Real Analysis: Part I

William G. Faris

February 2, 2004

Contents

Chapter 1

Mathematical proof

1.1 Logical language

There are many useful ways to present mathematics; sometimes a picture or a physical analogy produces more understanding than a complicated equation. However, the language of mathematical logic has a unique advantage: it gives a standard form for presenting mathematical truth. If there is doubt about whether a mathematical formulation is clear or precise, this doubt can be resolved by converting to this format. The value of a mathematical discovery is considerably enhanced if it is presented in a way that makes it clear that the result and its proof could be stated in such a rigorous framework.

Here is a somewhat simplified model of the language of mathematical logic. There may be *function symbols*. These may be 0-place function symbols, or constants. These stand for objects in some set. Example: 8. Or they may be 1-place functions symbols. These express functions from some set to itself, that is, with one input and one output. Example: square. Or they may be 2-place function symbols. These express functions with two inputs and one output. Example: $+$.

Once the function symbols have been specified, then one can form *terms*. The language also has a collection of variables $x, y, z, x', y', z', \ldots$. Each variable is a term. Each constant c is a term. If t is a term, and f is a 1-place function symbol, then $f(t)$ is a term. If s and t are terms, and g is a 2-place function symbol, then $g(s,t)$ or (sgt) is a term. Example: In an language with constant terms $1, 2, 3$ and 2-place function symbol $+$ the expression $(x + 2)$ is a term, and the expression $(3+(x+2))$ is a term. Note: Sometimes it is a convenient abbreviation to omit outer parentheses. Thus $3 + (x + 2)$ would be an abbreviation for $(3 + (x + 2))$.

The second ingredient is *predicate symbols*. These may be 0-place predicate symbols, or propositional symbols. They may stand for complete sentences. One useful symbol of this nature is $\bot$, which is interpreted as always false. Or they may be 1-place predicate symbols. These express properties. Example: even.

Or they may be 2-place predicate symbols. These express relations. Example: $<$.

Once the terms have been specified, then the *atomic formulas* are specified. A propositional symbol is an atomic formula. If p is a property symbol, and t is a term, then tp is an atomic formula. If s and t are terms, and r is a relation symbol, then srt is an atomic formula. Thus $(x + 2) < 3$ is an atomic formula. Note: This could be abbreviated $x + 2 < 3$.

Finally there are logical symbols $\wedge$, $\vee$, $\Rightarrow$, $\forall$, $\exists$, and parentheses.

Once the atomic formulas are specified, then the other *formulas* are obtained by logical operations. If A and B are formulas, then so are $(A \wedge B)$, $(A \vee B)$, and $(A \Rightarrow B)$. If x is a variable and $A(x)$ is a formula, then so are $\forall x\, A(x)$ and $\exists x\, A(x)$. Thus $\exists x\, x + 2 < 3$ is a formula.

We shall often abbreviate $(A \Rightarrow \bot)$ by $\neg A$. Thus facts about negation will be special cases of facts about implication. In writing a formula, we often omit the outermost parentheses. However this is just an abbreviation.

Another useful abbreviation is $(A \Leftrightarrow B)$ for $((A \Rightarrow B) \wedge (B \Rightarrow A))$.

Some of the logical operations deserve special comment. The implication $A \Rightarrow B$ is also written

if A, then B

A only if B

B if A.

The equivalence $A \Leftrightarrow B$ is also written

A if and only if B.

The *converse* of $A \Rightarrow B$ is $B \Rightarrow A$. The *contrapositive* of $A \Rightarrow B$ is $\neg B \Rightarrow \neg A$.

When A is defined by B, the definition is usually written in the form A if B. It has the logical force of $A \Leftrightarrow B$.

The universal quantified formula $\forall x\, A(x)$ is also written

for all x $A(x)$

for each x $A(x)$

for every x $A(x)$.

The existential quantified formula $\exists x\, A(x)$ is also written

there exists x with $A(x)$

for some x $A(x)$.

Note: Avoid at all cost expressions of the form "for any x $A(x)$." The word "any" does not function as a quantifier in the usual way. For example, if one says "z is special if and only if for any singular x it is the case that x is tied to z", it is not clear which quantifier on x might be intended.

Often a quantifier has a restriction. The restricted universal quantifier is $\forall x\, (C(x) \Rightarrow A(x))$. The restricted existential quantifier is $\exists x\, (C(x) \wedge A(x))$. Here $C(x)$ is a formula that places a restriction on the x for which the assertion is made.

It is common to have implicit restrictions. For example, say that the context of a discussion is real numbers x. There may be an implicit restriction $x \in \mathbf{R}$. Since the entire discussion is about real numbers, it may not be necessary to

make this explicit in each formula. This, instead of $\forall x\,(x \in \mathbf{R} \Rightarrow x^2 \geq 0)$ one would write just $\forall x\, x^2 \geq 0$.

Sometimes restrictions are indicated by use of special letters for the variables. Thus often i, j, k, l, m, n are used for integers. Instead of saying that m is odd if and only if $\exists y\,(y \in \mathbf{N} \wedge m = 2y + 1)$ one would just write that m is odd if and only if $\exists k\, m = 2k + 1$.

The letters ϵ, δ are used for strictly positive real numbers. The corresponding restrictions are $\epsilon > 0$ and $\delta > 0$. Thus instead of writing $\forall x\,(x > 0 \Rightarrow \exists y\,(y > 0 \wedge y < x))$ one would write $\forall \epsilon \exists \delta\, \delta < \epsilon;$.

Other common restrictions are to use f, g, h for functions or to indicate sets by capital letters. Reasoning with restricted variables should work smoothly, provided that one keeps the restriction in mind at the appropriate stages of the argument.

1.2 Free and bound variables

In a formula each occurrence of a variable is either free or bound. The occurrence of a variable x is *bound* if it is in a subformula of the form $\forall x\, B(x)$ or $\exists x\, B(x)$. (There may also be other operations, such as the set builder operation, that produce bound variables.) If the occurrence is not bound, then it is said to be *free*.

In general, a bound variable may be replaced by a new bound variable without changing the meaning of the formula. Thus, for instance, if y' is a variable that does not occur in the formula, one could replace the occurrences of y in the subformula $\forall y\, B(y)$ by y', so the new subformula would now be $\forall y'\, B(y')$. Of course if the variables are restricted, then the change of variable should respect the restriction.

Example: Let the formula be $\exists y\, x < y$. This says that there is a number greater than x. In this formula x is free and y is bound. The formula $\exists y'\, x < y'$ has the same meaning. In this formula x is free and y' is bound. On the other hand, the formula $\exists y\, x' < y$ has a different meaning. This formula says that there is a number greater than x'.

We wish to define *careful substitution* of a term t for the free occurrences of a variable x in $A(x)$. The resulting formula will be denoted $A(t)$ There is no particular problem in defining substitution in the case when the term t has no variables that already occur in $A(x)$. The care is needed when there is a subformula in which y is a bound variable and when the term t contains the variable y. Then mere substitution might produce an unwanted situation in which the y in the term t becomes a bound variable. So one first makes a change of bound variable in the subformula. Now the subformula contains a bound variable y' that cannot be confused with y. Then one substitutes t for the free occurrences of x in the modified formula. Then y will be a free variable after the substitution, as desired.

Example: Let the formula be $\exists y\, x < y$. Say that one wished to substitute $y+1$ for the free occurrences of x. This should say that there is a number greater

than $y + 1$. It would be wrong to make the careless substitution $\exists y\, y + 1 < y$. This statement is not only false, but worse, it does not have the intended meaning. The careful substitution proceeds by first changing the original formula to $\exists y'\, x < y'$. The careful substitution then produces $\exists y'\, y + 1 < y'$. This says that there is a number greater than $y + 1$, as desired.

The general rule is that if y is a variable with bound occurrences in the formula, and one wants to substitute a term t containing y for the free occurrences of x in the formula, then one should change the bound occurrences of y to bound occurrences of a new variable y' before the substitution. This gives the kind of careful substitution that preserves the intended meaning.

1.3 Proofs from analysis

The law of double negation states that $\neg\neg A \Leftrightarrow A$. De Morgan's laws for connectives state that $\neg(A \wedge B) \Leftrightarrow (\neg A \vee \neg B)$ and that $\neg(A \vee B) \Leftrightarrow (\neg A \wedge \neg B)$. De Morgan's laws for quantifiers state that $\neg\forall x\, A(x) \Leftrightarrow \exists x\, \neg A(x)$ and $\neg\exists x\, A(x) \Leftrightarrow \forall x\, \neg A(x)$. Since $\neg(A \Rightarrow B) \Leftrightarrow (A \wedge \neg B)$ and $\neg(A \wedge B) \Leftrightarrow (A \Rightarrow \neg B)$, De Morgan's laws continue to work with restricted quantifiers.

Examples:

1. The function f is continuous if $\forall a \forall \epsilon \exists \delta \forall x (|x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon)$. It is assumed that a, x, ϵ, δ are real numbers with $\epsilon > 0, \delta > 0$.

2. The function f is not continuous if $\exists a \exists \epsilon \forall \delta \exists x (|x - a| < \delta \wedge \neg|f(x) - f(a)| < \epsilon)$. This is a mechanical application of De Morgan's laws.

Similarly, the function f is uniformly continuous if $\forall \epsilon \exists \delta \forall a \forall x (|x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon)$. Notice that the only difference is the order of the quantifiers.

Examples:

1. Consider the proof that $f(x) = x^2$ is continuous. The heart of the proof is to prove the existence of δ. The key computation is $|x^2 - a^2| = |x + a||x - a| = |x - a + 2a||x - a|$. If $|x - a| < 1$ then this is bounded by $(2|a| + 1)|x - a|$.

   Here is the proof. Let $\epsilon > 0$. Suppose $|x - a| < \min(1, \epsilon/(2|a| + 1))$. From the above computation it is easy to see that $|x^2 - a^2| < \epsilon$. Hence $|x - a| < \min(1, \epsilon/(2|a| + 1)) \Rightarrow |x^2 - a^2| < \epsilon$. Since in this last statement x is arbitrary, $\forall x\, (|x - a| < \min(1, \epsilon/(2|a| + 1)) \Rightarrow |x^2 - a^2| < \epsilon)$. Hence $\exists \delta \forall x\, (|x - a| < \delta \Rightarrow |x^2 - a^2| < \epsilon)$. Since $\epsilon > 0$ and a are arbitrary, the final result is that $\forall a \forall \epsilon \exists \delta \forall x\, (|x - a| < \delta \Rightarrow |x^2 - a^2| < \epsilon)$.

2. Consider the proof that $f(x) = x^2$ is not uniformly continuous. Now the idea is to take $x - a = \delta/2$ and use $x^2 - a^2 = (x + a)(x - a) = (2a + \delta/2)(\delta/2)$.

   Here is the proof. With the choice of $x - a = \delta/2$ and with $a = 1/\delta$ we have that $|x - a| < \delta$ and $|x^2 - a^2| \geq 1$. Hence $\exists a \exists x\, (|x - a| < \delta \wedge |x^2 - a^2| \geq 1)$.

Since $\delta > 0$ is arbitrary, it follows that $\forall\delta\exists a\exists x\,(|x-a| < \delta \wedge |x^2-a^2| \geq 1)$. Finally we conclude that $\exists\epsilon\forall\delta\exists a\exists x\,(|x-a| < \delta \wedge |x^2-a^2| \geq \epsilon)$.

It is a general fact that f uniformly continuous implies f continuous. This is pure logic; the only problem is to interchange the $\exists\delta$ quantifier with the $\forall a$ quantifier. This can be done in one direction. Suppose that $\exists\delta\forall a\,A(\delta,a)$. Temporarily suppose that δ' is a name for the number that exists, so that $\forall a A(\delta',a)$. In particular, $A(\delta',a')$. It follows that $\exists\delta A(\delta,a')$. This conclusion does not depend on the name, so it follows from the original supposition. Since a' is arbitrary, it follows that $\forall a\exists\delta\,A(\delta,a)$.

What goes wrong with the converse argument? Suppose that $\forall a\exists\delta\,A(\delta,a)$. Then $\exists\delta\,A(\delta,a')$. Temporarily suppose $A(\delta',a')$. The trouble is that a' is not arbitrary, because something special has been supposed about it. So the generalization is not permitted.

Problems

1. A sequence of functions f_n converges pointwise (on some set of real numbers) to f as n tends to infinity if $\forall x\forall\epsilon\exists N\forall n(n \geq N \Rightarrow |f_n(x)-f(x)| < \epsilon)$. Here the restrictions are that x is in the set and $\epsilon > 0$. Show that for $f_n(x) = x^n$ and for suitable $f(x)$ there is pointwise convergence on the closed interval $[0,1]$.

2. A sequence of functions f_n converges uniformly (on some set of real numbers) to f as n tends to infinity if $\forall\epsilon\exists N\forall x\forall n(n \geq N \Rightarrow |f_n(x)-f(x)| < \epsilon)$. Show that for $f_n(x) = x^n$ and the same $f(x)$ the convergence is not uniform on $[0,1]$.

3. Show that uniform convergence implies pointwise convergence.

4. Show that if f_n converges uniformly to f and if each f_n is continuous, then f is continuous.

   Hint: The first hypothesis is $\forall\epsilon\exists N\forall x\forall n\,(n \geq N \Rightarrow |f_n(x)-f(x)| < \epsilon)$. Deduce that $\exists N\forall x\forall n\,(n \geq N \Rightarrow |f_n(x)-f(x)| < \epsilon'/3)$. Temporarily suppose $\forall x\forall n\,(n \geq N' \Rightarrow |f_n(x)-f(x)| < \epsilon'/3)$.

   The second hypothesis is $\forall n\forall a\forall\epsilon\exists\delta\forall x\,(|x-a| < \delta \Rightarrow |f_n(x)-f_n(a)| < \epsilon)$. Deduce that $\exists\delta\forall x\,(|x-a| < \delta \Rightarrow |f_{N'}(x)-f_{N'}(a)| < \epsilon'/3)$. Temporarily suppose that $\forall x\,(|x-a| < \delta' \Rightarrow |f_{N'}(x)-f_{N'}(a)| < \epsilon'/3)$.

   Suppose $|x-a| < \delta'$. Use the temporary suppositions above to deduce that $|f(x)-f(a)| < \epsilon'$. Thus $|x-a| < \delta' \Rightarrow |f(x)-f(a)| < \epsilon'$. This is well on the way to the desired conclusion. However be cautious: At this point x is arbitrary, but a is not arbitrary. (Why?) Explain in detail the additional arguments to reach the goal $\forall a\forall\epsilon\exists\delta\forall x(|x-a| < \delta \Rightarrow |f(x)-f(a)| < \epsilon)$.

1.4 Natural deduction

The way of formalizing the rules of logic that corresponds most closely to the practice of mathematical proof is *natural deduction*. Natural deduction proofs are constructed so that they may be read from the top down. (On the other hand, to construct a natural deduction proof, it is often helpful to work from the top down and the bottom up and try to meet in the middle.)

In natural deduction each **Suppose** introduces a new hypothesis to the set of hypotheses. Each matching **Thus** removes the hypothesis. Each line is a claim that the formula on this line follows logically from the hypotheses above that have been introduced by a **Suppose** and not yet eliminated by a matching **Thus** .

Here is an example of a natural deduction proof. Say that one wants to show that if one knows the algebraic fact $\forall x\,(x > 0 \Rightarrow (x + 1) > 0)$, then one is forced by pure logic to accept that $\forall y\,(y > 0 \Rightarrow ((y + 1) + 1) > 0)$. Here is the argument, showing every logical step. The comments on the right are not part of the proof.

Suppose $\forall x(x > 0 \Rightarrow (x + 1) > 0)$
  Suppose $z > 0$
  $z > 0 \Rightarrow (z + 1) > 0$ (specialize the hypothesis)
  $(z + 1) > 0$ (from the implication)
  $(z + 1) > 0 \Rightarrow ((z + 1) + 1) > 0$ (specialize the hypothesis again)
  $((z + 1) + 1) > 0$ (from the implication)
Thus $z > 0 \Rightarrow ((z + 1) + 1) > 0$ (introducing the implication)
$\forall y\,(y > 0 \Rightarrow ((y + 1) + 1) > 0$ (generalizing)

Notice that the indentation makes the hypotheses in force at each stage quite clear. On the other hand, the proof could also be written in narrative form. It could go like this.

Suppose that for all x, if $x > 0$ then $(x + 1) > 0$. **Suppose** $z > 0$. By specializing the hypothesis, obtain that if $z > 0$, then $(z + 1) > 0$. It follows that $(z+1) > 0$. By specializing the hypothesis again, obtain that if $(z+1) > 0$, then $((z + 1) + 1) > 0$. It follows that $((z + 1) + 1) > 0$. **Thus** if $z > 0$, then $((z + 1) + 1) > 0$. Since z is arbitrary, conclude that for all y, if $(y > 0$, then $((y + 1) + 1) > 0)$.

Mathematicians usually write in narrative form, but it is useful to practice proofs in outline form, with proper indentation to show the subarguments.

The following pages give the rules for natural deduction. In each rule there is a connective or quantifier that is the center of attention. It may be in the hypothesis or in the conclusion. The rule shows how to reduce an argument involving this logical operation to one without the logical operation. (To accomplish this, the rule needs to be used just once, except in two cases involving the substitution of terms. If it were not for these two exceptions, mathematics would be simple indeed.)

Conjunction rules:

$A \wedge B$
A and in hypothesis
B and in hypothesis

A

B
$A \wedge B$ and in conclusion

Universal rules:

$\forall x\, A(x)$
$A(t)$ all in hypothesis

Note: This rule may be used repeatedly with various terms.

If z is a variable that does not occur free in a hypothesis in force or in $\forall x\, A$, then

$A(z)$
$\forall x\, A(x)$ all in conclusion

Note: The restriction on the variable is usually signalled by an expression such as "since z is arbitrary, conclude $\forall x\, A(x)$."

Implication rules:

$A \Rightarrow B$

A
B implies in hypothesis

Note: This rule by itself is an incomplete guide to practice, since it may not be clear how to prove A. A template that always works is provided below.

Suppose A

B
Thus $A \Rightarrow B$ implies in conclusion

Negation rules:

$\neg A$

A
$\perp$ not in hypothesis

Note: This rule by itself is an incomplete guide to practice, since it may not be clear how to prove A. A template that always works is provided below.

Suppose A

$\perp$
Thus $\neg A$ not in conclusion

Disjunction rules:

$A \vee B$
Suppose A

C
Instead suppose B

C
Thus C or in hypothesis

A
$A \vee B$ or in conclusion

together with
B
$A \vee B$ or in conclusion

Note: This rule by itself is an incomplete guide to practice, since it may not be clear how to prove A or how to prove B. A template that always works is provided below.

Existential rules:

If z is a variable that does not occur free in a hypothesis in force, in $\exists x\, A$, or in C, then

$\exists x\, A(x)$
Suppose $A(z)$

C
Thus C exists in hypothesis

Note: The restriction on the variable could be signalled by an expression such as "since z is arbitrary, conclude C on the basis of the existential hypothesis $\exists x\, A(x)$."

$A(t)$
$\exists x\, A(x)$ exists in conclusion

Note: This rule by itself is an incomplete guide to practice, since it may not be clear how to prove $A(t)$. A template that always works is provided below. This template shows in particular how the rule may be used repeatedly with various terms.

Mathematicians tend not to use the exists in hypothesis rule explicitly. They simply suppose that some convenient variable may be used as a name for the thing that exists. They reason with this name up to a point at which they get a conclusion that no longer mentions it. At this point they just forget about their temporary supposition. One could try to formalize this procedure with a rule something like the following.

Abbreviated existential rule:

If z is a variable that does not occur free in a hypothesis in force, in $\exists x\, A$, or in C, then

$\exists x\, A(x)$
Temporarily suppose $A(z)$

C
From this point on treat C as a consequence of the existential hypothesis without the temporary supposition or its temporary consequences. In case of doubt, it is safer to use the original rule!

The rules up to this point are those of intuitionistic logic. This is a more flexible form of logic with a very interesting interpretation. Mathematicians find proofs of this type to be natural and direct. However in order to get the full force of classical logic one needs one more rule, the rule of contradiction. The section on natural deduction strategies will demonstrate how this rule may be used in a controlled way.

Contradiction rule:

Suppose $\neg C$

$\bot$

Thus C by contradiction

Note: The double negation law says that $\neg\neg A$ is logically equivalent to A. The rule for negation in conclusion and the double negation law immediately give the contradiction rule.

A natural deduction proof is read from top down. However it is often discovered by working simultaneously from the top and the bottom, until a meeting in the middle. The discoverer then obscures the origin of the proof by presenting it from the top down. This is convincing but often not illuminating.

Example: Here is a natural deduction proof of the fact that $\exists x\,(x \text{ happy} \wedge x \text{ rich})$ logically implies that $\exists x\, x \text{ happy} \wedge \exists x\, x \text{ rich}$.

Suppose $\exists x\,(x \text{ happy} \wedge x \text{ rich})$
  Suppose $z \text{ happy} \wedge z \text{ rich}$
  $z \text{ happy}$
  $z \text{ rich}$
  $\exists x\, x \text{ happy}$
  $\exists x\, x \text{ rich}$
  $\exists x\, x \text{ happy} \wedge \exists x\, x \text{ rich}$
Thus $\exists x\, x \text{ happy} \wedge \exists x\, x \text{ rich}$

Here is the same proof in narrative form.

Suppose $\exists x\,(x \text{ happy} \wedge x \text{ rich})$. **Suppose** $z \text{ happy} \wedge z \text{ rich}$. Then z happy and hence $\exists x\, x$ happy. Similarly, z rich and hence $\exists x\, x$ rich. It follows that $\exists x\, x \text{ happy} \wedge \exists x\, x \text{ rich}$. **Thus** (since z is an arbitrary name) it follows that $\exists x\, x \text{ happy} \wedge \exists x\, x \text{ rich}$ on the basis of the original supposition of existence.

Example: Here is a natural deduction proof of the fact that $\exists x\,(x \text{ happy} \wedge x \text{ rich})$ logically implies that $\exists x\, x \text{ happy} \wedge \exists x\, x \text{ rich}$ using the abbreviated existential rule.

Suppose $\exists x\,(x \text{ happy} \wedge x \text{ rich})$
Temporarily suppose $z \text{ happy} \wedge z \text{ rich}$
$z \text{ happy}$
$z \text{ rich}$
$\exists x\, x \text{ happy}$
$\exists x\, x \text{ rich}$
$\exists x\, x \text{ happy} \wedge \exists x\, x \text{ rich}$

Here is the same abbreviated proof in narrative form.

Suppose $\exists x\,(x \text{ happy} \wedge x \text{ rich})$. **Temporarily suppose** $z \text{ happy} \wedge z \text{ rich}$. Then z happy and hence $\exists x\, x$ happy. Similarly, z rich and hence $\exists x\, x$ rich. It follows that $\exists x\, x \text{ happy} \wedge \exists x\, x \text{ rich}$. Since z is an arbitrary name, this conclusion holds on the basis of the original supposition of existence.

Example: Here is a natural deduction proof that $\exists y \forall x\, x \leq y$ gives $\forall x \exists y\, x \leq y$.

Suppose $\exists y \forall x\, x \leq y$
  Suppose $\forall x\, x \leq y'$
  $x' \leq y'$
  $\exists y\, x' \leq y$
Thus $\exists y\, x' \leq y$
$\forall x \exists y\, x \leq y$

Here is the same proof in abbreviated narrative form.

Suppose $\exists y \forall x\, x \leq y$. **Temporarily suppose** $\forall x\, x \leq y'$. In particular, $x' \leq y'$. Therefore $\exists y\, x' \leq y$. In fact, since y' is arbitrary, this follows on

the basis of the original existential supposition. Finally, since x' is arbitrary, conclude that $\forall x \exists y\, x \leq y$.

The following problems are to be done using natural deduction. Indent. Justify every logical step. Each step involves precisely one logical operation. The logical operation must correspond to the logical type of the formula.

Problems

1. Prove that
$$\forall x\, (x \text{ rich} \Rightarrow x \text{ happy}) \Rightarrow (\forall x\, x \text{ rich} \Rightarrow \forall x\, x \text{ happy}). \tag{1.1}$$

2. Suppose
$$\forall z\, z^2 \geq 0, \forall x \forall y\, ((x-y)^2 \geq 0 \Rightarrow (2*(x*y)) \leq (x^2+y^2)). \tag{1.2}$$
Show that it follows logically that
$$\forall x \forall y\, (2*(x*y)) \leq (x^2+y^2). \tag{1.3}$$

3. Show that the hypotheses n odd $\Rightarrow n^2$ odd, n odd $\vee$ n even, $\neg(n^2$ odd $\wedge$ n^2 even) give the conclusion n^2 even $\Rightarrow n$ even.

4. Show that
$$\forall x\, x \text{ happy} \Rightarrow \neg \exists x \neg\, x \text{ happy}. \tag{1.4}$$

5. Show that
$$\forall x \exists y\, (x \text{ likes } y \Rightarrow x \text{ adores } y) \tag{1.5}$$
leads logically to
$$\exists x \forall y\, x \text{ likes } y \Rightarrow \exists x \exists y\, x \text{ adores } y. \tag{1.6}$$

1.5 Natural deduction strategies

A useful strategy for natural deduction is to begin with writing the hypotheses at the top and the conclusion at the bottom. Then work toward the middle. The most important point is to try to use the forall in conclusion rule and the exists in hypothesis rule early in this process of proof construction. This introduces new "arbitrary" variables. Then one uses the forall in hypothesis rule and the exists in conclusion rule with terms formed from these variables. So it is reasonable to use these latter rules later in the proof construction process. They may need to be used repeatedly.

The natural deduction rules as stated above do not have the property that they are reversible. The rules that are problematic are implies in hypothesis, not in hypothesis, or in conclusion, and exists in conclusion. So it is advisable to avoid or postpone the use of these rules.

However there are templates that may be used to overcome this difficulty. These have the advantage that they work in all circumstances.

$A \Rightarrow B$
Suppose $\neg C$

A
B implies in hypothesis

$\bot$
Thus C by contradiction

$\neg A$
Suppose $\neg C$

A
$\bot$ not in hypothesis
Thus C by contradiction

Note: The role of this rule to make use of a negated hypothesis $\neg A$. When the conclusion C has no useful logical structure, but A does, then the rule effectively switches A for C.

Suppose $\neg(A \vee B)$
  Suppose A
  $A \vee B$ or in conclusion
  $\bot$ not in hypothesis
Thus $\neg A$ not in conclusion
  Suppose B
  $A \vee B$ or in hypothesis
  $\bot$ not in hypothesis
Thus $\neg B$ not in conclusion

$\bot$
Thus $A \vee B$ by contradiction

Note: A shortcut is to use the DeMorgan's law that says that $A \vee B$ is logically equivalent to $\neg(\neg A \wedge \neg B)$. So if $\neg A \wedge \neg B$ leads to a contradiction, then conclude $A \vee B$.

Suppose $\neg\exists x\, A(x)$
  Suppose $A(t)$
  $\exists x\, A(x)$ exists in conclusion
  $\bot$ not in hypothesis
Thus $\neg A(t)$ not in conclusion
(may be repeated with various terms)

$\bot$
Thus $\exists x\, A(x)$ by contradiction

Note: A shortcut is to use the quantifier DeMorgan's law that says that $\exists x\, A(x)$ is logically equivalent to $\neg(\forall x\, \neg A(x))$. So if (possibly repeated) use of $\forall x\, \neg A(x)$ leads to a contradiction, then conclude $\exists x\, A(x)$.

Problems

1. Here is a mathematical argument that shows that there is no largest prime number. Assume that there were a largest prime number. Call it a. Then a is prime, and for every number j with $a < j$, j is not prime. However, for every number m, there is a number k that divides m and is prime. Hence there is a number k that divides $a! + 1$ and is prime. Call it b. Now every number $k > 1$ that divides $n! + 1$ must satisfy $n < k$. (Otherwise it would have a remainder of 1.) Hence $a < b$. But then b is not prime. This is a contradiction.

   Use natural deduction to prove that

   $$\forall m \exists k \, (k \text{ prime} \wedge k \text{ divides } m) \tag{1.7}$$

   $$\forall n \forall k \, (k \text{ divides } n! + 1 \Rightarrow n < k) \tag{1.8}$$

   logically imply

   $$\neg \exists n \, (n \text{ prime} \wedge \forall j \, (n < j \Rightarrow \neg \, j \text{ prime})). \tag{1.9}$$

2. It is a well-known mathematical fact that $\sqrt{2}$ is irrational. In fact, if it were rational, so that $\sqrt{2} = m/n$, then we would have $2n^2 = m^2$. Thus m^2 would have an even number of factors of 2, while $2n^2$ would have an odd number of factors of two. This would be a contradiction.

   Use natural deduction to show that

   $$\forall i \, i^2 \text{ even-twos} \tag{1.10}$$

   and

   $$\forall j \, (j \text{ even-twos} \Rightarrow \neg (2 * j) \text{ even-twos}) \tag{1.11}$$

   give

   $$\neg \exists m \exists n \, (2 * n^2) = m^2. \tag{1.12}$$

1.6 Equality

Often equality is thought of as a fundamental logical relation. Manipulations with this concept are very familiar, so there is no need to dwell on it here in detail. However it is worth noting that one could formulate equality rules for natural deduction.

Equality rules:

For a formula $A(z)$ with free variable z substitution of equals is permitted:

$s = t$

$A(s)$
$A(t)$ equality in hypothesis

Everything is equal to itself:

$t = t$ equality in conclusion

Problems

1. If X is a set, then $P(X)$ is the set of all subsets of X. If X is finite with n elements, then $P(X)$ is finite with 2^n elements. A famous theorem of Cantor states that there is no function f from X to $P(X)$ that is onto $P(X)$. Thus in some sense there are more elements in $P(X)$ than in X. This is obvious when X is finite, but the interesting case is when X is infinite.

   Here is an outline of a proof. Consider an arbitrary function f from X to $P(X)$. We want to show that there exists a set V such that for each x in X we have $f(x) \neq V$. Consider the condition that $x \notin f(x)$. This condition defines a set. That is, there exists a set U such that for all x, $x \in U$ is equivalent to $x \notin f(x)$. Call this set S. Let p be arbitrary. Suppose $f(p) = S$. Suppose $p \in S$. Then $p \notin f(p)$, that is, $p \notin S$. This is a contradiction. Thus $p \notin S$. Then $p \in f(p)$, that is, $p \in S$. This is a contradiction. Thus $f(p) \neq S$. Since this is true for arbitrary p, it follows that for each x in X we have $f(x) \neq S$. Thus there is a set that is not in the range of f.

   Prove using natural deduction that from

   $$\exists U \, \forall x \, ((x \in U \Rightarrow \neg x \in f(x)) \wedge (\neg x \in f(x) \Rightarrow x \in U)) \tag{1.13}$$

   one can conclude that

   $$\exists V \, \forall x \, \neg f(x) = V. \tag{1.14}$$

2. Here is an argument that if f and g are continuous functions, then the composite function $g \circ f$ defined by $(g \circ f)(x) = g(f(x))$ is a continuous function.

   Assume that f and g are continuous. Consider an arbitrary point a' and an arbitrary $\epsilon' > 0$. Since g is continuous at $f(a')$, there exists a $\delta > 0$ such that for all y the condition $|y - f(a')| < \delta$ implies that $|g(y) - g(f(a'))| < \epsilon'$. Call it δ_1. Since f is continuous at a', there exists a $\delta > 0$ such that for all x the condition $|x - a'| < \delta$ implies $|f(x) - f(a')| <$

δ_1. Call it δ_2. Consider an arbitrary x'. Suppose $|x' - a'| < \delta_2$. Then $|f(x') - f(a')| < \delta_1$. Hence $|g(f(x')) - g(f(a'))| < \epsilon'$. Thus $|x' - a'| < \delta_2$ implies $|g(f(x')) - g(f(a'))| < \epsilon'$. Since x' is arbitrary, this shows that for all x we have the implication $|x - a'| < \delta_2$ implies $|g(f(x)) - g(f(a'))| < \epsilon'$. It follows that there exists $\delta > 0$ such that all x we have the implication $|x - a'| < \delta$ implies $|g(f(x)) - g(f(a'))| < \epsilon'$. Since ϵ' is arbitrary, the composite function $g \circ f$ is continuous at a'. Since a' is arbitrary, the composite function $g \circ f$ is continuous.

In the following proof the restrictions that $\epsilon > 0$ and $\delta > 0$ are implicit. They are understood because this is a convention associated with the use of the variables ϵ and δ.

Prove using natural deduction that from

$$\forall a \forall \epsilon \exists \delta \forall x \, (|x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon) \tag{1.15}$$

and

$$\forall b \forall \epsilon \exists \delta \forall y \, (|y - b| < \delta \Rightarrow |g(y) - g(b)| < \epsilon) \tag{1.16}$$

one can conclude that

$$\forall a \forall \epsilon \exists \delta \forall x \, (|x - a| < \delta \Rightarrow |g(f(x)) - g(f(a))| < \epsilon). \tag{1.17}$$

1.7 Lemmas and theorems

In statements of mathematical theorems it is common to have implicit universal quantifiers. For example say that we are dealing with real numbers. Instead of stating the theorem that

$$\forall x \forall y \, 2xy \leq x^2 + y^2 \tag{1.18}$$

one simply claims that

$$2uv \leq u^2 + v^2. \tag{1.19}$$

Clearly the second statement is a specialization of the first statement. But it seems to talk about u and v, and it is not clear why this might apply for someone who wants to conclude something about p and q, such as $2pq \leq p^2 + w^2$. Why is this permissible?

The answer is that the two displayed statements are logically equivalent, provided that there is no hypothesis in force that mentions the variables u or v. Then given the second statement and the fact that the variables in it are arbitrary , the first statement is a valid generalization.

Notice that there is no similar principle for existential quantifiers. The statement

$$\exists x \, x^2 = x \tag{1.20}$$

is a theorem about real numbers, while the statement

$$u^2 = u \tag{1.21}$$

is a condition that is true for $u = 0$ or $u = 1$ and false for all other real numbers. It is certainly not a theorem about real numbers. It might occur in a context where there is a hypothesis that $u = 0$ or $u = 1$ in force, but then it would be incorrect to generalize.

One cannot be careless about inner quantifiers, even if they are universal. Thus there is a theorem

$$\exists x \, x < y. \tag{1.22}$$

This could be interpreted as saying that for each arbitrary y there is a number that is smaller than y. Contrast this with the statement

$$\exists x \forall y \, x < y \tag{1.23}$$

with an inner universal quantifier. This is clearly false for the real number system.

The proof rules provided here suffice for every proof in mathematics. This is the famous Gödel completeness theorem. This fact is less useful than one might think, because there is no upper limit to the number of terms that may be used to instantiate a universal hypothesis. Most instantiations are useless, and in complicated circumstances it may be difficult to know the correct one, or even to know that it exists.

In practice, the rules are useful only for the construction of small proofs and for verification of a proof after the fact. The way to make progress in mathematics is find concepts that have meaningful interpretations. In order to prove a major theorem, one prepares by proving smaller theorems or lemmas. Each of these may have a rather elementary proof. But the choice of the statements of the lemmas is crucial in making progress. So while the micro structure of mathematical argument is based on the rules of proof, the global structure is a network of lemmas, theorems, and theories based on astute selection of mathematical concepts.

1.8 More proofs from analysis

One of the most important concepts of analysis is the concept of open set. This makes sense in the context of the real line, or in the more general case of Euclidean space, or in the even more general setting of a metric space. Here we use notation appropriate to the real line, but little change is required to deal with the other cases.

For all subsets V, we say that V is open if $\forall a \, (a \in V \Rightarrow \exists \epsilon \forall x \, (|x - a| < \epsilon \Rightarrow x \in V))$.

Recall the definition of union of a collection Γ of subsets. This says that for all y we have $y \in \bigcup \Gamma$ if and only if $\exists W \, (W \in \Gamma \wedge y \in W)$.

Here is a proof of the theorem that for all collections of subsets Γ the hypothesis $\forall U \, (U \in \Gamma \Rightarrow U \text{ open})$ implies the conclusion $\bigcup \Gamma$ open. The style of the proof is a relaxed form of natural deduction in which some trivial steps are skipped.

Suppose $\forall U\,(U \in \Gamma \Rightarrow U \text{ open})$. **Suppose** $a \in \bigcup\Gamma$. By definition $\exists W(W \in \Gamma \wedge a \in W)$. **Temporarily suppose** $W' \in \Gamma \wedge a \in W'$. Since $W' \in \Gamma$ and $W' \in \Gamma \Rightarrow W'$ open, it follows that W' open. Since $a \in W'$ it follows from the definition that $\exists\epsilon\forall x\,(|x-a| < \epsilon \Rightarrow x \in W')$. **Temporarily suppose** $\forall x\,(|x-a| < \epsilon' \Rightarrow x \in W')$. **Suppose** $|x-a| < \epsilon'$. Then $x \in W'$. Since $W' \in \Gamma \wedge x \in W'$, it follows that $\exists W(W \in \Gamma \wedge x \in W')$. Then from the definition $x \in \bigcup\Gamma$. **Thus** $|x-a| < \epsilon' \Rightarrow x \in \bigcup\Gamma$. Since x is arbitrary, $\forall x\,(|x-a| < \epsilon' \Rightarrow x \in \bigcup\Gamma)$. So $\exists\epsilon\forall x\,(|x-a| < \epsilon \Rightarrow x \in \bigcup\Gamma)$. **Thus** $a \in \bigcup\Gamma \Rightarrow \exists\epsilon\forall x\,(|x-a| < \epsilon \Rightarrow x \in \bigcup\Gamma)$. Since a is arbitrary, $\forall a\,(a \in \bigcup\Gamma \Rightarrow \exists\epsilon\forall x\,(|x-a| < \epsilon \Rightarrow x \in \bigcup\Gamma))$. So by definition $\bigcup\Gamma$ open. **Thus** $\forall U\,(U \in \Gamma \Rightarrow U \text{ open}) \Rightarrow \bigcup\Gamma$ open.

Problems

1. Take the above proof that the union of open sets is open and put it in outline form, with one formula per line. Indent at every **Suppose** line. Remove the indentation at every **Thus** line. (However, do not indent at a **Temporarily suppose** line.)

2. Draw a picture to illustrate the proof in the preceding problem.

3. Prove that for all subsets U, V that $(U \text{ open} \wedge V \text{ open}) \Rightarrow U \cap V$ open. Recall that $U \cap V = \bigcap\{U, V\}$ is defined by requiring that for all y that $y \in U \cap V \Leftrightarrow (y \in U \wedge y \in V)$. It may be helpful to use the general fact that for all $t, \epsilon_1 > 0, \epsilon_2 > 0$ there is an implication $t < \min(\epsilon_1, \epsilon_2) \Rightarrow (t < \epsilon_1 \wedge t < \epsilon_2)$. Use a similar relaxed natural deduction format. Put in outline form, with one formula per line.

4. Draw a picture to illustrate the proof in the preceding problem.

5. Recall that for all functions f, sets W, and elements t we have $t \in f^{-1}[W] \Leftrightarrow f(t) \in W$. Prove that f continuous (with the usual ϵ-δ definition) implies $\forall U\,(U \text{ open} \Rightarrow f^{-1}[U] \text{ open})$. Use a similar relaxed natural deduction format.

6. It is not hard to prove a lemma that says that $\{y \mid |y-b| < \epsilon\}$ open. Use this lemma and the appropriate definitions to prove that $\forall U\,(U \text{ open} \Rightarrow f^{-1}[U] \text{ open})$ implies f continuous. Again present this in relaxed natural deduction format.

Chapter 2

Sets

2.1 Zermelo axioms

Mathematical objects include sets, functions, and numbers. It is natural to begin with sets. If A is a set, the expression

$$t \in A \tag{2.1}$$

can be read simply "t *in* A". Alternatives are "t is a member of A, or "t is an element of A", or "t belongs to A", or "t is in A". The expression $\neg t \in A$ is often abbreviated $t \notin A$ and read "t not in A".

If A and B are sets, the expression

$$A \subset B \tag{2.2}$$

is defined in terms of membership by

$$\forall t \, (t \in A \Rightarrow t \in B). \tag{2.3}$$

This can be read simply "A *subset* B." Alternatives are "A is included in B" or "A is a subset of B". (Some people write $A \subseteq B$ to emphasize that $A = B$ is allowed, but this is a less common convention.) It may be safer to avoid such phrases as "t is contained in A" or "A is contained in B", since here practice is ambiguous. Perhaps the latter is more common.

The following axioms are the starting point for Zermelo set theory. They will be supplemented later with the axiom of infinity and the axiom of choice. These axioms are taken by some to be the foundations of mathematics; however they also serve as a review of important constructions.

Extensionality A set is defined by its members. For all sets A, B

$$(A \subset B \wedge B \subset A) \Rightarrow A = B. \tag{2.4}$$

Empty set Nothing belongs to the empty set.

$$\forall y \, y \notin \emptyset. \tag{2.5}$$

Unordered pair For all objects a, b the unordered pair set $\{a, b\}$ satisfies

$$\forall y \, (y \in \{a, b\} \Leftrightarrow (y = a \vee y = b)). \tag{2.6}$$

Union If Γ is a set of sets, then its union $\bigcup \Gamma$ satisfies

$$\forall x \, (x \in \bigcup \Gamma \Leftrightarrow \exists A \, (A \in \Gamma \wedge x \in A)) \tag{2.7}$$

Power set If X is a set, the power set $P(X)$ is the set of all subsets of X, so

$$\forall A \, (A \in P(X) \Leftrightarrow A \subset X). \tag{2.8}$$

Selection Consider an arbitrary condition $p(x)$ expressed in the language of set theory. If B is a set, then the subset of B consisting of elements that satisfy that condition is a set $\{x \in B \mid p(x)\}$ satisfying

$$\forall y \, (y \in \{x \in B \mid p(x)\} \Leftrightarrow (y \in B \wedge p(y))). \tag{2.9}$$

2.2 Comments on the axioms

Usually in a logical language there is the logical relation symbol $=$ and a number of additional relation symbols and function symbols. The Zermelo axioms could be stated in an austere language in which the only non-logical relation symbol is $\in$, and there are no function symbols. The only terms are variables. While this is not at all convenient, it helps to give a more precise formulation of the selection axiom. The following list repeats the axioms in this limited language. However, in practice the other more convenient expressions for forming terms are used.

Extensionality

$$\forall A \forall B \, (\forall t \, (t \in A \Leftrightarrow t \in B) \Rightarrow A = B). \tag{2.10}$$

The axiom of extensionality says that a set is defined by its members. Thus, if A is the set consisting of the digits that occur at least once in my car's license plate 5373, and if B is the set consisting of the odd one digit prime numbers, then $A = B$ is the same three element set. All that matters are that its members are the numbers 7,3,5.

