such that $\Psi(x) \ge \Psi(x_0) + \alpha(x - x_0)$ for all x (Appendix C).

Let $\mathbb{P}$ be a probability measure, and X be an integrable random variable. Choose $x_0 := \mathbb{P}X$. From the inequality $\Psi(x) \ge -|\Psi(x_0)| - |\alpha|(|x| + |x_0|)$ we deduce that $\mathbb{P}\Psi(X)^- \le |\Psi(x_0)| + |\alpha|(\mathbb{P}|X| + |x_0|) < \infty$. Thus we should have no $\infty - \infty$ worries in taking expectations (that is, integrating with respect to $\mathbb{P}$) to deduce that $\mathbb{P}\Psi(X) \ge \Psi(\mathbb{P}X) + \alpha(\mathbb{P}X - x_0) = \Psi(\mathbb{P}X)$, a result known as ***Jensen's inequality***. One way to remember the direction of the inequality is to note that $0 \le \text{var}(X) = \mathbb{P}X^2 - (\mathbb{P}X)^2$, which corresponds to the case $\Psi(x) = x^2$. □

Integrals with respect to Lebesgue measure

Lebesgue measure $\mathfrak{m}$ on $\mathcal{B}(\mathbb{R})$ corresponds to length: $\mathfrak{m}[a,b] = b - a$ for each interval. I will occasionally revert to the traditional ways of writing such integrals,

$$\mathfrak{m}f = \int f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx \qquad \text{and} \qquad \mathfrak{m}^x\big(f(x)\{a \le x \le b\}\big) = \int_a^b f(x)dx.$$

Don't worry about confusing the Lebesgue integral with the Riemann integral over finite intervals. Whenever the Riemann is well defined, so is the Lebesgue, and the two sorts of integral have the same value. The Lebesgue is a more general concept. Indeed, facts about the Riemann are often established by an appeal to theorems about the Lebesgue. You do not have to abandon what you already know about integration over finite intervals.

The improper Riemann integral, $\int_{-\infty}^{\infty} f(x)\,dx = \lim_{n\to\infty} \int_{-n}^{n} f(x)\,dx$, also agrees with the Lebesgue integral provided $\mathfrak{m}|f| < \infty$. If $\mathfrak{m}|f| = \infty$, as in the case of the function $f(x) := \sum_{n=1}^{\infty}\{n \le x < n+1\}(-1)^n/n$, the improper Riemann integral might exist as a finite limit, while the Lebesgue integral $\mathfrak{m}f$ does not exist.

*4. Construction of integrals from measures

To construct the integral $\tilde{\mu}$ as a functional on $\mathcal{M}^+(\mathcal{X},\mathcal{A})$, starting from a measure μ on the sigma-field $\mathcal{A}$, we use approximation from below by means of simple functions.

First we must define $\tilde{\mu}$ on $\mathcal{M}^+_{\text{simple}}$. The representation of a simple function as a linear combination of indicator functions is not unique, but the additivity properties of the measure μ will let us use any representation to define the integral. For example, if $s := 3A_1 + 7A_2 = 3A_1A_2^c + 10A_1A_2 + 7A_1^cA_2$, then

$$3\mu(A_1) + 7\mu(A_2) = 3\mu(A_1A_2^c) + 10\mu(A_1A_2) + 7\mu(A_1^cA_2).$$

More generally, if $s := \sum_i \alpha_i A_i$ has another representation $s = \sum_j \beta_j B_j$, then $\sum_i \alpha_i \mu A_i = \sum_j \beta_j \mu B_j$. Proof? Thus we can uniquely define $\tilde\mu(s)$ for a simple function $s := \sum_i \alpha_i A_i$ by $\tilde\mu(s) := \sum_i \alpha_i \mu A_i$.

Define the increasing functional $\tilde\mu$ on $\mathcal{M}^+$ by

$$\tilde\mu(f) := \sup\{\tilde\mu(s) : f \geq s \in \mathcal{M}^+_{\text{simple}}\}.$$

That is, the integral of f is a supremum of integrals of nonnegative simple functions less than f.

From the representation of simple functions as linear combinations of disjoint sets in $\mathcal{A}$, it is easy to show that $\tilde\mu(\mathbb{I}_A) = \mu A$ for every A in $\mathcal{A}$. It is also easy to show that $\tilde\mu(0) = 0$, and $\tilde\mu(\alpha f) = \alpha\tilde\mu(f)$ for nonnegative real α, and

<17> $$\tilde\mu(f+g) \geq \tilde\mu(f) + \tilde\mu(g).$$

The last inequality, which is usually referred to as the superadditivity property, follows from the fact that if $f \geq u$ and $g \geq v$, and both u and v are simple, then $f + g \geq u + v$ with $u + v$ simple.

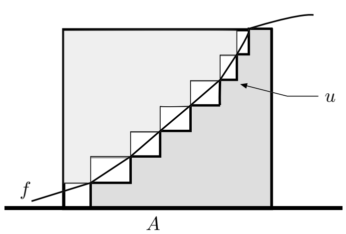
Only the Monotone Convergence property and the companion to <17>,

<18> $$\tilde\mu(f+g) \leq \tilde\mu(f) + \tilde\mu(g),$$

require real work. Here you will see why measurability is needed.

Proof of inequality <18>. Let s be a simple function $\leq f + g$, and let ϵ be a small positive number. It is enough to construct simple functions u, v with $u \leq f$ and $v \leq g$ such that $u + v \geq (1-\epsilon)s$. For then $\tilde\mu f + \tilde\mu g \geq \tilde\mu u + \tilde\mu v \geq (1-\epsilon)\tilde\mu s$, from which the subadditivity inequality <18> follows by taking a supremum over simple functions then letting ϵ tend to zero.

For simplicity of notation I will assume s to be very simple: $s := A$. You can repeat the argument for each A_i in a representation $\sum_i \alpha_i A_i$ with disjoint A_i to get the general result. Suppose $\epsilon = 1/m$ for some positive integer m. Write ℓ_j for j/m. Define simple functions

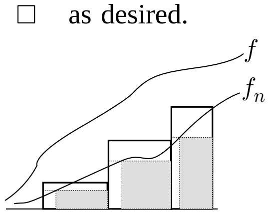


$$u := A\{f \geq 1\} + \sum_{j=1}^m A\left\{\ell_{j-1} \leq f < \ell_j\right\}\ell_{j-1},$$
$$v := \sum_{j=1}^m A\left\{\ell_{j-1} \leq f < \ell_j\right\}\left(1 - \ell_j\right).$$

The measurability of f ensures $\mathcal{A}$-measurability of all the sets entering into the definitions of u and v. For the inequality $v \leq g$, notice that $f + g \geq 1$ on A, so $g > 1 - \ell_j = v$ when $\ell_{j-1} \leq f < \ell_j$ on A. Finally, note that the simple functions were chosen so that

$$u + v = A\{f \geq 1\} + \sum_{j=1}^m A\left\{\ell_{j-1} \leq f < \ell_j\right\}(1-\epsilon) \geq (1-\epsilon)A,$$

□ as desired.



Proof of the Monotone Convergence property. Suppose $f_n \in \mathcal{M}^+$ and $f_n \uparrow f$. Suppose $f \geq s := \sum \alpha_i A_i$, with the A_i disjoint sets in $\mathcal{A}$ and $\alpha_i > 0$. Define approximating simple functions $s_n := \sum_i (1-\epsilon)\alpha_i A_i\{f_n \geq (1-\epsilon)\alpha_i\}$. Clearly $s_n \leq f_n$. The simple function s_n is one of those that enters into the supremum defining $\tilde{\mu} f_n$. It follows that

$$\tilde{\mu} f_n \geq \tilde{\mu}(s_n) = (1-\epsilon) \sum_i \alpha_i \mu \left(A_i \{ f_n \geq (1-\epsilon)\alpha_i \} \right).$$

On the set A_i the functions f_n increase monotonely to f, which is $\geq \alpha_i$. The sets $A_i\{f_n \geq (1-\epsilon)\alpha_i\}$ expand up to the whole of A_i. Countable additivity implies that the μ measures of those sets increase to μA_i. It follows that

$$\lim \tilde{\mu} f_n \geq \limsup \tilde{\mu} s_n \geq (1-\epsilon)\tilde{\mu} s.$$

□ Take a supremum over simple $s \leq f$ then let ϵ tend to zero to complete the proof.

5. Limit theorems

Theorem <13> identified an integral on $\mathcal{M}^+$ as an increasing linear functional with the Monotone Convergence property :

<19>
$$\text{if } 0 \leq f_n \uparrow \text{ then } \mu\left(\lim_{n\to\infty} f_n\right) = \lim_{n\to\infty} \mu f_n.$$

Two direct consequences of this limit property have important applications throughout probability theory. The first, ***Fatou's Lemma***, asserts a weaker limit property for nonnegative functions when the convergence and monotonicity assumptions are dropped. The second, ***Dominated Convergence***, drops the monotonicity and nonnegativity but imposes an extra domination condition on the convergent sequence $\{f_n\}$. I have slowly realized over the years that many simple probabilistic results can be established by Dominated Convergence arguments. The Dominated Convergence Theorem is the Swiss Army Knife of probability theory.

It is important that you understand why some conditions are needed before we can interchange integration (which is a limiting operation) with an explicit limit, as in <19>. Variations on the following example form the basis for many counterexamples.

<20> **Example.** Let μ be Lebesgue measure on $\mathcal{B}[0,1]$ and let $\{\alpha_n\}$ be a sequence of positive numbers. The function $f_n(x) := \alpha_n\{0 < x < 1/n\}$ converges to zero, pointwise, but its integral $\mu(f_n) = \alpha_n/n$ need not converge to zero. For example, $\alpha_n = n^2$ gives $\mu f_n \to \infty$; the integrals diverge. And

$$\alpha_n = \begin{cases} 6n & \text{for } n \text{ even} \\ 3n & \text{for } n \text{ odd} \end{cases} \qquad \text{gives} \qquad \mu f_n = \begin{cases} 6 & \text{for } n \text{ even} \\ 3 & \text{for } n \text{ odd.} \end{cases}$$

□ The integrals oscillate.

<21> **Fatou's Lemma.** *For every sequence $\{f_n\}$ in $\mathcal{M}^+$ (not necessarily convergent),* $\mu(\liminf_{n\to\infty} f_n) \leq \liminf_{n\to\infty} \mu(f_n)$.

Proof. Write f for $\liminf f_n$. Remember what a lim inf means. Define $g_n := \inf_{m\geq n} f_m$. Then $g_n \leq f_n$ for every n and the $\{g_n\}$ sequence increases monotonely to the function f. By Monotone Convergence, $\mu f = \lim_{n\to\infty} \mu g_n$. By the increasing property, $\mu g_n \leq \mu f_n$ for each n, and hence $\lim_{n\to\infty} \mu g_n \leq \liminf_{n\to\infty} \mu f_n$. □