Empty set

$$\exists N \forall y \, \neg y \in N. \tag{2.11}$$

By the axiom of extensionality there is only one empty set, and in practice it is denoted by the conventional name $\emptyset$.

Unordered pair

$$\forall a \forall b \exists E \forall y \, (y \in E \Leftrightarrow (y = a \vee y = b)). \tag{2.12}$$

By the axiom of extensionality, for each a, b there is only one unordered pair $\{a, b\}$. The unordered pair construction has this name because the order does not matter: $\{a, b\} = \{b, a\}$. Notice that this set can have either one or two elements, depending on whether $a = b$ or $a \neq b$. In the case when it has only one element, it is written $\{a\}$ and is called a singleton set.

If a, b, c are objects, then there is a set $\{a, b, c\}$ defined by the condition that for all y

$$y \in \{a, b, c\} \Leftrightarrow (y = a \vee y = b \vee y = c). \tag{2.13}$$

This is the corresponding unordered triple construction. The existence of this object is easily seen by noting that both $\{a, b\}$ and $\{b, c\}$ exist by the unordered pair construction. Again by the unordered pair construction the set $\{\{a, b\}, \{b, c\}\}$ exists. But then by the union construction the set $\bigcup\{\{a, b\}, \{b, c\}\}$ exists. A similar construction works for any finite number of objects.

Union

$$\forall \Gamma \exists U \forall x \, (x \in U \Leftrightarrow \exists A \, (A \in \Gamma \wedge x \in A)) \tag{2.14}$$

The standard name for the union is $\bigcup \Gamma$. Notice that $\bigcup \emptyset = \emptyset$ and $\bigcup P(X) = X$. A special case of the union construction is $A \cup B = \bigcup\{A, B\}$. This satisfies the property that for all x

$$x \in A \cup B \Leftrightarrow (x \in A \vee x \in B). \tag{2.15}$$

If $\Gamma \neq \emptyset$ is a set of sets, then the *intersection* $\bigcap \Gamma$ is defined by requiring that for all x

$$x \in \bigcap \Gamma \Leftrightarrow \forall A \, (A \in \Gamma \Rightarrow x \in A) \tag{2.16}$$

The existence of this intersection follows from the union axiom and the selection axiom: $\bigcap \Gamma = \{x \in \bigcup \Gamma \mid \forall A \, (A \in \Gamma \Rightarrow x \in A)\}$.

There is a peculiarity in the definition of $\bigcap \Gamma$ when $\Gamma = \emptyset$. If there is a context where X is a set and $\Gamma \subset P(X)$, then we can define

$$\bigcap \Gamma = \{x \in X \mid \forall A \, (A \in \Gamma \Rightarrow x \in A)\}. \tag{2.17}$$

If $\Gamma \neq \emptyset$, then this definition is independent of X and is equivalent to the previous definition. On the other hand, by this definition $\bigcap \emptyset = X$. This might seem strange, since the left hand side does not depend on X. However in most contexts there is a natural choice of X, and this is the definition that is appropriate to such contexts. There is a nice symmetry with the case of union, since for the intersection $\bigcap \emptyset = X$ and $\bigcap P(X) = \emptyset$.

A special case of the intersection construction is $A \cap B = \bigcap\{A, B\}$. This satisfies the property that for all x

$$x \in A \cap B \Leftrightarrow (x \in A \wedge A \in B). \tag{2.18}$$

If $A \subset X$, the *complement* $X \setminus A$ is characterized by saying that for all x

$$x \in X \setminus A \Leftrightarrow (x \in X \wedge x \notin A). \tag{2.19}$$

The existence again follows from the selection axiom: $X \setminus A = \{x \in X \mid x \notin A\}$. Sometimes the complement of A is denoted A^c when the set X is understood.

The constructions $A \cap B$, $A \cup B$, $\bigcap \Gamma$, $\bigcup \Gamma$, and $X \setminus A$ are means of producing objects that have a special relationship to the corresponding logical operations $\wedge, \vee, \forall, \exists, \neg$. A look at the definitions makes this apparent.

Two sets A, B are *disjoint* if $A \cap B = \emptyset$. (In that case it is customary to write the union of A and B as $A \sqcup B$.) More generally, a set $\Gamma \subset P(X)$ of sets is disjoint if for each A in Γ and $B \in \Gamma$ with $A \neq B$ we have $A \cap B = \emptyset$. A *partition* of X is a set $\Gamma \subset P(X)$ such that Γ is disjoint and $\emptyset \notin \Gamma$ and $\bigcup \Gamma = X$.

Power set

$$\forall X \exists P \forall A \, (A \in P \Leftrightarrow \forall t \, (t \in A \Rightarrow t \in X)). \tag{2.20}$$

The power set is the set of all subsets of X, and it is denoted $P(X)$. Since a large set has a huge number of subsets, this axiom has strong consequences for the size of the mathematical universe.

Selection The selection axiom is really an infinite family of axioms, one for each formula $p(x)$ expressed in the language of set theory.

$$\forall B \exists S \forall y \, (y \in S \Leftrightarrow (y \in B \wedge p(y))). \tag{2.21}$$

The selection axiom says that if there is a set B, then one may select a subset $\{x \in B \mid p(x)\}$ defined by a condition expressed in the language of set theory. The language of set theory is the language where the only non-logical relation symbol is $\in$. This is why it is important to realize that in principle the other axioms may be expressed in this limited language. The nice feature is that one can characterize the language as the one with just one non-logical relation symbol. However the fact that the separation axiom is stated in this linguistic way is troubling for one who believes that we are talking about a Platonic universe of sets.

Of course in practice one uses other ways of producing terms in the language, and this causes no particular difficulty. Often when the set B is understood the set is denoted more simply as $\{x \mid p(x)\}$. In the defining condition the quantified variable is implicitly restricted to range over B, so that the defining condition is that for all y we have $y \in \{x \mid p(x)\} \Leftrightarrow p(y)$.

The variables in the set builder construction are bound variables, so, for instance, $\{u \mid p(u)\}$ is the same set as $\{t \mid p(t)\}$.

The famous paradox of Bertrand Russell consisted of the discovery that there is no sensible way to define sets by conditions in a completely unrestricted way. Thus if there were a set $a = \{x \mid x \notin x\}$, then $a \in a$ would be equivalent to $a \notin a$, which is a contradiction.

Say that it is known that for every x in A there is another corresponding object $\phi(x)$ in B. Then another useful notation is

$$\{\phi(x) \in B \mid x \in A \wedge p(x)\}. \tag{2.22}$$

This can be defined to be the set

$$\{y \in B \mid \exists x\, (x \in A \wedge p(x) \wedge y = \phi(x)\}. \tag{2.23}$$

So it is a special case. Again, this is often abbreviated as $\{\phi(x) \mid p(x)\}$ when the restrictions on x and $\phi(x)$ are clear. In this abbreviated notion one could also write the definition as $\{y \mid \exists x\, (p(x) \wedge y = \phi(x))\}$.

Problems

1. Say X has n elements. How many elements are there in $P(X)$?

2. Say X has n elements. Denote the number of subsets of X with exactly k elements by $\binom{n}{k}$. Show that $\binom{n}{0} = 1$ and $\binom{n}{n} = 1$ and that

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}. \tag{2.24}$$

   Use this to make a table of $\binom{n}{k}$ up to $n = 7$.

3. Say that X has n elements. Denote the number of partitions of X into exactly k non-empty disjoint subsets by $S(n, k)$. This is a Stirling number of the second kind. Show that $S(n, 1) = 1$ and $S(n, n) = 1$ and

$$S(n, k) = S(n-1, k-1) + kS(n-1, k). \tag{2.25}$$

   Use this to make a table of $S(n, k)$ up to $n = 5$.

2.3 Ordered pairs and Cartesian product

There is also a very important *ordered pair* construction. If a, b are objects, then there is an object (a, b). This ordered pair has the following fundamental property: For all a, b, p, q we have

$$(a, b) = (p, q) \Leftrightarrow (a = p \wedge b = q). \tag{2.26}$$

If $y = (a, b)$ is an ordered pair, then the first coordinates of y is a and the second coordinate of y is b.

Some mathematicians like to think of the ordered pair (a, b) as the set $(a, b) = \{\{a\}, \{a, b\}\}$. The purpose of this rather artificial construction is to make it a mathematical object that is a set, so that one only needs axioms for sets, and not for other kinds of mathematical objects. However this definition does not play much of a role in mathematical practice.

There are also ordered triples and so on. The ordered triple (a, b, c) is equal to the ordered triple (p, q, r) precisely when $a = p$ and $b = q$ and $c = r$. If $z = (a, b, c)$ is an ordered triple, then the coordinates of z are a, b and c. One can construct the ordered triple from ordered pairs by $(a, b, c) = ((a, b), c)$. The ordered n-tuple construction has similar properties.

There are degenerate cases. There is an ordered 1-tuple (a). If $x = (a)$, then its only coordinate is a. Furthermore, there is an ordered 0-tuple $(\ ) = 0 = \emptyset$.

Corresponding to these constructions there is a set construction called *Cartesian product*. If A, B are sets, then $A \times B$ is the set of all ordered pairs (a, b) with $a \in A$ and $b \in B$. This is a set for the following reason. Let $U = A \cup B$. Then each of $\{a\}$ and $\{a, b\}$ belongs to $P(U)$. Therefore the ordered pair (a, b) belongs to $P(P(U))$. This is a set, by the power set axiom. So by the selection axiom $A \times B = \{(a, b) \in P(P(U)) \mid a \in A \wedge b \in B\}$ is a set.

One can also construct Cartesian products with more factors. Thus $A \times B \times C$ consists of all ordered triples (a, b, c) with $a \in A$ and $b \in B$ and $c \in C$.

The Cartesian product with only one factor is the set whose elements are the (a) with $a \in A$. There is a natural correspondence between this somewhat trivial product and the set A itself. The correspondence is that which associates to each (a) the corresponding coordinate a. The Cartesian product with zero factors is a set $1 = \{0\}$ with precisely one element $0 = \emptyset$.

There is a notion of sum of sets that is dual to the notion of product of sets. This is the *disjoint union* of two sets. The idea is to attach labels to the elements of A and B. Thus, for example, for each element a of A consider the ordered pair $(0, a)$, while for each element b of B consider the ordered pair $(1, b)$. Then even if there are elements common to A and B, their tagged versions will be distinct. Thus the sets $\{0\} \times A$ and $\{1\} \times B$ are disjoint. The disjoint union of A and B is the set $A + B$ such that for all y

$$y \in A + B \Leftrightarrow (y \in \{0\} \times A \vee y \in \{1\} \times B). \tag{2.27}$$

One can also construct disjoint unions with more summands in the obvious way.

2.4 Relations and functions

A *relation* R between sets A and B is a subset of $A \times B$. A *function* (or *mapping* F from A to B is a relation with the following two properties:

$$\forall x \exists y \, (x, y) \in F. \tag{2.28}$$

$$\forall y \forall y' \, (\exists x \, ((x, y) \in F \wedge (x, y') \in F) \Rightarrow y = y'). \tag{2.29}$$

In these statements the variable x is restricted to A and the variables y, y' are restricted to B. A function F from A to B is a *surjection* if

$$\forall y \exists x \, (x, y) \in F. \tag{2.30}$$

A function F from A to B is an *injection* if

$$\forall x \forall x' \, (\exists y \, ((x, y) \in F \wedge (x', y) \in F) \Rightarrow x = x'). \tag{2.31}$$

Notice the same pattern in these definitions as in the two conditions that define a function. As usual, if F is a function, and $(x, y) \in F$, then we write $F(x) = y$.

In this view a function is regarded as being identical with its graph as a subset of the Cartesian product. On the other hand, there is something to be said for a point of view that makes the notion of a function just as fundamental as the notion of set. In that perspective, each function from A to B would have a graph that would be a subset of $A \times B$. But the function would be regarded as an operation with an input and output, and the graph would be a set that is merely one means to describe the function.

There is a useful function builder notation that corresponds to the set builder notation. Say that it is known that for every x in A there is another corresponding object $\phi(x)$ in B. Then another useful notation is

$$[x \mapsto \phi(x) : A \to B] = \{(x, \phi(x)) \in A \times B \mid x \in A\}. \tag{2.32}$$

This is an explicit definition of a function from A to B. This could be abbreviated as $[x \mapsto \phi(x)]$ when the restrictions on x and $\phi(x)$ are clear. The variables in such an expression are of course bound variables, so, for instance, the squaring function $u \mapsto u^2$ is the same as the squaring function $t \mapsto t^2$.

Problems

1. How many functions are there from an n element set to a k element set?

2. How many injective functions are there from an n element set to a k element set?

3. How many surjective functions are there from an n element set to a k element set?

4. Show that $m^n = \sum_{k=0}^{m} \binom{m}{k} k! S(n, k)$.

5. Let $B_n = \sum_{k=0}^{n} S(n, k)$ be the number of partitions of an n element set. Show that B_n is equal to the expected number of functions from an n element set to an m element set, where m has a Poisson probability distribution with mean one. That is, show that

$$B_n = \sum_{m=0}^{\infty} m^n \frac{1}{m!} e^{-1}. \tag{2.33}$$

2.5 Number systems

The axiom of infinity states that there is an infinite set. In fact, it is handy to have a specific infinite set, the set of all natural numbers $\mathbf{N} = \{0, 1, 2, 3, \ldots\}$. The mathematician von Neumann gave a construction of the natural numbers that is perhaps too clever to be taken entirely seriously. He defined $0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$, $3 = \{0, 1, 2\}$, and so on. Each natural number is the set of all its predecessors. Furthermore, the operation s of adding one has a simple definition:

$$s(n) = n \cup \{n\}. \tag{2.34}$$

Thus $4 = 3 \cup \{3\} = \{0, 1, 2\} \cup \{3\} = \{0, 1, 2, 3\}$. Notice that each of these sets representing a natural number is a finite set. There is as yet no requirement that the natural numbers may be combined into a single set.

This construction gives one way of formulating the *axiom of infinity*. Say that a set I is inductive if $0 \in I$ and $\forall n\, (n \in I \Rightarrow s(n) \in I)$. The axiom of infinity says that there exists an inductive set. Then the set $\mathbf{N}$ of natural numbers may be defined as the intersection of the inductive subsets of this set.

According to this definition the natural number system $\mathbf{N}\{0, 1, 2, 3, \ldots\}$ has 0 as an element. It is reasonable to consider 0 as a natural number, since it is a possible result of a counting process. However it is sometimes useful to consider the set of natural numbers with zero removed. In this following we denote this set by by $\mathbf{N}_+ = \{1, 2, 3, \ldots\}$.

According to the von Neuman construction, the natural number n is defined by $n = \{0, 1, 2, \ldots, n-1\}$. This is a convenient way produce an n element index set, but in other contexts it can also be convenient to use $\{1, 2, 3, \ldots, n\}$.

This von Neumann construction is only one way of thinking of the set of natural numbers $\mathbf{N}$. However, once we have this infinite set, it is not difficult to construct a set $\mathbf{Z}$ consisting of all integers $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$. Furthermore, there is a set $\mathbf{Q}$ of rational numbers, consisting of all quotients of integers, where the denominator is not allowed to be zero. The next step after this is to construct the set $\mathbf{R}$ of real numbers. This is done by a process of completion, to be described later. The transition from $\mathbf{Q}$ to $\mathbf{R}$ is the transition from algebra to analysis. The result is that it is possible to solve equations by approximation rather than by algebraic means.

After that, next important number system is $\mathbf{C}$, the set of complex numbers. Each complex number is of the form $a + bi$, where a, b are real numbers, and $i^2 = -1$. Finally, there is $\mathbf{H}$, the set of quaternions. Each quaternion is of the form $t + ai + bj + ck$, where t, a, b, c are real numbers. Here $i^2 = -1, j^2 = -1, k^2 = -1$, $ij = k$, $jk = i$, $ki = j$, $ji = -k$, $kj = -i$, $ik = -j$. A pure quaternion is one of the form $ai + bj + ck$. The product of two pure quaternions is $(ai + bj + ck)(a'i + b'j + c'k) = -(aa' + bb' + cc') + (bc' - cb')i + (ca' - ac')j + (ab' - ba')k$. Thus quaternion multiplication includes both the dot product and the cross product in a single operation.

In summary, the number systems of mathematics are $\mathbf{N}, \mathbf{Z}, \mathbf{Q}, \mathbf{R}, \mathbf{C}, \mathbf{H}$. The systems $\mathbf{N}, \mathbf{Z}, \mathbf{Q}, \mathbf{R}$ each have a natural linear order, and there are natural order

preserving injective functions from $\mathbf{N}$ to $\mathbf{Z}$, from $\mathbf{Z}$ to $\mathbf{Q}$, and from $\mathbf{Q}$ to $\mathbf{R}$. The natural algebraic operations in $\mathbf{N}$ are addition and multiplication. In $\mathbf{Z}$ they are addition, subtraction, and multiplication. In $\mathbf{Q}, \mathbf{R}, \mathbf{C}, \mathbf{H}$ they are addition, subtraction, multiplication, and division by non-zero numbers. In $\mathbf{H}$ the multiplication and division are non-commutative. The number systems $\mathbf{R}, \mathbf{C}, \mathbf{H}$ have the completeness property, and so they are particularly useful for analysis.

Problems

1. A totally ordered set is densely ordered if between every two distinct points there is another point. Thus $\mathbf{Q}$ is densely ordered, and also $\mathbf{R}$ is densely ordered. Show that between every two distinct points of $\mathbf{Q}$ there is a point of $\mathbf{R}$ that is irrational.

2. Is it true that between every two distinct points of $\mathbf{R}$ there is a point of $\mathbf{Q}$? Discuss.

3. Define a map from $\mathbf{R}$ to $P(\mathbf{Q})$ by $j(x) = \{r \in \mathbf{Q} \mid r \leq x\}$. Prove that j is injective.

Chapter 3

Relations, Functions, Dynamical Systems

3.1 Identity, composition, inverse, intersection

A *relation* R between sets A and B is a subset of $A \times B$. In this context one often writes xRy instead of $(x, y) \in R$, and says that x is related to y by the relation. Often a relation between A and A is called a relation on the set A.

There is an important relation I_A on A, namely the *identity* relation consisting of all ordered pairs (x, x) with $x \in A$. That is, for x and y in A, the relation xI_Ay is equivalent to $x = y$.

Given an relation R between A and B and a relation S between B and C, there is a relation $S \circ R$ between A and C called the *composition*. It is defined in such a way that $x(S \circ R)z$ is equivalent to the existence of some y in B such that xRy and ySz. Thus if R relates A to B, and S relates B to C, then $S \circ R$ relates A to C. In symbols,

$$S \circ R = \{(x, z) \mid \exists y \, (xRy \wedge ySz)\}. \tag{3.1}$$

Notice the order in which the factors occur, which accords with the usual convention for functions. For functions it is usual to use such a notation to indicate that R acts first, and then S. This is perhaps not the most natural convention for relations, so in some circumstances it might be convenient to define another kind of composition in which the factors are written in the opposite order.

There are two more useful operations on relations. If R is a relation between A and B, then there is an *inverse* relation R^{-1} between B and A. It consists of all the (y, x) such that (x, y) is in R. That is, $yR^{-1}x$ is equivalent to xRy.

Finally, if R and S are relations between A and B, then there is a relation $R \cap S$. This is also a useful operation. Notice that $R \subset S$ is equivalent to $R \cap S = R$.

Sometimes if $X \subset A$ one writes $R[X]$ for the image of X under R, that is,

$$R[X] = \{y \mid \exists x \, (x \in X \wedge xRy)\}. \tag{3.2}$$

Also, if a is in A, it is common to write $R[a]$ instead of $R[\{a\}]$. Thus y is in $R[a]$ if aRy.

3.2 Picturing relations

There are two common ways of picturing a relation R between A and B. One way is to draw the product space $A \times B$ and sketch the set of points (x, y) in R. This is the *graph* of the relation. The other way is to draw the disjoint union $A + B$ and for each (x, y) in R sketch an arrow from x to y. This is the *cograph* of the relation.

3.3 Equivalence relations

Consider a relation R on A. The relation R is *reflexive* if $I_A \subset R$. The relation R is *symmetric* if $R = R^{-1}$. The relation R is *transitive* if $R \circ R \subset R$. A relation that is reflexive, symmetric, and transitive (RST) is called an *equivalence relation*.

Theorem 3.1 *Consider a set A. Let Γ be a partition of A. Then there is a corresponding equivalence relation E, such that $(x, y) \in E$ if and only if for some subset U in Γ both x in U and y in U. Conversely, for every equivalence relation E on A there is a unique partition Γ of A that gives rise to the relation in this way.*

The sets in the partition defined by the equivalence relation are called the *equivalence classes* of the relation.

Problems

1. Show that a relation is reflexive if and only if $\forall x\, xRx$.
2. Show that a relation is symmetric if and only if $\forall x \forall y\, (xRy \Rightarrow yRx)$.
3. Here are two possible definitions of a transitive relation. This first is $\forall x \forall y \forall z\, ((xRy \wedge yRz) \Rightarrow xRz)$. The second is $\forall x \forall z\, (\exists y(xRy \wedge yRz) \Rightarrow xRz)$. Which is correct? Discuss.

3.4 Generating relations

Theorem 3.2 *For every relation R on A, there is a smallest transitive relation R^T such that $R \subset R^T$. This is the transitive relation generated by R.*

Theorem 3.3 *For every relation R on A, there is a smallest symmetric and transitive relation R^{ST} such that $R \subset R^{ST}$. This is the symmetric and transitive relation generated by R.*

Theorem 3.4 *For every relation R on A, there is a smallest equivalence relation $E = R^{RST}$ such that $R \subset E$. This is the equivalence relation generated by R.*

Proof: The proofs of these theorems all follow the same pattern. Here is the proof of the last one. Let R be a relation on A, that is, let R be a subset of $A \times A$. Let Δ be the set of all equivalence relations S with $R \subset S$. Then since $A \times A \in \Delta$, it follows that Δ is non-empty. Let $E = \bigcap \Delta$. Now note three facts. The intersection of a set of transitive relations is transitive. The intersection of a set of symmetric relations is symmetric. The intersection of a set of reflexive relations is reflexive. It follows that E is transitive, reflexive, and symmetric. This is the required equivalence relation. □

This theorem shows that by specifying a relation R one also specifies a corresponding equivalence relation E. This can be a convenient way of describing an equivalence relation.

3.5 Ordered sets

A relation R on A is *antisymmetric* if $R \cap R^{-1} \subset I_A$. This just says that $\forall x \forall y \left((x \leq y \wedge y \leq x) \Rightarrow x = y \right)$. A *ordering* of A is a relation that is reflexive, antisymmetric, and transitive (RAT). Ordered sets will merit further study. Here is one theorem about how to describe them.

Theorem 3.5 *Consider a relation R such that there exists an order relation S with $R \subset S$. Then there exists a smallest order relation $P = R^{RT}$ with $R \subset P$.*

Proof: Let R be a relation on A that is a subset of some order relation. Let Δ be the set of all such order relations S with $R \subset S$. . By assumption $\Delta \neq \emptyset$. Let $P = \bigcap \Delta$. Argue as in the case of an equivalence relation. A subset of an antisymmetric relation is antisymmetric. (Note that for an non-empty set of sets the intersection is a subset of the union.) The relation P is the required order relation. □

The above theorem gives a convenient way of specifying an order relation P. For example, if A is finite, then P is generated by the successor relation R.

A *linearly ordered* (or *totally ordered*) set is an ordered set such that the order relation satisfies $R \cup R^{-1} = A \times A$. Thus just says that $\forall x \forall y \left(x \leq y \vee y \leq x \right)$. A *well-ordered set* is a linearly ordered set with the property that each non-empty subset has a least element.

A *rooted tree* is an ordered set with a least element, the root, such that for each point in the set, the elements below the point form a well-ordered set.

3.6 Functions

A relation F from A to B is a *total relation* if $I_A \subset F^{-1} \circ F$. It is a *partial function* if $F \circ F^{-1} \subset I_B$. It is a *function* if it is both a total relation and a partial function (that is, it is a total function).

A function F is an *injective function* if it is a function and F^{-1} is a partial function. A function F is a *surjective function* if it is a function and also F^{-1} is a total relation. It is a bijective function if it is both an injective function and a surjective function. For a bijective function F the inverse relation F^{-1} is a function from B to A, in fact a bijective function.

Problems

1. Let F be a function. Describe $F^T[a]$ (the forward orbit of a under F).
2. Let F be a function. Describe $F^{RT}[a]$ (the orbit of a under F).
3. Let F be a function. Is it possible that $F^T[a] = F^{RT}[a]$? Discuss in detail.

3.7 Relations inverse to functions

Lemma 3.6 *Let F be a relation that is a function from A to B, and let F^{-1} be the inverse relation. Then the sets $F^{-1}[b]$ for b in the range of F form a partition of A, and $F^{-1}[b] = \emptyset$ for b not in the range of F. If V is a subset of B, then $F^{-1}[V]$ is the union of the disjoint sets $F^{-1}[b]$ for b in V.*

This lemma immediately gives the following remarkable and important theorem.

Theorem 3.7 *Let F be a relation that is a function from A to B, and let F^{-1} be the inverse relation. Then F^{-1} respects the set operations of union, intersection, and complement. Thus:*

1. *If Γ is a set of subsets of B, then $F^{-1}[\bigcup \Gamma] = \bigcup \{F^{-1}[V] \mid V \in \Gamma\}$.*
2. *If Γ is a set of subsets of B, then $F^{-1}[\bigcap \Gamma] = \bigcap \{F^{-1}[V] \mid V \in \Gamma\}$.*
3. *If V is a subset of B, then $F^{-1}[B \setminus V] = A \setminus F^{-1}[V]$.*

3.8 Dynamical systems

Consider a function F from A to A. Such a function is often called a *dynamical system*. Thus if a is the present state of the system, at the next stage the state is $f(a)$, and at the following stage after that the state is $f(f(a))$, and so on.

The orbit of a point a in A is $F^{RT}[a]$, the image of a under the relation F^{RT}. This is the entire future history of the system (including the present), when it is started in the state a. Each orbit S is invariant under F, that is, $F[S] \subset S$. If b is in the orbit of a, then we say that a leads to b.

The simplest way to characterize the orbit of a is as the set $\{a, f(a), f(f(a)), f(f(f(a))), \ldots\}$, that is, the set of $f^{(n)}(a)$ for $n \in \mathbf{N}$, where $f^{(n)}$ is the nth iterate of f. (The nth iterate of f is the composition of f with itself n times.)

Theorem 3.8 *Let $F : A \to A$ be a function. Each orbit of a under F is either finite and consists of a sequence of points that eventually enters a periodic cycle, or it is an infinite sequence of distinct points.*

In the finite case the orbit may be described as having the form of a lasso. Special cases of the lasso are a cycle and a single point.

3.9 Picturing dynamical systems

Since a dynamical system is a function $F : A \to A$, there is a peculiarity that the domain and the target are the same space. However this gives a nice way of picturing orbits.

One method is to plot the graph of F as a subset of $A \times A$, and use this to describe the dynamical system as acting on the diagonal. For each x in the orbit, start with the point (x, x) on the diagonal. Draw the vertical line from (x, x) to $(x, f(x))$ on the graph, and then draw the horizontal line from $(x, f(x))$ to $(f(x), f(x))$ back on the diagonal. This process gives a broken line curve that gives a picture of the dynamical system acting on the diagonal.

A method that is more compatible with the cograph point of view is to look at the set A and draw an arrow from x to $f(x)$ for each x in the orbit.

3.10 Structure of dynamical systems

Let $F : A \to A$ be a function. Then A is a disjoint union of equivalence classes under the equivalence relation F^{RST} generated by F. The following theorem gives a more concrete way of thinking about this equivalence relation.

Theorem 3.9 *Let $F : A \to A$ be a function. Say that aEb if and only if the orbit of a under F has a non-empty intersection with the orbit of b under F. Then E is an equivalence relation, and it is the equivalence relation generated by F.*

Proof: To show that E is an equivalence relation, it is enough to show that it is reflexive, symmetric, and transitive. The first two properties are obvious. To prove that it is transitive, consider points a, b, c with aEb and bEc. Then there are m, n with $f^{(m)}(a) = f^{(n)}(b)$ and there are r, s with $f^{(r)}(b) = f^{(s)}(c)$. Suppose that $n \leq r$. Then $f^{(m+r-n)}(a) = f^{(r)}(b) = f^{(s)}(c)$. Thus in that case aEc. Instead suppose that $r \leq n$. A similar argument shows that aEc. Thus it follows that aEc.

It is clear that E is an equivalence relation with $F \subset E$. Let E' be an arbitrary equivalence relation with $F \subset E'$. Say that aEb. Then there is a c with $aF^{RT}c$ and $bF^{RT}c$. Then $aE'c$ and $bE'c$. Since E' is an equivalence relation, it follows that $cE'b$ and hence $aE'b$. So $E \subset E'$. This shows that E is the smallest equivalence relation E' with $F \subset E'$. That is, E is the equivalence relation generated by F. $\square$

Each equivalence class of a dynamical system F is invariant under F. Thus to study a dynamical system one needs only to look at what happens on each equivalence class.

One can think of a dynamical system as reversible if the function is bijective, as conservative if the function is injective, and as dissipative in the general case. The following theorem describes the general case. There are two possibilities. Either there is eventually stabilization at a periodic cycle. Or the dissipation goes on forever.

Theorem 3.10 *Let $F : A \to A$ be a function. Then on each equivalence class F acts in one of two possible ways. Case 1. Each point in the class has a finite orbit. In this case there is a unique cycle with some period $n \geq 1$ included in the class. Furthermore, the class itself is partitioned into n trees, each rooted at a point of the cycle, such that the points in each tree lead to the root point without passing through other points of the cycle. Case 2. Each point in the class has an infinite orbit. Then the points that lead to a given point in the class form a tree rooted at the point.*

Proof: If a and b are equivalent, then they each lead to some point c. If a leads to a cycle, then c leads to a cycle. Thus b leads to a cycle. So if one point in the equivalence class leads to a cycle, then all points lead to a cycle. There can be only one cycle in an equivalence class.

In this case, consider a point r on the cycle. Say that a point leads directly to r if it leads to r without passing through other points on the cycle. The point r together with the points that lead directly to r form a set $T(r)$ with r as the root. A point q in $T(r)$ is said to be below a point p in $T(r)$ when p leads to q. There cannot be distinct points p, q on $T(r)$ with q below p and p below q, since then there would be another cycle. Therefore $T(r)$ is an ordered set. If p is in $T(r)$, the part of $T(r)$ below p is a finite linearly ordered set, so $T(r)$ is a tree. Each point a in the equivalence class leads directly to a unique point r on the cycle. It follows that the trees $T(r)$ for r in the cycle form a partition of the equivalence class.

The other case is when each point in the class has an infinite orbit. There can be no cycle in the equivalence class. Consider a point r in the class. The same kind of argument as in the previous case shows that the set $T(r)$ of points that lead to r is a tree. □

The special case of conservative dynamical systems given by an injective function is worth special mention. In that case there can be a cycle, but no tree can lead to the cycle. In the case of infinite orbits, the tree that leads to a point has only one branch (infinite or finite).

Corollary 3.11 *Let $F : A \to A$ be an injective function. Then on each equivalence class F acts either like a shift on $\mathbf{Z}_n$ for some $n \geq 1$ (a periodic cycle) or a shift on $\mathbf{Z}$ or a right shift on $\mathbf{N}$.*

The above theorem shows exactly how an injection F can fail to be a bijection. A point p is not in the range of F if and only if it is an initial point for one of the right shifts.

Finally, the even more special case of a reversible dynamical systems given by a bijective function is worth recording. In that case there can be a cycle, but no tree can lead to the cycle. In the case of infinite orbits, the tree that leads to a point has only one branch, and it must be infinite.

Corollary 3.12 *Let $F : A \to A$ be a bijective function. Then on each equivalence class F acts either like a shift on $\mathbf{Z}_n$ for some $n \geq 1$ (a periodic cycle) or a shift on $\mathbf{Z}$.*

A final corollary of this last result is that every permutation of a finite set is a product of disjoint cycles.

The following problems use the concept of cardinal number. A countable infinite set has cardinal number ω_0. A set that may be placed in one-to-one correspondence with an interval of real numbers has cardinal number c.

Problems

1. My social security number is 539681742. This defines a function defined on 123456789. It is a bijection from a nine point set to itself. What are the cycles? How many are they? How many points in each cycle?

2. Let $f : \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x + 1$. What are the equivalence classes? How many are they (cardinal number)? How many points in each equivalence class (cardinal number)?

3. Let $f : \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = 2\arctan(x)$. (Recall that the derivative of $f(x)$ is $f'(x) = 2/(1 + x^2) > 0$, so f is strictly increasing.) What is the range of f? How many points are there in the range of f (cardinal number)? What are the equivalence classes? How many are there (cardinal number)? How many points in each equivalence class (cardinal number)? Hint: It may help to use a calculator or draw graphs.

4. Let $f : A \to A$ be an injection with range $R \subset A$. Let R' be a set with $R \subset R' \subset A$. Show that there is an injection $j : A \to A$ with range R'. Hint: Use the structure theorem for injective functions.

5. Bernstein's theorem. Let $g : A \to B$ be an injection, and let $h : B \to A$ be an injection. Prove that there is a bijection $k : A \to B$. Hint: Use the result of the previous problem.

Chapter 4

Functions, Cardinal Number

4.1 Functions

A *function* (or *mapping*) $f : A \to B$ with *domain* A and *target* (or *codomain*) B assigns to each element x of A a unique element $f(x)$ of B.

The set of values $f(x)$ for x in A is called the *range* of f or the image of A under f. In general for $S \subset A$ the set $f[S]$ of values $f(x)$ in B for x in A is called the *image* of S under f. On the other hand, for $T \subset B$ the set $f^{-1}[T]$ consisting of all x in A with $f(x)$ in T is the *inverse image* of T under f. In this context the notation f^{-1} does not imply that f has an inverse function; instead it refers to the inverse relation.

The function is *injective* (or one-to-one) if $f(x)$ uniquely determines x, and it is *surjective* (or onto) if each element of B is an $f(x)$ for some x, that is, the range is equal to the target. The function is *bijective* if it is both injective and surjective. In that case it has an *inverse function* $f^{-1} : B \to A$.

If $f : A \to B$ and $g : B \to C$ are functions, then the *composition* $g \circ f : A \to C$ is defined by $(g \circ f)(x) = g(f(x))$ for all x in A.

Say that $r : A \to B$ and $s : B \to A$ are functions and that $r \circ s = I_B$, the identity function on B. That is, say that $r(s(b)) = b$ for all b in B. In this situation when r is a left inverse of s and s is a right inverse of r, the function r is called a *retraction* and the function s is called a *section*.

Theorem 4.1 *If r has a right inverse, then r is a surjection.*

Theorem 4.2 *If s has a left inverse, then s is an injection.*

Theorem 4.3 *Suppose $s : B \to A$ is an injection. Assume that $B \neq \emptyset$. Then there exists a function $r : A \to B$ that is a left inverse to s.*

Suppose $r : A \to B$ is a surjection. The *axiom of choice* says that there is a function s that is a right inverse to r. Thus for every b in b there is a set of

x with $r(x) = b$, and since s is a surjection, each such set is non-empty. The function s makes a choice $s(b)$ of an element in each set.

4.2 Picturing functions

Each function $f : A \to B$ has a *graph* which is a subset of $A \times B$ and a *cograph* illustrated by the disjoint union $A + B$ and an arrow from each element of A to the corresponding element of B.

Sometimes there is a function $f : I \to B$, where I is an index set or parameter set that is not particularly of interest. Then the function is called a *parameterized family* of elements of B. In that case it is common to draw the image of I under f as a subset of B.

Another situation is when there is a function $f : A \to J$, where J is an index set. In that case it might be natural to call A a *classified set*. The function induces a partition of A. In many cases these partitions may be called contour sets. Again it is common to picture a function through its contour sets.

4.3 Indexed sums and products

Let A be a set-valued function defined on an index set I. Then the union of A is the union of the range of A and is written $\bigcup_{t\in I} A_t$. Similarly, when $I \neq \emptyset$ the intersection of A is the intersection of the range of A and is written $\bigcap_{t\in I} A_t$.

Let A be a set-valued function defined on an index set I. Let $S = \bigcup_{t\in I} A_t$. The disjoint union or sum of A is

$$\sum_{t\in I} A_t = \{(t, a) \in I \times S \mid a \in A_t\}. \tag{4.1}$$

For each $j \in I$ there is a natural mapping $[a \mapsto (j, a) : A_j \to \sum_t A_t]$. This is the injection of the jth summand into the disjoint union. Notice that the disjoint union may be pictured as something like the union, but with the elements labelled to show where they come from.

Similarly, there is a natural Cartesian product of A given by

$$\prod_{t\in I} A_t = \{f \in S^I \mid \forall t\, f(t) \in A_t\}. \tag{4.2}$$

For each j in I there is a natural mapping $[f \mapsto f(j) : \prod_t A_t \to A_j]$. This is the projection of the product onto the jth factor. The Cartesian product should be thought of as a kind of rectangular box in a high dimensional space, where the dimension is the number of points in the index set I. The jth side of the box is the set A_j.

Theorem 4.4 *The product of an indexed family of non-empty sets is non-empty.*

This theorem is another version of the axiom of choice. Suppose that each $A_t \neq \emptyset$. The result says that there is a function f such that for each t it makes an arbitrary choice of an element $f(t) \in A_t$.

Proof: Define a function $r : \sum_{t \in I} A_t \to I$ by $r((t, a)) = t$. Thus r takes each point in the disjoint union and maps it to its label. The condition that each $A_t \neq \emptyset$ guarantees that r is a surjection. By the axiom of choice r has a right inverse s with $r(s(t)) = t$ for all t. Thus s takes each label into some point of the disjoint union corresponding to that label. Let $f(t)$ be the second component of the ordered pair $s(t)$. Then $f(t) \in A_t$. Thus f takes each label to some point in the set corresponding to that label. □

Say that f is a function such that $f(t) \in A_t$ for each $t \in I$. Then the function may be pictured as a single point in the product space $\prod_{t \in I} A_t$. This geometric picture of a function as a single point in a space of high dimension is a powerful conceptual tool.

4.4 Cartesian powers

The set of all functions from A to B is denoted B^A. In the case when $A = I$ is an index set, the set B^I is called a *Cartesian power*. This is the special case of Cartesian product when the indexed family of sets always has the same value B. This is a common construction in mathematics. For instance, $\mathbf{R}^n$ is a Cartesian power.

Write $2 = \{0, 1\}$. Each element of 2^A is the *indicator function* of a subset of A. There is a natural bijective correspondence between the 2^A and $P(A)$. If χ is an element of 2^A, then $\chi^{-1}[1]$ is a subset of A. On the other hand, if X is a subset of A, then the indicator function 1_X that is 1 on X and 0 on $A \setminus X$ is an element of 2^A.

4.5 Cardinality

Say that a set A is *countable* if A is empty or if there is a surjection $f : \mathbf{N} \to A$.

Theorem 4.5 *If A is countable, then there is an injection from $A \to \mathbf{N}$.*

Proof: This can be proved without the axiom of choice. For each $a \in A$, define $g(a)$ to be the least element of $\mathbf{N}$ such that $f(g(a)) = a$. Then g is the required injection. □

There are sets that are not countable. For instance, $P(\mathbf{N})$ is such a set. This follows from the following theorem.

Theorem 4.6 (Cantor) *Let X be a set. There is no surjection from X to $P(X)$.*

The proof that follows is a diagonal argument. Suppose that $f : X \to P(X)$. Form an array of ordered pairs (a, b) with a, b in X. One can ask whether

$b \in f(a)$ or $b \notin f(a)$. The trick is to look at the diagonal $a = b$ and construct the set of all a where $a \notin f(a)$.

Proof: Assume that $f : X \to P(X)$. Let $S = \{x \in X \mid x \notin f(x)\}$. Suppose that S were in the range of f. Then there would be a point a in X with $f(a) = S$. Suppose that $a \in S$. Then $a \notin f(a)$. But this means that $a \notin S$. This is a contradiction. Thus $a \notin S$. This means $a \notin f(a)$. Hence $a \in S$. This is a contradiction. Thus S is not in the range of f. □

One idea of Cantor was to associate to each set A, finite or infinite, a cardinal number $\#A$. The important thing is that if there is a bijection between two sets, then they have the same cardinal number. If there is no bijection, then the cardinal numbers are different. That is, the statement that $\#A = \#B$ means simply that there is a bijection from A to B.

The two most important infinite cardinal numbers are $\omega_0 = \#\mathbf{N}$ and $c = \#P(\mathbf{N})$. The Cantor theorem shows that these are different cardinal numbers.

If there is an injection $f : A \to B$, then it is natural to say that $\#A \leq \#B$. Thus, for example, it is easy to see that $\omega_0 \leq c$. In fact, by Cantor's theorem $\omega_0 < c$. The following theorem was proved in an earlier chapter as an exercise.

Theorem 4.7 (Bernstein) *If there is an injection $f : A \to B$ and there is an injection $g : B \to A$, then there is a bijection $h : A \to B$.*

It follows from Bernstein's theorem that $\#A \leq \#B$ and $\#B \leq \#A$ together imply that $\#A = \#B$. This result gives a way of calculating the cardinalities of familiar sets.

Theorem 4.8 *The set $\mathbf{N}^2 = \mathbf{N} \times \mathbf{N}$ has cardinality ω_0.*

Proof: It is sufficient to construct a bijection $f : \mathbf{N}^2 \to \mathbf{N}$. Let

$$f(m, n) = \frac{r(r+1)}{2} + m, \quad r = m + n. \tag{4.3}$$

The inverse function $g(s)$ is given by finding the largest value of $r \geq 0$ with $r(r+1)/2 \leq s$. Then $m = s - r(r+1)/2$ and $n = r - m$. Clearly $0 \leq m$. Since $s < (r+1)(r+2)/2$, it follows that $m < r + 1$, that is, $m \leq r$. Thus also $0 \leq n$. There is a lovely picture that makes this all obvious and that justifies the expression "diagonal argument". □

Corollary 4.9 *A countable union of countable sets is countable.*

Proof: Let Γ be a countable collection of countable sets. Then there exists a surjection $u : \mathbf{N} \to \Gamma$. For each $S \in \Gamma$ there is a non-empty set of surjections from $\mathbf{N}$ to S. By the axiom of choice, there is a function that assigns to each S in Γ a surjection $v_S : \mathbf{N} \to S$. Let $w(m, n) = v_{u(m)}(n)$. Then v is a surjection from $\mathbf{N}^2$ to $\bigcup \Gamma$. It is a surjection because each element q of $\bigcup \Gamma$ is an element of some S in Γ. There is an m such that $u(m) = S$. Furthermore, there is an n such that $v_S(n) = q$. It follows that $w(m, n) = q$. However once we have the surjection $w : \mathbf{N}^2 \to \bigcup \Gamma$ we also have a surjection $\mathbf{N} \to \mathbf{N}^2 \to \bigcup \Gamma$. □

Theorem 4.10 *The set* $\mathbf{Z}$ *of integers has cardinality* ω_0.

Proof: There is an obvious injection from $\mathbf{N}$ to $\mathbf{Z}$. On the other hand, there is also a surjection $(m, n) \mapsto m - n$ from $\mathbf{N}^2$ to $\mathbf{Z}$. There is a bijection from $\mathbf{N}$ to $\mathbf{N}^2$ and hence a surjection from $\mathbf{N}$ to $\mathbf{Z}$. Therefore there is an injection from $\mathbf{Z}$ to $\mathbf{N}$. This proves that $\#\mathbf{Z} = \omega_0$. □

Theorem 4.11 *The set* $\mathbf{Q}$ *of rational numbers has cardinality* ω_0.

Proof: There is an obvious injection from $\mathbf{Z}$ to $\mathbf{Q}$. On the other hand, there is also a surjection from $\mathbf{Z}^2$ to $\mathbf{Q}$ given by $(m, n) \mapsto m/n$ when $n \neq 0$ and $(m, 0) \mapsto 0$. There is a bijection from $\mathbf{Z}$ to $\mathbf{Z}^2$. (Why?) Therefore there is a surjection from $\mathbf{Z}$ to $\mathbf{Q}$. It follows that there is an injection from $\mathbf{Q}$ to $\mathbf{Z}$. (Why?) This proves that $\#\mathbf{Q} = \omega_0$. □

Theorem 4.12 *The set* $\mathbf{R}$ *of real numbers has cardinality* c.

Proof: First we give an injection $f : \mathbf{R} \to P(\mathbf{Q})$. In fact, we let $f(x) = \{q \in \mathbf{Q} \mid q \leq x\}$. This maps each real number x to a set of rational numbers. If $x < y$ are distinct real numbers, then there is a rational number r with $x < r < y$. This is enough to establish that f is an injection. From this it follows that there is an injection from $\mathbf{R}$ to $P(\mathbf{N})$.

Recall that there is a natural bijection between $P(\mathbf{N})$ (all sets of natural numbers) and $2^{\mathbf{N}}$ (all sequences of zeros and ones). For the other direction, we give an injection $g : 2^{\mathbf{N}} \to \mathbf{R}$. Let

$$g(s) = \sum_{n=0}^{\infty} \frac{2s_n}{3^{n+1}}. \tag{4.4}$$

This maps $2^{\mathbf{N}}$ as an injection with range equal to the Cantor middle third set. This completes the proof that $\#\mathbf{R} = c$. □

Theorem 4.13 *The set* $\mathbf{R}^{\mathbf{N}}$ *of infinite sequences of real numbers has cardinality* c.

Proof: Map $\mathbf{R}^{\mathbf{N}}$ to $(2^{\mathbf{N}})^{\mathbf{N}}$ to $2^{\mathbf{N}\times\mathbf{N}}$ to $2^{\mathbf{N}}$. □

Problems

1. What is the cardinality of the set $\mathbf{N}^{\mathbf{N}}$ of all infinite sequences of natural numbers? Prove that your answer is correct.

2. What is the cardinality of the set of all finite sequences of natural numbers? Prove that your answer is correct.

3. Define the function $g : 2^{\mathbf{N}} \to \mathbf{R}$ by

$$g(s) = \sum_{n=0}^{\infty} \frac{2s_n}{3^{n+1}}. \tag{4.5}$$

Prove that it is an injection.

4. Define the function $g : 2^{\mathbf{N}} \to \mathbf{R}$ by

$$g(s) = \sum_{n=0}^{\infty} \frac{s_n}{2^{n+1}}. \tag{4.6}$$

   What is its range? Is it an injection?

5. Let A be a set and let $f : A \to A$ be a function. Then f is a relation on A that generates an equivalence relation. Can there be uncountably many equivalence classes? Explain. Can an equivalence class be uncountable? Explain. What is the situation if the function is an injection? How about if it is a surjection?

Chapter 5

Ordered sets and completeness

5.1 Ordered sets

The main topic of this chapter is ordered sets and order completeness. The motivating example is the example of the set P of rational numbers r such that $0 \leq r \leq 1$. Consider the subset S of rational numbers r that also satisfy $r^2 < 1/2$. The upper bounds of S consist of rational numbers s that also satisfy $s^2 > 1/2$. (There is no rational number whose square is $1/2$.) There is no least upper bound of S.

Contrast this with the example of the set L of real numbers x such that $0 \leq x \leq 1$. Consider the subset T of real numbers x that also satisfy $x^2 < 1/2$. The upper bounds of T consists of real numbers y that also satisfy $y^2 \geq 1/2$. The number $\sqrt{2}$ is the least upper bound of T. So know whether you have an upper bound of T is equivalent to knowing whether you have an upper bound of $\sqrt{2}$. As far as upper bounds are concerned, the set T is represented by a single number.

Completeness is equivalent to the existence of least upper bounds. This is the property that says that there are no missing points in the ordered set. The theory applies to many other ordered sets other than the rational and real number systems. So it is worth developing in some generality.

An *ordered set* is a set P and a binary relation $\leq$ that is a subset of $P \times P$. The order relation $\leq$ must satisfy the following properties:

1. $\forall p\, p \leq p$ (reflexivity)
2. $\forall p \forall q \forall r((p \leq q \wedge q \leq r) \Rightarrow p \leq r)$ (transitivity)
3. $\forall p \forall q\, ((p \leq q \wedge q \leq p) \Rightarrow p = q)$. (antisymmetry)

An ordered set is often called a *partially ordered set* or a *poset*. In an ordered set we write $p < q$ if $p \leq q$ and $p \neq q$. Once we have one ordered set, we have

many related order sets, since each subset of an ordered set is an ordered set in a natural way.

In an ordered set we say that p, q are *comparable* if $p \leq q$ or $q \leq p$. An ordered set is *totally ordered* (or *linearly ordered*) if each two points are comparable. (Sometime a totally ordered set is also called a *chain*.)

Examples:

1. The number systems $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}$, and $\mathbf{R}$ are totally ordered sets.
2. Let I be a set and let P be an ordered set. Then P^I with the pointwise ordering is an ordered set.
3. In particular, $\mathbf{R}^I$, the set of all real functions on I, is an ordered set.
4. In particular, $\mathbf{R}^n$ is an ordered set.
5. If X is a set, the power set $P(X)$ with the subset relation is an ordered set.
6. Since $2 = \{0, 1\}$ is an ordered set, the set 2^X with pointwise ordering is an ordered set. (This is the previous example in a different form.)

Let S be a subset of P. We write $p \leq S$ to mean $\forall q\,(q \in S \Rightarrow p \leq q)$. In this case we say that p is a *lower bound* for S. Similarly, $S \leq q$ means $\forall p\,(p \in S \Rightarrow p \leq q)$. Then q is an *upper bound* for S.

We write $\uparrow S$ for the set of all upper bounds for S. Similarly, we write $\downarrow S$ for the set of all lower bounds for S. If $S = \{r\}$ consists of just one point we write the set of upper bounds for r as $\uparrow r$ and the set of lower bounds for r as $\downarrow r$.

An element p of S is the *least* element of S if $p \in S$ and $p \leq S$. Equivalently, $p \in S$ and $S \subset \uparrow p$. An element q of S is the *greatest* element of S if $q \in S$ and $S \leq q$. Equivalently, $q \in S$ and $S \subset \downarrow q$.

An element p of S is a *minimal* element of S if $\downarrow p \cap S = \{p\}$. An element q of S is a *maximal* element of S if $\uparrow p \cap S = \{p\}$.

Theorem 5.1 *If p is the least element of S, then p is a minimal element of S. If q is the greatest element of S, then a is a maximal element of S.*

In a totally ordered set a minimal element is a least element and a maximal element is a greatest element.

5.2 Order completeness

A point p is the *infimum* or *greatest lower bound* of S if $\downarrow S = \downarrow p$. The infimum of S is denoted $\inf S$ or $\bigwedge S$. A point q is the *supremum* or *least upper bound* of S if $\uparrow S = \uparrow q$. The supremum of S is denoted $\sup S$ or $\bigvee S$. The reader should check that $p = \inf S$ if and only if p is the greatest element of $\downarrow S$. Thus $p \in \downarrow S$

and $\downarrow S \leq p$. Similarly, $q = \sup S$ if and only if q is the least element of $\uparrow S$. Thus $q \in \uparrow S$ and $q \leq \uparrow S$.

An ordered set L is a *lattice* if every pair of points p, q has an infimum $p \wedge q$ and a supremum $p \vee q$. An ordered set L is a *complete lattice* if every subset S of L has an infimum $\bigwedge S$ and a supremum $\bigvee S$. The most important example of a totally ordered complete lattice is the closed interval $[-\infty, +\infty]$ consisting of all extended real numbers. An example that is not totally ordered is the set $P(X)$ of all subsets of a set X. In this case the infimum is the intersection and the supremum is the union.

Examples:

1. If $[a, b] \subset [-\infty, +\infty]$ is a closed interval, then $[a, b]$ is a complete lattice.
2. Let I be a set and let P be a complete lattice. Then P^I with the pointwise ordering is a complete lattice.
3. In particular, $[a, b]^I$, the set of all extended real functions on I with values in the closed interval $[a, b]$ is an complete lattice.
4. In particular, $[a, b]^n$ is a complete lattice.
5. If X is a set, the power set $P(X)$ with the subset relation is a complete lattice.
6. Since $2 = \{0, 1\}$ is a complete lattice, the set 2^X with pointwise ordering is a complete lattice. (This is the previous example in a different form.)

Problems

1. Show that $S \neq \emptyset$ implies $\inf S \leq \sup S$.
2. Show that $\sup S \leq \inf T$ implies $S \leq T$ (every element of S is $\leq$ every element of T).
3. Show that $S \leq T$ implies $\sup S \leq \inf T$.

5.3 Sequences in a complete lattice

Let $r : \mathbf{N} \to L$ be a sequence of points in a complete lattice L. Let $s_n = \sup_{k \geq n} r_k$. Then the decreasing sequence s_n itself has an infimum. Thus there is an element

$$\limsup_{k \to \infty} r_k = \inf_n \sup_{k \geq n} r_k. \tag{5.1}$$

Similarly, the increasing sequence $s_n = \inf_{k \geq n} r_k$ has a supremum, and there is always an element

$$\liminf_{k \to \infty} r_k = \sup_n \inf_{k \geq n} r_k. \tag{5.2}$$

It is not hard to see that $\liminf_{k \to \infty} r_k \leq \limsup_{k \to \infty} r_k$.

The application of this construction to the extended real number system is discussed in a later section. However here is another situation where it is important. This situation is quite common in probability. Let Ω be a set, and let $P(\Omega)$ be the set of all subsets. Now sup and inf are union and intersection. Let $A : \mathbf{N} \to P(\Omega)$ be a sequence of subsets. Then $\liminf_{k\to\infty} A_k$ and $\limsup_{k\to\infty} A_k$ are subsets of Ω, with the first a subset of the second. The interpretation of the first one is that a point $\omega \in \liminf_{k\to\infty} A_k$ if and only if ω is eventually in the sets A_k as k goes to infinity. The interpretation of the second one is ω is in $\limsup_{k\to\infty} A_k$ if and only if ω is in A_k infinitely often as k goes to infinity.

5.4 Order completion

For each subset S define its downward closure as $\downarrow\uparrow S$. These are the points that are below every upper bound for S. Thus $S \subset \downarrow\uparrow S$, that is, S is a subset of its downward closure. A subset A is a lower Dedekind cut if it is its own downward closure: $A = \downarrow\uparrow A$. This characterizes a lower Dedekind cut A by the property that if a point is below every upper bound for A, then it is in A.

Lemma 5.2 *For each subset S the subset $\downarrow S$ is a lower Dedekind cut. In fact $\downarrow\uparrow\downarrow S = \downarrow S$.*

Proof: Since for all sets T we have $T \subset \downarrow\uparrow T$, it follows by taking $T = \downarrow S$ that $\downarrow S \subset \downarrow\uparrow\downarrow S$. Since for all sets $S \subset T$ we have $\downarrow T \subset \downarrow S$, we can take $T = \uparrow\downarrow S$ and get $\downarrow\uparrow\downarrow S \subset \downarrow S$. □

Theorem 5.3 *If L is an ordered set in which each subset has a supremum, then L is a complete lattice.*

Proof: Let S be a subset of L. Then $\downarrow S$ is another subset of L. Let r be the supremum of $\downarrow S$. This says that $\uparrow\downarrow S = \uparrow r$. It follows that $\downarrow\uparrow\downarrow S = \downarrow\uparrow r$. This is equivalent to $\downarrow S = \downarrow r$. Thus r is the infimum of S. □

Theorem 5.4 *A lattice L is complete if and only if for each lower Dedekind cut A there exists a point p with $A = \downarrow p$.*

Proof: Suppose L is complete. Let A be a lower Dedekind cut and p be the infimum of $\uparrow A$. Then $\downarrow\uparrow A = \downarrow p$. Thus $A = \downarrow p$.

On the other hand, suppose that for every lower Dedekind cut A there exists a point p with $A = \downarrow p$. Let S be a subset. Then $\downarrow S$ is a lower Dedekind cut. It follows that $\downarrow S = \downarrow p$. Therefore p is the infimum of S. □

The above theorem might justify the following terminology. Call a lower Dedekind cut a virtual point. Then the theorem says that a lattice is complete if and only if every virtual point is given by a point. This is the sense in which order completeness says that there are no missing points.

Theorem 5.5 *Let P be an ordered set. Let L be the ordered set of all subsets of P that are lower Dedekind cuts. The ordering is set inclusion. Then L is a complete lattice. Furthermore, the map $p \mapsto \downarrow p$ is an injection from P to L that preserves the order relation.*

Proof: To show that L is a complete lattice, it is sufficient to show that every subset Γ of L has a supremum. This is not so hard: the supremum is the downward closure of $\bigcup \Gamma$. To see this, we must show that for every lower Dedekind cut B we have $\downarrow\uparrow \bigcup \Gamma \subset B$ if and only if for every A in Γ we have $A \subset B$. The only if part is obvious from the fact that each $A \subset \bigcup \Gamma \subset \downarrow\uparrow \bigcup \Gamma$. For the if part, suppose that $A \subset B$ for all A in Γ. Then $\bigcup A \subset B$. It follows that $\downarrow\uparrow \bigcup A \subset \downarrow\uparrow B = B$. The properties of the injection are easy to verify. □

Examples:

1. Here is a simple example of an ordered set that is not a lattice. Let P be the ordered set four points. There are elements b, c each below each of x, y. Then P is not complete. The reason is that if $S = \{b, c\}$, then $\downarrow S = \emptyset$ and $\uparrow S = \{x, y\}$.

2. Here is an example of a completion of an ordered set. Take the previous example. The Dedekind lower cuts are $A = \emptyset$, $B = \{b\}$, $C = \{c\}$, $M = \{b, c\}$, $X = \{b, c, x\}$, $Y = \{b, c, y\}$, $Z = \{b, c, x, y\}$. So the completion L consists of seven points A, B, C, M, X, Y, Z. This lattice is complete. For example, the set $\{B, C\}$ has infimum A and supremum M.

5.5 The Knaster-Tarski fixed point theorem

Theorem 5.6 (Knaster-Tarski) *Let L be a complete lattice and $f : L \to L$ be an increasing function. Then f has a fixed point a with $f(a) = a$.*

Proof: Let $S = \{x \mid f(x) \leq x\}$. Let $a = \inf S$. Since a is a lower bound for S, it follows that $a \leq x$ for all x in S. Since f is increasing, it follows that $f(a) \leq f(x) \leq x$ for all x in S. It follows that $f(a)$ is a lower bound for S. However a is the greatest lower bound for S. Therefore $f(a) \leq a$.

Next, since f is increasing, $f(f(a)) \leq f(a)$. This says that $f(a)$ is in S. Since a is a lower bound for S, it follows that $a \leq f(a)$. □

5.6 The extended real number system

The extended real number system $[-\infty, +\infty]$ is a complete lattice. In fact, one way to construct the extended real number system is to define it as the order completion of the ordered set $\mathbf{Q}$ of rational numbers. That is, the definition of the extended real number system is as the set of all lower Dedekind cuts of rational numbers. (Note that in many treatments Dedekind cuts are defined

in a slightly different way, so that they never have a greatest element. The definition used here seems most natural in the case of general lattices.)

The extended real number system is a totally ordered set. It follows that the supremum of a set $S \subset [-\infty, +\infty]$ is the number p such that $S \leq p$ and for all $a < p$ there is an element q of S with $a < q$. There is a similar characterization of infimum.

Let $s : \mathbf{N} \to [-\infty, +\infty]$ be a sequence of extended real numbers. Then s is said to be increasing if $m \leq n$ implies $s_m \leq s_n$. For an increasing sequence the limit exists and is equal to the supremum. Similarly, for a decreasing sequence the limit exists and is equal to the infimum.

Now consider an arbitrary sequence $r : \mathbf{N} \to [-\infty, \infty]$. Then $\limsup_{k\to\infty} r_k$ and $\liminf_{k\to\infty} r_k$ are defined.

Theorem 5.7 *If* $\liminf_{k\to\infty} r_k = \limsup_{k\to\infty} r_k = a$*, then* $\lim_{k\to\infty} r_k = a$.

Theorem 5.8 *If* $r : \mathbf{N} \to \mathbf{R}$ *is a Cauchy sequence, then* $\liminf_{k\to\infty} r_k = \limsup_{k\to\infty} r_k = a$*, where* a *is in* $\mathbf{R}$*. Hence in this case* $\lim_{k\to\infty} r_k = a$*. Every Cauchy sequence of real numbers converges to a real number.*

This result shows that the order completeness of $[-\infty, +\infty]$ implies the metric completeness of $\mathbf{R}$.

Problems

1. Let L be a totally ordered complete lattice. Show that p is the supremum of S if and only if p is an upper bound for S and for all $r < p$ there is an element q of S with $r < q$.

2. Let L be a complete lattice. Suppose that p is the supremum of S. Does it follow that for all $r < p$ there is an element q of S with $r < q$? Give a proof or a counterexample.

3. Let S_n be the set of symmetric real n by n matrices. Each A in S_n defines a real quadratic form $x \mapsto x^T Ax : \mathbf{R}^n \to \mathbf{R}$. Here x^T is the row vector that is the transpose of the column vector x. Since the matrix A is symmetric, it is its own transpose: $A^T = A$. The order on S_n is the pointwise order defined by the real quadratic forms. Show that S_2 is not a lattice. Hint: Let P be the matrix with 1 in the upper left corner and 0 elsewhere. Let Q be the matrix with 1 in the lower right corner and 0 elsewhere. Let $I = P + Q$. Show that $P \leq I$ and $Q \leq I$. Show that if $P \vee Q$ exists, then $P \vee Q = I$. Let W be the symmetric matrix that is $4/3$ on the diagonal and $2/3$ off the diagonal. Show that $P \leq W$ and $Q \leq W$, but $I \leq W$ is false.

4. Let $L = [0, 1]$ and let $f : L \to L$ be an increasing function. Can a fixed point be found by iteration? Discuss.

Chapter 6

Metric spaces

6.1 Metric space notions

A *metric space* is a set M together with a function $d : M \times M \to [0, +\infty)$ with the following three properties:

1. For all x, y we have $d(x, y) = 0$ if and only if $x = y$
2. For all x, y, z we have $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality)
3. For all x, y we have $d(x, y) = d(y, x)$.

Proposition 6.1 *For all x, y, z we have $|d(x, z) - d(y, z)| \leq d(x, y)$.*

Proof: From the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$ we obtain $d(x, z) - d(y, z) \leq d(x, y)$. On the other hand, from the triangle inequality we also have $d(y, z) \leq d(y, x) + d(x, z)$ which implies $d(y, z) - d(x, z) \leq d(y, x) = d(x, y)$. □

In a metric space M the *open ball* centered at x of radius $\epsilon > 0$ is defined to be $B(x, \epsilon) = \{y \mid d(x, y) < \epsilon\}$. The *closed ball* centered at x of radius $\epsilon > 0$ is defined to be $\overline{B(x, \epsilon)} = \{y \mid d(x, y) \leq \epsilon\}$. The *sphere* centered at x of radius $\epsilon > 0$ is defined to be $S(x, \epsilon) = \{y \mid d(x, y) = \epsilon\}$.

6.2 Normed vector spaces

One common way to get a metric is to have a *norm* on a vector space. A *norm* on a real vector space V is a function from V to $[0, +\infty)$ with the following three properties:

1. For all x we have $\|x\| = 0$ if and only if $x = 0$.
2. For all x, y we have $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).
3. For all x and real t we have $\|tx\| = |t|\|x\|$.

The corresponding metric is then $d(x, y) = \|x - y\|$.

The classic example, of course, is Euclidean space $\mathbf{R}^n$ with the usual square root of sum of squares norm. In the following we shall see that this ℓ_n^2 norm is just one possibility among many.

6.3 Spaces of finite sequences

Here are some possible metrics on $\mathbf{R}^n$. The most geometrical metric is the ℓ_n^2 metric given by the ℓ_n^2 norm. This is $d_2(x, y) = \|x - y\|_2 = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$. It is the metric with the nicest geometric properties. A sphere in this metric is a nice round sphere.

Sometimes in subjects like probability one wants to look at the sum of absolute values instead of the sum of squares. The ℓ_n^1 metric is $d_1(x, y) = \|x - y\|_1 = \sum_{k=1}^n |x_k - y_k|$. A sphere in this metric is actually a box with corners on the coordinate axes.

In other areas of mathematics it is common to look at the biggest or worst case. The ℓ_n^∞ metric is $d_\infty(x, y) = \|x - y\|_\infty = \max_{1 \le k \le n} |x_k - y_k|$. A sphere in this metric is a box with the flat sides on the coordinate axes.

Comparisons between these metrics are provided by

$$d_\infty(x, y) \le d_2(x, y) \le d_1(x, y) \le n d_\infty(x, y). \tag{6.1}$$

The only one of these comparisons that is not immediate is $d_2(x, y) \le d_1(x, y)$. But this follows from $d_2(x, y) \le \sqrt{d_1(x, y) d_\infty(x, y)} \le d_1(x, y)$.

6.4 Spaces of infinite sequences

The ℓ^2 metric is defined on the set of all infinite sequences such that $\|x\|_2^2 = \sum_{k=1}^\infty |x_k|^2 < \infty$. The metric is $d_2(x, y) = \|x - y\|_2 = \sqrt{\sum_{k=1}^\infty (x_k - y_k)^2}$. This is again a case with wonderful geometric properties. It is a vector space with a norm called real Hilbert space. The fact that the norm satisfies the triangle inequality is the subject of the following digression.

Lemma 6.2 (Schwarz inequality) *Suppose the inner product of two real sequences is to be defined by*

$$\langle x, y \rangle = \sum_{k=1}^\infty x_k y_k. \tag{6.2}$$

If the two sequences x, y are in ℓ^2, then this inner product is absolutely convergent and hence well-defined, and it satisfies

$$|\langle x, y \rangle| \le \|x\|_2 \|y\|_2. \tag{6.3}$$

This well-known lemma says that if we define the cosine of the angle between two non-zero vectors by $\langle x, y\rangle = \|x\|_2\|y\|_2 \cos(\theta)$, then $-1 \le \cos(\theta) \le 1$, and so the cosine has a reasonable geometrical interpretation. If we require the angle to satisfy $0 \le \theta \le \pi$, then the angle is also well-defined and makes geometrical sense.

The Schwarz inequality is just what is needed to prove the triangle inequality. The calculation is

$$\|x+y\|_2^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \le \|x\|_2^2 + 2\|x\|_2\|y\|_2 + \|y\|_2^2 = (\|x\|_2 + \|y\|_2)^2. \tag{6.4}$$

The ℓ^1 metric is defined on the set of all infinite sequences x with $\|x\|_1 = \sum_{k=1}^{\infty} |x_k| < \infty$. The metric is $d_1(x, y) = \sum_{k=1}^{\infty} |x_k - y_k|$. This is the natural distance for absolutely convergence sequences. It is again a vector space with a norm. In this case it is not hard to prove the triangle inequality for the norm using elementary inequalities.

The ℓ^∞ metric is defined on the set of all bounded sequences. The metric is $d_\infty(x, y) = \sup_{1 \le k < \infty} |x_k - y_k|$. That is, $d_\infty(x, y)$ is the least upper bound (supremum) of the $|x_k - y_k|$ for $1 \le k < \infty$. This is yet one more vector space with a norm. The fact that this is a norm requires a little thought. The point is that for each k we have $|x_k + y_k| \le |x_k| + |y_k| \le \|x\|_\infty + \|y\|_\infty$, so that $\|x\|_\infty + \|y\|_\infty$ is an upper bound for the set of numbers $|x_k + y_k|$. Since $\|x+y\|_\infty$ is the least upper bound for these numbers, we have $\|x + y\|_\infty \le \|x\|_\infty + \|y\|_\infty$.

Comparisons between two of these these metrics are provided by

$$d_\infty(x, y) \le d_2(x, y) \le d_1(x, y). \tag{6.5}$$

As sets $\ell^1 \subset \ell^2 \subset \ell^\infty$. This is consistent with the fact that every absolutely convergent sequence is bounded. The proof of these inequalities is almost the same as for the finite-dimensional case. However the ℓ^∞ metric is defined by a supremum rather than by a maximum. So to bound it, one finds an upper bound for the set of all $|x_k - y_k$ and then argues that $\|x - y\|_\infty$ is the least such upper bound.

Yet another possibility is to try to define a metric on the product space $\mathbf{R}^{\mathbf{N}}$ of all sequences of real numbers. We shall often refer to this space as $\mathbf{R}^\infty$. This is the biggest possible metric space of sequences. In order to do this, it is helpful to first define a somewhat unusual metric on $\mathbf{R}$ by $d_b(s, t) = |s-t|/(1+|s-t|)$. We shall see below that the space $\mathbf{R}$ with this new metric is uniformly equivalent to the space $\mathbf{R}$ with its usual metric $d(s, t) = |s-t|$. However d_b has the advantage that it is a metric that is bounded by one.

The metric on $\mathbf{R}^\infty$ is $d_p(x, y) = \sum_{k=1}^{\infty} \frac{1}{2^k} d_b(x_k, y_k)$. This is called the product metric. The comparison between these metrics is given by the inequality

$$d_p(x, y) \le d_\infty(x, y). \tag{6.6}$$

The d_p metric is an example of a metric on a vector space that is not given by a norm.

In all these examples the sequences have been indexed by $\mathbf{N}_+ = \mathbf{N} \setminus \{0\}$. There are variations in which the index set is $\mathbf{N}$ or even $\mathbf{Z}$. All that really matters is that it is a countable infinite set.

6.5 Spaces of bounded continuous functions

Here is another example. Let X be a set. Let $B(X)$ be the set of all bounded functions on X. If f and g are two such functions, then $|f - g|$ is bounded, and so we can define the uniform metric

$$d_{\sup}(f, g) = \|f - g\|_{\sup} = \sup_s |f(s) - g(s)|. \tag{6.7}$$

This again is a normed vector space.

Suppose that X is a metric space. let $C(X)$ be the set of real continuous functions on X. Let $BC(X)$ be the set of bounded continuous real functions on X. This is the appropriate metric space for formulating the concept of uniform convergence of a sequence of continuous functions to a continuous function. Thus, the uniform convergence of f_n to g as $n \to \infty$ is equivalent to the condition $\lim_{n\to\infty} d_{\sup}(f_n, g) = 0$.

It should be remarked that all these examples have complex versions, where the only difference is that sequences of real numbers are replaced by sequences of complex numbers. So there is a complex Hilbert space, a space of bounded continuous complex functions, and so on.

6.6 Open and closed sets

A subset U of a metric space M is *open* if $\forall x \, (x \in U \Rightarrow \exists \epsilon \, B(x, \epsilon) \subset U)$. The following results are well-known facts about open sets.

Theorem 6.3 *Let Γ be a set of open sets. Then $\bigcup \Gamma$ is open.*

Proof: Let x be a point in $\bigcup \Gamma$. Then there exists some S in Γ such that $x \in S$. Since S is open there exists $\epsilon > 0$ with $B(x, \epsilon) \subset S$. However $S \subset \bigcup \Gamma$. So $B(x, \epsilon) \subset \bigcup \Gamma$. Hence $\bigcup \Gamma$ is open. □

Notice that $\bigcup \emptyset = \emptyset$, so the empty set is open.

Theorem 6.4 *Let Γ be a finite set of open sets. Then $\bigcap \Gamma$ is open.*

Proof: Let x be a point in $\bigcap \Gamma$. Then x is in each of the sets S_k in Γ. Since each set S_k is open, for each S_k there is an $\epsilon_k > 0$ such that $B(x, \epsilon_k) \subset S_k$. Let ϵ be the minimum of the ϵ_k. Since Γ is finite, this number $\epsilon > 0$. Furthermore, $B(x, \epsilon) \subset S_k$ for each k. It follows that $B(x, \epsilon) \subset \bigcap \Gamma$. Hence $\bigcap \Gamma$ is open. □

Notice that under our conventions $\bigcap \emptyset = M$, so the entire space M is open. A subset F of a metric space is *closed* if $\forall x \, (\forall \epsilon B(x, \epsilon) \cap F \neq \emptyset \Rightarrow x \in F)$. Here are some basic facts about closed sets.

Theorem 6.5 *The closed subsets are precisely the complements of the open subsets.*

Proof: Let U be a set and $F = M \setminus U$ be its complement. Then $x \in U \Rightarrow \exists \epsilon \, B(x, \epsilon) \subset U$ is logically equivalent to $\forall \epsilon \neg B(x, \epsilon) \subset U \Rightarrow x \notin U$. But this says $\forall \epsilon \, B(x, \epsilon) \cap F \neq \emptyset \Rightarrow x \in F$. From this it is evident that F is closed precisely when U is open. □

Theorem 6.6 *A set F in a metric space is an closed subset if and only if every convergent sequence $s : \mathbf{N} \to M$ with values $s_n \in F$ has limit $s_\infty \in F$.*

Proof: Suppose that F is closed. Let s be a convergent sequence with $s_n \in F$ for each n. Let $\epsilon > 0$. Then for n sufficiently large $d(s_n, s_\infty) < \epsilon$, that is, $s_n \in B(s_\infty, \epsilon)$. This shows that $B(s_\infty, \epsilon) \cap F \neq \emptyset$. Since $\epsilon > 0$ is arbitrary, it follows that $s_\infty \in F$.

For the other direction, suppose that F is not closed. Then there is a point $x \notin F$ such that $\forall \epsilon B(x, \epsilon) \cap F \neq \emptyset$. Then for each n we have $B(x, 1/n) \cap F \neq \emptyset$. By the axiom of choice, we can choose $s_n \in B(x, 1/n) \cap F$. Clearly s_n converges to $s_\infty = x$ as $n \to \infty$. Yet s_∞ is not in F. □

Given an arbitrary subset A of M, the *interior* A^o of A is the largest open subset of A. Similarly, the *closure* $\bar{A}$ of A is the smallest closed superset of A. The set A is *dense* in M if $\bar{A} = M$.

6.7 Continuity

Let f be a function from a metric space A to another metric space B. Then f is said to be *continuous* at a if for every $\epsilon > 0$ there exists $\delta > 0$ such that for all x we have that $d(x, a) < \delta$ implies $d(f(x), f(a)) < \epsilon$.

Let f be a function from a metric space A to another metric space B. Then there are various notions of how the function can respect the metric.

1. f is a *contraction* if for all x, y we have $d(f(x), f(y)) \leq d(x, y)$.

2. f is *Lipschitz* (bounded slope) if there exists $M < \infty$ such that for all x, y we have $d(f(x), f(y)) \leq M d(x, y)$.

3. f is *uniformly continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that for all x, y we have that $d(x, y) < \delta$ implies $d(f(x), f(y)) < \epsilon$.

4. f is *continuous* if for every y and every $\epsilon > 0$ there exists $\delta > 0$ such that for all x we have that $d(x, y) < \delta$ implies $d(f(x), f(y)) < \epsilon$.

Clearly contraction implies Lipschitz implies uniformly continous implies continuous. The converse implications are false.

Let A and B be metric spaces. Suppose that there is a function $f : A \to B$ with inverse function $f^{-1} : B \to A$. There are various notions of equivalence of metric spaces.

1. The metric spaces are *isometric* if f and f^{-1} are both contractions.

2. The metric spaces are *Lipschitz equivalent* if f and f^{-1} are both Lipschitz.

3. The metric spaces are *uniformly equivalent* if f and f^{-1} are both uniformly continuous.

4. The metric spaces are *topologically equivalent* (or *homeomorphic*) if f and f^{-1} are both continuous.

Again there is a chain of implications for the various kinds of equivalence: isometric implies Lipschitz implies uniform implies topological.

The following theorem shows that the notion of continuity depends only on the collection of open subsets of the metric space, and makes no other use of the metric. It follows that the property of topological equivalence also depends only on a specification of the collection of open subsets for each metric space.

Theorem 6.7 *Let A and B be metric spaces. Then $f : A \to B$ is continuous if and only if for each open set $V \subset B$, the set $f^{-1}[V] = \{x \in A \mid f(x) \in V\}$ is open.*

Proof: Suppose f continuous. Consider an open set V. Let x be in $f^{-1}[V]$. Since V is open, there is a ball $B(f(x), \epsilon) \subset V$. Since f is continuous, there is a ball $B(x, \delta)$ such that $B(x, \delta) \subset f^{-1}[B(f(x), \epsilon)] \subset f^{-1}[V]$. Since for each x there is such a δ, it follows that $f^{-1}[V]$ is open.

Suppose that the relation f^{-1} maps open sets to open sets. Consider an x and let $\epsilon > 0$. The set $B(f(x), \epsilon)$ is open, so the set $f^{-1}[B(f(x), \epsilon)$ is open. Therefore there is a $\delta > 0$ such that $B(x, \delta) \subset f^{-1}[B(f(x), \epsilon)]$. This shows that f is continuous at x.

□

Problems

1. Show that if $s : \mathbf{N} \to \mathbf{R}^\infty$ is a sequence of infinite sequences, then the product space distance $d_p(s_n, x) \to 0$ as $n \to \infty$ if and only if for each k, $s_{nk} \to x_k$ as $n \to \infty$ with respect to the usual metric on $\mathbf{R}$. This shows that the product metric is the metric for pointwise convergence.

2. Regard ℓ^2 as a subset of $\mathbf{R}^\infty$. Find a sequence of points in the unit sphere of ℓ^2 that converges in the $\mathbf{R}^\infty$ sense to zero.

3. Let X be a metric space. Give a careful proof using precise definitions that $BC(X)$ is a closed subset of $B(X)$.

4. Give four examples of bijective functions from $\mathbf{R}$ to $\mathbf{R}$: an isometric equivalence, a Lipschitz but not isometric equivalence, a uniform but not Lipschitz equivalence, and a topological but not uniform equivalence.

5. Show that for F a linear transformation of a normed vector space to itself, F continuous at zero implies F Lipschitz (bounded slope).

6.8 Uniformly equivalent metrics

Consider two metrics on the same set A. Then the identity function from A with the first metric to A with the second metric may be a contraction, Lipschitz, uniformly continuous, or continuous. There are corresponding notions of equivalence of metrics: the metrics may be the same, they may be Lipschitz equivalent, they may be uniformly equivalent, or they may be topologically equivalent.

For metric spaces the notion of uniform equivalence is particularly important. The following result shows that given a metric, there is a bounded metric that is uniformly equivalent to it. In fact, such a metric is

$$d_b(x, y) = \frac{d(x, y)}{1 + d(x, y)}. \tag{6.8}$$

The following theorem puts this in a wider context.

Theorem 6.8 *Let $\phi : [0, +\infty) \to [0, +\infty)$ be a continuous function that satisfies the following three properties:*

1. *ϕ is increasing: $s \leq t$ implies $\phi(s) \leq \phi(t)$*

2. *ϕ is subadditive: $\phi(s + t) \leq \phi(s) + \phi(t)$*

3. *$\phi(t) = 0$ if and only if $t = 0$.*

Then if d is a metric, the metric d' defined by $d'(x, y) = \phi(d(x, y))$ is also a metric. The identity map from the set with metric d to the set with metric d' is uniformly continuous with uniformly continuous inverse.

Proof: The subadditivity is what is needed to prove the triangle inequality. The main thing to check is that the identity map is uniformly continuous in each direction.

Consider $\epsilon > 0$. Since ϕ is continuous at 0, it follows that there is a $\delta > 0$ such that $t < \delta$ implies $\phi(t) < \epsilon$. Hence if $d(x, y) < \delta$ it follows that $d'(x, y) < \epsilon$. This proves the uniform continuity in one direction.

The other part is also simple. Let $\epsilon > 0$. Let $\delta = \phi(\epsilon) > 0$. Since ϕ is increasing, $t \geq \epsilon \Rightarrow \phi(t) \geq \delta$, so $\phi(t) < \delta \Rightarrow t < \epsilon$. It follows that if $d'(x, y) < \delta$, then $d(x, y) < \epsilon$. This proves the uniform continuity in the other direction. □

In order to verify the subadditivity, it is sufficient to check that $\phi'(t)$ is decreasing. For in this case $\phi'(s + u) \leq \phi'(s)$ for each $u \geq 0$, so

$$\phi(s + t) - \phi(s) = \int_0^t \phi'(s + u)\, du \leq \int_0^t \phi'(u)\, du = \phi(t). \tag{6.9}$$

This works for the example $\phi(t) = t/(1+t)$. The derivative is $\phi'(t) = 1/(1+t)^2$, which is positive and decreasing.

6.9 Sequences

A *sequence* is often taken to be a function defined on $\mathbf{N} = \{0, 1, 2, 3, \ldots\}$, but it is sometimes also convenient to regard a sequence as defined on $\mathbf{N}_+ = \{1, 2, 3, \ldots\}$. Consider a sequence $s : \mathbf{N}_+ \to B$, where B is a metric space. Then the *limit* of s_n as $n \to \infty$ is s_∞ provided that $\forall \epsilon > 0 \exists N \forall n \, (n \geq N \Rightarrow d(s_n, s_\infty) < \epsilon)$.

Theorem 6.9 *If A and B are metric spaces, then $f : A \to B$ is continuous if and only if whenever s is a sequence in A converging to s_∞, it follows that $f(s)$ is a sequence in B converging to $f(s_\infty)$.*

Proof: Suppose that $f : A \to B$ is continuous. Suppose that s is a sequence in A converging to s_∞. Consider arbitrary $\epsilon > 0$. Then there is a $\delta > 0$ such that $d(x, s_\infty) < \delta$ implies $d(f(x), f(s_\infty)) < \epsilon$. Then there is an N such that $n \geq N$ implies $d(s_n, s_\infty) < \delta$. It follows that $d(f(s_n), f(s_\infty)) < \epsilon$. This is enough to show that $f(s)$ converges to $f(s_\infty)$.

The converse is not quite so automatic. Suppose that for every sequence s converging to some s_∞ the corresponding sequence $f(s)$ converges to $f(s_\infty)$. Suppose that f is not continuous at some point a. Then there exists $\epsilon > 0$ such that for every $\delta > 0$ there is an x with $d(x, a) < \delta$ and $d(f(x), f(a)) \geq \epsilon$. In particular, the set of x with $d(x, a) < 1/n$ and $d(f(x), f(a)) \geq \epsilon$ is non-empty. By the axiom of choice, for each n there is an s_n in this set. Let $s_\infty = a$. Then $d(s_n, s_\infty) < 1/n$ and $d(f(s_n), f(s_\infty)) \geq \epsilon$. This contradicts the hypothesis that f maps convergent sequences to convergent sequences. Thus f is continuous at every point. □

One way to make this definition look like the earlier definitions is to define a metric on $\mathbf{N}_+$. Set

$$d^*(m, n) = \left| \frac{1}{m} - \frac{1}{n} \right|. \tag{6.10}$$

We may extend this to a metric on $\mathbf{N}_+ \cup \{\infty\}$ if we set $1/\infty = 0$.

Theorem 6.10 *With the metric d^* on $\mathbf{N}_+ \cup \{\infty\}$ defined above, the limit of s_n as $n \to \infty$ is s_∞ if and only if the function s is continuous from the metric space $\mathbf{N}_+ \cup \{\infty\}$ to B.*

Proof: The result is obvious if we note that $n > N$ is equivalent to $d^*(n, \infty) = 1/n < \delta$, where $\delta = 1/N$. □

Another important notion is that of *Cauchy sequence*. A sequence $s : \mathbf{N}_+ \to B$ is a Cauchy sequence if $\forall \epsilon \exists N \forall m \forall n \, ((m \geq N \wedge n \geq N) \Rightarrow d(s_m, s_n) < \epsilon)$.

Theorem 6.11 *If we use the d^* metric on $\mathbf{N}_+$ defined above, then for every sequence $s : \mathbf{N}_+ \to B$, s is a Cauchy sequence if and only if s is uniformly continuous.*

Proof: Suppose that s is uniformly continuous. Then $\forall \epsilon > 0 \exists \delta > 0 (|1/m - 1/n| < \delta \Rightarrow d(s_m, s_n) < \epsilon)$. Temporarily suppose that δ' is such that $|1/m - 1/n| < \delta \Rightarrow d(s_m, s_n) < \epsilon)$. Take N with $2/\delta' < N$. Suppose $m \geq N$ and

$n \geq N$. Then $|1/m - 1/n| \leq 2/N < \delta'$. Hence $d(s_m, s_n) < \epsilon$. Thus $(m \geq N \wedge n \geq N) \Rightarrow d(s_m, s_n) < \epsilon$. From this it is easy to conclude that s is a Cauchy sequence.

Suppose on the other hand that s is a Cauchy sequence. This means that $\forall \epsilon > 0 \exists N \forall m \forall n \left((m \geq N \wedge n \geq N) \Rightarrow d(s_m, s_n) < \epsilon\right)$. Temporarily suppose that N' is such that $\forall m \forall n \left((m \geq N' \wedge n \geq N') \Rightarrow d(s_m, s_n) < \epsilon\right)$. Take $\delta = 1/(N'(N'+1))$. Suppose that $|1/m - 1/n| < \delta$. Either $m < n$ or $n < m$ or $m = n$. In the first case, $1/(m(m+1)) = 1/m - 1/(m+1) < 1/m - 1/n < 1/(N'(N'+1))$, so $m > N'$, and hence also $n > N'$. So $d(s_m, s_n) < \epsilon$ Similarly, in the second case both $m > N'$ and $n > N'$, and again $d(s_m, s_n) < \epsilon$. Finally, in the third case $m = n$ we have $d(s_m, s_n) = 0 < \epsilon$. So we have shown that $|1/m - 1/n| < \delta \Rightarrow d(s_m, s_n) < \epsilon$. $\square$

Problems

1. Let K be an infinite matrix with $\|K\|_{1,\infty} = \sup_n \sum_m |K_{mn}| < \infty$. Show that $F(x)_m = \sum_n K_{mn} x_n$ defines a Lipschitz function from ℓ^1 to itself.

2. Let K be an infinite matrix with $\|K\|_{\infty,1} = \sup_m \sum_n |K_{mn}| < \infty$. Show that $F(x)_m = \sum_n K_{mn} x_n$ defines a Lipschitz function from ℓ^∞ to itself.

3. Let K be an infinite matrix with $\|K\|_{2,2}^2 = \sum_m \sum_n |K_{mn}|^2 < \infty$. Show that $F(x)_m = \sum_n K_{mn} x_n$ defines a Lipschitz function from ℓ^2 to itself.

4. Let K be an infinite matrix with $\|K\|_{1,\infty} < \infty$ and $\|K\|_{\infty,1} < \infty$. Show that $F(x)_m = \sum_n K_{mn} x_n$ defines a Lipschitz function from ℓ^2 to itself.

Chapter 7

Metric spaces and completeness

7.1 Completeness

Let A be a metric space. Then A is *complete* means that every Cauchy sequence with values in A converges. In this section we give an alternative perspective on completeness that makes this concept seem particularly natural.

If z is a point in a metric space A, then z defines a function $f_z : A \to [0, +\infty)$ by

$$f_z(x) = d(z, x). \tag{7.1}$$

This function has the following three properties:

1. $f_z(y) \leq f_z(x) + d(x, y)$
2. $d(x, y) \leq f_z(x) + f_z(y)$
3. $\inf f_z = 0$.

Say that a function $f : A \to [0, +\infty)$ is a virtual point if it has the three properties:

1. $f(y) \leq f(x) + d(x, y)$
2. $d(x, y) \leq f(x) + f(y)$
3. $\inf f = 0$.

We shall see that a metric space is complete if and only if every virtual point is a point. That is, it is complete iff whenever f is a virtual point, there is a point z in the space such that $f = f_z$.

It will be helpful later on to notice that the first two conditions are equivalent to $|f(y) - d(x, y)| \leq f(x)$. Also, it follows from the first condition and symmetry that $|f(x) - f(y)| \leq d(x, y)$. Thus virtual points are contractions, and in particular they are continuous.

Theorem 7.1 *A metric space is complete if and only if every virtual point is given by a point.*

Proof: Suppose that every virtual point is a point. Let s be a Cauchy sequence of points in A. Then for each x in A, $d(s_n, x)$ is a Cauchy sequence in $\mathbf{R}$. This is because $|d(s_m, x) - d(s_n, x)| \leq d(s_m, s_n)$. However every Cauchy sequence in $\mathbf{R}$ converges. Define $f(x) = \lim_{n\to\infty} d(s_n, x)$. It is easy to verify that f is a virtual point. By assumption it is given by a point z, so $f(x) = f_z(x) = d(z, x)$. But $d(s_n, z)$ converges to $f(z) = d(z, z) = 0$, so this shows that $s_n \to z$ as $n \to \infty$.

Suppose on the other hand that every Cauchy sequence converges. Let f be a virtual point. Let s_n be a sequence of points such that $f(s_n) \to 0$ as $n \to \infty$. Then $d(s_m, s_n) \leq f(s_m) + f(s_n) \to 0$ as $m, n \to \infty$, so s_n is a Cauchy sequence. Thus it must converges to a limit z. Since f is continuous, $f(z) = 0$. Furthermore, $|f(y) - d(z, y)| \leq f(z) = 0$, so $f = f_z$. □

Theorem 7.2 *Let A be a dense subset of the metric space $\bar{A}$. Let M be a complete metric space. Let $f : A \to M$ be uniformly continuous. Then there exists a unique uniformly continuous function $\bar{f} : \bar{A} \to M$ that extends f.*

Proof: Regard the function f as a subset of $\bar{A} \times M$. Define the relation $\bar{f}$ to be the closure of f. If x is in $\bar{A}$, let $s_n \in A$ be such that $s_n \to x$ as $n \to \infty$. Then s_n is a Cauchy sequence in A. Since f is uniformly continuous, it follows that $f(s_n)$ is a Cauchy sequence in M. Therefore $f(s_n)$ converges to some y in M. This shows that (x, y) is the relation $\bar{f}$. So the domain of $\bar{f}$ is $\bar{A}$.

Let $\epsilon > 0$. By uniform continuity there is a $\delta > 0$ such that for all x, u' in A we have that $d(x', u') < \delta$ implies $d(f(x'), f(u')) < \epsilon/3$.

Now let $(x, y) \in \bar{f}$ and $(u, v) \in \bar{f}$ with $d(x, u) < \delta/3$. There exists x' in A such that $f(x') = y'$ and $d(x', x) < \delta/3$ and $d(y', y) < \epsilon/3$. Similarly, there exists u' in A such that $f(u') = v'$ and $d(u', u) < \delta/3$ and $d(v', v) < \epsilon/3$. It follows that $d(x', u') \leq d(x', x) + d(x, u) + d(u, u') < \delta$. Hence $d(y, v) \leq d(y, y') + d(y', v') + d(v', v) < \epsilon$. Thus $d(x, u) < \delta/3$ implies $d(y, v) < \epsilon$. This is enough to show that $\bar{f}$ is a function and is uniformly continuous. □

A *Banach space* is a vector space with a norm that is a complete metric space. Here are examples of complete metric spaces. All of them except for $\mathbf{R}^\infty$ are Banach spaces. Notice that ℓ^∞ is the special case of $B(X)$ when X is countable. For $BC(X)$ we take X to be a metric space, so that the notion of continuity is defined.

Examples:

1. $\mathbf{R}^n$ with either the ℓ^1_n, ℓ^2_n, or ℓ^∞_n metric.
2. ℓ^1.
3. ℓ^2.
4. ℓ^∞.

5. $\mathbf{R}^\infty$ with the product metric.

6. $B(X)$ with the uniform metric.

7. $BC(X)$ with the uniform metric.

In these examples the points of the spaces are real functions. There are obvious modifications where one instead uses complex functions. Often the same notation is used for the two cases, so one must be alert to the distinction.

7.2 Uniform equivalence of metric spaces

Theorem 7.3 *Let A be a metric space, and let M be a complete metric space. Suppose that there is a uniformly continuous bijection $f : A \to M$ such that f^{-1} is continuous. Then A is complete.*

Proof: Suppose that $n \mapsto s_n$ is a Cauchy sequence with values in A. Since f is uniformly continuous, the composition $n \mapsto f(s_n)$ is a Cauchy sequence in M. Since M is complete, there is a y in M such that $f(s_n) \to y$ as $n \to \infty$. Let $x = f^{-1}(y)$. Since f^{-1} is continuous, it follows that $s_n \to x$ as $n \to \infty$. □

Corollary 7.4 *The completeness property is preserved under uniform equivalence.*

It is important to understand that completeness is not a topological invariant. For instance, take the function $g : \mathbf{R} \to (-1, 1)$ defined by $g(x) = (2/\pi) \arctan(x)$. This is a topological equivalence. Yet $\mathbf{R}$ is complete, while $(-1, 1)$ is not complete.

7.3 Completion

Theorem 7.5 *Every metric space is densely embedded in a complete metric space.*

This theorem says that if A is a metric space, then there is a complete metric space F and an isometry from A to F with dense range.

Proof: Let F consist of all the virtual points of A. These are continuous functions on A. The distance $\bar{d}$ between two such functions is the usual sup norm $\bar{d}(f, g) = \sup_{x \in A} d(f(x), g(x))$. It is not hard to check that the virtual points form a complete metric space of continuous functions. The embedding sends each point z in a into the corresponding f_z. Again it is easy to verify that this embedding preserves the metric, that is, that $\bar{d}(f_z, f_w) = d(z, w)$. Furthermore, the range of this embedding is dense. The reason for this is that for each virtual point f and each $\epsilon > 0$ there is an x such that $f(x) < \epsilon$. Then $|f(y) - f_x(y)| = |f(y) - d(x, y)| \leq f(x) < \epsilon$. This shows that $\bar{d}(f, f_x) \leq \epsilon$. □

The classic example is the completion of the rational number system $\mathbf{Q}$. A virtual point of $\mathbf{Q}$ is a function whose graph is in the general shape of a letter V.

When the bottom tip of the V is at a rational number, then the virtual point is already a point. However most of these V functions have tips that point to a gap in the rational number system. Each such gap in the rational number system corresponds to the position of an irrational real number in the completion.

7.4 The Banach fixed point theorem

If f is a Lipschitz function from a metric space to another metric space, then there is a constant $C < +\infty$ such that for all x and y we have $d(f(x), f(y)) \leq Cd(x, y)$. The set of all C is a set of upper bounds for the quotients, and so there is a least such upper bound. This is called the least Lipschitz constant of the function.

A Lipschitz function is a contraction if its least Lipschitz constant is less than or equal to one. It is a *strict contraction* if its least Lipschitz constant is less than one.

Theorem 7.6 (Banach) *Let A be a complete metric space. Let $f : A \to A$ be a strict contraction. Then f has a unique fixed point. For each point in A, its orbit converges to the fixed point.*

Proof: Let a be a point in A, and let $s_k = f^{(k)}(a)$. Then by induction $d(s_k, s_{k+1}) \leq M^k d(s_0, s_1)$. Then again by induction $d(s_m, s_{m+p})) \leq \sum_{k=m}^{p-1} M^k d(s_0, s_1) \leq K^m/(1 - K)d(s_0, s_1)$. This is enough to show that s is a Cauchy sequence. By completeness it converges to some s_∞. Since f is continuous, this is a fixed point. □

Recall that a Banach space is a complete normed vector space. The Banach fixed point theorem applies in particular to a linear transformations of a Banach space to itself that is a strict contraction.

For instance, consider one of the Banach spaces of sequences. Let $f(x) = Kx + u$, where K is a matrix, and where u belongs to the Banach space. The function f is Lipschitz if and only if multiplication by K is Lipschitz. If the Lipschitz constant is strictly less than one, then the Banach theorem gives the solution of the linear system $x - Kx = u$.

To apply this, first look at the Banach space ℓ^∞. Define $\|K\|_{\infty\to\infty}$ to be the least Lipschitz constant. Define

$$\|K\|_{\infty,1} = \sup_m \sum_{n=1}^{\infty} |K_{mn}|. \tag{7.2}$$

Then it is not difficult to see that $\|K\|_{\infty\to\infty} = \|K\|_{\infty,1}$.

For another example, consider the Banach space ℓ^1. Define $\|K\|_{1\to1}$ to be the least Lipschitz constant. Define

$$\|K\|_{1,\infty} = \sup_n \sum_{m=1}^{\infty} |K_{mn}|. \tag{7.3}$$

Then it is not difficult to see that $\|K\|_{1\to 1} = \|K\|_{1,\infty}$.

The interesting case is the Hilbert space ℓ^2. Define $\|K\|_{2\to 2}$ to be the least Lipschitz constant. Define

$$\|K\|_{2,2} = \sqrt{\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} K_{mn}^2}. \tag{7.4}$$

Then an easy application of the Schwarz inequality will show that $\|K\|_{2\to 2} \leq \|K\|_{2,2}$. However this is usually not an equality!. A somewhat more clever application of the Schwarz inequality will show that $\|K\|_{2\to 2} \leq \sqrt{\|K\|_{1,\infty}\|K\|_{\infty,1}}$. Again this is not in general an equality. Finding the least Lipschitz constant is a non-trivial task. However one or the other of these two results will often give useful information.

Problems

1. Show that a closed subset of a complete metric space is complete.

2. Let c_0 be the subset of ℓ^∞ consisting of all sequences that converge to zero. Show that c_0 is a complete metric space.

3. Let A be a dense subset of the metric space $\bar{A}$. Let M be a complete metric space. Let $f : A \to M$ be continuous. It does not follow in general that there is a continuous function $\bar{f} : \bar{A} \to M$ that extends f. (a) Give an example of a case when the closure $\bar{f}$ of the graph is a function on A but is not defined on $\bar{A}$. (b) Give an example when the closure $\bar{f}$ of the graph is a relation defined on $\bar{A}$ but is not a function.

4. Let $C([0,1])$ be the space of continuous real functions on the closed unit interval. Give it the metric $d_1(f,g) = \int_0^1 |f(x) - g(x)|\, dx$. Let h be a discontinuous step function equal to 0 on half the interval and to 1 on the other half. Show that the map $f \mapsto \int_0^1 |f(x) - h(x)|\, dx$ is a virtual point of $C([0,1])$ (with the d_1 metric) that does not come from a point of $C([0,1])$.

5. Let E be a complete metric space. Let $f : E \to E$ be a strict contraction with constant $C < 1$. Consider z in E and r with $r \geq d(f(z), z)/(1 - C)$. Then f has a fixed point in the ball consisting of all x with $d(x, z) \leq r$. Hint: First show that this ball is a complete metric space.

7.5 Coerciveness

A continuous function defined on a compact space assumes its minimum (and its maximum). This result is both simple and useful. However in general the point where the minimum is assumed is not unique. Furthermore, the condition that the space is compact is too strong for many applications. A result that only uses completeness could be helpful, and the following is one of the most useful results of this type.

Theorem 7.7 *Let M be a complete metric space. Let f be a continuous real function on M that is bounded below. Let $a = \inf\{f(x) \mid x \in M\}$. Suppose that there is an increasing function ϕ from $[0, +\infty)$ to itself such that $\phi(t) = 0$ only for $t = 0$ with the coercive estimate*

$$a + \phi(d(x, y)) \leq \frac{f(x) + f(y)}{2}. \tag{7.5}$$

Then there is a unique point p where $f(p) = a$. That is, there exists a unique point p where F assumes its minimum value.

Proof: Let s_n be a sequence of points such that $f(s_n) \to a$ as $n \to \infty$. Consider $\epsilon > 0$. Let $\delta = \phi(\epsilon) > 0$. Since ϕ is increasing, $\phi(t) < \delta$ implies $t < \epsilon$. For large enough m, n we can arrange that $\phi(d(s_m, s_n)) < \delta$. Hence $d(s_m, s_n) < \epsilon$. Thus s_n is a Cauchy sequence. Since M is complete, the sequence converges to some p in M. By continuity, $f(p) = a$. Suppose also that $f(q) = a$. Then from the inequality $d(p, q) = 0$, so $p = q$. □

This theorem looks impossible to use in practice, because it seems to requires a knowledge of the infimum of the function. However the following result shows that there is a definite possibility of a useful application.

Corollary 7.8 *Let M be a closed convex subset of a Banach space. Let f be a continuous real function on M. Say that $a = \inf_{x \in M} f(x)$ is finite and that there is a $c > 0$ such that the strict convexity condition*

$$c\|x - y\|^2 \leq \frac{f(x) + f(y)}{2} - f(\frac{x + y}{2}) \tag{7.6}$$

is satisfied. Then there is a unique point p in M with $f(p) = a$.

Proof: Since M is convex, $(x + y)/2$ is in M, and so $a \leq f((x + y)/2)$. □

Chapter 8

Metric spaces and compactness

8.1 Total boundedness

The notion of compactness is meaningful and important in general topological spaces. However it takes a quantitative form in metric spaces, and so it is worth making a special study in this particular setting. A metric space is complete when it has no nearby missing points (that is, when every virtual point is a point). It is compact when, in addition, it is well-approximated by finite sets. The precise formulation of this approximation property is in terms of the following concept.

A metric space M is *totally bounded* if for every $\epsilon > 0$ there exists a finite subset F of M such that the open ϵ-balls centered at the points of F cover M.

We could also define M to be totally bounded if for every $\epsilon > 0$ the space M is the union of finitely many sets each of diameter at most 2ϵ. For some purposes this definition is more convenient, since it does not require the sets to be balls.

The notion of total boundedness is quantitative. If M is a metric space, then there is a function that assigns to each $\epsilon > 0$ the smallest number N such that M is the union of N sets each of diameter at most 2ϵ. The slower the growth of this function, the better the space is approximated by finitely many points.

For instance, consider a box of side $2L$ in a Euclidean space of dimension k. Then the N is roughly $(L/\epsilon)^k$. This shows that the covering becomes more difficult as the size L increases, but also as the dimension k increases.

Theorem 8.1 *Let $f : K \to M$ be a uniformly continuous surjection. If K is totally bounded, then M is totally bounded.*

Corollary 8.2 *Total boundedness is invariant under uniform equivalence of metric spaces.*

8.2 Compactness

For metric spaces we can say that a metric space is *compact* if it is both complete and totally bounded.

Lemma 8.3 *Let K be a metric space. Let F be a subset of K. If F is complete, then F is a closed subset of K. Suppose in addition that K is complete. If F is a closed subset of K, then F is complete.*

Proof: Suppose F is complete. Say that s is a sequence of points in F that converges to a limit a in K. Then s is a Cauchy sequence in F, so it converges to a limit in F. This limit must be a, so a is in F. This proves that F is a closed subset of K. Suppose for the converse that K is complete and F is closed in K. Let s be a Cauchy sequence in F. Then it converges to a limit a in K. Since F is closed, the point a must be in F. This proves that F is complete. □

Lemma 8.4 *Let K be a totally bounded metric space. Let F be a subset of K. Then F is totally bounded.*

Proof: Let $\epsilon > 0$. Then K is the union of finitely many sets, each of diameter bounded by 2ϵ. Then F is the union of the intersections of these sets with F, and each of these intersections has diameter bounded by 2ϵ. □

Theorem 8.5 *Let K be a compact metric space. Let F be a subset of K. Then F is compact if and only if it is a closed subset of K.*

Proof: Since K is compact, it is complete and totally bounded. Suppose F is compact. Then it is complete, so it is a closed subset of K. For the converse, suppose F is a closed subset of K. It follows that F is complete. Furthermore, from the last lemma F is totally bounded. It follows that F is compact. □

Examples:

1. The unit sphere (cube) in ℓ^∞ is not compact. In fact, the unit basis vectors δ_n are spaced by 1.

2. The unit sphere in ℓ^2 is not compact. The unit basis vectors δ_n are spaced by $\sqrt{2}$.

3. The unit sphere in ℓ^1 is not compact. The unit basis vectors δ_n are spaced by 2.

Examples:

1. Let $c_k \geq 1$ be a sequence that increases to infinity. The squashed solid rectangle of all x with $c_k |x_k| \leq 1$ for all k is compact in ℓ^∞.

2. Let $c_k \geq 1$ be a sequence that increases to infinity. The squashed solid ellipsoid of all x with $\sum_{k=1}^{\infty} c_k x_k^2 \leq 1$ is compact in ℓ^2.

3. Let $c_k \geq 1$ be a sequence that increases to infinity. The squashed region of all x with $\sum_{k=1}^{\infty} c_k |x_k| \leq 1$ is compact in ℓ^1.

Problems

1. Let $c_k \geq 1$ be a sequence that increases to infinity. Show that the squashed solid ellipsoid of all x with $\sum_{k=1}^{\infty} c_k x_k^2 \leq 1$ is compact in ℓ^2.

2. Prove that the squashed solid ellipsoid in ℓ^2 is not homeomorphic to the closed unit ball in ℓ^2.

3. Let $c_k \geq 1$ be a sequence that increases to infinity. Is the squashed ellipsoid of all x with $\sum_{k=1}^{\infty} c_k x_k^2 = 1$ compact in ℓ^2?

4. Is the squashed ellipsoid in ℓ^2 homeomorphic to the unit sphere in ℓ^2?

8.3 Countable product spaces

Let M_j for $j \in \mathbf{N}_+$ be a sequence of metric spaces. Let $\prod_j M_j$ be the product space consisting of all functions f such that $f(j) \in M_j$. Let $\phi(t) = t/(1+t)$. Define the product metric by

$$d(f,g) = \sum_{j=1}^{\infty} \frac{1}{2^j} \phi(d(f(j), g(j))). \tag{8.1}$$

The following results are elementary.

Lemma 8.6 *If each M_j is complete, then $\prod_j M_j$ is complete.*

Lemma 8.7 *If each M_j is totally bounded, then $\prod_j M_j$ is totally bounded.*

Theorem 8.8 *If each M_j is compact, then $\prod_j M_j$ is compact.*

Examples:

1. The product space $\mathbf{R}^\infty$ is complete but not compact.

2. The closed unit ball (solid cube) in ℓ^∞ is a compact subset of $\mathbf{R}^\infty$ with respect to the $\mathbf{R}^\infty$ metric. In fact, it is a product of compact spaces. What makes this work is that the $\mathbf{R}^\infty$ metric measures the distances for various coordinates in increasingly less stringent ways.

3. The unit sphere (cube) in ℓ^∞ is not compact with respect to the $\mathbf{R}^\infty$ metric, in fact, it is not even closed. The sequence δ_n converges to zero. The zero sequence is in the closed ball (solid cube), but not in the sphere.

8.4 Compactness and continuous functions

Theorem 8.9 *A metric space M is compact if and only if every sequence with values in M has a subsequence that converges to a point of M.*

Proof: Suppose that M is compact. Thus it is totally bounded and complete. Let s be a sequence with values in M. Since M is bounded, it is contained in a ball of radius C.

By induction construct a sequence of balls B_j of radius $C/2^j$ and a decreasing sequence of infinite subsets N_j of the natural numbers such that for each k in N_j we have s_k in B_j. For $j = 0$ this is no problem. If it has been accomplished for j, cover B_j by finitely many balls of radius $C/2^{j+1}$. Since N_j is infinite, there must be one of these balls such that s_k is in it for infinitely many of the k in N_j. This defines B_{j+1} and N_{j+1}.

Let r be a strictly increasing sequence of numbers such that r_j is in N_j. Then $j \mapsto s_{r_j}$ is a subsequence that is a Cauchy sequence. By completeness it converges.

The converse proof is easy. The idea is to show that if the space is either not complete or not totally bounded, then there is a sequence without a convergent subsequence. In the case when the space is not complete, the idea is to have the sequence converge to a point in the completion. In the case when the space is not totally bounded, the idea is to have the terms in the sequence separated by a fixed distance. □

The theorem shows that for metric spaces the concept of compactness is invariant under topological equivalence. In fact, it will turn out that compactness is a purely topological property.

Theorem 8.10 *Let K be a compact metric space. Let L be another metric space. Let $f : K \to L$ be a continuous function. Then f is uniformly continuous.*

Proof: Suppose f were not uniformly continuous. Then there exists $\epsilon > 0$ such that for each $\delta > 0$ the set of pairs (x, y) with $d(x, y) < \delta$ and $d(f(x), f(y)) \geq \epsilon$ is not empty. Consider the set of pairs (x, y) with $d(x, y) < 1/n$ and $d(f(x), f(y)) \geq \epsilon$. Choose s_n and t_n with $d(s_n, t_n) < 1/n$ and $d(f(s_n), f(t_n)) \geq \epsilon$. Since K is compact, there is a subsequence $u_k = s_{r_k}$ that converges to some limit a. Then also $v_k = t_{r_k}$ converges to a. But then $f(u_k) \to f(a)$ and $f(v_k) \to f(a)$ as $k \to \infty$. In particular, $d(f(u_k), f(v_k)) \to d(a, a) = 0$ as $k \to \infty$. This contradicts the fact that $d(f(u_k), f(v_k)) \geq \epsilon$. □

A corollary of this result is that for compact metric spaces the concepts of uniform equivalence and topological equivalence are the same.

Theorem 8.11 *Let K be a compact metric space. Let L be another metric space. Let $f : K \to L$ be continuous. Then $f[K]$ is compact.*

Proof: Let t be a sequence with values in $f[K]$. Choose s_k with $f(s_k) = t_k$. Then there is a subsequence $u_j = s_{r_j}$ with $u_j \to a$ as $j \to \infty$. It follows that

$t_{r_j} = f(s_{r_j}) = f(u_j) \to f(a)$ as $j \to \infty$. This shows that t has a convergence subsequence. □

The classic application of this theorem is to the case when $f : K \to \mathbf{R}$, where K is a non-empty metric space. Then $f[K]$ is a non-empty compact subset of $\mathbf{R}$. However, a non-empty compact set of real numbers has a least element and a greatest element. Therefore there is a p in K where f assumes its minimum value, and there is a q in K where f assumes its maximum value.

8.5 Semicontinuity

A function from a metric space M to $[-\infty, +\infty)$ is said to be *upper semicontinuous* if for every u and every $r > f(u)$ there is a $\delta > 0$ such that all v with $d(u, v) < \delta$ satisfy $f(v) < r$. An example of an upper semicontinuous function is one that is continuous except where it jumps up at a single point. It is easy to fall from this peak. The indicator function of a closed set is upper semicontinuous. The infimum of a non-empty collection of upper semicontinuous functions is upper semicontinuous. This generalizes the statement that the intersection of a collection of closed sets is closed.

There is a corresponding notion of lower semicontinuous function. A function from a metric space M to $(-\infty, +\infty]$ is said to be *lower semicontinuous* if for every u and every $r < f(u)$ there is a $\delta > 0$ such that all v with $d(u, v) < \delta$ satisfy $f(v) > r$. An example of a lower semicontinuous function is one that is continuous except where jumps down at a single point. The indicator function of an open set is lower semicontinuous. The supremum of a non-empty collection of lower semicontinuous functions is lower semicontinuous. This generalizes the fact that the union of a collection of open sets is open.

Theorem 8.12 *Let K be compact and not empty. Let $f : K \to (-\infty, +\infty]$ be lower semicontinuous. Then there is a point p in K where f assumes its minimum value.*

Proof: Let a be the infimum of the range of f. Suppose that s is a sequence of points in K such that $f(s_n) \to a$. By compactness there is a strictly increasing sequence g of natural numbers such that the subsequence $j \mapsto s_{g_j}$ converges to some p in K. Consider $r < f(p)$. The lower semicontinuity implies that for sufficiently large j the values $f(s_{g_j}) > r$. Hence $a \geq r$. Since $r < f(p)$ is arbitrary, we conclude that $a \geq f(p)$. □

There is a corresponding theorem for the maximum of an upper semicontinuous function on a compact space that is not empty.

8.6 Compact sets of continuous functions

Let A be a family of functions on a metric space M to another metric space. Then A is *equicontinuous* if for every x and every $\epsilon > 0$ there is a $\delta > 0$ such

that for all f in A the condition $d(x, y) < \delta$ implies $d(f(x), f(y)) < \epsilon$. Thus the δ does not depend on the f in A.

Similarly, A is *uniformly equicontinuous* if for every $\epsilon > 0$ there is a $\delta > 0$ such that for all f in A the condition $d(x, y) < \delta$ implies $d(f(x), f(y)) < \epsilon$. Thus the δ does not depend on the f in A or on the point in the domain.

Finally, A is *equiLipschitz* if there is a constant C such that for all f in A the condition $d(x, y) < \delta$ implies $d(f(x), f(y)) < Cd(x, y)$ is satisfied.

It is clear that equiLipschitz implies uniformly equicontinuous implies equicontinuous.

Lemma 8.13 *Let K be a compact metric space. If A is an equicontinuous set of functions on K, then A is a uniformly equicontinuous set of functions on K.*

Let K, M be metric spaces, and let $BC(K \to M)$ be the metric space of all bounded continuous functions from K to M. The distance between two functions is given by the supremum over K of the distance of their values in the M metric. When M is complete, this is a complete metric space. When K is compact or M is bounded, this is the same as the space $C(K \to M)$ of all continuous functions from K to M. A common case is when $M = [-m, m] \subset \mathbf{R}$, a closed bounded interval of real numbers.

Theorem 8.14 (Arzelà-Ascoli) *Let K and M be totally bounded metric spaces. Let A be a subset of $C(K \to M)$. If A is uniformly equicontinuous, then A is totally bounded.*

Proof: Let $\epsilon > 0$. By uniform equicontinuity there exists a $\delta > 0$ such that for all f in A and all x, y the condition $d(x, y) < \delta$ implies that $|f(x) - f(y)| < \epsilon/4$. Furthermore, there is a finite set $F \subset K$ such that every point in K is within δ of a point of F. Finally, there is a finite set G of points in M that are within $\epsilon/4$ of every point in M. The set G^F is finite.

For each h in G^F let D_h be the set of all g in A such that g is within $\epsilon/4$ of h on F. Every g is in some D_h. Each x in K is within δ of some a in F. Then for g in D_h we have

$$|g(x) - h(a)| \leq |g(x) - g(a)| + |g(a) - h(a)| < \epsilon/4 + \epsilon/4 = \epsilon/2. \tag{8.2}$$

We conclude that each pair of functions in D_h is within ϵ of each other. Thus A is covered by finitely many sets of diameter ϵ. □

In practice the way to prove that A is uniformly equicontinuous is to prove that A is equiLipschitz with constant C. Then the theorem shows in a rather explicit way that A is totally bounded. In fact, the functions are parameterized to within a tolerance ϵ by functions from the finite set F of points spaced by $\delta = \epsilon/(4C)$ to the finite set G of points spaced by $\epsilon/4$.

Corollary 8.15 (Arzelà-Ascoli) *Let K, M be a compact metric spaces. Let A be a subset of $C(K \to M)$. If A is equicontinuous, then its closure $\bar{A}$ is compact.*

Proof: Since K is compact, the condition that A is equicontinuous implies that A is uniformly equicontinuous. By the theorem, A is totally bounded. It follows easily that the closure $\bar{A}$ is totally bounded. Since M is compact and hence complete, $C(K \to M)$ is complete. Since $\bar{A}$ is a closed set of a complete space, it is also complete. The conclusion is that $\bar{A}$ is compact. □

The theorem has consequences for existence results. Thus every sequence of functions in A has a subsequence that converges in the metric of $C(K \to M)$ to a function in the space.

Problems

1. Consider a metric space A with metric d. Say that there is another metric space B with metric d_1. Suppose that $A \subset B$, and that $d_1 \leq d$ on $A \times A$. Finally, assume that there is a sequence f_n in A that approaches h in $B \setminus A$ with respect to the d_1 metric. Show that A is not compact with respect to the d metric. (Example: Let A be the unit sphere in ℓ^2 with the ℓ^2 metric, and let B be the closed unit ball in ℓ^2, but with the $\mathbf{R}^\infty$ metric.)

2. Is the metric space of continuous functions on $[0, 1]$ to $[-1, 1]$ with the sup norm compact? Prove or disprove. (Hint: Use the previous problem.)

3. Consider the situation of the Arzelà-Ascoli theorem applied to a set $A \subset C(K)$ with bound m and Lipschitz constant C. Suppose that the number of δ sets needed to cover K grows like $(L/\delta)^k$, a finite dimensional behavior (polynomial in $1/\delta$). What is the growth of the number of ϵ sets needed to cover $A \subset C(K)$? It this a finite dimensional rate?

8.7 Curves of minimum length

The following is an application of the ideas of this section. Let M be a compact metric space. Fix points p and q in M. Consider the metric space of all continuous functions $\phi : [0, 1] \to M$ with $\phi(0) = p$ and $\phi(1) = q$. An element of this space is called a curve from p to q.

If $\tau_0, \ldots, \tau_n$ is a strictly increasing sequence of points in $[0, 1]$ with $\tau_0 = 0$ and $\tau_n = 1$, define

$$F_\tau(\phi) = \sum_{i=1}^{n} d(\phi(\tau_{i-1}), \phi(\tau_i)). \tag{8.3}$$

Then F_τ is a continuous real function on the space of all curves from p to q. It is a function that computes an approximation to the length of the curve ϕ.

Define a function F on the path space with values in $[0, +\infty]$ by $F(\phi) = \sup_\tau F_\tau(\phi)$. Since each F_τ is continuous, it follows that F is lower semicontinuous. This function may be thought of as the length of the curve ϕ. This length may be infinite.

The interesting feature of the definition of length is that the length of a curve need not be a continuous function of the curve. The reason is that one could

take a smooth curve ϕ and approximate it very well in the uniform sense by a very irregular curve ψ. So while ψ is uniformly close to ψ, it can have a length much greater than the length of ϕ.

Theorem 8.16 *Let M be a compact metric space. Let p and q be points in M. Suppose there is a curve from p to q of finite length. Then there is a curve from p to a of minimum length.*

Proof: For each curve of finite length L there is a representation of the curve as a function of arc length along the curve. Such a representation gives the curve as a function on $[0, L]$ with Lipschitz constant 1. By changing scale one gets the curve as a function on $[0, 1]$ with Lipschitz constant L. So in searching for a curve of minimum length we may as well use Lipschitz curves.

Suppose that there is at least one curve of finite length L^*. Consider the non-empty set of all curves from p to q with length bounded by L^* and with Lipschitz constant L^*. The Arzelá-Ascoli theorem shows that this is a compact metric space. Since the length function F is a lower semicontinuous function on a non-empty compact space, the conclusion is that there is a curve ϕ from p to q of minimum length. □

Chapter 9

Vector lattices

9.1 Positivity

In the following we shall refer to a real number $x \geq 0$ as *positive*, and a number $x > 0$ as *strictly positive*. A sequence s of real numbers is *increasing* if $m \leq n$ implies $s_m \leq s_n$, while it is *strictly increasing* if $m < n$ implies $s_m < s_n$. Note that many authors prefer the terminology non-negative or non-decreasing for what is here called positive or increasing. In the following we shall often write $s_n \uparrow$ to indicate that s_n is increasing in our sense.

The terminology for real functions is more complicated. A function with $f(x) \geq 0$ for all x is called *positive* (more specifically, pointwise positive), and we write $f \geq 0$. Correspondingly, a function f with $f \geq 0$ that is not the zero function is called *positive non-zero*. While it is consistent with the conventions for ordered sets to write $f > 0$, this may risk confusion. Sometimes a term like positive semi-definite is used. In other contexts, one needs another ordering on functions. Thus the condition that either f is the zero function or $f(x) > 0$ for all x might be denoted $f \geq\geq 0$, though this is far from being a standard notation. The corresponding condition that $f(x) > 0$ for all x is called *pointwise strictly positive*, and a suitable notation might be $f >> 0$. An alternative is to say that $f > 0$ *pointwise* or $f > 0$ *everywhere*. Sometimes a term like positive definite is used.

The main use of the term positive definite is in connection with quadratic forms. A quadratic form is always zero on the zero vector, so it is reasonable to restrict attention to non-zero vectors. Then according to the writer semi-definite can mean positive or positive non-zero, while positive definite would ordinarily mean pointwise strictly positive. However some authors use the word positive definite in the least restrictive sense, that is, to indicate merely that the quadratic form is positive. A reader must remain alert to the definition in use on a particular occasion.

A related notion that will be important in the following is the pointwise ordering of functions. We write $f \leq g$ to mean that for all x there is an

inequality $f(x) \leq g(x)$. Similarly, we write $f_n \uparrow$ to indicate an increasing sequence of functions, that is, $m \leq n$ implies $f_m \leq f_n$. Also, $f_n \uparrow f$ means that $f_n \uparrow$ and f_n converges to f pointwise.

9.2 Integration of regulated functions

Perhaps the simplest definition of integral is based on the sup norm, that is, on the notion of uniform convergence. The functions that are integrable by this definition are known as regulated functions. Each continuous function is regulated, so this notion of integral is good for many calculus application. Furthermore, it works equally well for integrals with values in a Banach space.

Let $[a, b] \subset \mathbf{R}$ be a closed interval. Consider a partition $a \leq a_0 < a_1 < \ldots < a_n = b$ of the interval. A *general step function* is a function f from $[a, b]$ to $\mathbf{R}$ that is constant on each open interval (a_i, b_{i+1}) of such a partition. For each general step function f there is an integral $\lambda(f)$ that is the sum

$$\lambda(f) = \int_a^b f(x)\,dx = \sum_{i=0}^{n-1} f(c_i)(a_{i+1} - a_i), \tag{9.1}$$

where $a_i < c_i < a_{i+1}$.

Let $R([a, b])$ be the closure of the space S of general step functions in the space $B([a, b])$ of all bounded functions. This called the space of *regulated functions*. Since every continuous function is a regulated function, we have $C([a, b]) \subset R([a, b])$.

The function λ defined on the space S of general step functions is a Lipschitz function with Lipschitz constant $b - a$. In particular it is uniformly continuous, and so it extends by continuity to a function on the closure $R([a, b])$. This extended function is also denoted by λ and is the regulated integral. In particular, the regulated integral is defined on $C([a, b])$ and agrees with the integral for continuous functions that is used in elementary calculus.

9.3 The Riemann integral

The Riemann integral, by contrast, is based on the idea of order. Let f be a bounded function on the interval $[a, b]$ of real numbers. Define the lower integral by

$$\underline{\lambda}(g) = \sup\{\lambda(f) \mid f \in S, f \leq g\}. \tag{9.2}$$

Similarly, define the upper integral by

$$\overline{\lambda}(g) = \inf\{\lambda(h) \mid f \in S, g \leq h\}. \tag{9.3}$$

Then g is Riemann integrable if

$$\underline{\lambda}(g) = \overline{\lambda}(g). \tag{9.4}$$

In that case the integral is denoted

$$\lambda(g) = \int_a^b g(t)\,dt. \tag{9.5}$$

The Riemann integral is somewhat more general than the regulated integral. However the Lebesgue integral is much more powerful, and it has made the Riemann integral obsolete. The reason for the improvement is that the Riemann integral is defined by a one stage approximation procedure. The function to be integrated is approximated by step functions. The Lebesgue integral is defined by a two stage approximation procedure. First the integral is extended to lower functions and to upper functions. This is the first stage. These lower and upper functions are much more complicated than step functions. Then the function to be integrated is approximated by lower and upper functions. This is the second stage. The remarkable fact is that two stages is sufficient to produce a theory that is stable under many limiting processes.

9.4 Step functions

A general step function can have arbitrary values at the end points of the intervals. It is sometimes nicer to make a convention that makes the step functions left continuous (or right continuous). This will eventually make things easier when dealing with more general integrals where individual points count.

A *rectangular function* is an indicator function of an interval $(a, b]$ of real numbers. Here a and b are real numbers, and the interval $(a, b]$ consists of all real numbers x with $a < x \leq b$. The convention that the interval is open on the left and closed on the right is arbitrary but convenient. The nice thing about these intervals is that their intersection is an interval of the same type. Furthermore, the union of two such intervals is a finite union of such intervals. And the relative complement of two such intervals is a finite union of such intervals.

A *step function* is a finite linear combination of rectangular functions. In fact, each step function may be represented as a finite linear combination of rectangular functions that correspond to disjoint subsets.

Another important rectangular function is the *binary function* $f_{n;k}$ defined for $0 \leq n$ and $0 \leq k < 2^n$. Consider the 2^n disjoint intervals of length $1/2^n$ partitioning the interval $(0, 1]$. Number them from 0 to $2^n - 1$. Then $f_{n;k}$ is the indicator function of the kth interval. Fix n. Then the $f_{n;k}$ for $0 \leq k < 2^n$ form a basis for a 2^n dimensional vector space $\mathcal{F}_n$.

An important step function is the *Bernoulli function* b_n defined for $n \geq 1$. This is defined by partitioning $(0, 1]$ into 2^n disjoint intervals of length $1/2^n$. Number the intervals from 0 to 2^{n-1}. The Bernoulli function is the function that is 0 on the even numbered intervals and 1 on the odd numbered intervals.

There is a perhaps unexpected relation between the Bernoulli functions and the binary functions $f_{n;k}$. Write the number k in binary form. Look at the

corresponding subset S of $\{1, \ldots, n\}$ and its complement S^c in $\{1, \ldots, n\}$. Then

$$f_{n;k} = \prod_{j \in S} b_j \prod_{j \in S^c} (1 - b_j). \tag{9.6}$$

Notice that if $n = 0$ this is an empty product, and so it has the value 1.

A step function that is closely related to the Bernoulli function is the *Rademacher function* $r_n = 1 - 2b_n$ defined for $n \geq 1$. Again consider intervals of length $1/2^n$. The Rademacher function is the function that is 1 on the even numbered intervals and -1 on the odd numbered intervals.

A *Walsh function* is a product of Rademacher functions. . Let $S \subset \{1, 2, 3, \ldots\}$ be a finite set of strictly positive natural numbers. Let the Walsh function be defined by

$$w_S = \prod_{j \in S} r_j. \tag{9.7}$$

The Walsh functions w_S for $S \subset \{1, \ldots, n\}$ form another basis for $\mathcal{F}_n$. Notice that when S is empty the product is 1.

The Walsh functions may be generated from the Rademacher functions in a systematic way. At stage zero start with the function 1. At stage one take also r_1. At stage two take r_2 times each of the functions from the previous stages. This gives also r_2 and $r_1 r_2$. At stage three take r_3 times each of the functions from the previous stages. This gives also r_3 and $r_1 r_3$ and $r_2 r_3$ and $r_1 r_2 r_3$. It is clear how to continue.

A *Haar function* is a multiple of a product of a binary rectangular function with a Rademacher function. Then for $n \geq 0$ and $0 \leq k < 2^n$ define the Haar function to be

$$h_{n;k} = c_n f_{n;k} r_{n+1}, \tag{9.8}$$

and define $h_{-1;0} = 1$. For $n \geq 0$ the coefficient $c_n > 0$ is determined by $c_n^2 = 1/2^n$. The function $h_{-1;0}$ together with the other Haar functions $h_{j;k}$ for $j = 0$ to $n - 1$ and $0 \leq k < 2^j$ form a basis for $\mathcal{F}_n$. Note that the number of such functions is $1 + \sum_{j=0}^{n-1} 2^j = 2^n$.

The Haar functions may be generated in a systematic way. At stage zero start with the function 1. At stage one take also r_1. At stage two take also $f_{1;0} r_2$ and $f_{1;1} r_2$. At stage three take also $f_{2;0} r_3$ and $f_{2;1} r_3$ and $f_{2;2} r_3$ and $f_{2;3} r^3$.

An important infinite dimensional vector space is

$$L = \bigcup_n \mathcal{F}_n. \tag{9.9}$$

This is the space of all step functions on $(0, 1]$ that have end points that are multiples of $1/2^n$ for some n. This space is spanned by the finite linear combinations of Walsh functions (or of Haar functions) of arbitrary order.

Theorem 9.1 *The Walsh functions form an orthonormal family of vectors with respect to the inner product*

$$\langle f, g \rangle = \lambda(fg) = \int_0^1 f(x) g(x)\, dx. \tag{9.10}$$

For an arbitrary continuous function f on $[0,1]$ the Walsh expansion

$$f(x) = \sum_S \langle w_S, f\rangle w_S(x) \tag{9.11}$$

of f converges uniformly to f on $(0,1]$.

Theorem 9.2 *The Haar functions form an orthonormal family of vectors with respect to the inner product*

$$\langle f, g\rangle = \lambda(fg) = \int_0^1 f(x)g(x)\,dx. \tag{9.12}$$

For an arbitrary continuous function f on $[0,1]$ the Haar expansion

$$f(x) = \sum_{n=-1}^{\infty} \sum_{0\le k<2^n} \langle h_{n;k}, f\rangle h_{n;k}(x) \tag{9.13}$$

of f converges uniformly to f on $(0,1]$.

Proof: The important thing to notice is that these theorems are the same theorem. The partial sum of the Walsh series that is in $\mathcal{F}_n$ is the same as the partial sum of the Haar series that is in $\mathcal{F}_n$. In fact, this is the same as the partial sum of the expansion into rectangular functions that is in $\mathcal{F}_n$. Each of these partial sums can be characterized as the projection of f onto $\mathcal{F}_n$. So the only problem is to demonstrate that this last partial sum converges uniformly to the continuous function f.

However the rectangular functions $f_{n;k}$ for fixed n form an orthogonal basis for $\mathcal{F}_n$. So the error in the expansion is just

$$f(x) - \sum_k 2^n \langle f, f_{n;k}\rangle f_{n;k}(x) = \sum_k 2^n \int_{I_{n;k}} (f(x) - f(t))\,dt f_{n;k}(x), \tag{9.14}$$

where $f_{n;k}$ is the indicator function of $I_{n;k}$. Let $\epsilon > 0$. By uniform continuity, there is an δ so that if $|x-t| < \delta$, then the values $|f(x) - f(t)| < \epsilon$. Take n such that $2^{-n} < \delta$. Then the absolute values of the error is bounded for each x by ϵ. □

9.5 Coin tossing

Consider the set $\Omega = 2^{\mathbf{N}_+}$ of all sequences of zeros and ones indexed by $\mathbf{N}_+ = \{1, 2, 3, \ldots\}$. This is thought of as a sequences of tails and heads, or of failures and successes. Each element ω is called an *outcome* of the coin tossing experiment. For $n \ge 0$, let $\mathcal{F}_n$ be the set of real functions on Ω that depend at most on the first n coordinates.

A *random variable* is a function f from Ω to $\mathbf{R}$ (satisfying certain technical conditions, to be specified later). A random variable is a prescription for determining an experimental number, since the number $f(\omega)$ depends on the actual

result ω of the experiment. Each function in $\mathcal{F}_n$ is a random variable. These are the random variables that may be determined only knowing the results of the first n coin tosses.

One important function in $\mathcal{F}_n$ is the *binary function* $f_{n;k}$ defined for $0 \leq n$ and $0 \leq k < 2^n$. Write k in binary notation. Then $f_{n;k}$ is equal to one on every sequence ω that agrees with the binary digits of k in the first n places. Then the $f_{n;k}$ for $0 \leq k < 2^n$ form a basis for the 2^n dimensional vector space $\mathcal{F}_n$.

Another function is the *Bernoulli function* b_n. This is defined by $b_n(\omega) = \omega_n$. In other words, it is the nth coordinate function. It just measures failure or success on the nth toss of the coin.

The relation between the Bernoulli functions and the binary functions f_{nk} is the following. Write the number k in binary form. Look at the corresponding subset S of $\{1, \ldots, n\}$ and its complement S^c in $\{1, \ldots, n\}$. Then

$$f_{n;k} = \prod_{j \in S} b_j \prod_{j \in S^c} (1 - b_j). \tag{9.15}$$

That is, the binary function is one precisely for those coin tosses that have a particular pattern of successes and failures in the first n trials, without regard to what happens in later trials.

Another important function is the Rademacher function $r_n = 1 - 2b_n$. This is the function that is 1 when the nth trial results in failure and -1 when the nth trial results in success.

A *Walsh function* is a product of Rademacher functions. Let $S \subset \{1, 2, 3, \ldots\}$ be a finite set of strictly positive natural numbers. Let the Walsh function be defined by

$$w_S = \prod_{j \in S} r_j. \tag{9.16}$$

The Walsh functions w_S for $S \subset \{1, \ldots, n\}$ form another basis for $\mathcal{F}_n$. The probability interpretation is that w_S is 1 if there are an even of successes in the subset of trials specified by the set S; otherwise w_S is -1 if there is an odd number of successes in the trials specified by S. If we think of Ω as a commutative group, then the Walsh functions are the homomorphisms of this group to the group $\{1, -1\}$.

A *Haar function* is a product of a binary rectangular function with a Rademacher function. Let $S \subset \{1, \ldots, n\}$. Then for $n \geq 0$ and $0 \leq k < 2^n$ define the Haar function to be

$$h_{n;k} = f_{n;k} r_{n+1}, \tag{9.17}$$

and define $h_{-1} = 1$. The function h_{-1} together with the other Haar functions $h_{j;k}$ for $j = 0$ to $n - 1$ and $0 \leq k < 2^j$ form a basis for $\mathcal{F}_n$. Note that the number of such functions is $1 + \sum_{j=0}^{n-1} 2^j = 2^n$. The probability interpretation is that the Haar function $h_{j;k}$ is non-zero only for a particular pattern in the first j trials (the pattern determined by k), and its sign depends on failure or success in the $j + 1$st trial.

An important infinite dimensional vector space is

$$L = \bigcup_{n=0}^{\infty} \mathcal{F}_n. \tag{9.18}$$

This is the space of all random variables that depend only on some finite number of trials.

The reader will have noticed the close relation between step functions on the interval $(0, 1]$ and functions on the space of coin tossing outcomes. The relation is the following. Let $g : \Omega \to [0, 1]$ be defined by

$$g(\omega) = \sum_{n=1}^{\infty} \frac{\omega_n}{2^n}. \tag{9.19}$$

Then g is a function that is a surjection, but not quite an injection. The points where it fails to be an injection correspond to the end points of the intervals on which the rectangular functions are defined. Away from these points, there is a perfect correspondence between the examples of step functions and the corresponding examples of coin tossing functions.

For instance, a natural example on the coin tossing side is the function $s_n = r_1 + \cdots + r_n$. This is the number of failures minus the number of successes in n trials. This family of random variables as a function of n is sometimes called *random walk*. On the step function side this is a rather complicated function.

9.6 Vector lattices

A set of real functions L is called a *vector space* of functions if the zero function is in L, f in L and $g \in L$ imply that $f + g$ is in L, and a in $\mathbf{R}$ and f in L imply that af is in L. A set of real functions L is called a *lattice* of functions if f in L and g in L imply that the infimum $f \wedge g$ is in L and that the supremum $f \vee g$ is in L. The set L is called a *vector lattice* of functions if it is both a vector space and a lattice.

Notice that if f is in a vector lattice L, then the absolute value given by the formula $|f| = f \vee 0 - f \wedge 0$ is in L.

Examples:

1. The space of real continuous functions defined on an interval $[a, b]$ is a vector lattice.

2. The space of step functions (piecewise constant real functions) defined on an interval $(a, b]$ is a vector lattice.

3. The space of step functions defined on $(0, 1]$ and with end points that are multiples of $1/2^n$ for some n.

4. The space of functions on the coin tossing space that depend only on finitely many coordinates.

As we have seen, the last two examples are intimately related.

9.7 Elementary integrals

Let X be a non-empty set. Let L be a vector lattice of real functions on X. Then μ is an *elementary integral* on L provided that

1. $\mu : L \to \mathbf{R}$ is linear;

2. $\mu : L \to \mathbf{R}$ is order preserving;

3. μ satisfies monotone convergence within L.

To say that μ satisfies monotone convergence within L is to say that if each f_n is in L, and $f_n \uparrow f$, and f is in L, then $\mu(f_n) \uparrow \mu(f)$.

Proposition 9.3 *Suppose that g_n in L and $g_n \downarrow 0$ imply $\mu(g_n) \downarrow 0$. Then μ satisfies monotone convergence within L.*

Proof: Suppose that f_n is in L and $f_n \uparrow f$ and f is in L. Since L is a vector space, it follows that $g_n = f - f_n$ is in L. Furthermore $g_n \downarrow 0$. Therefore $\mu(g_n) \downarrow 0$. This says that $\mu(f_n) \uparrow \mu(f)$. □

9.8 Integration on a product of finite spaces

Let $\Omega = \{0,1\}^{\mathbf{N}_+}$ be the set of all infinite sequences of zeros and ones indexed by $\mathbf{N}_+ = \{1, 2, 3, \ldots\}$. For each $k = 0, 1, 2, 3, \ldots$ consider the set $\mathcal{F}_k$ of functions f on Ω that depend only on the first k elements of the sequence, that is, such that $f(\omega) = g(\omega_1, \ldots, \omega_k)$ for some function g on $\mathbf{R}^k$. This is a vector lattice with dimension 2^k. The vector lattice under consideration will be the space L that is the union of all the $\mathcal{F}_k$ for $k = 0, 1, 2, 3, \ldots$. In the following, we suppose that we have an elementary integral μ on L.

A subset A of Ω is said to be an $\mathcal{F}_k$ set when its indicator function 1_A is in $\mathcal{F}_k$. In such a case we write $\mu(A)$ for $\mu(1_A)$ and call $\mu(A)$ the measure of A. Thus measure is a special case of integral.

In the following we shall need a few simple properties of measure. First, note that $\mu(\emptyset) = 0$. Second, the additivity of the integral implies the corresponding property $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$. In particular, if $A \cap B = \emptyset$, then $\mu(A \cup B) = \mu(A) + \mu(B)$. This is called the additivity of measure. Finally, the order preserving property implies that $A \subset B$ implies $\mu(A) \leq \mu(B)$.

Here is an example. If the function f is in $\mathcal{F}_k$, let

$$\mu(f) = \sum_{\omega_1=0}^{1} \cdots \sum_{\omega_k=0}^{1} f(\omega) \frac{1}{2^k}. \tag{9.20}$$

This is a consistent definition, since if f is regarded as being in $\mathcal{F}_j$ for $k < j$, then the definition involves sums over 2^j sequences, but the numerical factor is $1/2^j$, and the result is the same. This example describes the expectation for independent of tosses of a fair coin. Suppose A is a subset of Ω whose

definition depends only on finitely many coordinates. Then A defines an event that happens or does not happen according to information about finitely many tosses of the coin. The measure $\mu(A) = \mu(1_A)$ is the probability of this event.

The following results shows that such an example automatically satisfies the monotone convergence property and thus gives an elementary integral. The remarkable thing about the proof that follows is that it uses no notions of topology: it is pure measure theory.

Lemma 9.4 *Suppose that L is a vector lattice consisting of bounded functions. Suppose that 1 is an element of L. Suppose furthermore that for each f in L and each real α the indicator function of the set where $f \geq \alpha$ is in L. Suppose that $\mu : L \to \mathbf{R}$ is linear and order preserving. If μ satisfies monotone convergence for sets, then μ satisfies monotone convergence for functions.*

Proof: Suppose that μ satisfies monotone convergence for sets, that is, suppose that $A_n \downarrow \emptyset$ implies $\mu(A_n) \downarrow 0$. Suppose that $f_n \downarrow 0$. Say $f_1 \leq M$. Let $\epsilon > 0$. Choose $\alpha > 0$ so that $\alpha\mu(1) < \epsilon/2$. Let A_n be the set where $f_n \geq \alpha > 0$. Then $f_n \leq \alpha + M1_{A_n}$. Hence $\mu(f_n) \leq \alpha\mu(1) + M\mu(A_n)$. Since $A_n \downarrow \emptyset$, we can choose n so that $M\mu(A_n) < \epsilon/2$. Then $\mu(f_n) < \epsilon$. Since $\epsilon > 0$ is arbitrary, this shows that $\mu(f_n) \downarrow 0$. Thus μ satisfies monotone convergence for functions. □

Theorem 9.5 *Let $\Omega = \{0,1\}^{\mathbf{N}_+}$ be the set of all infinite sequences of zeros and ones. Let $L = \bigcup_{k=0}^{\infty} \mathcal{F}_k$ be the vector lattice of all functions f on Ω that each depend only on the first k elements of the sequence for some k. Suppose that $\mu : L \to \mathbf{R}$ is linear and order preserving. Then μ satisfies monotone convergence within L.*

Proof: By the lemma, it is enough to show that if $A_n \downarrow \emptyset$ is a sequence of sets, each of which is an $\mathcal{F}_k$ set for some k, then $\mu(A_n) \downarrow 0$. The idea is to prove the contrapositive. Suppose then that there is an $\epsilon > 0$ such that $\mu(A_n) \geq \epsilon$ for all n.

Let $\bar{\omega}[k] = (\bar{\omega}_1, \ldots, \bar{\omega}_k)$ be a finite sequence of k zeros and ones. Let

$$B_{\bar{\omega}[k]} = \{\omega \mid \omega_1 = \bar{\omega}_1, \ldots, \omega_k = \bar{\omega}_k\} \tag{9.21}$$

This is the binary set of all sequences in Ω that agree with $\bar{\omega}[k]$ in the first k places. It is an $\mathcal{F}_k$ set. (For $k = 0$ we may regard this as the set of all sequences in Ω.)

The main step in the proof is to show that there is a consistent family of sequences $\bar{\omega}[k]$ such that for each n

$$\mu(A_n \cap B_{\bar{\omega}[k]}) \geq \epsilon \frac{1}{2^k}. \tag{9.22}$$

The proof is by induction. The statement is true for $k = 0$. Suppose the statement is true for k. By additivity

$$\mu(A_n \cap B_{\bar{\omega}[k]}) = \mu(A_n \cap B_{\bar{\omega}[k]0}) + \mu(A_n \cap B_{\bar{\omega}[k]1}). \tag{9.23}$$

Here $\bar{\omega}[k]0$ is the sequence of length $k+1$ consisting of $\bar{\omega}[k]$ followed by a 0. Similarly, $\bar{\omega}[k]1$ is the sequence of length $k+1$ consisting of $\bar{\omega}[k]$ followed by a 0. Suppose that there is an n_1 such that the first term on the right is less than $\epsilon/2^{k+1}$. Suppose also that there is an n_2 such that the second term on the right is less than $\epsilon/2^{k+1}$. Then, since the sets are decreasing with n, there exists an n such that both terms are less than $\epsilon/2^{k+1}$. But then the measure on the left would be less than $\epsilon/2^k$ for this n. This is a contradiction. Thus one of the two suppositions must be false. This says that one can choose $\bar{\omega}[k+1]$ with $\bar{\omega}_{k+1}$ equal to 1 or to 0 so that for all n we have $\mu(A_n \cap B_{\bar{\omega}[k+1]}) \geq \epsilon/2^{k+1}$. This completes the inductive proof of the main step.

The consistent family of finite sequences $\omega[k]$ defines an infinite sequence $\bar{\omega}$. This sequence $\bar{\omega}$ is in each A_n. The reason is that for each n there is a k such that A_n is an $\mathcal{F}_k$ set. Each $\mathcal{F}_k$ set is a disjoint union of a collection of binary sets, each of which consists of the set of all sequences where the first k elements have been specified in some way. The set $B_{\bar{\omega}[k]}$ is such a binary set. Hence either $A_n \cap B_{\bar{\omega}[k]} = \emptyset$ or $B_{\bar{\omega}[k]} \subset A_n$. Since $\mu(A_n \cap B_{\bar{\omega}[k]}) > 0$ the first possibility is ruled out. We conclude that

$$\bar{\omega} \in B_{\bar{\omega}[k]} \subset A_n. \tag{9.24}$$

The last argument proves that there is a sequence $\bar{\omega}$ that belongs to each A_n. Thus it is false that $A_n \downarrow \emptyset$. This completes the proof of the contrapositive. □

Problems

1. It is known that a random walk is an inefficient way to travel. In fact, in n steps a typical amount of progress is on the order of $\sqrt{n}$. This can be made precise as follows. Let $s_n = r_1 + \cdots + r_n$ be the random walk. Express s_n^2 as a linear combination of Walsh functions. What is the constant term in this combination? What is the integral of s_n^2?

2. Let $Q = \{0, 1, 2, \ldots, q-1\}$ and let $\Omega = Q^{\mathbf{N}_+}$. Let L be the subset of $\mathbf{R}^\Omega$ consisting of functions that depend only on finitely many coordinates. Let $\mu : L \to \mathbf{R}$ be linear and order preserving. Show that μ is an elementary integral.

3. Let K be compact. Let A_n be a decreasing sequence of non-empty closed subsets of K. Show that $\bigcap_n A_n$ is non-empty.

4. Prove Dini's theorem. Suppose K is compact. If f_n is a sequence of continuous functions on K and $f_n \downarrow 0$ pointwise as $n \to \infty$, then $f_n \to 0$ uniformly. Hint: Consider $\epsilon > 0$. Let $A_n = \{x \in K \mid f_n(x) \geq \epsilon\}$.

5. Let K be compact, and let $L = C(K)$. Let $\mu : L \to \mathbf{R}$ be linear and order preserving. Show that μ is an elementary integral.

Chapter 10

The integral

10.1 The Daniell construction

This section is an outline of the Daniell construction of the integral. This is a two stage process.

Let X be a non-empty set. Let L be a vector lattice of real functions on X. Then μ is an *elemantary integral* on L provided that

1. $\mu : L \to \mathbf{R}$ is linear;

2. $\mu : L \to \mathbf{R}$ is order preserving;

3. μ satisfies monotone convergence within L.

Let $L\uparrow$ consist of the functions $h : X \to (-\infty, +\infty]$ such that there exists a sequence h_n in L with $h_n \uparrow h$ pointwise. These are the *upper functions*. Similarly, let $L\downarrow$ consist of the functions $f : X \to [-\infty, +\infty)$ such that there exists a sequence f_n in L with $f_n \downarrow f$ pointwise. These are the lower functions. The first stage of the construction is to extend the integral to upper functions and to lower functions.

This terminology of upper functions and lower functions is quite natural, but it may not be ideal in all respects. If L is vector lattice of continuous functions, then the upper functions are lower semicontinuous, while the lower functions are upper semicontinuous.

Lemma 10.1 *There is a unique extension of μ from L to the upper functions $L\uparrow$ that satisfies the upward monotone convergence property: if h_n is in $L\uparrow$ and $h_n \uparrow h$, then h is in $L\uparrow$ and $\mu(h_n) \uparrow \mu(h)$. Similarly, there is a unique extension of μ from L to the lower functions $L\downarrow$ that satisfies the corresponding downward monotone convergence property.*

The second stage of the process is to extend the integral to functions that are approximated by upper and lower functions in a suitable sense. Let g be a

real function on X. Define the *upper integral*

$$\mu^*(g) = \inf\{\mu(h) \mid h \in L\uparrow, g \le h\}. \tag{10.1}$$

Similarly, define the *lower integral*

$$\mu_*(g) = \sup\{\mu(f) \mid f \in L\downarrow, f \le g\}. \tag{10.2}$$

Lemma 10.2 *The upper integral is order preserving and subadditive: $\mu^*(g_1 + g_2) \le \mu^*(g_1) + \mu^*(g_2)$. Similarly, the lower integral is order preserving and superadditive: $\mu_*(g_1 + g_2) \ge \mu_*(g_1) + \mu_*(g_2)$. Furthermore, $\mu_*(g) \le \mu^*(g)$ for all g.*

Define $\mathcal{L}^1(X, \mu)$ to be the set of all $g : X \to \mathbf{R}$ such that both $\mu_*(g)$ and $\mu^*(g)$ are real, and

$$\mu_*(g) = \mu^*(g). \tag{10.3}$$

Let their common value be denoted $\tilde{\mu}(g)$. This $\tilde{\mu}$ is the *integral* on the space $\mathcal{L}^1 = \mathcal{L}^1(X, \mu)$ of μ integrable functions.

We shall see that this extended integral satisfies a remarkable monotone convergence property. The upward version says that if f_n is a sequence in $\mathcal{L}^1$ and $f_n \uparrow f$ pointwise and the $\tilde{\mu}(f_n)$ are bounded above, then f is in $\mathcal{L}^1$ and $\tilde{\mu}(f_n) \uparrow \tilde{\mu}(f)$. There is a similar downward version. The remarkable thing is that the fact that the limiting function f is in $\mathcal{L}^1$ is not a hypothesis but a conclusion.

Theorem 10.3 (Daniell) *Let μ be an elementary integral on a vector lattice L of functions on X. Then the corresponding space $\mathcal{L}^1 = \mathcal{L}^1(X, \mu)$ of μ integrable functions is a vector lattice, and the extension $\tilde{\mu}$ is an elementary integral on it. Furthermore, the integral $\tilde{\mu}$ on $\mathcal{L}^1$ satisfies the monotone convergence property.*

If an indicator function 1_A is in $\mathcal{L}^1$, then $\tilde{\mu}(1_A)$ is written $\tilde{\mu}(A)$ and is called the *measure* of the set A. In the following we shall often write the integral of f in $\mathcal{L}^1$ as $\mu(f)$ and the measure of A with 1_A in $\mathcal{L}^1$ as $\mu(A)$.

In the following corollary we consider a vector lattice L. Let $L\uparrow$ consist of pointwise limits of increasing limits from L, and let $L\downarrow$ consist of pointwise limits of decreasing sequences from L. Similarly, let $L\uparrow\downarrow$ consist of pointwise limits of decreasing sequences from $L\uparrow$, and let $L\downarrow\uparrow$ consist of pointwise limits of increasing sequences from $L\downarrow$.

Corollary 10.4 *Let L be a vector lattice and let μ be an elementary integral. Consider its extension $\tilde{\mu}$ to $\mathcal{L}^1$. Then for every g in $\mathcal{L}^1$ there is a f in $L\downarrow\uparrow$ and an h in $L\uparrow\downarrow$ with $f \le g \le h$ and $\tilde{\mu}(g - f) = 0$ and $\tilde{\mu}(h - g) = 0$.*

This corollary says that if we identify functions in $\tilde{L}^1$ when the integral of the absolute value of the difference is zero, then all the functions that we ever will need may be taken, for instance, from $L\uparrow\downarrow$. However this class is not closed under pointwise limits.

The proof of the theorem has a large number of routine verifications. However there are a few key steps. These will be outlined in the following sections. For more detailed accounts there are several excellent references. One is Lynn H. Loomis, Abstract Harmonic Analysis, van Nostrand, New York, 1953, Chapter III. Another more recent account with more of a probability flavor is Daniel W. Stroock, A Concise Introduction to the Theory of Integration, 3rd edition, Birkhäuser, Boston, 1999.

10.2 Stage one

Begin with a vector lattice L and an elementary integral μ. Let $L\uparrow$ be the set of all pointwise limits of increasing sequences of elements of L. These functions are allowed to take on the value $+\infty$. Similarly, let $L\downarrow$ be the set of all pointwise limits of decreasing sequences of L. These functions are allowed to take on the value $-\infty$. Note that the functions in $L\downarrow$ are the negatives of the functions in $L\uparrow$.

For h in $L\uparrow$, take $h_n \uparrow h$ with h_n in L and define $\mu(h) = \lim_n \mu(h_n)$. The limit of the integral exists because this is a monotone sequence of numbers. Similarly, if f in $L\downarrow$, take $f_n \downarrow f$ with f_n in L and define $\mu(f) = \lim_n \mu(f_n)$.

Lemma 10.5 *The definition of $\mu(h)$ for h in $L\uparrow$ is independent of the sequence. There is a similar conclusion for $L\downarrow$.*

Proof: Say that h_m is in L with $h_m \uparrow h$ and k_n is in L with $k_n \uparrow k$ and $h \le k$. We will show that $\mu(h) \le \mu(k)$. This general fact is enough to establish the uniqueness. To prove it, fix m and notice that $\mu(h_m \wedge k_n) \le \mu(k_n) \le \mu(k)$. However $h_m \wedge k_n \uparrow h_m$ as $n \to \infty$, so $\mu(h_m) \le \mu(k)$. Now take $n \to \infty$; it follows that $\mu(h) \le \mu(k)$. □

Lemma 10.6 *Upward monotone convergence holds for $L\uparrow$. Similarly, downward monotone convergence holds for $L\downarrow$.*

Proof: Here is the argument for upward monotone convergence. Say that the h_n are in $L\uparrow$ and $h_n \uparrow h$ as $n \to \infty$. For each n, let g_{nm} be a sequence of functions in L such that $g_{nm} \uparrow h_n$ as $m \to \infty$. Let $u_n = g_{1n} \vee g_{2n} \vee \cdots \vee g_{nn}$. Then u_n is in L and we have the squeeze inequality

$$g_{in} \le u_n \le h_n \tag{10.4}$$

for $1 \le i \le n$. As $n \to \infty$ the $g_{in} \uparrow h_i$ and the $h_n \uparrow h$. Furthermore, as $i \to \infty$ the $h_i \uparrow h$. By the squeeze inequality $u_n \uparrow h$. From the squeeze inequality we get

$$\mu(g_{in}) \le \mu(u_n) \le \mu(h_n) \tag{10.5}$$

for $1 \le i \le n$. By definition of the integral on $L\uparrow$ we can take $n \to \infty$ and get $\mu(h_i) \le \mu(h) \le \lim_n \mu(h_n)$. Then we can take $i \to \infty$ and get $\lim_i \mu(h_i) \le \mu(h) \le \lim_n \mu(h_n)$. This shows that the integrals converge to the correct value. □

10.3 Stage two

The integral $\mu(g)$ is the supremum of all the $\mu(f)$ for f in $L\downarrow$ with $f \leq g$ and is also the infimum of all the $\mu(h)$ for h in $L\uparrow$ with $g \leq h$. Alternatively, a function g is in $\mathcal{L}^1$ if for every $\epsilon > 0$ there is a function f in $L\downarrow$ and a function h in $L\uparrow$ such that $f \leq g \leq h$, $\mu(f)$ and $\mu(h)$ are finite, and $\mu(h) - \mu(f) < \epsilon$.

It is not hard to show that the set $\mathcal{L}^1$ of absolutely summable functions is a vector lattice and that μ is a positive linear functional on it. The crucial point is that there is also a monotone convergence theorem. This theorem says that if the g_n are absolutely summable functions with $\mu(g_n) \leq M < \infty$ and if $g_n \uparrow g$, then g is absolutely summable with $\mu(g_n) \uparrow \mu(g)$.

Lemma 10.7 *The integral on $\mathcal{L}^1$ satisfies the monotone convergence property.*

Proof: We may suppose that $g_0 = 0$. Since the absolutely summable functions $\mathcal{L}^1$ are a vector space, each $g_n - g_{n-1}$ for $n \geq 1$ is absolutely summable. Consider $\epsilon > 0$. Choose h_n in $L\uparrow$ for $n \geq 1$ such that $g_n - g_{n-1} \leq h_n$ and such that

$$\mu(h_n) \leq \mu(g_n - g_{n-1}) + \frac{\epsilon}{2^n}. \tag{10.6}$$

Let $s_n = \sum_{i=1}^n h_i$ in $L\uparrow$. Then $g_n \leq s_n$ and

$$\mu(s_n) \leq \mu(g_n) + \epsilon \leq M + \epsilon. \tag{10.7}$$

Also $s_n \uparrow s$ in $L\uparrow$ and $g \leq s$. and so by monotone convergence for $L\uparrow$

$$\mu(s) \leq \lim_n \mu(g_n) + \epsilon \leq M + \epsilon. \tag{10.8}$$

Now pick m so large that $g_m \leq g$ satisfies $\mu(s) < \mu(g_m) + \frac{3}{2}\epsilon$. Then pick r in $L\downarrow$ with $r \leq g_m$ so that $\mu(g_m) \leq \mu(r) + \frac{1}{2}\epsilon$. Then $r \leq g \leq s$ with $\mu(s) - \mu(r) < 2\epsilon$. Since ϵ is arbitrary, this proves that g is absolutely summable. Since $g_n \leq g$, it is clear that $\lim_n \mu(g_n) \leq \mu(g)$. On the other hand, the argument has shown that for each $\epsilon > 0$ we can find s in $L\uparrow$ with $g \leq s$ and $\mu(g) \leq \mu(s) \leq \lim_n \mu(g_n) + \epsilon$. Since ϵ is arbitrary, we conclude that $\mu(g) \leq \lim_n \mu(g_n)$. □

The proof of the monotone convergence theorem for the functions in $L\uparrow$ and for the functions in $L\downarrow$ is routine. However the proof of the monotone convergence theorem for the functions in $\mathcal{L}^1$ is deeper. In particular, it uses in a critical way the fact that the sequence of functions is indexed by a countable set of n. Thus the errors in the approximations can be estimated by $\epsilon/2^n$, and these sum to the finite value ϵ.

10.4 Example: Coin tossing

An example to which this result applies is the space Ω of the coin tossing example. Recall that the elementary integral is defined on the space $L = \bigcup_{n=0}^{\infty} \mathcal{F}_n$, where $\mathcal{F}_n$ consists of the functions that depend only on the first n coordinates.

Thus L consists of functions each of which depends only on finitely many coordinates. A subset S of Ω is said to be an $\mathcal{F}_n$ set if its indicator function 1_S belongs to $\mathcal{F}_n$. This just means that the definition of the set depends only on the first n coordinates. In the same way, S is said to be an L set if 1_S is in L.

Consider the elementary integral for fair coin tossing. The elementary integral $\mu(f)$ of a function f in $\mathcal{F}_n$ may be calculated by a finite sum involving at most 2^n terms. It is just the sum of the values of the function for all of the 2^n possibilities for the first n coin flips, divided by 2^n. Similarly, the elementary measure $\mu(S)$ of an $\mathcal{F}_n$ set is the number among the 2^n possibilities of the first n coin flips that are satisfied by S, again weighted by $1/2^n$.

Thus consider for example the measure of the uncountable set S consisting of all ω such that $\omega_1 + \omega_2 + \omega_3 = 2$. If we think of S as an $\mathcal{F}_3$ set, its measure is $3/2^3 = 3/8$. If we think of S as an $\mathcal{F}_4$ set, its measure is still $6/2^4 = 3/8$.

The elementary integral on L extends to an integral on $\mathcal{L}^1$. The integral of a function f in $\mathcal{L}^1$ is then denoted $\mu(f)$. This is interpreted as the expectation of the random variable f. Consider a subset S of Ω such that its indicator function 1_S is in $\mathcal{L}^1$. (Every set that one will encounter in practical computations will have this property.) The measure $\mu(S)$ of S is the integral $\mu(1_S)$ of its indicator function 1_S. This is interpreted as the probability of the event S in the coin tossing experiment.

Proposition 10.8 *Consider the space Ω for infinitely many tosses of a coin, and the associated integral for tosses of a fair coin. Then each subset with exactly one point has measure zero.*

Proof: Consider such a set $\{\bar{\omega}\}$. Let B_k be the set of all ω in Ω such that ω agrees with $\bar{\omega}$ in the first k places. The indicator function of B_k is in L. Since $\{\bar{\omega}\} \subset B_k$, we have $0 \leq \mu(\{\bar{\omega}\}) \leq \mu(B_k) = 1/2^k$ for each k. Hence $\mu(\{\bar{\omega}\}) = 0$. □

Proposition 10.9 *Consider the space Ω for infinitely many tosses of a coin, and the associated integral that gives the expectation for tosses of a fair coin. Let $S \subset \Omega$ be a countable subset. Then the measure of S is zero.*

Proof: Here is a proof from the definition of the integral. Let $j \mapsto \omega^{(j)}$ be an enumeration of S. Let $\epsilon > 0$. For each j let $B^{(j)}$ be a set with indicator function in L such that $\omega^{(j)} \in B^{(j)}$ and $\mu(B^{(j)}) < \frac{\epsilon}{2^j}$. For instance, one can take $B^{(j)}$ to be the set of all ω that agree with $\omega^{(j)}$ in the first k places, where $1/2^k \leq \epsilon/2^j$. Then

$$0 \leq 1_S \leq 1_{\bigcup_j B^{(j)}} \leq \sum_j 1_{B^{(j)}}. \tag{10.9}$$

The right hand side of this equation is in $L^\uparrow$ and has integral bounded by ϵ. Hence $0 \leq \mu(S) \leq \epsilon$. It follows that $\mu(S) = 0$. □

Proof: Here is a proof from the monotone convergence theorem. Let $j \mapsto \omega^{(j)}$ be an enumeration of S. Then

$$\sum_j 1_{\omega^{(j)}} = 1_S. \tag{10.10}$$

By the previous proposition each term in the integral has integral zero. Hence each partial sum has integral zero. By the monotone convergence theorem the sum has integral zero. Hence $\mu(S) = 0$. □

Corollary 10.10 *Consider the space Ω for infinitely many tosses of a coin, and the associated integral that gives the expectation for tosses of a fair coin. Let $S \subset \Omega$ be the set of all sequences that are eventually either all zeros or all ones. Then the measure of S is zero.*

Examples:

1. As a first practical example, consider the function b_j on Ω defined by $b_j(\omega) = \omega_j$, for $j \geq 1$. This scores one for a success in the jth trial. It is clear that b_j is in $\mathcal{F}_j$ and hence in L. It is easy to compute that $\mu(b_j) = 1/2$ for the fair coin μ.

2. A more interesting example is $c_n = b_1 + \cdots + b_n$, for $n \geq 0$. This random variable counts the number of successes in the first n trials. It is a function in $\mathcal{F}_n$ and hence in L. The fair coin expectation of c_n is $n/2$. In n coin tosses the expected number of successes is $n/2$.

3. Consider the set defined by the condition $c_n = k$ for $0 \leq k \leq n$. This is an $\mathcal{F}_n$ set, and its probability is $\mu(c_n = k) = \binom{n}{k} 1/2^n$. This is the famous binomial probability formula. These probabilities add to one:

$$\sum_{k=0}^{n} \binom{n}{k} \frac{1}{2^n} = 1. \tag{10.11}$$

   This formula has a combinatorial interpretation: the total number of subsets of an n element set is 2^n. However the number of subsets with k elements is $\binom{n}{k}$. The formula for the expectation of c_n gives another identity:

$$\sum_{k=0}^{n} k \binom{n}{k} \frac{1}{2^n} = \frac{1}{2} n. \tag{10.12}$$

   This also has a combinatorial interpretation: the total number of ordered pairs consisting of a subset and a point within it is the same as the number of ordered pairs consisting of a point and a subset of the complement, that is, $n2^{n-1}$. However the number of ordered pairs consisting of a k element set and a point within it is $\binom{n}{k} k$.

4. Let $u_1(\omega)$ be the first k such that $\omega_k = 1$. This waiting time random variable is not in L, but for each m with $1 \leq m < \infty$ the event $u_1 = m$ is an $\mathcal{F}^m$ set and hence an L set. The probability of $u_1 = m$ is $1/2^m$. The event $u_1 = \infty$ is not an L set, but it is a one point set, so it has zero probability. This is consistent with the fact that the sum of the probabilities is a geometric series with $\sum_{m=1}^{\infty} 1/2^m = 1$.

5. The random variable $u_1 = \sum_{m=1}^{\infty} m 1_{u_1=m}$ is in $L\uparrow$. Its expectation is $\mu(u_1) = \sum_{m=1}^{\infty} m/2^m = 2$. This says that the expected waiting time to get a success is two tosses.

6. Let $t_n(\omega)$ for $n \geq 0$ be the nth value of k such that $\omega_k = 1$. (Thus $t_0 = 0$ and $t_1 = u_1$.) Look at the event that $t_n = k$ for $1 \leq k \leq n$, which is an $\mathcal{F}_k$ set. This is the same as the event $c_{k-1} = n - 1, b_k = 1$ and so has probability $\binom{k-1}{n-1}1/2^{k-1}1/2 = \binom{k-1}{n-1}1/2^k$. These probabilities add to one, but this is already not such an elementary fact. However the event $t_n = \infty$ is a countable set and thus has probability zero. So in fact

$$\sum_{k=n}^{\infty} \binom{k-1}{n-1} \frac{1}{2^k} = 1. \tag{10.13}$$

This is an infinite series; a combinatorial interpretation is not apparent.

7. For $n \geq 1$ let $u_n = t_n - t_{n-1}$ be the nth waiting time. It is not hard to show that the event $t_{n-1} = k, u_n = m$ has probability $\mu(t_{n-1} = k)1/2^m$, and hence that the event $u_n = m$ has probability $1/2^m$. So u_n also is a geometric waiting time random variable, just like u_1. In particular, it has expectation 2.

8. We have $t_n = u_1 + \cdots + u_n$. Hence the expectation $\mu(t_n) = 2n$. The expected total time to wait until the nth success is $2n$. This gives another remarkable identity

$$\sum_{k=n}^{\infty} k \binom{k-1}{n-1} \frac{1}{2^k} = 2n. \tag{10.14}$$

It would not make much sense without the probability intuition.

10.5 Example: Lebesgue measure

Let Ω be the space of all infinite sequences of zeros and ones. Let $g : \Omega \to [0, 1]$ be the function given by

$$g(s) = \sum_{k=1}^{\infty} \frac{s_k}{2^k}. \tag{10.15}$$

It is then reasonable to define the Lebesgue integral $\lambda(f)$ of a function f on $[0, 1]$ as the fair coin expectation of $f \circ g$ on Ω, provided that the latter integral exists. For this integral we write as usual

$$\lambda(f) = \int_0^1 f(t)\, dt. \tag{10.16}$$

Let S be the subset of all ω in Ω such that the sequence ω is eventually all zeros or all ones. Then g maps S to the binary division points in $[0, 1]$. It also gives a bijection of the complement of S in Ω to the complement of the set of

binary division points in $[0,1]$. The values of f on the binary division points will not matter in determining its integral, since these all are determined by values of $f \circ g$ on the set S, which is of measure zero.

Here is another approach to defining the integral. Consider the vector lattice L of step functions on $[0,1]$ that have the property that for some $k = 1, 2, 3, \ldots$ the function has the form

$$f(x) = c_0 1_{[0,1/2^k]} + \sum_{j=1}^{2^k-1} c_j 1_{(j/2^k,(j+1)/2^k]}(x). \tag{10.17}$$

These are step functions based on binary intervals. (However the first interval includes the left end point.) The integral of such a function is determined by the corresponding coin tossing expectation. Since the points in S do not matter, the resulting elementary integral is

$$\lambda(f) = \sum_{j=0}^{2^k-1} c_j \frac{1}{2^k}. \tag{10.18}$$

Since there is a monotone convergence theorem for the coin tossing integral, there is a corresponding monotone convergence theorem for this elementary integral. Then one can extend the elementary integral to $\mathcal{L}^1([0,1],\lambda)$ by the standard construction.

In this context one can show directly from the definition of the integral that that the Lebesgue measure of a countable set Q is 0. This will involve a two-stage process. Let q_j, $j = 1, 2, 3, \ldots$ be an enumeration of the points in Q. Fix $\epsilon > 0$. For each j, find a binary interval B_j of length less than $\epsilon/2^j$ such that q_j is in the interval. The indicator function 1_{B_j} of each such interval is in L. Let $h = \sum_j 1_{B_j}$. Then h is in $L\uparrow$ and $\lambda(h) \leq \epsilon$. Furthermore, $0 \leq 1_Q \leq h$. This is the first stage of the approximation. Now consider a sequence of $\epsilon > 0$ values that approach zero, and construct in the same way a sequence of h_ϵ such that $0 \leq 1_Q \leq h_\epsilon$ and $\lambda(h_\epsilon) \leq \epsilon$. This is the second stage of the approximation. This shows that the integral of 1_Q is zero.

Notice that this could not have been done in one stage. There is no way to cover Q by finitely many binary intervals of small total length. It was necessary first find infinitely many binary intervals that cover Q and have small total length, and only then let this length approach zero.

One way to define the Lebesgue integral of f on $\mathbf{R}$ is to say that $\mathcal{L}^1(\mathbf{R},\lambda)$ consists of all f such that the restriction of f to each interval $(n, n+1]$ is absolutely summable and

$$\lambda(|f|) = \int_{-\infty}^{\infty} |f(t)|\, dt = \sum_n \int_n^{n+1} |f(t)|\, dt < \infty. \tag{10.19}$$

Then

$$\lambda(f) = \int_{-\infty}^{\infty} f(t)\, dt = \sum_n \int_n^{n+1} f(t)\, dt. \tag{10.20}$$

Problems

1. Let $k \to r_k$ be an enumeration of the rational points in $[0,1]$. Define $g(x) = \sum_k 2^k 1_{\{r_k\}}(x)$. Evaluate the Lebesgue integral of g directly from the definition in terms of integrals of step functions, integrals of lower and upper functions, and integrals of functions squeezed between lower and upper functions.

2. The Cantor set C is the subset of $[0,1]$ that is the image of $\Omega = \{0,1\}^{\mathbf{N}_+}$ under the injection
$$c(\omega) = \sum_{n=1}^{\infty} \frac{2\omega_n}{3^n}. \tag{10.21}$$
The complement of the Cantor set in $[0,1]$ is an open set obtained by removing middle thirds. Show that the indicator function of the complement of the Cantor set is a function in $L\uparrow$. Find the Lebesgue measure of the complement of the Cantor set directly from the definition. Then find the Lebesgue measure of the Cantor set.

3. Let c be the cardinality of the continuum. Show that the cardinality of the set of all real functions on $[0,1]$ is c^c. Show that $c^c = 2^c$.

4. Show that the cardinality of the set of real functions on $[0,1]$ with finite Lebesgue integral is 2^c. Hint: Think about the Cantor set.

5. The Lebesgue integral may be defined starting with the regulated elementary integral λ defined on $L = C([0,1])$. Show that $L\uparrow$ consists of lower semicontinuous functions, and $L\downarrow$ consists of upper semicontinuous functions.

Chapter 11

Measurable functions

11.1 Monotone classes

A set of real functions $\mathcal{F}$ is a monotone class if it satisfies the following two properties. Whenever $f_n \uparrow f$ is an increasing sequence of functions f_n in $\mathcal{F}$ with pointwise limit f, then f is also in $\mathcal{F}$. Whenever $f_n \downarrow f$ is a decreasing sequence of functions f_n in $\mathcal{F}$ with pointwise limit f, then f is also in $\mathcal{F}$.

Theorem 11.1 *Let L be a vector lattice of real functions. Let $\mathcal{F}$ be the smallest monotone class of which L is a subset. Then $\mathcal{F}$ is a vector lattice.*

Proof: The task is to show that $\mathcal{F}$ is closed under addition, scalar multiplication, sup, and inf. Begin with addition. Let f be in L. Consider the set $M(f)$ of functions g such that $f + g$ is in $\mathcal{F}$. This set includes L and is closed under monotone limits. So $\mathcal{F} \subset M(f)$. Thus f in L and g in $\mathcal{F}$ imply $f + g \in \mathcal{F}$. Now let g be in $\mathcal{F}$. Consider the set $\tilde{M}(g)$ of functions f such that $f + g$ is in $\mathcal{F}$. This set includes L and is closed under monotone limits. So $\mathcal{F} \subset \tilde{M}(g)$. Thus f and g in $\mathcal{F}$ implies $f + g$ in $\mathcal{F}$. The proof is similar for the other operations. □

A set of real functions $\mathcal{F}$ is a *Stone vector lattice* of functions if it is a vector lattice and satisfies the property: $f \in \mathcal{F}$ and $a > 0$ imply $f \wedge a \in \mathcal{F}$. The following theorem may be proved by the same monotone class technique.

Theorem 11.2 *Let L be a Stone vector lattice of real functions. Let $\mathcal{F}$ be the smallest monotone class of which L is a subset. Then $\mathcal{F}$ is a Stone vector lattice.*

A set of real functions $\mathcal{F}$ is a *vector lattice with constants* of functions if it is a vector lattice and each constant function belongs to $\mathcal{F}$. The following theorem is trivial, but it may be worth stating the obvious.

Theorem 11.3 *Let L be a vector lattice with constants. Let $\mathcal{F}$ be the smallest monotone class of which L is a subset. Then $\mathcal{F}$ is a vector lattice with constants.*

We shall now see that a monotone class is closed under all pointwise limits.

Theorem 11.4 *Let $\mathcal{F}$ be a monotone class of functions. Let f_n be in $\mathcal{F}$ for each n. Suppose that $\liminf_n f_n$ and $\limsup f_n$ are finite. Then they are also in $\mathcal{F}$.*

Proof: Let $n < m$ and let $h_{nm} = f_n \wedge f_{n+1} \wedge \cdots \wedge f_m$. Then $h_{nm} \downarrow h_n$ as $m \to \infty$, where h_n is the infimum of the f_k for $k \geq n$. However $h_n \uparrow \liminf_n f_n$. □

The trick in this proof is to write a general limit as an increasing limit followed by a decreasing limit. We shall see in the following that this is a very important idea in integration.

11.2 Generating monotone classes

The following theorem says that if L is a vector lattice that generates $\mathcal{F}$ by monotone limits, then the positive functions L^+ generate the positive functions $\mathcal{F}^+$ by monotone limits.

Theorem 11.5 *Let L be a vector lattice of real functions. Suppose that $\mathcal{F}$ is the smallest monotone class that includes L. Let L^+ be the positive elements of L, and let $\mathcal{F}^+$ be the positive elements of $\mathcal{F}$. Then $\mathcal{F}^+$ is the smallest monotone class that includes L^+.*

Proof: It is clear that $\mathcal{F}^+$ includes L^+. Furthermore, $\mathcal{F}^+$ is a monotone class. So all that remains to show is that if $\mathcal{G}$ is a monotone class that includes L^+, then $\mathcal{F}^+$ is a subset of $\mathcal{G}$. For that it is sufficient to show that for each f in $\mathcal{F}$ the positive part $f \vee 0$ is in $\mathcal{G}$.

Consider the set M of f in $\mathcal{F}$ such that $f \vee 0$ is in $\mathcal{G}$. The set L is a subset of M, since f in L implies $f \vee 0$ in L^+. Furthermore, M is a monotone class. To check this, note that if each f_n is in M and $f_n \uparrow f$, then $f_n \vee 0$ is in $\mathcal{G}$ and $f_n \vee 0 \uparrow f \vee 0$, and so $f \vee 0$ is also in $\mathcal{G}$, that is, f is in M. The argument is the same for downward convergence. Hence $\mathcal{F} \subset M$. □

A real function f is said to be *L-bounded* if there is a function g in L^+ with $|f| \leq g$. Say that L consists of bounded functions. Then if f is L-bounded, then f is also bounded. Say on the other hand that the constant functions are in L. Then if f is bounded, it follows that f is L-bounded. However there are also cases when L consists of bounded functions, but the constant functions are not in L. In such cases, being L-bounded is more restrictive.

A set of real functions $\mathcal{H}$ is an *L-bounded monotone class* if it satisfies the following two properties. Whenever $f_n \uparrow f$ is an increasing sequence of L-bounded functions f_n in $\mathcal{H}$ with pointwise limit f, then f is also in $\mathcal{H}$. Whenever $f_n \downarrow f$ is a decreasing sequence of L-bounded functions f_n in $\mathcal{H}$ with pointwise limit f, then f is also in $\mathcal{H}$. Notice that the functions in $\mathcal{H}$ do not have to be L-bounded.

The following theorem says that if L^+ generates $\mathcal{F}^+$ by monotone limits, then L^+ generates $\mathcal{F}^+$ using only monotone limits of L-bounded functions.

Theorem 11.6 *Let L be a vector lattice of bounded real functions that includes the constant functions. Let $\mathcal{F}^+$ be the smallest monotone class of which L^+ is a subset. Let $\mathcal{H}$ be the smallest L-bounded monotone class of which L^+ is a subset. Then $\mathcal{H} = \mathcal{F}^+$.*

Proof: It is clear that $\mathcal{H} \subset \mathcal{F}^+$. The task is to prove that $\mathcal{F}^+ \subset \mathcal{H}$.

Consider $g \geq 0$ be in L^+. Let $M(g)$ be the set of all f in $\mathcal{F}^+$ such that $f \wedge g$ is in $\mathcal{H}$. It is clear that $L^+ \subset M(g)$. If $f_n \uparrow f$ and each f_n is in $M(g)$, then $f_n \wedge g \uparrow f \wedge g$. Since each $f_n \wedge g$ is in $\mathcal{H}$ and is L-bounded, it follows that $f \wedge g$ is in $\mathcal{H}$. Thus $M(g)$ is closed under upward monotone convergence. Similarly, $M(g)$ is closed under downward monotone convergence. Therefore $\mathcal{F}^+ \subset M(g)$. This establishes that for each f in $\mathcal{F}_+$ and g in L^+ it follows that $f \wedge g$ is in $\mathcal{H}$.

Now consider the set of all f in $\mathcal{F}$ such that there exists h in $L\uparrow$ with $f \leq h$. Certainly L belongs to this set. Furthermore, this set is monotone. This is obvious for downward monotone convergence. For upward monotone convergence, it follows from the fact that $L\uparrow$ is closed under upward monotone convergence. It follows that every element in $\mathcal{F}$ is in this set.

Let f be in $\mathcal{F}^+$. Then there exists h in $L\uparrow$ such that $f \leq h$. There exists h_n in L^+ with $h_n \uparrow h$. Then $f \wedge h_n$ is in $\mathcal{H}$, by the first part of the proof. Furthermore, $f \wedge h_n \uparrow f$. It follows that f is in $\mathcal{H}$. This completes the proof that $\mathcal{F}^+ \subset \mathcal{H}$. □

11.3 Sigma-algebras of functions

A *σ-algebra of functions* $\mathcal{F}$ is a vector lattice of functions that is a monotone class and that includes the constant functions. A function f in $\mathcal{F}$ is said to be *measurable*. The reason for this terminology will be discussed below in connection with the generation of σ-algebras.

Let a be a real number and let f be a function. Define the function $1_{f>a}$ to have the value 1 at all points where $f > a$ and to have the value 0 at all points where $f \leq a$.

Theorem 11.7 *Let $\mathcal{F}$ be a σ-algebra of functions. If f is in $\mathcal{F}$ and a is real, then $1_{f>a}$ is in $\mathcal{F}$.*

Proof: The function $f - f \wedge a$ is in $\mathcal{F}$. It is strictly greater than zero at precisely the points where $f > a$. The sequence of functions $g_n = n(f - f \wedge a)$ satisfies $g_n \uparrow \infty$ for points where $f > a$ and $g_n = 0$ at all other points. Hence the family of functions $g_n \wedge 1$ increases pointwise to $1_{f>a}$. □

Theorem 11.8 *Let $\mathcal{F}$ be a σ-algebra of functions. If f is a function, and if for each real number $a > 0$ the function $1_{f>a}$ is in $\mathcal{F}$, then f is in $\mathcal{F}$.*

Proof: First note that for $0 < a < b$ the function $1_{a<f\leq b} = 1_{f>a} - 1_{f>b}$ is also in $\mathcal{F}$. Next, consider the numbers $c_{nk} = k/2^n$ for $n \in \mathbf{N}$ and $k \in \mathbf{Z}$. This

divides the real axis into intervals of length $1/2^n$. Then

$$f_n = \sum_{k=-\infty}^{\infty} a_{nk} 1_{c_{nk} < f \leq c_{n\,k+1}} \tag{11.1}$$

is in $\mathcal{F}$. However $f_n \uparrow f$ as $n \to \infty$. □

Theorem 11.9 *If f is in $\mathcal{F}$, then so is f^2.*

Proof: Since $\mathcal{F}$ is a lattice, f in $\mathcal{F}$ implies $|f|$ in $\mathcal{F}$. For $a \geq 0$ the condition $f^2 > a$ is equivalent to the condition $|f| > \sqrt{a}$. On the other hand, for $a < 0$ the condition$f^2 > a$ is always satisfied. □

Theorem 11.10 *Let $\mathcal{F}$ be a σ-algebra of functions. If f, g are in $\mathcal{F}$, then so is the pointwise product fg.*

Proof: Since $\mathcal{F}$ is a vector space, it follows that $f + g$ and $f - g$ are in $\mathcal{F}$. However $4fg = (f + g)^2 - (f - g)^2$. □

This last theorem shows that $\mathcal{F}$ is not only closed under addition, but also under multiplication. Thus $\mathcal{F}$ deserves to be called an algebra. It is called a σ-algebra because of the closure under monotone limits.

Examples:

1. An important example of a σ-algebra of functions is the monotone class of functions on $[0, 1]$ generated by the step functions. This is the same as the monotone class generated by the continuous functions. In order to prove that these are the same, one should prove that each step function can be obtained by monotone limits of continuous functions and that each continuous function is a monotone limit of step functions. It is clear that the constant functions belong to the monotone class. This σ algebra is known as the Borel σ-algebra $\mathcal{B}$ of functions on $[0, 1]$.

2. Another example of a σ-algebra of functions is the monotone class of functions on $\mathbf{R}$ generated by the step functions with compact support. This is the same as the monotone class of functions on $\mathbf{R}$ generated by the continuous functions with compact support. These provide examples where the generating classes do not contain the constant functions, but the corresponding monotone class does contain the constant functions. This σ algebra is known as the Borel σ-algebra $\mathcal{B}$ of functions on $\mathbf{R}$.

Such a Borel σ algebra $\mathcal{B}$ of functions is huge; in fact, it is difficult to think of a real function that does not belong to $\mathcal{B}$. However it is possible to show that the cardinality of $\mathcal{B}$ is c, while the cardinality of the σ algebra of all real functions is 2^c.

11.4 Generating sigma-algebras

If we are given a set S of functions, then the σ-algebra of functions $\sigma(S)$ generated by this set is the smallest σ-algebra of functions that contains the original set. The Borel σ-algebra $\mathcal{B}$ of functions on $\mathbf{R}$ is generated by the single function x. Similarly, the Borel σ-algebra of functions on $\mathbf{R}^k$ is generated by the coordinates $x_1, \ldots, x_k$. The following theorem shows that measurable functions are closed under nonlinear operations in a very strong sense.

Theorem 11.11 *Let $f_1, \ldots, f_k$ be functions on X. Let $\mathcal{B}$ be the σ-algebra of Borel functions on $\mathbf{R}^k$. Let*

$$\mathcal{G} = \{\phi(f_1, \ldots, f_k) \mid \phi \in \mathcal{B}\}. \tag{11.2}$$

The conclusion is that $\sigma(f_1, \ldots, f_k) = \mathcal{G}$. That is, the σ-algebra of functions generated by $f_1, \ldots, f_k$ consists of the Borel functions of the functions in the generating set.

Proof: First we show that $\mathcal{G} \subset \sigma(f_1, \ldots, f_k)$. Let $\mathcal{B}'$ be the set of functions ϕ such that $\phi(f_1, \ldots, f_k) \in \sigma(f_1, \ldots, f_k)$. Each coordinate function x_j of $\mathbf{R}^n$ is in $\mathcal{B}'$, since this just says that f_j is in $\sigma(f_1, \ldots, f_k)$. Furthermore, $\mathcal{B}'$ is a σ-algebra. This is a routine verification. For instance, here is how to check upward monotone convergence. Suppose that ϕ_n is in $\mathcal{B}'$ for each n. Then $\phi_n(f_1, \ldots, f_k) \in \sigma(f_1, \ldots, f_k)$ for each n. Suppose that $\phi_n \uparrow \phi$ pointwise. Then $\phi_n(f_1, \ldots, f_k) \uparrow \phi(f_1, \ldots, f_k)$, so $\phi(f_1, \ldots, f_k) \in \sigma(f_1, \ldots, f_k)$. Thus ϕ is in $\mathcal{B}'$. Since $\mathcal{B}'$ is a σ-algebra containing the coordinate functions, it follows that $\mathcal{B} \subset \mathcal{B}'$. This shows that $\mathcal{G} \subset \sigma(f_1, \ldots, f_k)$.

Now we show that $\sigma(f_1, \ldots, f_k) \subset \mathcal{G}$. It is enough to show that $\mathcal{G}$ contains $f_1, \ldots, f_k$ and is a σ-algebra of functions. The first fact is obvious. To show that $\mathcal{G}$ is a σ-algebra of functions, it is necessary to verify that it is a vector lattice with constants and is closed under monotone convergence. The only hard part is the monotone convergence. Suppose that $\phi_n(f_n, \ldots, f_k) \uparrow g$ pointwise. The problem is to find a Borel function ϕ such that $g = \phi(f_1, \ldots, f_k)$. There is no way of knowing whether the Borel functions ϕ_n converge on all of $\mathbf{R}^k$. However let G be the subset of $\mathbf{R}^k$ on which ϕ_n converges. Then G also consists of the subset of $\mathbf{R}^k$ on which ϕ_n is a Cauchy sequence. So

$$G = \bigcap_j \bigcup_N \bigcap_{m \geq N} \bigcap_{n \geq N} \{x \mid |\phi_m(x) - \phi_n(x)| < 1/j\} \tag{11.3}$$

is a Borel set. Let ϕ be the limit of the ϕ_n on G and $\phi = 0$ on the complement of G. Then ϕ is a Borel function. Next note that the range of $f_1, \ldots, f_k$ is a subset of G. So $\phi_n(f_1, \ldots, f_k) \uparrow \phi(f_1, \ldots, f_k) = g$. $\square$

Corollary 11.12 *Let $f_1, \ldots, f_n$ be in a σ-algebra $\mathcal{F}$ of measurable functions. Let ϕ be a Borel function on $\mathbf{R}^n$. Then $\phi(f_1, \ldots, f_n)$ is also in $\mathcal{F}$.*

Proof: From the theorem $\phi(f_1, \ldots, f_n) \in \sigma(f_1, \ldots, f_n)$. Since $\mathcal{F}$ is a σ-algebra and $f_1, \ldots, f_n$ are in $\mathcal{F}$, it follows that $\sigma(f_1, \ldots, f_n) \subset \mathcal{F}$. Thus $\phi(f_1, \ldots, f_n) \in \mathcal{F}$. □

This discussion illuminates the use of the term measurable for elements of a σ-algebra. The idea is that there is a starting set of functions S that are regarded as those quantities that may be directly measured in some experiment. The σ-algebra $\sigma(S)$ consists of all functions that may be computed as the result of the direct measurement and other mathematical operations. Thus these are all the functions that are measurable. Notice that the idea of what is possible in mathematical computation is formalized by the concept of Borel function.

This situation plays a particularly important role in probability theory. For instance, consider the σ-algebra of functions $\sigma(S)$ generated by the functions in S. There is a concept of conditional expectation of a random variable f given S. This is a numerical prediction about f when the information about the values of the functions in S is available. This conditional expectation will be a function in $\sigma(S)$, since it is computed by the mathematical theory of probability from the data given by the values of the functions in S.

Problems

1. Let $\mathcal{B}$ be the smallest σ-algebra of real functions on $\mathbf{R}$ containing the function x. This is called the σ-algebra of Borel functions. Show by a direct construction that every continuous function is a Borel function.

2. Show that every monotone function is a Borel function.

3. Can a Borel function be discontinuous at every point?

4. Let $\sigma(x^2)$ be the smallest σ-algebra of functions on $\mathbf{R}$ containing the function x^2. Show that $\sigma(x^2)$ is not equal to $\mathcal{B} = \sigma(x)$. Which algebra of measurable functions is bigger (that is, which one is a subset of the other)?

5. Consider the σ-algebras of functions generated by $\cos(x)$, $\cos^2(x)$, and $\cos^4(x)$. Compare them with the σ-algebras in the previous problem and with each other. (Thus specify which ones are subsets of other ones.)

11.5 Sigma-rings of functions

This section may be omitted on a first reading.

Occasionally one needs a slightly more general concept of measurable function. A set of real functions $\mathcal{F}$ is a *σ-ring of functions* if it is a Stone vector lattice of functions that is also a monotone class.

Every σ-algebra of functions is a σ-ring of functions. A simple example of a σ-ring of functions that is not a σ-algebra of functions is given by the set of all real functions on X that are each non-zero on a countable set. If X is uncountable, then the constant functions do not belong to this σ-ring.

Theorem 11.13 *Let L be a Stone vector lattice of real functions. Let $\mathcal{F}$ be the smallest monotone class of which F is a subset. Then $\mathcal{F}$ is a σ-ring of functions.*

Let a be a real number and let f be a function. Define the function $1_{f>a}$ to have the value 1 at all points where $f > a$ and to have the value 0 at all points where $f \leq a$.

Theorem 11.14 *Let $\mathcal{F}$ be a σ-ring of functions. If f is in $\mathcal{F}$ and $a > 0$ is real, then $1_{f>a}$ is in $\mathcal{F}$.*

Proof: The function $f - f \wedge a$ is in $\mathcal{F}$. It is strictly greater than zero at precisely the points where $f > a$. The sequence of functions $g_n = n(f - f \wedge a)$ satisfies $g_n \uparrow \infty$ for points where $f > a$ and $g_n = 0$ at all other points. Hence the family of functions $g_n \wedge 1$ increases pointwise to $1_{f>a}$. □

Theorem 11.15 *Let $\mathcal{F}$ be a σ-ring of functions. If $f \geq 0$ is a function, and if for each real number $a > 0$ the function $1_{f>a}$ is in $\mathcal{F}$, then f is in $\mathcal{F}$.*

Proof: First note that for $0 < a < b$ the function $1_{a<f\leq b} = 1_{f>a} - 1_{f>b}$ is also in $\mathcal{F}$. Next, consider the numbers $c_{nk} = k/2^n$ for $n \in \mathbf{N}$ and $k \in \mathbf{Z}$. This divides the real axis into intervals of length $1/2^n$. Let $a_{nk} = \exp(c_{nk})$. This is a corresponding division of the strictly positive half line. Then

$$f_n = \sum_{k=-\infty}^{\infty} a_{nk} 1_{a_{nk}<f\leq a_{n\,k+1}} \tag{11.4}$$

is in $\mathcal{F}$. However $f_n \uparrow f$ as $n \to \infty$. □

Theorem 11.16 *If $f \geq 0$ is in $\mathcal{F}$, then so is f^2.*

Proof: For $a > 0$ the condition $f^2 > a$ is equivalent to the condition $f > \sqrt{a}$. □

Theorem 11.17 *Let $\mathcal{F}$ be a σ-ring of functions. If f, g are in $\mathcal{F}$, then so is the pointwise product fg.*

Proof: Since $\mathcal{F}$ is a vector space, it follows that $f + g$ and $f - g$ are in $\mathcal{F}$. Since $\mathcal{F}$ is a lattice, it follows that $|f + g|$ and $|f - g|$ are in $\mathcal{F}$. However $4fg = (f + g)^2 - (f - g)^2 = |f + g|^2 - |f - g|^2$. □

This last theorem shows that $\mathcal{F}$ is not only closed under addition, but also under multiplication. Thus $\mathcal{F}$ deserves to be called a ring. It is called a σ-ring because of the closure under monotone limits.

Theorem 11.18 *Let $\mathcal{F}_0$ be a σ-ring of functions that is not a σ-algebra of functions. Then the set $\mathcal{F}$ of all functions $f + a$, where f is in $\mathcal{F}_0$ and a is a constant function, is the smallest σ-algebra including $\mathcal{F}_0$.*

Proof: Suppose $\mathcal{F}_0$ is not a σ-algebra. The problem is that the only constant function in $\mathcal{F}_0$ is 0. While f in $\mathcal{F}_0$ implies that the indicator function $1_{f \neq 0}$ is in $\mathcal{F}_0$, it will not be the case that the indicator function $1_{f=0}$ is in $\mathcal{F}_0$.

In this case define $\mathcal{F}$ to consists of all sums $f + a$ of elements f of $\mathcal{F}_0$ with constant functions. If $f + a = g + b$, then $f - g = b - a$, and so $a = b$. Thus f and a are uniquely determined by $f + a$.

The next task is to show that $\mathcal{F}$ is indeed a σ-algebra of functions. It is easy to see that it is a vector space. To verify that it is a lattice, it is necessary to check that $(f + a) \wedge (g + b)$ and $(f + a) \vee (g + b)$ are in $\mathcal{F}$. The indicator functions $1_{f \neq 0}$ and $1_{g \neq 0}$ are in $\mathcal{F}_0$. Let h be the supremum of these two indicator functions, so $h = 1$ precisely where $f \neq 0$ or $g \neq 0$. The indicator function h is also in $\mathcal{F}_0$. Then $(f + a) \wedge (g + b) = [(f + ah) \wedge (g + bh) - (a + b)h] + (a + b)$. So $(f + a) \wedge (g + b)$ is in $\mathcal{F}$. The argument is the same for $(f + a) \vee (g + b)$.

The remaining thing to check is that $\mathcal{F}$ is closed under monotone convergence. If f_n is a countable family of functions in $\mathcal{F}_0$, then the set where some $f_n \neq 0$ cannot be the whole space. It follows that if each f_n is in $\mathcal{F}_0$ and $f_n + a_n \uparrow f + a$, then $f_n \uparrow f$ and $a_n \to a$. So f is in $\mathcal{F}_0$ and $f + a$ is in $\mathcal{F}$. □

11.6 Rings and algebras of sets

This section is mainly for reference. It may be omitted on a first reading.

This will be a brief description of the ideas under consideration in the language of sets. Let X be a set. A *ring of sets* is a collection of subsets $\mathcal{R}$ such that the empty set is in $\mathcal{R}$ and such that $\mathcal{R}$ is closed under the operations of union and relative complement.

A ring of sets $\mathcal{A}$ is an *algebra of sets* if in addition the set X belongs to $\mathcal{A}$. Thus the empty set belongs to $\mathcal{A}$ and it is closed under the operations of union and complement. To get from a ring of sets to an algebra of sets, it is enough to put in the complements of the sets in the ring.

An example of a ring of sets is the ring $\mathbf{R}$ of subsets of $\mathbf{R}$ generated by the intervals $(a, b]$ with $a < b$. This consists of the collection of sets that are finite unions of such intervals. Another example is the ring $\mathbf{R}_0$ of sets generated by the intervals $(a, b]$ such that either $a < b < 0$ or $0 < a < b$. None of the sets in this ring have the number 0 as a member.

Proposition 11.19 *Let $\mathcal{R}$ be a ring of sets. Then the set of finite linear combinations of indicator functions 1_A with A in $\mathcal{R}$ is a Stone vector lattice.*

Proposition 11.20 *Let $\mathcal{A}$ be a algebra of sets. Then the set of finite linear combinations of indicator functions 1_A with A in $\mathcal{A}$ is a vector lattice including the constant functions.*

A ring of sets is a *σ-ring of sets* if it is closed under countable unions. Similarly, an algebra of sets is a *σ-algebra of sets* if it is closed under countable unions.

An example of a σ-ring of sets that is not a σ-algebra of sets is the set of all countable subsets of an uncountable set X. The smallest σ-algebra including this σ-ring consists of all subsets that are either countable or have countable complement.

A standard example of a σ-algebra of sets is the Borel σ-algebra $\mathcal{B}$ of subsets of $\mathbf{R}$ generated by the intervals $(a, +\infty)$ with $a \in \mathbf{R}$. A corresponding standard example of a σ-ring that is not a σ-algebra is the σ-ring $\mathcal{B}_0$ consisting of all Borel sets A such that $0 \notin A$.

Recall that a σ-ring of functions is a Stone vector lattice of functions that is also a monotone class. Similarly, a σ-algebra of functions is a vector lattice of functions including the constant functions that is also a monotone class. Recall also that a σ-ring of functions or a σ-algebra of functions is automatically closed not only under the vector space and lattice operations, but also under pointwise multiplication. In addition, there is closure under pointwise limits (not necessarily monotone).

Proposition 11.21 *Let $\mathcal{F}_0$ be a σ-ring of real functions on X. Then the sets A such that 1_A are in $\mathcal{F}_0$ form a σ-ring $\mathcal{R}_0$ of subsets of X.*

Proposition 11.22 *Let $\mathcal{F}$ be a σ-algebra of real functions on X. Then the sets A such that 1_A are in $\mathcal{F}$ form a σ-algebra $\mathcal{R}$ of subsets of X.*

Let $\mathcal{R}_0$ be a σ-ring of subsets of X. Let $f : X \to \mathbf{R}$ be a function. Then f is said to be *measurable* with respect to $\mathcal{R}_0$ if for each B in $\mathcal{B}_0$ the inverse image $f^{-1}[B]$ is in $\mathcal{R}_0$.

Similarly, let $\mathcal{R}$ be a σ-algebra of subsets of X. Let $f : X \to \mathbf{R}$ be a function. Then f is said to be *measurable* with respect to $\mathcal{R}$ if for each B in $\mathcal{B}$ the inverse image $f^{-1}[B]$ is in $\mathcal{R}$.

To check that a function is measurable, it is enough to check the inverse image property with respect to a generating class. For $\mathcal{B}$ this could consist of the intervals $(a, +\infty)$ where a is in $\mathbf{R}$. Thus to prove a function f is measurable with respect to a σ-algebra $\mathcal{R}$, it would be enough to show that for each $a > 0$ the set where $f > a$ is in $\mathcal{R}$. For $\mathcal{B}_0$ a generating class could consist of the intervals $(a, +\infty)$ with $a > 0$ together with the intervals $(-\infty, a)$ with $a < 0$.

Proposition 11.23 *Let $\mathcal{R}_0$ be a σ-ring of subsets of X. Then the collection $\mathcal{F}_0$ of real functions on X that are measurable with respect to $\mathcal{R}_0$ is a σ-ring of functions on X.*

Proposition 11.24 *Let $\mathcal{R}$ be a σ-algebra of subsets of X. Then the collection $\mathcal{F}_0$ of real functions on X that are measurable with respect to $\mathcal{R}$ is a σ-algebra of functions on X.*

Notice that there are really two concepts: measurability with respect to a σ-ring of sets and measurability with respect to a σ-algebra of sets. The latter is by far the most commonly encountered.

Chapter 12

The integral on measurable functions

12.1 Integration

For integration it is often convenient to extend the real number system to include extra points $+\infty$ and $-\infty$. This is natural from the point of view of order, but not so pleasant in terms of algebra. The biggest problem is with addition. For a an extended real number with $a \neq \mp\infty$ we can define $a + (\pm\infty) = (\pm\infty) + a = \pm\infty$. There is however a fundamental problem with $(+\infty) + (-\infty)$ and with $(-\infty) + (+\infty)$, which are undefined.

For $a > 0$ we have $a \cdot (\pm\infty) = (\pm\infty) \cdot a = \pm\infty$, and for $a < 0$ we have $a \cdot (\pm\infty) = (\pm\infty) \cdot a = \mp\infty$. In particular $-(\pm\infty) = \mp\infty$. Thus the addition problem translates to a subtraction problem for expressions like $(+\infty) - (+\infty)$, which are undefined.

For multiplication there is a special problem with a product such as $0 \cdot (\pm\infty)$ or $(\pm\infty) \cdot 0$. In some contexts in the theory of measure and integral this is regarded as the limit of $0 \cdot n$ as $n \to \pm\infty$, so it has value 0. However in general it is an ambiguous expression, and appropriate care is necessary. Certainly there is a lack of continuity, since with this convention the limit of $(1/n) \cdot (\pm)\infty = \pm\infty$ as $n \to \infty$ is not equal to $0 \cdot (\pm\infty) = 0$.

The starting point is a σ-algebra $\mathcal{F}$ of real functions on a non-empty set X. A real function g is said to be measurable with respect to $\mathcal{F}$ if $g \in \mathcal{F}$. A set A is said to be measurable with respect to $\mathcal{F}$ if 1_A is in $\mathcal{F}$.

[It must be said that many treatments refer instead to a σ-algebra of subsets $\mathcal{F}$. In that treatment a set A is measurable if it belongs to $\mathcal{F}$. A real function g is measurable with respect to $\mathcal{F}$ if the inverse image of each set in the Borel σ-algebra of sets is in the σ-algebra $\mathcal{F}$ of sets. Needless to say, these are equivalent points of view.]

An *integral* is a function $\mu : \mathcal{F}^+ \to [0, +\infty]$ defined on the positive elements $\mathcal{F}^+$ of a σ-algebra $\mathcal{F}$ of functions on X. It must satisfy the following properties:

1. μ is additive and respects scalar multiplication by positive scalars.

2. μ satisfies the upward monotone convergence property.

The first property says that $\mu(0) = 0$ and that for $f \geq 0$ and $g \geq 0$ we always have $\mu(f + g) = \mu(f) + \mu(g)$. Furthermore, for $a > 0$ and $f \geq 0$ we have $\mu(af) = a\mu(f)$.

The equation $\mu(af) = a\mu(f)$ is also true for $a = 0$, but this requires a convention that $0(+\infty) = 0$ in this context.

The second property says that if each $f_n \geq 0$ and $f_n \uparrow f$ as $n \to \infty$, then $\mu(f_n) \uparrow \mu(f)$ as $n \to \infty$. This is usually called the *upward monotone convergence theorem*.

Measure is a special case of integral. If 1_A is in $\mathcal{F}$, then the measure of A is the number $\mu(A)$ with $0 \leq \mu(A) \leq +\infty$ defined to be equal to the integral $\mu(1_A)$. Because of the intimate connection between measure and integral, some people refer to an integral as a measure. This is most common in situations when there is more than one integral involved. For instance, for each elementary integral on a suitable space of continuous functions there is an associated integral. The elementary integral is usually called a measure, though it has nothing particular to do with sets.

[It should also be said that many treatments begin with a measure defined on a σ-algebra of subsets of X. It is then shown that such a measure defines an integral on positive measurable functions on X. Thus the notions of measure on measurable subsets and of integral on measurable functions are entirely equivalent.]

Theorem 12.1 *If $0 \leq f \leq g$, then $0 \leq \mu(f) \leq \mu(g)$.*

Proof: Clearly $(g-f)+f = g$. So $\mu(g-f)+\mu(f) = \mu(g)$. But $\mu(g-f) \geq 0$. □

If f in $\mathcal{F}$ is a measurable function, then its positive part $f_+ = f \vee 0 \geq 0$ and its negative part $f_- = -f \wedge 0 \geq 0$. So they each have integrals. If either $\mu(f_+) < +\infty$ or $\mu(f_-) < +\infty$, then we may define the integral of $f = f_+ - f_-$ to be

$$\mu(f) = \mu(f_+) - \mu(f_-). \tag{12.1}$$

The possible values for this integral are real, $+\infty$, or $-\infty$. However, if both $\mu(f_+) = +\infty$ and $\mu(f_-) = +\infty$, then the integral is not defined. The expression $(+\infty) - (+\infty) = (+\infty) + (-\infty)$ is ambiguous! This is the major flaw in the theory, and it is responsible for most challenges in applying the theory of integration.

Theorem 12.2 *Suppose that $\mu(f_-) < +\infty$ and $\mu(g_-) < +\infty$. Then $\mu(f+g) = \mu(f) + \mu(g)$.*

Proof: Let $h = f + g$. Then $h_+ \leq f_+ + g_+$ and $h_- \leq f_- + g_-$. So under the hypothesis of the theorem $\mu(h_-) < +\infty$. Furthermore, from $h_+ - h_- = f_+ - f_- + g_+ - g_-$ it follows that $h_+ + f_- + g_- = h_- + f_+ + g_+$. Since these

are all positive functions $\mu(h_+) + \mu(f_-) + \mu(g_-) = \mu(h_-) + \mu(f_+) + \mu(g_+)$. However then $\mu(h_+) - \mu(h_-) = \mu(f_+) - \mu(f_-) + \mu(g_+) - \mu(g_-)$. This is allowed because the terms that are subtracted are not infinite. The conclusion is that $\mu(h) = \mu(f) + \mu(g)$. □

Theorem 12.3 *If f is in $\mathcal{F}$ and $\mu(|f|) = \mu(f_+) + \mu(f_-) < \infty$, then $\mu(f) = \mu(f_+) - \mu(f_-)$ is defined, and*

$$|\mu(f)| \leq \mu(|f|). \tag{12.2}$$

If f is in $\mathcal{F}$, then f is said to be *absolutely summable* with respect to μ if $\mu(|f|) = \mu(f_+) + \mu(f_-) < \infty$. The space $\mathcal{L}^1(X, \mathcal{F}, \mu)$ is defined as the space of functions in $\mathcal{F}$ that are absolutely summable with respect to μ. So μ is defined on all of $\mathcal{L}^1(X, \mathcal{F}, \mu)$.

Theorem 12.4 *(improved monotone convergence) If $\mu(f_1) > -\infty$ and $f_n \uparrow f$, then $\mu(f_n) \uparrow \mu(f)$. Similarly, if $\mu(h_1) < +\infty$ and $h_n \downarrow h$, then $\mu(h_n) \downarrow \mu(h)$.*

Proof: For the first apply monotone convergence to $f_n - f_1$. For the second let $f_n = -h_n$. □

It is very common to denote

$$\mu(f) = \int f \, d\mu \tag{12.3}$$

or even

$$\mu(f) = \int f(x) \, d\mu(x). \tag{12.4}$$

This notation is suggestive in the case when there is more than one integral in play. Say that ν is an integral, and $w \geq 0$ is a measurable function. Then the integral $\mu(f) = \nu(fw)$ is defined. We would write this as

$$\int f(x) \, d\mu(x) = \int f(x) \, w(x) d\nu(x). \tag{12.5}$$

So the relation between the two integrals would be $d\mu(x) = w(x)d\nu(x)$. This suggests that $w(x)$ plays the role of a derivative of one integral with respect to the other.

12.2 Uniqueness of integrals

Theorem 12.5 *Let L be a vector lattice. Let m be an elementary integral on L. Let $\mathcal{F}$ be the monotone-class generated by L. Suppose that $\mathcal{F}$ contains the constant functions, so that $\mathcal{F}$ is a σ-algebra of functions. If there is an integral μ on $\mathcal{F}^+$ such that $\mu = m$ on L^+, then this extension is unique.*

Proof: Let μ_1 and μ_2 be two integrals on $\mathcal{F}^+$ that each agree with m on L^+. Let $\mathcal{H}$ be the smallest L-monotone class such that $L^+ \subset \mathcal{H}$. Let $\mathcal{G}$ be the set of all functions in $\mathcal{F}^+$ on which μ_1 and μ_2 agree. The main task is to show that $\mathcal{H} \subset \mathcal{G}$. It is clear that $L \subset \mathcal{G}$. Suppose that h_n is in $\mathcal{G}$ and $h_n \uparrow h$. If $\mu_1(h_n) = \mu_2(h_n)$ for each n, then $\mu_1(h) = \mu_2(h)$. Suppose that f_n is in $\mathcal{G}$ and is L-bounded for each n and $f_n \downarrow f$. If $\mu_1(f_n) = \mu_2(f_n)$ for all n, then by improved monotone convergence $\mu_1(f) = \mu_2(f)$. This shows that $\mathcal{G}$ is a L-monotone class such that $L^+ \subset \mathcal{G}$. It follows that $\mathcal{H} \subset \mathcal{G}$. However the earlier result on L-monotone classes showed that $\mathcal{H} = \mathcal{F}^+$. So $\mathcal{F}^+ \subset \mathcal{G}$. □

12.3 Existence of integrals

Theorem 12.6 *Let L be a vector lattice. Let m be an elementary integral on L. Let $\mathcal{F}$ be the monotone-class generated by L. Suppose that $\mathcal{F}$ contains the constant functions, so that $\mathcal{F}$ is a σ-algebra of functions. Then there is an integral μ on $\mathcal{F}^+$ that agrees with m on L^+.*

Proof: Consider the space $\mathcal{L}^1 = \mathcal{L}^1(X, m) \cap \mathcal{F}$. These are the functions that are integrable as before, and also belong to the σ-algebra. Consider $g \geq 0$ in $\mathcal{L}^1$. First we show that if $h \geq 0$ is in $\mathcal{F}^+$, then $h \wedge g$ is in $\mathcal{L}^1$. This follows from the fact that the class of h in $\mathcal{F}_0^+$ such $h \wedge g$ is in $\mathcal{L}^1$ includes L^+ and is a monotone class. Therefore this class includes all of $\mathcal{F}^+$.

It follows that if f and g are in $\mathcal{F}$ and $0 \leq f \leq g$, and g in in $\mathcal{L}^1$, then f is in $\mathcal{L}^1$. This is just because $f = f \wedge g$.

We conclude that for $f \geq 0$ in $\mathcal{F}^+$ we can define $\mu(f)$ to be as before for f in $\mathcal{L}^1$ and $\mu(f) = +\infty$ if f is not in $\mathcal{L}^1$. Then we have that $0 \leq f \leq g$ implies $0 \leq \mu(f) \leq \mu(g)$. The other properties of the integral are easy to check. □

12.4 Probability and expectation

An integral is a *probability integral* (or *expectation*) provided that $\mu(1) = 1$. This of course implies that $\mu(c) = c$ for every real constant c. In this context there is a special terminology. The set on which the functions are defined is called Ω. A point ω in Ω is called an *outcome*.

A measurable function $f : \Omega \to \mathbf{R}$ is called a *random variable*. The value $f(\omega)$ is regarded as an experimental number, the value of the random variable when the outcome of the experiment is ω. The integral $\mu(f)$ is the expectation of the random variable, provided that the integral exists. For a bounded measurable function f the expectation $\mu(f)$ always exists.

A subset $A \subset \Omega$ is called an *event*. When the outcome $\omega \in A$, the event A is said to happen. The measure $\mu(A)$ of an event is called the *probability* of the event. The probability $\mu(A)$ of an event A is the expectation $\mu(1_A)$ of the random variable 1_A that is one if the event happens and is zero if the event does not happen.

Theorem 12.7 *Let $\Omega = \{0,1\}^{\mathbf{N}_+}$ be the set of all infinite sequences of zeros and ones. Fix p with $0 \le p \le 1$. If the function f on Ω is in the space $\mathcal{F}_k$ of functions that depend only on the first k values of the sequence, let $f(\omega) = h(\omega_1, \ldots, \omega_k)$ and define*

$$\mu_p(f) = \sum_{\omega_1=0}^{1} \cdots \sum_{\omega_k=0}^{1} h(\omega_1, \ldots, \omega_k) p^{\omega_1}(1-p)^{1-\omega_1} \cdots p^{\omega_k}(1-p)^{1-\omega_k}. \quad (12.6)$$

This defines an elementary integral μ_p on the vector lattice L that is the union of the $\mathcal{F}_k$ for $k = 0, 1, 2, 3, \ldots$. Let $\mathcal{F}$ be the σ-algebra generated by L. Then the elementary integral extends to an integral μ_p on $\mathcal{F}^+$, and this integral is uniquely defined.

This theorem describes the expectation for a sequence of independent coin tosses where the probability of heads on each toss is p and the probability of tails on each toss is $1 - p$. The special case $p = 1/2$ describes a fair coin. The proof of the theorem follows from previous considerations. It is not difficult to calculate that μ is consistently defined on L. It is linear and order preserving on the coin tossing vector lattice L, so it is automatically an elementary integral. Since L contains the constant functions, the integral extends uniquely to the σ-algebra $\mathcal{F}$.

This family of integrals has a remarkable property. For each p with $0 \le p \le 1$ let $F_p \subset \Omega$ be defined by

$$F_p = \{\omega \in \Omega \mid \lim_{n\to\infty} \frac{\omega_1 + \cdots + \omega_n}{n} = p\}. \quad (12.7)$$

It is clear that for $p \neq p'$ the sets F_p and $F_{p'}$ are disjoint. This gives an uncountable family of disjoint measurable subsets of Ω. The remarkable fact is that for each p we have that the probability $\mu_p(F_p) = 1$. (This is the famous strong law of large numbers.) It follows that for $p' \neq p$ we have that the probability $\mu_p(F_{p'}) = 0$. Thus there are uncountably many expectations μ_p. These are each defined with the same set Ω of outcomes and the same σ-algebra $\mathcal{F}$ of random variables. Yet they are concentrated on uncountably many different sets.

12.5 Image integrals

There are several ways of getting new integrals from old ones. One is by using a weight function. For instance, if

$$\lambda(f) = \int_{-\infty}^{\infty} f(x)\, dx \quad (12.8)$$

is the Lebesgue integral defined for Borel functions f, and if $w \ge 0$ is a Borel function, then

$$\mu(f) = \int_{-\infty}^{\infty} f(x)w(x)\, dx \quad (12.9)$$

is another integral. In applications w can be a mass density, a probability density, or the like.

A more important method is to map the integral forward. For instance, let $y = \phi(x) = x^2$. Then the integral μ described just above maps to an integral $\nu = \phi[\mu]$ given by

$$\nu(g) = \int_{-\infty}^{\infty} g(x^2) w(x)\, dx. \tag{12.10}$$

This is a simple and straightforward operation. Notice that the forward mapped integral lives on the range of the mapping, that is, in this case, the positive real axis. The trouble begins only when one wants to write this new integral in terms of the Lebesgue integral. Thus we may also write

$$\nu(g) = \int_{0}^{\infty} g(y) \frac{1}{2\sqrt{y}} [w(\sqrt{y}) + w(-\sqrt{y})]\, dy. \tag{12.11}$$

Here is the same idea in a general setting. Let $\mathcal{F}$ be a σ-algebra of measurable functions on X. Let $\mathcal{G}$ be a σ-algebra of measurable functions on Y. A function $\phi : X \to Y$ is called a *measurable map* if for every g in $\mathcal{G}$ the composite function $g \circ \phi$ is in $\mathcal{F}$.

Given an integral μ defined on $\mathcal{F}$, and given a measurable map $\phi : X \to Y$, there is an integral $\phi[\mu]$ defined on $\mathcal{G}$. It is given by

$$\phi[\mu](g) = \mu(g \circ \phi). \tag{12.12}$$

It is called the *image* of the integral μ under ϕ.

This construction is important in probability theory. Let Ω be a measure space equipped with a σ-algebra of functions $\mathcal{F}$ and an expectation μ defined on $\mathcal{F}^+$. If ϕ is a random variable, that is, a measurable function from Ω to $\mathbf{R}$ with the Borel σ-algebra, then it may be regarded as a measurable map. The image of the expectation μ under X is an integral $\nu = \phi[\mu]$ on the Borel σ-algebra called the *distribution* of ϕ. We have the identity.

$$\mu(h(\phi)) = \nu(h) = \int_{-\infty}^{\infty} h(x)\, d\nu(x). \tag{12.13}$$

Sometimes the calculations do not work so smoothly. The reason is that while an integral maps forward, a differential form maps backward. For instance, the differential form $g(y)\, dy$ maps backward to the differential form $g(\phi(x))\phi'(x)\, dx$. Thus a differential form calculation like

$$\int_a^b g(\phi(x))\phi'(x)\, dx = \int_{\phi(a)}^{\phi(b)} g(y)\, dy \tag{12.14}$$

works very smoothly. On the other hand, the calculation of an integral with a change of variable with $\phi'(x) > 0$ gives

$$\int_a^b g(\phi(x))\, dx = \int_{\phi(a)}^{\phi(b)} g(y) \frac{1}{\phi'(\phi^{-1}(y))}\, dy \tag{12.15}$$

which is unpleasant. The problem is not with the integral, which is perfectly well defined by the left hand side with no restrictions on the function ϕ. The difficulty comes when one tries to express the integral in terms of a Lebesgue integral with a weight function, and it is only at this stage that the differential form calculations play a role.

The ultimate source of this difficulty is that integrals (or measures) and differential forms are different kinds of objects. An integral assigns a number to a function. Functions map backward, so integrals map forward. Thus g pulls back to $g \circ \phi$, so μ pushes forward to $\phi[\mu]$. The value of $\phi[\mu]$ on g is the value of μ on $g \circ \phi$. (It makes no difference if we think instead of measures defined on subsets, since subsets map backwards and measures map forward.) A differential form assigns a number to an oriented curve. Curves map forward, so differential forms map backward. Thus a curve from a to b pushes forward to a curve from $\phi(a)$ to $\phi(b)$. The differential form $g(y)\,dy$ pulls back to the differential form $g(\phi(x))\phi'(x)\,dx$. The value of $g(\phi(x)\phi'(x)\,dx$ over the curve from a to b is the value of $g(y)\,dy$ over the curve from $\phi(a)$ to $\phi(b)$.

12.6 The Lebesgue integral

The image construction may be used to define Lebesgue measure and other measures.

Theorem 12.8 *Let $0 \leq p \leq 1$. Define the expectation μ_p for coin tossing on the set Ω of all infinite sequences $\omega : \mathbf{N}_+ \to \{0, 1\}$ as in the theorem. Here p is the probability of heads on each single toss. Let*

$$\phi(\omega) = \sum_{k=1}^{\infty} \omega(k) \frac{1}{2^k}. \tag{12.16}$$

Then the image expectation $\phi[\mu]$ is an expectation ν_p defined for Borel functions on the unit interval $[0, 1]$.

The function ϕ in this case is a random variable that rewards the nth coin toss by $1/2^n$ if it results in heads, and by zero if it results in tails. The random variable is the sum of all these rewards. Thus ν_p is the distribution of this random variable.

When $p = 1/2$ (the product expectation for tossing of a fair coin) the expectation $\lambda = \nu_{\frac{1}{2}}$ is the *Lebesgue integral* for functions on $[0, 1]$. However note that there are many other integrals, for the other values of p. We have the following amazing fact. For each p there is an integral ν_p defined for functions on the unit interval. If $p \neq p'$ are two different parameters, then there is a measurable set that has measure 1 for the ν_p measure and measure 0 for the $\nu_{p'}$ measurable. The set comes from the set of coin tosses for which the sample means converge to the number p. This result shows that these measures each live in a different world.

From now on we take this as the definition of the Lebesgue integral for Borel functions on the unit interval $[0, 1]$ and use standard notation, such as

$$\lambda(f) = \int_0^1 f(u)\, du. \tag{12.17}$$

Consider the map $x = \psi(u) = \ln(u/(1-u))$ from the open interval $(0, 1)$ to $\mathbf{R}$. This is a bijection. It has derivative $dx/du = 1/(u(1-u))$. The inverse is $u = 1/(1 - e^{-x})$ with derivative $u(1-u) = 1/(2 + 2\cosh(x))$. It is a transformation that is often used in statistics to relate problems on the unit interval $(0, 1)$ and on the line $(-\infty, +\infty)$. The image of Lebesgue measure for $[0, 1]$ under this map is also a probability integral. It is given by

$$\psi[\lambda](f) = \int_0^1 f(\ln\left(\frac{u}{1-u}\right))\, du = \int_{-\infty}^{\infty} f(x)\, \frac{1}{2}\frac{1}{1+\cosh(x)}\, dx. \tag{12.18}$$

A variation of this idea may be used to define the Lebesgue integral for Borel functions defined on the real line $\mathbf{R}$. Let

$$\sigma(h) = \int_0^1 h(u)\frac{1}{u(1-u)}\, du. \tag{12.19}$$

This is not a probability integral. The image under ψ is

$$\psi[\sigma](f) = \int_0^1 f(\ln(\frac{u}{1-u}))\frac{1}{u(1-u)}\, du = \int_{-\infty}^{\infty} f(x)\, dx = \lambda(f). \tag{12.20}$$

This calculation shows that the dx integral is the image of the $1/(u(1-u))\, du$ integral under the transformation $x = \ln(u/(1-u))$. It could be taken as a perhaps somewhat unusual definition of the Lebesgue integral λ for functions on the line.

12.7 The Lebesgue-Stieltjes integral

Once we have the Lebesgue integral defined for Borel functions on the line, we can construct a huge family of other integrals, also defined on Borel functions on the line. These are called Lebesgue-Stieltjes integrals. Often when several integrals are being discussed, the integrals are referred to as measures. Of course an integral defined on functions does indeed define a measure on subsets.

The class of measures under consideration consists of those measures defined on Borel functions on the line (or on Borel subsets of the line) that give finite measure to bounded Borel sets. Such a measure will be called a *regular* Borel measure.

Examples:

1. The first example is given by taking a function $w \geq 0$ such that w in integrable over each bounded Borel set. The measure is then $\mu(f) = \lambda(fw) = \int_{\infty}^{\infty} f(x)w(x)\,dx$. Such a measure is called absolutely continuous with respect to Lebesgue measure. Often the function w is called the *density* (of mass or probability).

2. Another kind of example is of the form $\mu(f) = \sum_{p \in S} c_p f(p)$, where S is a countable subset of the line, and each $c_p > 0$. This is called the measure that assigns point mass c_p to each point p in S. We require that $\sum_{a<p\leq b} c_p < \infty$ for each a, b with $-\infty < a < b < +\infty$. Often the measure that assigns mass one to a point p is denoted δ_p, so $\delta_p(f) = f(p)$. With this notation the measure μ of this example is $\mu = \sum_{p \in S} c_p \delta_p$.

Suppose that μ is a regular Borel measure, so that the measure of each set $(a, b]$ for $a \leq b$ real is finite. Define $F(x) = \mu((0, x])$ for $x \geq 0$ and $F(x) = -\mu((x, 0])$ for $x \leq 0$. Then $F((a, b]) = F(b) - F(a)$ for all $a \leq b$. The function F is increasing and right continuous. With this normalization $F(0) = 0$, but one can always add a constant to F and still get the property that $F((a, b]) = F(b) - F(a)$. This nice thing about this is that the increasing right continuous function F gives a rather explicit description of the measure μ.

Examples:

1. For the absolutely continuous measure $F(b) - F(a) = \int_a^b w(x)\,dx$. The function F is a continuous function. However not every continuous function is absolutely continuous.

2. For the point mass measure $F(b) - F(a) = \sum_{p \in (a,b]} c_p$. The function F is continuous except for jumps at the points p in S.

Theorem 12.9 *Let F be an increasing right continuous function on $\mathbf{R}$. Then there exists a regular measure μ_F defined on the Borel σ-algebra $\mathcal{B}$ such that*

$$\mu_F((a, b]) = F(b) - F(a). \tag{12.21}$$

Furthermore, this measure may be obtained as the image of Lebesgue measure on an interval under a map G.

Proof: Let $m = \inf F$ and let $M = \sup F$. For $m < y < M$ let

$$G(y) = \sup\{x \mid F(x) < y\}. \tag{12.22}$$

We can compare the least upper bound $G(y)$ with an arbitrary upper bound c. Thus $G(y) \leq c$ is equivalent to the condition that for all x, $F(x) < y$ implies $x \leq c$. This in turn is equivalent to the condition that for all x, $c < x$ implies $y \leq F(x)$. Since F is increasing and right continuous, it follows that this in turn is equivalent to the condition that $y \leq F(c)$.

It follows that $a < G(y) \leq b$ is equivalent to $F(a) < y \leq F(b)$. Thus G is a kind of inverse to F.

Let λ be Lebesgue measure on the interval (m, M). Let $\mu_F = G[\lambda]$ be the image of this Lebesgue measure under G. Then

$$\mu_F((a,b]) = \lambda(\{y \mid a < G(y) \le b\}) = \lambda(\{y \mid F(a) < y \le F(b)\}) = F(b) - F(a), \tag{12.23}$$

so μ_F is the desired measure. □

The above proof says that every regular Borel measure with a certain total mass $M - m$ may be obtained from Lebesgue measure with the same mass. The forward mapping G just serves to redistribute the mass.

Often the Lebesgue-Stieltges integral is written

$$\mu_F(h) = \int_{-\infty}^{\infty} h(x)\, dF(x). \tag{12.24}$$

This is reasonable, since if F were smooth with smooth inverse G we would have $F(G(y)) = y$ and $F'(G(y))G'(y) = 1$ and so

$$\mu_F(h) = \lambda(h \circ G) = \int_m^M h(G(y))\, dy = \int_m^M h(G(y))F'(G(y))G'(y)\, dy = \int_{-\infty}^{\infty} h(x)F'(x)\, dx. \tag{12.25}$$

However in general it is not required that F be smooth, or that it have an inverse function.

These increasing functions F give a relatively concrete description of the regular Borel measures. There are three qualitatively different situations. If the function $F(x)$ is the indefinite integral of a function $w(x)$, then F and μ_F are said to be *absolutely continuous* with respect to Lebesgue measure. (In this case, it is reasonable to write $w(x) = F'(x)$. However $F(x)$ need not be differentiable at each point x. The example when $w(x)$ is a rectangle function provides an example.) If the function F is constant except for jumps at a countable set of points, F and μ_F are said to have *point masses*. The third situation is intermediate and rather strange: the function F has no jumps, but it is constant except on a set of measure zero. In this case F and μ_F are said to be *singular continuous*.

Here is an example of the singular continuous case. Let $\mu_{\frac{1}{2}}$ be the fair coin measure on the space Ω of all sequences of zeros and ones. Let $\chi(\omega) = \sum_{n=1}^{\infty} 2\omega_n/3^n$. Then χ maps Ω bijectively onto the Cantor set. Thus $\chi[\mu_{\frac{1}{2}}]$ is a probability measure on the line that assigns all its weight to the Cantor set. The function F that goes with this measure is called the Cantor function. It is a continuous function that increases from zero to one, yet it is constant except on the Cantor set, which has measure zero.

The Cantor function F is the distribution function of the random variable χ, that is, $F(x) = \mu_{\frac{1}{2}}(\chi \le x)$. It also has a simple non-probabilistic description. To see this, recall that ϕ defined by $\phi(\omega) = \sum_{n=1}^{\infty} \omega_n/2^n$ has a uniform distribution on $[0, 1]$. This says that $\mu_{\frac{1}{2}}(\phi \le y) = y$ for all y in $[0, 1]$. For x in the Cantor set there is a unique $\chi^{-1}(x)$ in Ω and a corresponding $y = \phi(\chi^{-1}(x))$ in $[0, 1]$. The set where $\phi \le y$ is the same as the set where $\chi \le x$, up to a set of measure

zero. Therefore $F(x) = \mu_{\frac{1}{2}}(\chi \le x) = \mu_{\frac{1}{2}}(\phi \le y) = y$. The conclusion is that F restricted to the Cantor set is $\phi \circ \chi^{-1}$, and F is constant elsewhere.

The conclusion of this discussion is that there are many regular Borel measures on Borel subsets of the line. However there is a kind of unity, since each of these is an image of Lebesgue measure on some interval.

Problems

1. Suppose the order preserving property $f \le g$ implies $\mu(f) \le \mu(g)$ is known for positive measurable functions. Show that it follows for all measurable functions, provided that the integrals exist. Hint: Decompose the functions into positive and negative parts.

2. Consider the space $\Omega = \{0,1\}^{\mathbf{N}_+}$ with the measure μ that describes fair coin tossing. Let S_3 be the random variable given by $S_3(\omega) = \omega_1 + \omega_2 + \omega_3$ that describes the number of heads in the first three tosses. Draw the graph of the corresponding function on the unit interval. Find the area under the graph, and check that this indeed gives the expectation of the random variable.

3. Let
$$\mu(g) = \int_{-\infty}^{\infty} g(t)w(t)\,dt \tag{12.26}$$
be a integral defined by a density $w(t) \ge 0$ with respect to Lebesgue measure dt. Let $\phi(t)$ be a suitable smooth function that is increasing or decreasing on certain intervals, perhaps constant on other intervals. Show that the image integral
$$\phi[\mu](f) = \int_{-\infty}^{\infty} f(s)h(s)\,ds + \sum_{s^*} f(s^*)c(s^*) \tag{12.27}$$
is given by a density $h(s)$ and perhaps also some point masses $c(s^*)\delta_{s^*}$. Here
$$h(s) = \sum_{\phi(t)=s} w(t)\frac{1}{|\phi'(t)|} \tag{12.28}$$
and
$$c(s^*) = \int_{\phi(t)=s^*} w(t)\,dt. \tag{12.29}$$

4. What is the increasing right continuous function that defines the integral
$$\mu(g) = \int_{-\infty}^{\infty} g(x)\frac{1}{\pi}\frac{1}{1+x^2}\,dx \tag{12.30}$$
involving the Cauchy density?

5. What is the increasing right continuous function that defines the δ_a integral given by $\delta_a(g) = g(a)$?

12.8 Integrals on a σ-ring

This entire section may be omitted on a first reading.

Theorem 12.10 *Let L be a Stone vector lattice. Let m be an elementary integral on L. Let $\mathcal{F}_0$ be the monotone-class generated by L. Then $\mathcal{F}_0$ is a σ-ring of functions. If there is an extension of μ to an integral on $\mathcal{F}_0^+$ such that $\mu = m$ on L^+, then this extension is unique.*

Theorem 12.11 *Let L be a Stone vector lattice. Let m be an elementary integral on L. Let $\mathcal{F}_0$ be the monotone-class generated by L. Then $\mathcal{F}_0$ is a σ-ring of functions. If $\mathcal{F}_0$ is a σ-algebra of functions, set $\mathcal{F} = \mathcal{F}_0$. If $\mathcal{F}$ is not a σ-algebra of functions, then set $\mathcal{F}$ to be all sums of a function of $\mathcal{F}_0$ with a constant function. Then $\mathcal{F}$ is a σ-algebra of functions, and there is an integral μ on $\mathcal{F}^+$ that agrees with m on L^+.*

Proof: The construction used before gives the desired integral on $\mathcal{F}_0$. . The only problem comes in the (rather unusual) case when $\mathcal{F}_0$ is not a σ-algebra. In this case define $\mathcal{F}$ to consists of all sums $f + a$ of elements f in $\mathcal{F}_0$ with constant functions. The integral may now be defined on $\mathcal{F}^+$ by defining $\mu(f + a) = +\infty$ if $a > 0$. In some cases there are other possible extensions, but this one always works. □

Problems

1. Let X be a set. Let L be the vector lattice of functions that are non-zero only on finite sets. The elementary integral m is defined by $m(f) = \sum_{x \in S} f(x)$ if $f \neq 0$ on S. Find the σ-ring of functions $\mathcal{F}_0$ generated by L. When is it a σ-algebra? Extend m to an integral μ on the smallest σ-algebra generated by L. Is the value of μ on the constant functions uniquely determined?

2. Consider the previous problem. The largest possible σ-algebra of functions on X consists of all real functions on X. For $f \geq 0$ in this largest σ-algebra define the integral μ by $\mu(f) = \sum_{x \in S} f(x)$ if f is non-zero on a countable set S. Otherwise define $\mu(f) = +\infty$. Is this an integral?

3. Let X be a set. Let A be a countable subset of S, and let p be a function on A with $p(x) \geq 0$ and $\sum_{x \in A} p(x) = 1$. Let L be the vector lattice of functions that are non-zero only on finite sets. The probability sum is defined for f in L by $\mu(f) = \sum_{x \in A \cap S} f(x)p(x)$ if $f \neq 0$ on S. Let $\mathcal{F}_0$ be the σ-ring of functions generated by L. Show that if X is uncountable, then μ has more than one extension to the σ-algebra $\mathcal{F}$ consisting of the sum of functions in $\mathcal{F}_0$ with constant functions. Which extension is natural for probability theory?

Chapter 13

Integrals and measures

13.1 Terminology

Here is a terminology review. A measurable space is a space X with a given σ-algebra $\mathcal{F}$ of real functions or the corresponding σ-algebra of subsets. A function or subset is called measurable if it belongs to the appropriate σ-algebra. Suppose X is a metric space. Then the Borel σ-algebra of functions on X is generated by the continuous real functions, while the Borel σ-algebra of subsets of X is generated by the open subsets. A function or subset that is measurable with respect to the appropriate Borel σ-algebra is a Borel function or a Borel subset.

An integral is defined as a function $\mu : \mathcal{F}^+ \to [0, +\infty]$ from the positive functions in a sigma-algebra $\mathcal{F}$ of functions to positive extended real numbers that satisfies certain algebraic conditions and upward monotone convergence. Given an integral, there is an associated function $\mu : \mathcal{L}^1(X, \mathcal{F}, \mu) \to \mathbf{R}$ defined on absolutely integrable functions. A measure is defined as a function from a sigma-algebra of subsets to $[0, +\infty]$ that satisfies countable additivity. There is a one-to-one natural correspondence between integrals and measures, so the concepts are interchangeable. Sometimes this general concept of integration is called the Lebesgue theory.

Probability theory is the special case $\mu(1) = 1$. A function in the σ-algebra is called a random variable, an integral is an expectation, a subset in the sigma-algebra is called an event, a measure is a probability measure.

The integral for Borel functions on the line or the measure on Borel subsets of the line that is translation invariant and has the customary normalization is called the Lebesgue integral or Lebesgue measure and denoted λ, so

$$\lambda(f) = \int_{-\infty}^{\infty} f(x)\, dx. \tag{13.1}$$

There is a similar Lebesgue integral and Lebesgue measure for $\mathbf{R}^n$.

An integral μ_F on Borel functions on the line or the measure on Borel subsets of the line that is given by an increasing right continuous function F is called a

Lebesgue-Stieltjes integral or measure. Thus

$$\mu_F(f) = \int_{-\infty}^{\infty} f(x)\, dF(x). \tag{13.2}$$

These are the measures on Borel subsets of the line that are finite on compact sets. Thus Lebesgue measure is the special case corresponding to the function $F(x) = x$.

There are also integrals and measures defined on larger classes of functions or subsets. Often one refers to the completion of Lebesgue measure on $\mathbf{R}^n$. Sometimes this is called Lebesgue measure on the Lebesgue measurable subsets of $\mathbf{R}^n$.

The Daniell-Stone construction starts from an elementary integral on a Stone vector lattice L and constructs an integral associated with a sigma-algebra of functions $\mathcal{F}$. The Caratheodory construction starts with a countably additive set function on a ring of subsets and constructs a measure on a sigma-algebra of subsets.

A Radon measure is an elementary integral defined on the space L of continuous functions with compact support. For the case of functions on $\mathbf{R}^n$ the Radon measures correspond to the measures on Borel subsets that are finite on compact sets. So for functions on the line they coincide with the Lebesgue-Stieltjes integrals.

13.2 Convergence theorems

The most fundamental convergence theorem is improved monotone convergence. This was proved in the last chapter, but it is well to record it again here.

Theorem 13.1 *(improved monotone convergence) If $\mu(f_1) > -\infty$ and $f_n \uparrow f$, then $\mu(f_n) \uparrow \mu(f)$. Similarly, if $\mu(h_1) < +\infty$ and $h_n \downarrow h$, then $\mu(h_n) \downarrow \mu(h)$.*

The next theorem is a consequence of monotone convergence that applies to a sequence of functions that is not monotone.

Theorem 13.2 *(Fatou) Suppose each $f_n \geq 0$. Let $f = \liminf_{n\to\infty} f_n$. Then*

$$\mu(f) \leq \liminf_{n\to\infty} \mu(f_n). \tag{13.3}$$

Proof: Let $r_n = \inf_{k\geq n} f_k$. It follows that $0 \leq r_n \leq f_k$ for each $k \geq n$. So $0 \leq \mu(r_n) \leq \mu(f_k)$ for each $k \geq n$. This gives the inequality

$$0 \leq \mu(r_n) \leq \inf_{k\geq n} \mu(f_k). \tag{13.4}$$

However $0 \leq r_n \uparrow f$. By monotone convergence $0 \leq \mu(r_n) \uparrow \mu(f)$. Therefore passing to the limit in the inequality gives the result. □

Fatou's lemma says that in the limit one can lose positive mass density, but one cannot gain it.

Examples:

1. Consider functions $f_n = n1_{(0,1/n)}$ on the line. It is clear that $\lambda(f_n) = 1$ for each n. On the other hand, $f_n \to 0$ pointwise, and $\lambda(0) = 0$. The density has formed a spike near the origin, and this does not produce a limiting density.

2. Consider functions $f_n = 1_{(n,n+1)}$. It is clear that $\lambda(f_n) = 1$ for each n. On the other hand, $f_n \to 0$ pointwise, and $\lambda(0) = 0$. The density has moved off to $+\infty$ and is lost in the limit.

It is natural to ask where the mass has gone. The only way to answer this is to reinterpret the problem as a problem about measure. Define the measure $\nu_n(\phi) = \lambda(\phi f_n)$. Take ϕ bounded and continuous. Then it is possible that $\nu_n(\phi) \to \nu(\phi)$ as n to ∞. If this happens, then ν may be interpreted as a limiting measure that contains the missing mass. However this measure need not be given by a density.

Examples:

1. Consider functions $nf_n = 1_{(0,1/n)}$ on the line. In this case $\nu_n(\phi) = \lambda(\phi f_n) \to \phi(0) = \delta_0(\phi)$. The limiting measure is a point mass at the origin.

2. Consider functions $f_n = 1_{(n,n+1)}$. Suppose that we consider continuous functions with right and left hand limits at $+\infty$ and $-\infty$. In this case $\nu_n(\phi) = \lambda(\phi f_n) \to \phi(+\infty) = \delta_{+\infty}(\phi)$. The limiting measure is a point mass at $+\infty$.

Theorem 13.3 *(dominated convergence) Let $|f_n| \leq g$ for each n, where g is in $\mathcal{L}^1(X, \mathcal{F}, \mu)$, that is, $\mu(g) < \infty$. Suppose $f_n \to f$ pointwise as $n \to \infty$. Then f is in $\mathcal{L}^1(X, \mathcal{F}, \mu)$ and $\mu(f_n) \to \mu(f)$ as $n \to \infty$.*

This theorem is amazing because it requires only pointwise convergence. The only hypothesis is the existence of the dominating function

$$\forall n \forall x \, |f_n(x)| \leq g(x) \tag{13.5}$$

with

$$\int g(x) \, d\mu(x) < +\infty. \tag{13.6}$$

Then pointwise convergence

$$\forall x \lim_{n \to \infty} f_n(x) = f(x) \tag{13.7}$$

implies convergence of the integrals

$$\lim_{n \to \infty} \int f_n(x) \, d\mu(x) = \int f(x) \, d\mu(x). \tag{13.8}$$

Proof: We have $|f_k| \le g$, so $-g \le f_k \le g$. Let $r_n = \inf_{k \ge n} f_k$ and $s_n = \sup_{k \ge n} f_k$. Then

$$-g \le r_n \le f_n \le s_n \le g. \tag{13.9}$$

This gives the inequality

$$-\infty < -\mu(g) \le \mu(r_n) \le \mu(f_n) \le \mu(s_n) \le \mu(g) < +\infty. \tag{13.10}$$

However $r_n \uparrow f$ and $s_n \downarrow f$. It follows from improved monotone convergence that $\mu(r_n) \uparrow \mu(f)$ and $\mu(s_n) \downarrow \mu(f)$. It follows from the inequality that $\mu(f_n) \to \mu(f)$. □

Corollary 13.4 *Let $|f_n| \le g$ for each n, where g is in $\mathcal{L}^1(X, \mathcal{F}, \mu)$. It follows that each f_n is in $\mathcal{L}^1(X, \mathcal{F}, \mu)$. Suppose $f_n \to f$ pointwise as $n \to \infty$. Then f is in $\mathcal{L}^1(X, \mathcal{F}, \mu)$ and $f_n \to f$ in the sense that $\mu(|f_n - f|) \to 0$ as $n \to \infty$.*

Proof: It suffices to apply the dominated convergence theorem to $|f_n - f| \le 2g$. □

In applying the dominated convergence theorem, the function $g \ge 0$ must be independent of n and have finite integral. However there is no requirement that the convergence be uniform or monotone.

Here is a simple example. Consider the sequence of functions $f_n(x) = \cos^n(x)/(1 + x^2)$. The goal is to prove that $\lambda(f_n) = \int_{-\infty}^{\infty} f_n(x)\, dx \to 0$ as $n \to \infty$. Note that $f_n \to 0$ as $n \to \infty$ pointwise, except for points that are a multiple of π. At these points one can redefine each f_n to be zero, and this will not change the integrals. Apply the dominated convergence to the redefined f_n. For each n we have $|f_n(x)| \le g(x)$, where $g(x) = 1/(1 + x^2)$ has finite integral. Hence $\lambda(f_n) \to \lambda(0) = 0$ as $n \to \infty$.

The following examples show what goes wrong when the condition that the dominating function has finite integral is not satisfied.

Examples:

1. Consider functions $f_n = n 1_{(0,1/n)}$ on the line. These are dominated by $g(x) = 1/x$ on $0 < x \le 1$, with $g(x) = 0$ for $x \ge 0$. This is independent of n. However $\lambda(g) = \int_0^1 1/x\, dx = +\infty$. The dominated convergence does not apply, and the integral of the limit is not the limit of the integrals.

2. Consider functions $f_n = 1_{(n,n+1)}$. Here the obvious dominating function is $g = 1_{(0,+\infty)}$. However again $\lambda(g) = +\infty$. Thus there is nothing to prevent mass density being lost in the limit.

13.3 Measure

If E is a subset of X, then 1_E is the indicator function of E. Its value is 1 for every point in E and 0 for every point not in E. The set E is said to be measurable if the function 1_E is measurable. The *measure* of such an E is $\mu(1_E)$. This is often denoted $\mu(E)$.

Theorem 13.5 *An integral is uniquely determined by the corresponding measure.*

Proof: Let $f \geq 0$ be a measurable function. Define

$$f_n = \sum_{k=0}^{\infty} \frac{k}{2^n} 1_{\frac{k}{2^n} < f \leq \frac{k+1}{2^n}}. \tag{13.11}$$

The integral of f_n is determined by the measures of the sets where $\frac{k}{2^n} < f \leq \frac{k+1}{2^n}$. However $f_n \uparrow f$, and so the integral of f is determined by the corresponding measure. □

This theorem justifies a certain amount of confusion between the notion of measure and the notion of integral. In fact, this whole subject is sometimes called measure theory.

Sometimes we denote a subset of X by a condition that defines the subset. Thus, for instance, $\{x \mid f(x) > a\}$ is denoted $f > a$, and its measure is $\mu(f > a)$.

Theorem 13.6 *If the set where $f \neq 0$ has measure zero, then $\mu(|f|) = 0$.*

Proof: For each n the function $|f| \wedge n \leq n 1_{|f|>0}$ and so has integral $\mu(|f| \wedge n) \leq n \cdot 0 = 0$. However $|f| \wedge n \uparrow |f|$ as $n \to \infty$. So from monotone convergence $\mu(|f|) = 0$. □

The preceding theorem shows that changing a function on a set of measure zero does not change its integral. Thus, for instance, if we change g_1 to $g_2 = g_1 + f$, then $|\mu(g_2) - \mu(g_1)| = |\mu(f)| \leq \mu(|f|) = 0$, so $\mu(g_1) = \mu(g_2)$.

There is a terminology that is standard in this situation. If a property of points is true except on a subset of μ measure zero, then it is said to hold *almost everywhere* with respect to μ. Thus the theorem would be stated as saying that if $f = 0$ almost everywhere, then its integral is zero. Similarly, if $g = h$ almost everywhere, then g and h have the same integral.

In probability the terminology is slightly different. Instead of saying that a property holds almost everywhere, on says that the event happens *almost surely* or *with probability one*.

The convergence theorems hold even when the hypotheses are violated on a set of measure zero. For instance, the dominated convergence theorem can be stated: If $|g| \leq g$ almost everywhere with respect to μ and $\mu(g) < +\infty$, then $f_n \to f$ almost everywhere with respect to μ implies $\mu(f_n) \to \mu(f)$.

Theorem 13.7 *(Chebyshev inequality) Let f be a real measurable function and a be a real number. Let ϕ be an increasing real function on $[a, +\infty)$ with $\phi(a) > 0$ and $\phi \geq 0$ on the range of f. Then*

$$\mu(f \geq a) \leq \frac{1}{\phi(a)} \mu(\phi(f)). \tag{13.12}$$

Proof: This follows from the pointwise inequality

$$1_{f \geq a} \leq 1_{\phi(f) \geq \phi(a)} \leq \frac{1}{\phi(a)} \phi(f). \tag{13.13}$$

At the points where $f \geq a$ we have $\phi(f) \geq \phi(a)$ and so the right hand size is one or greater. In any case the right hand size is positive. Integration preserves the inequality. □

The Chebyshev inequality is used in practice mainly in certain important special cases. Thus for $a > 0$ we have

$$\mu(|f| \geq a) \leq \frac{1}{a}\mu(|f|) \tag{13.14}$$

and

$$\mu(|f| \geq a) \leq \frac{1}{a^2}\mu(f^2). \tag{13.15}$$

Another important case is when $t > 0$ and

$$\mu(f \geq a) \leq \frac{1}{e^{ta}}\mu(e^{tf}). \tag{13.16}$$

Theorem 13.8 *If $\mu(|f|) = 0$, then the set where $f \neq 0$ has measure zero.*

Proof: By the Chebyshev inequality, for each n we have $\mu(1_{|f|>1/n}) \leq n\mu(|f|) = n \cdot 0 = 0$. However as $n \to \infty$, the functions $1_{|f|>1/n} \uparrow 1_{|f|>0}$. So $\mu(1_{|f|>0}) = 0$. □

The above theorem also has a statement in terms of an almost everywhere property. It says that if $|f|$ has integral zero, then $f = 0$ almost everywhere.

13.4 Extended real valued measurable functions

In connection with Tonelli's theorem it is natural to look at functions with values in the set $[0, +\infty]$. This system is well behaved under addition. In the context of measure theory it is useful to define $0 \cdot (+\infty) = (+\infty) \cdot 0 = 0$. It turns out that this is the most useful definition of multiplication.

Let X be a non-empty set, and let $\mathcal{F}$ be a σ-algebra of real functions on X. A function $f : X \to [0, +\infty]$ is said to be measurable with respect to $\mathcal{F}$ if there is a sequence f_n of functions in $\mathcal{F}$ with $f_n \uparrow f$ pointwise. A function is measurable in this sense if and only if there is a measurable set A with $f = +\infty$ on A and f coinciding with a function in $\mathcal{F}$ on the complement A^c.

An integral $\mu : \mathcal{F}^+ \to [0, +\infty]$ is extended to such measurable functions f by monotone convergence. Notice that if A is the set where $f = +\infty$, then we can set $f_n = n$ on A and f on A_c. Then $\mu(f_n) = n\mu(A) + \mu(f1_{A_c})$. If we take $n \to \infty$, we get $\mu(f) = (+\infty)\mu(A) + \mu(f1_{A^c})$. For the monotone convergence theorem to hold we must interpret $(+\infty) \cdot 0 = 0$. Notice that if $\mu(f) < +\infty$, then it follows that $\mu(A) = 0$.

13.5 Fubini's theorem for sums and integrals

Theorem 13.9 *(Tonelli for positive sums) If $w_k \geq 0$, then*

$$\mu(\sum_{k=1}^{\infty} w_k) = \sum_{k=1}^{\infty} \mu(w_k). \tag{13.17}$$

Proof: This theorem says that for positive functions integrals and sums may be interchanged. This is the monotone convergence theorem in disguise. That is, let $f_n = \sum_{k=1}^n w_k$. Then $f_n \uparrow f = \sum_{k=1}^\infty w_k$. Hence $\mu(f_n) = \sum_{k=1}^n \mu(w_k) \uparrow \mu(f)$. □

Theorem 13.10 *(Fubini for absolutely convergent sums) Suppose that the condition* $\sum_{k=1}^\infty \mu(|w_k|) < +\infty$ *is satisfied. Set* $g = \sum_{k=1}^\infty |w_k|$. *Then* g *is in* $\mathcal{L}^1(X, \mathcal{F}, \mu)$ *and so the set* Λ *where* $g < +\infty$ *has* $\mu(\Lambda^c) = 0$. *On this set* Λ *let*

$$f = \sum_{k=1}^\infty w_k \tag{13.18}$$

and on Λ^c *set* $f = 0$. *Then* f *is in* $\mathcal{L}^1(X, \mathcal{F}, \mu)$ *and*

$$\mu(f) = \sum_{k=1}^\infty \mu(w_k). \tag{13.19}$$

In other words,

$$\int_\Lambda \sum_{k=1}^\infty w_k \, d\mu = \sum_{k=1}^\infty \int w_k \, d\mu. \tag{13.20}$$

Proof: This theorem says that absolute convergence implies that integrals and sums may be interchanged. Here is a first proof. By the hypothesis and Tonelli s theorem $\mu(g) < +\infty$. It follows that $g < +\infty$ on a set Λ whose complement has measure zero. Let $f_n = \sum_{k=1}^n 1_\Lambda w_k$. Then $|f_n| \leq g$ for each n. Furthermore, the series defining f is absolutely convergent on Λ and hence convergent on Λ. Thus $f_n \to f$ as $n \to \infty$. Furthermore $\mu(f_n) = \sum_{k=1}^n \mu(1_\Lambda w_k) = \sum_{k=1}^n \mu(w_k)$. The conclusion follows by the dominated convergence theorem. □

Proof: Here is a second proof. Decompose each $w_j = w_j^+ - w_j^-$ into a positive and negative part. Then by Tonelli's theorem $\mu(\sum_{j=1}^\infty w_j^\pm) < +\infty$. Let Λ be the set where both sums $\sum_{j=1}^\infty w_j^\pm < +\infty$. Then $\mu(\Lambda^c) = 0$. Let $f = \sum_{j=1}^\infty w_j$ on Λ and $f = 0$ on Λ^c. Then $f = \sum_{j=1}^\infty 1_\Lambda w_j^+ - \sum_{j=1}^\infty 1_\Lambda w_j^-$. Therefore $\mu(f) = \mu(\sum_{j=1}^\infty 1_\Lambda w_j^+) - \mu(\sum_{j=1}^\infty 1_\Lambda w_j^-) = \sum_{j=1}^\infty \mu(w_j^+) - \sum_{j=1}^\infty \mu(w_j^-) = \sum_{j=1}^\infty (\mu(w_j^+) - \mu(w_j^-)) = \sum_{j=1}^\infty \mu(w_j)$. The hypothesis guarantees that there is never a problem with $(+\infty) - (+\infty)$. □

13.6 Fubini's theorem for sums

The following two theorems give conditions for when sums may be interchanged. Usually these results are applied when the sums are both over countable sets. However the case when one of the sums is uncountable also follows from the corresponding theorems in the preceding section.

Theorem 13.11 *(Tonelli for positive sums) If $w_k(x) \geq 0$, then*

$$\sum_x \sum_{k=1}^{\infty} w_k(x) = \sum_{k=1}^{\infty} \sum_x w_k(x). \tag{13.21}$$

Theorem 13.12 *(Fubini for absolutely convergent sums) Suppose that the condition $\sum_{k=1}^{\infty} \sum_x |w_k(x)| < +\infty$ is satisfied. Then for each x the series $\sum_{k=1}^{\infty} w_k(x)$ is absolutely convergent, and*

$$\sum_x \sum_{k=1}^{\infty} w_k(x) = \sum_{k=1}^{\infty} \sum_x w_k(x). \tag{13.22}$$

Here is an example that shows why absolute convergence is essential. Let $g : \mathbf{N} \times \mathbf{N} \to \mathbf{R}$ be defined by $g(m,n) = 1$ if $m = n$ and $g(m,n) = -1$ if $m = n + 1$. Then

$$\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} g(m,n) = 0 \neq 1 = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} g(m,n). \tag{13.23}$$

Problems

1. This problem is to show that one can get convergence theorems when the family of functions is indexed by real numbers. Prove that if $f_t \to f$ pointwise as $t \to t_0$, $|f_t| \leq g$ pointwise, and $\mu(g) < \infty$, then $\mu(f_t) \to \mu(f)$ as $t \to t_0$.

2. Show that if f is a Borel function and $\int_{-\infty}^{\infty} |f(x)|\, dx < \infty$, then $F(b) = \int_{-\infty}^{b} f(x)\, dx$ is continuous.

3. Must the function F in the preceding problem be differentiable at every point? Discuss.

4. Show that
$$\int_0^{\infty} \frac{\sin(e^x)}{1 + nx^2}\, dx \to 0 \tag{13.24}$$
as $n \to \infty$.

5. Show that
$$\int_0^{1} \frac{n \cos(x)}{1 + n^2 x^{\frac{3}{2}}}\, dx \to 0 \tag{13.25}$$
as $n \to \infty$.

6. Evaluate
$$\lim_{n \to \infty} \int_a^{\infty} \frac{n}{1 + n^2 x^2}\, dx \tag{13.26}$$
as a function of a.

7. Consider the integral
$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{1+nx^2}}\,dx. \tag{13.27}$$
Show that the integrand is monotone decreasing and converges pointwise as $n \to \infty$, but the integral of the limit is not equal to the limit of the integrals. How does this relate to the monotone convergence theorem?

8. Let g be a Borel function with
$$\int_{-\infty}^{\infty} |g(x)|\,dx < \infty \tag{13.28}$$
and
$$\int_{-\infty}^{\infty} g(x)\,dx = 1 \tag{13.29}$$
Let
$$g_\epsilon(x) = g(\frac{x}{\epsilon})\frac{1}{\epsilon}. \tag{13.30}$$
Let ϕ be bounded and continuous. Show that
$$\int_{-\infty}^{\infty} g_\epsilon(y)\phi(y)\,dy \to \phi(0) \tag{13.31}$$
as $\epsilon \to 0$. This problem gives a very general class of functions $g_\epsilon(x)$ such that integration with $g_\epsilon(x)\,dx$ converges to the Dirac delta integral δ_0 given by $\delta_0(\phi) = \phi(0)$.

9. Let f be bounded and continuous. Show that for each x the convolution
$$\int_{-\infty}^{\infty} g_\epsilon(x-z)f(z)\,dz \to f(x) \tag{13.32}$$
as $\epsilon \to 0$.

10. Prove *countable subadditivity*:
$$\mu(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n). \tag{13.33}$$
Show that if the A_n are disjoint this is an equality (countable additivity). Hint: $1_{\bigcup_{n=1}^{\infty} A_n} \leq \sum_{n=1}^{\infty} 1_{A_n}$.

Chapter 14

Fubini's theorem

14.1 Introduction

As an introduction, consider the Tonelli and Fubini theorems for Borel functions of two variables.

Theorem 14.1 (Tonelli) *If $f(x,y) \geq 0$, then*

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dx \right] dy = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dy \right] dx. \tag{14.1}$$

Theorem 14.2 (Fubini) *If*

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} |f(x,y)|\,dx \right] dy < +\infty, \tag{14.2}$$

then

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dx \right] dy = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x,y)\,dy \right] dx. \tag{14.3}$$

A slightly more careful statement of Fubini's theorem would acknowledge that the inner integrals may not be defined. However let

$$\Lambda_1 = \{x \mid \int_{-\infty}^{\infty} |f(x,y)|\,dy < +\infty\} \tag{14.4}$$

and

$$\Lambda_2 = \{y \mid \int_{-\infty}^{\infty} |f(x,y)|\,dx < +\infty\} \tag{14.5}$$

Then the inner integrals are well-defined on these sets. Furthermore, by the hypothesis of Fubini's theorem and by Tonelli's theorem, the complements of these sets have measure zero. So a more precise statement of the conclusion of Fubini's theorem is that

$$\int_{\Lambda_2} \left[\int_{-\infty}^{\infty} f(x,y)\,dx \right] dy = \int_{\Lambda_1} \left[\int_{-\infty}^{\infty} f(x,y)\,dy \right] dx. \tag{14.6}$$

This just amounts to replacing the undefined inner integrals by zero on the troublesome sets that are the complements of Λ_1 and Λ_2. It is quite fortunate that these sets are of measure zero.

The Tonelli and Fubini theorems may be formulated in a way that does not depend on writing the variables of integration explicitly. Consider for example Tonelli's theorem, which applies to a positive measurable function f on the plane. Let $f^{|1}$ be the function on the line whose value at a real number is obtained by holding the first variable fixed at this number and looking at f as a function of the second variable. Thus the value $f^{|1}(x)$ is the function $y \mapsto f(x,y)$. Similarly, let $f^{|2}$ be the function on the line whose value at a real number is obtained by holding the second variable fixed at this number and looking at f as a function of the first variable. The value $f^{|2}(y)$ is the function $x \mapsto f(x,y)$. Then the inner integrals are $(\lambda \circ f^{|2})(y) = \lambda(f^{|2}(y)) = \int_{-\infty}^{\infty} f(x,y)\,dx$ and $(\lambda \circ f^{|1})(x) = \lambda(f^{|1}(x)) = \int_{-\infty}^{\infty} f(x,y)\,dy$. So $\lambda \circ f^{|2}$ and $\lambda \circ f^{|1}$ are each a positive measurable function on the line. The conclusion of Tonelli's theorem may then be stated as the equality $\lambda(\lambda \circ f^{|2}) = \lambda(\lambda \circ f^{|1})$.

Here is rather interesting example where the hypothesis and conclusion of Fubini's theorem are both violated. Let $\sigma^2 > 0$ be a fixed diffusion constant. Let

$$u(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp(-\frac{x^2}{2\sigma^2 t}). \tag{14.7}$$

This describes the diffusion of a substance that has been created at time zero at the origin. For instance, it might be a broken bottle of perfume, and the molecules of perfume each perform a kind of random walk, moving in an irregular way. The motion is so irregular that the average squared distance that a particle moves in time t is only $x^2 = \sigma^2 t$.

As time increases the density gets more and more spread out. Then u satisfies

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2}. \tag{14.8}$$

Note that

$$\frac{\partial u}{\partial x} = -\frac{x}{\sigma^2 t} u \tag{14.9}$$

and

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{\sigma^2 t}(\frac{x^2}{\sigma^2 t} - 1)u. \tag{14.10}$$

This says that u is increasing in the space time region $x^2 > \sigma^2 t$ and decreasing in the space-time region $x^2 < \sigma^2 t$.

Fix $s > 0$. It is easy to compute that

$$\int_s^\infty \int_{-\infty}^\infty \frac{\partial u}{\partial t}\,dx\,dt = \frac{\sigma^2}{2} \int_s^\infty \int_{-\infty}^\infty \frac{\partial^2 u}{\partial x^2}\,dx\,dt = 0 \tag{14.11}$$

and

$$\int_{-\infty}^\infty \int_s^\infty \frac{\partial u}{\partial t}\,dt\,dx = -\int_{-\infty}^\infty u(x,s)\,dx = -1. \tag{14.12}$$

One can stop at this point, but it is interesting to look at the mechanism of the failure of the Fubini theorem. It comes from the fact that the time integral is extended to infinity, and in this limit the density spreads out more and more and approaches zero pointwise. So mass is lost in this limit, at least if one tries to describe it as a density. A description of the mass as a measure might lead instead to the conclusion that the mass is sitting at $x = \pm\infty$ in the limit $t \to \infty$. Even this does not capture the essence of the situation, since the diffusing particles do not go to infinity in any systematic sense; they just wander more and more.

14.2 Sigma-finite integrals

The proof of general Tonelli and Fubini theorems turns out to depend on the σ-finiteness condition. The role of this condition is give a unique determination of an integral. Thus we begin with a review of this subject.

Recall that a Stone vector lattice L of real functions on X is a vector lattice that satisfies the Stone condition: If f is in L, then $f \wedge 1$ is in L. Furthermore, recall that a monotone class is a class of real functions closed under upward and downward pointwise convergence.

Let L be a Stone vector lattice of real functions on X. Let $\mathcal{F}$ be the smallest monotone class including L. Then $\mathcal{F}$ is itself a Stone vector lattice. Furthermore, the smallest monotone class including L^+ is $\mathcal{F}^+$.

Recall that a function f is L-bounded if there exists a g in L^+ such that $|f| \leq g$. Furthermore, an L-monotone class is a class of functions that is closed under upward and downward pointwise convergence, provided that each function in the sequence is L-bounded. It was proved that the smallest L-monotone class including L^+ is $\mathcal{F}^+$.

If $\mathcal{F}$ includes all constant functions, then $\mathcal{F}$ is a σ-algebra of real functions. Furthermore, each elementary integral $m : L \to \mathbf{R}$ on L uniquely determines an integral $\mu : \mathcal{F}^+ \to [0, +\infty]$. The reason that it is uniquely determined is that the improved monotone convergence theorem applies to sequences of positive L-bounded functions under both upward and downward convergence.

As always, $\mathcal{L}^1(X, \mathcal{F}, \mu)$ consists of all f in $\mathcal{F}$ with $\mu(|f|) < +\infty$. Then μ also defines an integral $\mu : \mathcal{L}^1(X, \mathcal{F}, \mu) \to \mathbf{R}$ with $|\mu(f)| \leq \mu(|f|)$.

Consider a set X and a σ-algebra $\mathcal{F}$ of real functions on X. Consider an integral μ defined on $\mathcal{F}^+$, or the corresponding measure μ defined by $\mu(A) = \mu(1_A)$. The integral is called *finite* if the integral $\mu(1)$ is finite. This is the same as requiring that the measure $\mu(X)$ is finite.

An integral is *σ-finite* if there is a sequence $0 \leq u_n \uparrow 1$ of measurable functions with each $\mu(u_n) < +\infty$. If this is the case, define E_n as the set where $u_n \geq 1/2$. By Chebyshev's inequality the measure $\mu(E_n) \leq 2\mu(u_n) < +\infty$. Furthermore, $E_n \uparrow X$ as $n \to \infty$. Suppose on the other hand that there exists an increasing sequence E_n of measurable subsets of X such that each $\mu(E_n) < +\infty$ and $X = \bigcup_n E_n$. Then it is not difficult to show that μ is σ-finite. In fact, it suffices to take u_n to be the indicator function of E_n.

Theorem 14.3 *Let $\mathcal{F}$ be a σ-algebra of real functions on X. Let $\mu : \mathcal{F}^+ \to [0, +\infty]$ be an integral. Then μ is σ-finite if and only if there exists a vector lattice L such that the restriction of μ to L has only finite values and such that the smallest monotone class including L is $\mathcal{F}$.*

Proof: Suppose that μ is σ-finite. Let $L = \mathcal{L}^1(X, \mathcal{F}, \mu)$. Consider the monotone class generated by L. Since μ is σ-finite, the constant functions belong to this monotone class. So it is a σ-algebra. In fact, this monotone class is equal to $\mathcal{F}$. To see this, let E_n be a family of finite measure sets that increase to X. Consider a function g in $\mathcal{F}$. For each n the function $g_n = g 1_{E_n} 1_{|g| \leq n}$ is in L. Then $g = \lim_n g_n$ is in the monotone class generated by L.

Suppose on the other hand that there exists such a vector lattice L. Consider the class of functions f for which there exists h in $L\uparrow$ with $f \leq h$. This class includes L and is monotone, so it is includes all of $\mathcal{F}$.

Take f in $\mathcal{F}^+$. Then there exists h in $L\uparrow$ with $f \leq h$. Take $h_n \in L^+$ with $h_n \uparrow h$. Then $u_n = f \wedge h_n \uparrow f$. Thus there is a sequence of L-bounded functions u_n in $\mathcal{F}^+$ such that $u_n \uparrow f$. Each of these functions u_n has finite integral. In the present case $\mathcal{F}$ is a σ-algebra, so we may take take $f = 1$. This completes the proof that μ is σ-finite. □

14.3 Summation

Summation is a special case of integration. Let X be a set. Then there is an integral $\sum : [0, +\infty)^X \to [0, +\infty]$. It is defined for $f \geq 0$ by $\sum f = \sup_W \sum_{j \in W} f(j)$, where the supremum is over all finite subsets $W \subset X$. Since each $f(j) \geq 0$, the result is a number in $[0, +\infty]$. As usual, the sum is also defined for functions that are not positive, but only provided that there is no $(+\infty) - (+\infty)$ problem.

Suppose $f \geq 0$ and $\sum f < +\infty$. Let S_k be the set of j in X such that $f(j) \geq 1/k$. Then S_k is a finite set. Let S be the set of j in X such that $f(j) > 0$. Then $S = \bigcup_k S_k$, so S is countable. This argument proves that the sum is infinite unless f vanishes off a countable set. So a finite sum is just the usual sum over a countable index set.

The $\sum$ integral is σ-finite if and only if X is countable. This is because whenever $f \geq 0$ and $\sum f < +\infty$, then f vanishes off a countable set S. So if each f_n vanishes off a countable set S_n, and $f_n \uparrow f$, then f vanishes off $S = \bigcup S_n$, which is also a countable set. This shows that f cannot be a constant function $a > 0$ unless X is a countable set.

One could define $\sum$ on a smaller σ-algebra of functions. The smallest one that seems natural consists of all functions of the form $f = g + a$, where the function g is zero on the complement of some countable subset of X, and a is constant. If $f \geq 0$ and $a = 0$, then $\sum f = \sum g$ is a countable sum. On the other hand, if $f \geq 0$ and $a > 0$ then $\sum f = +\infty$.

The Tonelli and Fubini theorems are true for the Lebesgue integral defined for Borel functions. However they are not true for arbitrary integrals that are not

required to be σ-finite. Here is an example based on the example of summation over an uncountable set.

Let $\lambda(g) = \int_0^1 g(x)\,dx$ be the usual uniform Lebesgue integral on the interval $[0,1]$. Let $\sum h = \sum_y h(y)$ be summation indexed by the points in the interval $[0,1]$. The measure $\sum$ is not σ-finite, since there are uncountably many points in $[0,1]$. Finally, let $\delta_{xy} = 1$ if $x = y$, and $\delta_{xy} = 0$ for $x \neq y$. Now for each x, the sum $\sum_y \delta_{xy} = 1$. So the integral over x is also 1. On the other hand, for each y the integral $\int_0^1 \delta_{xy}\,dx = 0$, since the integrand is zero except for a single point of λ measure zero, where it has the value one. So the sum over y is also zero. Thus the two orders of integration give different results.

14.4 Product sigma-algebras

This section defines the product σ-algebra. Let X_1 and X_2 be non-empty sets. Then their product $X_1 \times X_2$ is another non-empty set. There are projections $\pi_1 : X_1 \times X_2 \to X_1$ and $\pi_2 : X_1 \times X_2 \to X_2$. These are of course define by $\pi_1(x,y) = x$ and $\pi_2(x,y) = y$.

Suppose that $\mathcal{F}_1$ is a σ-algebra of real functions on X_1 and $\mathcal{F}_2$ is a σ-algebra of real functions on X_2. Then there is a *product* σ-algebra $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ of real functions on $X_1 \times X_2$. This is the smallest σ-algebra of functions on $\mathcal{F}_1 \times \mathcal{F}_2$ such that the projections π_1 and π_2 are measurable maps.

The condition that the projections π_1 and π_2 are measurable maps is the same as saying that for each g in $\mathcal{F}_1$ the function $g \circ \pi_1$ is measurable and for each h in $\mathcal{F}_2$ the function $h \circ \pi_2$ is measurable. In other words, the functions $(x,y) \mapsto g(x)$ and $(x,y) \mapsto h(y)$ are required to be measurable functions. This condition determines the σ-algebra of measurable functions $\mathcal{F}_1 \bigotimes \mathcal{F}_2$.

If g is a real function on X_1 and h is a real function on X_2, then there is a real function $g \otimes h$ on X defined by

$$(g \otimes h)(x,y) = g(x)h(y). \tag{14.13}$$

This is sometimes called the tensor product of the two functions. Such functions are called *decomposable*. Another term is *separable*, as in "separation of variables." The function $g \otimes h$ could be define more abstractly as $g \otimes h = (g \circ \pi_1)(h \circ \pi_2)$. This identity could also be stated as $g \otimes h = (g \otimes 1)(1 \otimes h)$. It is easy to see that $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ may also be characterized as the σ-algebra generated by the functions $g \otimes h$ with g in $\mathcal{F}_1$ and h in $\mathcal{F}_2$.

Examples:

1. If $\mathcal{B}$ is the Borel σ-algebra of functions on the line, then $\mathcal{B} \bigotimes \mathcal{B}$ is the Borel σ-algebra of functions on the plane.

2. Take the two sigma-algebras to be the Borel σ-algebra of real functions on $[0,1]$ and the σ-algebra $\mathbf{R}^{[0,1]}$ of all real functions on $[0,1]$. These are the σ-algebras relevant to the counterexample with λ and $\sum$. The product σ-algebra then consists of all functions f on the square such that $x \mapsto f(x,y)$

is a Borel function for each y. The diagonal function δ is measurable, but $\sum$ is not σ-finite, so Tonelli's theorem does not apply.

3. Take the two sigma-algebras to be the Borel σ-algebra of real functions on $[0,1]$ and the σ-algebra consisting of all real functions $y \mapsto a+h(y)$ on $[0,1]$ that differ from a constant function a on a countable set. These are the σ-algebras relevant to the counterexample with λ and $\sum$, but in the case when we restrict $\sum$ to the smallest σ-algebra for which it makes sense. The product σ-algebra is generated by functions of the form $(x,y) \mapsto g(x)$ and $(x,y) \mapsto a+h(y)$, where h vanishes off a countable set. This is a rather small σ-algebra; the diagonal function δ used in the counterexample does not belong to it. Already for this reason Tonelli's theorem cannot be used.

Lemma 14.4 *Let X_1 be a set with σ-algebra $\mathcal{F}_1$ of functions and σ-finite integral μ_1. Let X_2 be another set with a σ-algebra $\mathcal{F}_2$ of functions and σ-finite integral μ_2. Let $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ be the product σ-algebra of functions on $X_1 \times X_2$. Let L consist of finite linear combinations of indicator functions of products of sets of finite measure. Then L is a vector lattice, and the smallest monotone class including L is $\mathcal{F}_1 \bigotimes \mathcal{F}_2$.*

Proof: Let $L \subset \mathcal{F}_1 \bigotimes \mathcal{F}_2$ be the set of all finite linear combinations

$$f = \sum_i c_i 1_{A_i \times B_i} = \sum_i c_i 1_{A_i} \otimes 1_{B_i}, \tag{14.14}$$

where A_i and B_i each have finite measure. The space L is obviously a vector space. The proof that it is a lattice is found in the last section of the chapter.

Let E_n be a sequence of sets of finite measure that increase to X_1. Let F_n be a sequence of sets of finite measure that increase to X_2. Then the $E_n \times F_n$ increase to $X_1 \times X_2$. This is enough to show that the the constant functions belong to the monotone class generated by L. Since L is a vector lattice and the monotone class generated by L has all constant functions, it follows that the monotone class generated by L is a σ-algebra. To show that this σ-algebra is equal to all of $\mathcal{F}_1 \bigotimes \mathcal{F}_2$, it is sufficient to show that each $g \otimes h$ is in the σ-algebra generated by L. Let $g_n = g1_{E_n}$ and $h_n = h1_{F_n}$. It is sufficient to show that each $g_n \otimes h_n$ is in this σ-algebra. However g_n may be approximated by functions of the form $\sum_i a_i 1_{A_i}$ with A_i of finite measure, and h_n may also be approximated by functions of the form $\sum_j b_j 1_{B_j}$ with B_j of finite measure. So $g_n \otimes h_n$ is approximated by $\sum_i \sum_j a_i b_j 1_{A_i} \otimes 1_{B_j} = \sum_i \sum_j a_i b_j 1_{A_i \times B_j}$. These are indeed functions in L. □

14.5 The product integral

This section gives a proof of the uniqueness of the product of two σ-finite integrals.

Theorem 14.5 *Let $\mathcal{F}_1$ be a σ-algebra of measurable functions on X_1, and let $\mathcal{F}_2$ be a σ-algebra of measurable functions on X_2. Let $\mu_1 : \mathcal{F}_1^+ \to [0, +\infty]$ and $\mu_2 : \mathcal{F}_2^+ \to [0, +\infty]$ be corresponding σ-finite integrals. Consider the product space $X_1 \times X_2$ and the product σ-algebra of functions $\mathcal{F}_1 \bigotimes \mathcal{F}_2$. Then there exists at most one σ-finite integral $\nu : (\mathcal{F}_1 \bigotimes \mathcal{F}_2)^+ \to [0, +\infty]$ with the property that if A and B each have finite measure, then $\nu(A \times B) = \mu_1(A)\mu_2(B)$.*

Proof: Let L be the vector lattice of the preceding lemma. The integral ν is uniquely defined on L by the explicit formula. Since the smallest monotone class including L is $\mathcal{F}_1 \bigotimes \mathcal{F}_2$, it follows that the smallest L-monotone class including L^+ is $(\mathcal{F}_1 \bigotimes \mathcal{F}_2)^+$. Say that ν and ν' were two such integrals. Then they agree on L, since they are given by an explicit formula. However the set of functions on which they agree is an L-monotone class. Therefore the integral is uniquely determined on all of $\mathcal{F}^+$. □

The integral ν described in the above theorem is called the *product integral* and denoted $\mu_1 \times \mu_2$. The corresponding measure is called the *product measure*. The existence of the product of σ-finite integrals will be a byproduct of the Tonelli theorem. This product integral ν has the more general property that if $g \geq 0$ is in $\mathcal{F}_1$ and $h \geq 0$ is in $\mathcal{F}_2$, then

$$\nu(g \otimes h) = \mu_1(g)\mu_2(h). \tag{14.15}$$

The product of integrals may be of the form $0 \cdot (+\infty)$ or $(+\infty) \cdot 0$. In that case the multiplication is performed using $0 \cdot (+\infty) = (+\infty) \cdot 0 = 0$. The characteristic property $(\mu_1 \times \mu_2)(g \otimes h) = \mu_1(g)\mu_2(h)$ may also be written in the more explicit form

$$\int g(x)h(y)\, d(\mu_1 \times \mu_2)(x, y) = \int g(x)\, d\mu_1(x) \int h(y)\, d\mu_2(y). \tag{14.16}$$

The definition of product integral does not immediately give a useful way to compute the integral of functions that are not written as sums of decomposable functions. For this we need Tonelli's theorem and Fubini's theorem.

14.6 Tonelli's theorem

Let X_1 and X_2 be two sets. Let $f : X_1 \times X_2 \to \mathbf{R}$ be a function on the product space. Then there is a function $f^{|1}$ from X_1 to $\mathbf{R}^{X_2}$ defined by saying that the value $f^{|1}(x)$ is the function $y \mapsto f(x, y)$. In other words, $f^{|1}$ is f with the first variable temporarily held constant.

Similarly, there is a function $f^{|2}$ from X_2 to $\mathbf{R}^{X_1}$ defined by saying that the value $f^{|2}(y)$ is the function $x \mapsto f(x, y)$. In other words, $f^{|2}$ is f with the second variable temporarily held constant.

Lemma 14.6 *Let $f : X_1 \times X_2 \to [0, +\infty]$ be a $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ measurable function. Then for each x the function $f^{|1}(x)$ is a $\mathcal{F}_2$ measurable function on X_2. Also, for each y the function $f^{|2}(y)$ is a $\mathcal{F}_1$ measurable function on X_1.*

Explicitly, this lemma says that the functions

$$y \mapsto f(x, y) \tag{14.17}$$

with fixed x and

$$x \mapsto f(x, y) \tag{14.18}$$

with fixed y are measurable functions.

Proof: Let L be the space of finite linear combinations of indicator functions of products of sets of finite measure. Consider the class S of functions f for which the lemma holds. If f is in L, then $f = \sum_i c_i 1_{A_i \times B_i}$, where each A_i is an $\mathcal{F}_1$ set and each B_i is a $\mathcal{F}_2$ set. Then for fixed x consider the function $y \mapsto \sum_i c_i 1_{A_i}(x) 1_{B_i}(y)$. This is clearly in $\mathcal{F}_2$. This shows that $L \subset S$. Now suppose that $f_n \uparrow f$ and each f_n is in S. Then for each x we have that $f_n(x, y)$ is measurable in y and increases to $f(x, y)$ pointwise in y. Therefore $f(x, y)$ is measurable in y. This proves S is closed under upward monotone convergence. The argument for downward monotone convergence is the same. Thus S is a monotone class. Since $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ is the smallest monotone class including L, this establishes the result. □

Lemma 14.7 *Let μ_1 be a σ-finite integral defined on $\mathcal{F}_1^+$. Also let μ_2 be a σ-finite integral defined on $\mathcal{F}_2^+$. Let $f : X_1 \times X_2 \to [0, +\infty]$ be a $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ measurable function. Then the function $\mu_2 \circ f^{|1}$ is an $\mathcal{F}_1$ measurable function on X_1 with values in $[0, +\infty]$. Also the function $\mu_1 \circ f^{|2}$ is an $\mathcal{F}_2$ measurable function on X_2 with values in $[0, +\infty]$.*

Explicitly, this lemma says that the functions

$$x \mapsto \int f(x, y)\, d\mu_2(y) \tag{14.19}$$

and

$$y \mapsto \int f(x, y)\, d\mu_1(y) \tag{14.20}$$

are measurable functions.

Proof: The previous lemma shows that the integrals are well defined. Consider the class S of functions f for which the first assertion of the lemma holds. If f is in L^+, then $f = \sum_i c_i 1_{A_i \times B_i}$, where each A_i is an $\mathcal{F}_1$ set and each B_i is a $\mathcal{F}_2$ set. Then for fixed x consider the function $y \mapsto \sum_i c_i 1_{A_i}(x) 1_{B_i}(y)$. Its μ_2 integral is $\sum_i c_i 1_{A_i}(x) \mu(B_i)$. This is clearly in $\mathcal{F}_1$ as a function of x. This shows that $L \subset S$. Now suppose that f_n is a sequence of L-bounded functions, that $f_n \uparrow f$, and each f_n is in S. Then we have that $\int f_n(x, y)\, d\mu_2(y)$ is measurable in x. Furthermore, for each x it increases to $\int f(x, y)\, d\mu_2(y)$, by the monotone convergence theorem. Therefore $\int f(x, y)\, d\mu_2(y)$ is measurable in x. This proves S is closed under upward monotone convergence of L-bounded functions. The argument for downward monotone convergence uses the improved monotone convergence theorem; here it is essential that each f_n be an L-bounded function. Thus S is an L-bounded monotone class including L^+. It follows that $(\mathcal{F}_1 \bigotimes \mathcal{F}_2)^+ \subset S$. □

Lemma 14.8 *Let $f : X_1 \times X_2 \to [0, +\infty]$ be a $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ measurable function . Then $\nu_{12}(f) = \mu_2(\mu_1 \circ f^{|2})$ defines an integral ν_{12}. Also $\nu_{21}(f) = \mu_1(\mu_2 \circ f^{|1})$ defines an integral ν_{21}.*

Explicitly, this lemma says that the iterated integrals

$$\nu_{12}(f) = \int \left(\int f(x, y)\, d\mu_1(x) \right) d\mu_2(y) \tag{14.21}$$

and

$$\nu_{21}(f) = \int \left(\int f(x, y)\, d\mu_2(y) \right) d\mu_1(x) \tag{14.22}$$

are defined.

Proof: The previous lemma shows that the integral ν_{12} is well defined. It is easy to see that ν_{12} is linear and order preserving. The remaining task is to prove upward monotone convergence. Say that $f_n \uparrow f$ pointwise. Then by the monotone convergence theorem for μ_1 we have that for each y the integral $\int f_n(x, y)\, d\mu_1(x) \uparrow \int f(x, y)\, d\mu_1(x)$. Hence by the monotone convergence theorem for μ_2 we have that $\int \int f_n(x, y)\, d\mu_1(x)\, d\mu_2(y) \uparrow \int f(x, y)\, d\mu_1(x)\, d\mu_2(y)$. This is the same as saying that $\nu_{12}(f_n) \uparrow \mu_{12}(f)$. □

Theorem 14.9 (Tonelli's theorem) *. Let $\mathcal{F}_1$ be a σ-algebra of real functions on X_1, and let $\mathcal{F}_2$ be a σ-algebra of real functions on X_2. Let $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ be the product σ-algebra of real functions on $X_1 \times X_2$. Let $\mu_1 : \mathcal{F}_1^+ \to [0, +\infty]$ and $\mu_2 : \mathcal{F}_2^+ \to [0, +\infty]$ be σ-finite integrals. Then there is a unique σ-finite integral*

$$\mu_1 \times \mu_2 : (\mathcal{F}_1 \bigotimes \mathcal{F}_2)^+ \to [0, +\infty] \tag{14.23}$$

such that $(\mu_1 \times \mu_2)(g \otimes h) = \mu_1(g)\mu_2(h)$ for each g in $\mathcal{F}_1^+$ and h in $\mathcal{F}_2^+$. Furthermore, for f in $(\mathcal{F}_1 \bigotimes \mathcal{F}_2)^+$ we have

$$(\mu_1 \times \mu_2)(f) = \mu_2(\mu_1 \circ f^{|2}) = \mu_1(\mu_2 \circ f^{|1}). \tag{14.24}$$

In this statement of the theorem $f^{|2}$ is regarded as a function on X_2 with values that are functions on X_1. Similarly, $f^{|1}$ is regarded as a function on X_1 with values that are functions on X_2. Thus the composition $\mu_1 \circ f^{|2}$ is a function on X_2, and the composition $\mu_2 \circ f^{|1}$ is a function on X_1.

The theorem may be also be stated in a version with bound variables:

$$\int f(x, y)\, d(\mu_1 \times \mu_2)(x, y) = \int \left[\int f(x, y)\, d\mu_1(x) \right] d\mu_2(y) = \int \left[\int f(x, y)\, d\mu_2(y) \right] d\mu_1(x). \tag{14.25}$$

Proof: The integrals ν_{12} and ν_{21} agree on L^+. Consider the set S of $f \in (\mathcal{F}_1 \bigotimes \mathcal{F}_2)^+$ such that $\nu_{12}(f) = \nu_{21}(f)$. The argument of the previous lemma shows that this is an L-monotone class. Hence S is all of $(\mathcal{F}_1 \bigotimes \mathcal{F}_2)^+$. Define $\nu(f)$ to be the common value $\nu_{12}(f) = \nu_{21}(f)$. Then ν is uniquely defined by its values on L^+. This ν is the desired product measure $\mu_1 \times \mu_2$. □

The integral ν is called the product integral and is denoted by $\mu_1 \times \mu_2$. Let $F^2 : \mathbf{R}^{X_1 \times X_2} \to (\mathbf{R}^{X_1})^{X_2}$ be given by $f \mapsto f^{|2}$, that is, F_2 says to hold the second second variable constant. Similarly, let $F^1 : \mathbf{R}^{X_1 \times X_2} \to (\mathbf{R}^{X_2})^{X_1}$ be given by $f \mapsto f^{|1}$, that is, F^1 says to hold the first variable constant. Then the Tonelli theorem says that the product integral $\mu_1 \times \mu_2 : (\mathcal{F}_1 \times \mathcal{F}_2)^+ \to [0, +\infty]$ satisfies

$$\mu_1 \times \mu_2 = \mu_2 \circ \mu_1 \circ F^2 = \mu_1 \circ \mu_2 \circ F^1. \tag{14.26}$$

14.7 Fubini's theorem

Recall that for an arbitrary non-empty set X, σ-algebra of functions $\mathcal{F}$, and integral μ, the space $\mathcal{L}^1(X, \mathcal{F}, \mu)$ consists of all real functions f in $\mathcal{F}$ such that $\mu(|f|) < +\infty$. For such a function $\mu(|f|) = \mu(f_+) + \mu(f_-)$, and $\mu(f) = \mu(f_+) - \mu(f_-)$ is a well-defined real number.

Let f be in $\mathcal{L}^1(X \times Y, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$. Let Λ_1 be the set of all x with $f^{|1}(x)$ in $\mathcal{L}^1(X_2, \mathcal{F}_2, \mu_2)$ and let Λ_2 be the set of all y with $f^{|2}(y)$ in $\mathcal{L}^1(X_1, \mathcal{F}_1, \mu_1)$. Then $\mu_1(\Lambda_1^c) = 0$ and $\mu_2(\Lambda_2^c) = 0$. Define the *partial integral* $\mu_2(f \mid 1)$ by $\mu_2(f \mid 1)(x) = \mu_2(f^{|1}(x))$ for $x \in \Lambda_1$ and $\mu_2(f \mid 1)(x) = 0$ for $x \in \Lambda_1^c$. Define the partial integral $\mu_1(f \mid 2)$ by $\mu_1(f \mid 2)(y) = \mu_1(f^{|2}(y))$ for $y \in \Lambda_2$ and $\mu_1(f \mid 2)(y) = 0$ for $y \in \Lambda_2^c$.

Theorem 14.10 *Let $\mathcal{F}_1$ be a σ-algebra of real functions on X_1, and let $\mathcal{F}_2$ be a σ-algebra of real functions on X_2. Let $\mathcal{F}_1 \bigotimes \mathcal{F}_2$ be the product σ-algebra of real functions on $X_1 \times X_2$. Let μ_1 and μ_2 be σ-finite integrals, and consider the corresponding functions*

$$\mu_1 : \mathcal{L}^1(X, \mathcal{F}_1, \mu_1) \to \mathbf{R} \tag{14.27}$$

and

$$\mu_2 : \mathcal{L}^1(X_2, \mathcal{F}_2, \mu_2) \to \mathbf{R}. \tag{14.28}$$

The product integral $\mu_1 \times \mu_2$ defines a function

$$\mu_1 \times \mu_2 : \mathcal{L}^1(X_1 \times X_2, \mathcal{F}_1 \bigotimes F_2, \mu_1 \times \mu_2) \to \mathbf{R}. \tag{14.29}$$

Let f be in $\mathcal{L}^1(X \times Y, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$. Then the partial integral $\mu_2(f \mid 1)$ is in $\mathcal{L}^1(X_1, \mathcal{F}_1, \mu_1)$, and the partial integral $\mu_1(f \mid 2)$ is in $\mathcal{L}^1(X_2, \mathcal{F}_2, \mu_2)$. Finally,

$$(\mu_1 \times \mu_2)(f) = \mu_1((\mu_2(f \mid 1)) = \mu_2(\mu_1(f \mid 2)). \tag{14.30}$$

In this statement of the theorem $\mu_2(f \mid 1)$ is the μ_2 partial integral with the first variable fixed, regarded after integration as a function on X_1. Similarly, $\mu_1(f \mid 2)$ is the μ_1 partial integral with the second variable fixed, regarded after integration as a function on X_2.

Fubini's theorem may also be stated with bound variables:

$$\int f(x,y)\, d(\mu_1 \times \mu_2)(x,y) = \int_{\Lambda_1} \left[\int f(x,y)\, d\mu_2(x) \right] d\mu_1(x) = \int_{\Lambda_2} \left[\int f(x,y)\, d\mu_1(x) \right] d\mu_2(y). \tag{14.31}$$

Here as before Λ_1 and Λ_2 are sets where the inner integral converges absolutely. The complement of each of these sets has measure zero.

Proof: By Tonelli's theorem we have that $\mu_2 \circ |f^{|1}|$ is in $\mathcal{L}^1(X_1, \mathcal{F}_1, \mu_1)$ and that $\mu_1 \circ |f^{|2}|$ is in $\mathcal{L}^2(X_2, \mathcal{F}_2, \mu_2)$. This is enough to show that $\mu_2(\Lambda_1^c) = 0$ and $\mu_1(\Lambda_2^c) = 0$. Similarly, by Tonelli's theorem we have

$$(\mu_1 \times \mu_2)(f) = (\mu_1 \times \mu_2)(f_+) - (\mu_1 \times \mu_2)(f_-) = \mu_1(\mu_2 \circ f_+^{|1}) - \mu_1(\mu_2 \circ f_-^{|1}). \quad (14.32)$$

Since Λ_1 and Λ_2 are sets whose complements have measure zero, we can also write this as

$$(\mu_1 \times \mu_2)(f) = \mu_1(1_{\Lambda_1}(\mu_2 \circ f_+^{|1})) - \mu_1(1_{\Lambda_1}(\mu_2 \circ f_-^{|1})). \quad (14.33)$$

Now for each fixed x in Λ_1 we have

$$\mu_2(f^{|1}(x)) = \mu_2(f_+^{|1}(x)) - \mu_2(f_-^{|1}(x)). \quad (14.34)$$

This says that

$$\mu_2(f \mid 1) = 1_{\Lambda_1}(\mu_2 \circ f_+^{|1}) - 1_{\Lambda_1}(\mu_2 \circ f_-^{|1}). \quad (14.35)$$

Each function on the right hand side is a real function in $\mathcal{L}^1(X_1, \mathcal{F}_1, \mu_1)$. So

$$(\mu_1 \times \mu_2)(f) = \mu_1(\mu_2(f \mid 1)). \quad (14.36)$$

□

Tonelli's theorem and Fubini's theorem are often used together to justify an interchange of order of integration. Here is a typical pattern. Say that one can show that the iterated integral with the absolute value converges:

$$\int \left[\int |h(x,y)| \, d\nu(y) \right] d\mu(x) < \infty. \quad (14.37)$$

By Tonelli's theorem the product integral also converges:

$$\int |h(x,y)| \, d(\mu \times \nu)(x,y) < \infty. \quad (14.38)$$

Then from Fubini's theorem the integrated integrals are equal:

$$\int \left[\int h(x,y) \, d\nu(y) \right] d\mu(x) = \int \left[\int h(x,y) \, d\mu(x) \right] d\nu(y). \quad (14.39)$$

The outer integrals are each taken over a set for which the inner integral converges absolutely; the complement of this set has measure zero.

14.8 Semirings and rings of sets

This section supplies the proof that finite linear combinations of indicator functions of rectangles form a vector lattice. It may be omitted on a first reading.

The first and last results in this section are combinatorial lemmas that are proved in books on measure theory. See R. M. Dudley, Real Analysis and Probability, Cambridge University Press, Cambridge, 2002, Chapter 3.

Let X be a set. A ring $\mathcal{R}$ of subsets of X is a collection such that $\emptyset$ is in $\mathcal{R}$ and such that A and B in $\mathcal{R}$ imply $A \cap B$ is in $\mathcal{R}$ and such that A and B in $\mathcal{R}$ imply that $A \setminus B$ is in $\mathcal{R}$.

A semiring $\mathcal{D}$ of subsets of X is a collection such that $\emptyset$ is in $\mathcal{D}$ and such that A and B in $\mathcal{D}$ imply $A \cap B$ is in $\mathcal{D}$ and such that A and B in $\mathcal{D}$ imply that $A \setminus B$ is a finite union of disjoint members of $\mathcal{D}$.

Proposition 14.11 *Let $\mathcal{D}$ be a semiring of subsets of X. Let $\mathcal{R}$ be the ring generated by $\mathcal{D}$. Then $\mathcal{R}$ consists of all finite unions of members of $\mathcal{D}$.*

Proposition 14.12 *Let $\mathcal{D}$ be a semiring of subsets of a set X. Let Γ be a finite collection of subsets in $\mathcal{D}$. Then there exists a finite collection Δ of disjoint subsets in $\mathcal{D}$ such that each set in Γ is a finite union of some subcollection of Δ.*

Proof: For each non-empty subcollection Γ' of Γ consider the set $A_{\Gamma'}$ that is the intersection of the sets in Γ' with the intersection of the complements of the sets in $\Gamma \setminus \Gamma'$. The sets $A_{\Gamma'}$ are in $\mathcal{R}$ and are disjoint. Furthermore, each set C in Γ is the finite disjoint union of the sets $A_{\Gamma'}$ such that $C \in \Gamma'$. The proof is completed by noting that by the previous proposition each of these sets $A_{\Gamma'}$ is itself a finite disjoint union of sets in $\mathcal{D}$. □

Theorem 14.13 *Let $\mathcal{D}$ be a semiring of subsets of X. Let L be the set of all finite linear combinations of indicator functions of sets in $\mathcal{D}$. Then L is a vector lattice.*

Proof: The problem is to prove that L is closed under the lattice operations. Let f and g be in L. Then f is a finite linear combination of indicator functions of sets in $\mathcal{D}$. Similarly, g is a finite linear combination of indicator functions of sets in $\mathcal{D}$. Take the union Γ of these two collections of sets. These sets may not be disjoint, but there is a collection Δ of disjoint sets in $\mathcal{D}$ such that each set in the union is a disjoint union of sets in Δ. Then f and g are each linear combinations of indicator functions of disjoint sets in Δ. It follows that $f \wedge g$ and $f \vee g$ also have such a representation. □

Theorem 14.14 *Let X_1 and X_2 be non-empty sets, and let $\mathcal{D}_1$ and $\mathcal{D}_2$ be semirings of subsets. Then the set of all $A \times B$ with $A \in \mathcal{D}_1$ and $B \in \mathcal{D}_2$ is a semiring of subsets of $X_1 \times X_2$.*

In the application to product measures the sets $\mathcal{D}_1$ and $\mathcal{D}_2$ consist of sets of finite measure. Thus each of $\mathcal{D}_1$ and $\mathcal{D}_2$ is a ring of subsets. It follows from the last theorem that the product sets form a semiring of subsets of the product space. The previous theorem then shows that the finite linear combinations form a vector lattice.

Chapter 15

Probability

15.1 Coin-tossing

A basic probability model is that for coin-tossing. The set of outcomes of the experiment is $\Omega = 2^{\mathbf{N}_+}$. Let b_j be the jth coordinate function. Let f_{nk} be the indicator function of the set of outcomes that have the k pattern in the first n coordinates. Here $0 \leq k < 2^n$, and the pattern is given by the binary representation of k. If S is the subset of $\{1, \ldots, n\}$ where the 1s occur, and S^c is the subset where the 0s occur, then

$$f_{nk} = \prod_{j \in S} b_j \prod_{j \in S^c} (1 - b_j). \tag{15.1}$$

The expectation μ is determined by

$$\mu(f_{nk}) = p^j q^{n-j}, \tag{15.2}$$

where j is the number of 1s in the binary expansion of k, or the number of points in S. It follows that if S and T are disjoint subsets of $\{1, \ldots, n\}$, then

$$\mu(\prod_{j \in S} b_j \prod_{j \in T} (1 - b_j)) = p^j q^\ell, \tag{15.3}$$

where j is the number of elements in S, and ℓ is the number of elements in T.

It follows from these formulas that the probability of success on one trial is $\mu(b_j) = p$ and the probability of failure on one trial is $\mu(1-b_j) = q$. Similarly, for two trials $i < j$ the probabilities of two successes is $\mu(b_i b_j) = p^2$, the probability of success followed by failure is $\mu(b_i)(1 - b_j)) = pq$, the probability of failure followed by success is $\mu((1 - b_i)b_j = qp$, and the probability of two failures is $\mu((1 - b_i)(1 - b_j)) = q^2$.

15.2 Weak law of large numbers

Theorem 15.1 (Weak law of large numbers) *Let*

$$s_n = b_1 + \cdots + b_n \tag{15.4}$$

be the number of successes in the first n trials. Then

$$\mu(s_n) = np \tag{15.5}$$

and

$$\mu((s_n - np)^2) = npq. \tag{15.6}$$

Proof: Expand $(s_n - np)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} (b_i - p)(b_j - p)$. The expectation of each of the cross terms vanishes. The expectation of each of the diagonal terms is $(1-p)^2 p + (0-p)^2 q = q^2 p + p^2 q = pq$. □

Corollary 15.2 (Weak law of large numbers) *Let*

$$f_n = \frac{b_1 + \cdots + b_n}{n} \tag{15.7}$$

be the proportion of successes in the first n trials. Then

$$\mu(f_n) = p \tag{15.8}$$

and

$$\mu((f_n - p)^2) = \frac{pq}{n} \leq \frac{1}{4n}. \tag{15.9}$$

The quantity that is usually used to evaluate the error is the standard deviation, which is the square root of this quantity. The version that should be memorized is thus

$$\sqrt{\mu((f_n - p)^2)} = \frac{\sqrt{pq}}{\sqrt{n}} \leq \frac{1}{2\sqrt{n}}. \tag{15.10}$$

This $1/\sqrt{n}$ factor is what makes probability theory work (in the sense that it is internally self-consistent).

Corollary 15.3 *Let*

$$f_n = \frac{b_1 + \cdots + b_n}{n} \tag{15.11}$$

be the proportion of successes in the first n trials. Then

$$\mu(|f_n - p| \geq \epsilon) = \frac{pq}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}. \tag{15.12}$$

This corollary follows immediately from Chebyshev's inequality. It gives a perhaps more intuitive picture of the meaning of the weak law of large numbers. Consider a tiny $\epsilon > 0$. Then it says that if n is sufficiently large, then, with probability very close to one, the experimental proportion f_n differs from p by less than ϵ.

15.3 Strong law of large numbers

Theorem 15.4 *Let*

$$s_n = b_1 + \cdots + b_n \tag{15.13}$$

be the number of successes in the first n trials. Then

$$\mu(s_n) = np \tag{15.14}$$

and

$$\mu((s_n - np)^4) = n(pq^4 + qp^4) + 3n(n-1)(pq)^2. \tag{15.15}$$

This is bounded by $(1/4)n^2$ for $n \geq 4$.

Proof: Expand $(s_n - np)^4 = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n (b_i - p)(b_j - p)(b_k - p)(b_l - p)$. The expectation of each of the terms vanishes unless all four indices coincide or there are two pairs of coinciding indices. The expectation for the case when all four indices coincide is $(1-p)^4 p + (0-p)^4 q = q^4 p + p^4 q = pq(q^3 + p^3)$. There are n such terms. The expectation when there are two pairs of coinciding indices works out to be $(pq)^2$. There are $3n(n-1)$ such terms.

The inequality then follows from $npq(q^3+p^3)+3n^2(pq)^2 \leq n/4+3/(16)n^2 \leq (1/4)n^2$ for $n \geq 4$. □

Corollary 15.5 *Let*

$$f_n = \frac{b_1 + \cdots + b_n}{n} \tag{15.16}$$

be the proportion of successes in the first n trials. Then

$$\mu(f_n) = p \tag{15.17}$$

and

$$\mu((f_n - p)^4) \leq \frac{1}{4n^2} \tag{15.18}$$

for $n \geq 4$.

Corollary 15.6 (Strong law of large numbers) *Let*

$$f_n = \frac{b_1 + \cdots + b_n}{n} \tag{15.19}$$

be the proportion of successes in the first n trials. Then

$$\mu(\sum_{n=k}^{\infty} (f_n - p)^4) \leq \frac{1}{4(k-1)} \tag{15.20}$$

for $k \geq 4$.

This corollary has a remarkable consequence. Fix k. The fact that the expectation is finite implies that the sum converges almost everywhere. In particular, the terms of the sum approach zero almost everywhere. This means that $f_n \to p$ as $n \to \infty$ almost everywhere. This is the traditional formulation of the strong law of large numbers.

Corollary 15.7 (Strong law of large numbers) *Let*

$$f_n = \frac{b_1 + \cdots + b_n}{n} \tag{15.21}$$

be the proportion of successes in the first n trials. Then for $k \geq 4$

$$\mu(\sup_{n \geq k} |f_n - p| \geq \epsilon) \leq \frac{1}{4(k-1)\epsilon^4}. \tag{15.22}$$

Proof: This corollary follows from the trivial fact that $\sup_{n \geq k} |f_n - p|^4 \leq \sum_{n=k}^{\infty} (f_n - p)^4$ and Chebyshev's inequality. □

This corollary give a perhaps more intuitive picture of the meaning of the strong law of large numbers. Consider a tiny $\epsilon > 0$. Then it says that if k is sufficiently large, then, with probability very close to one, for the entire future history of $n \geq k$ the experimental proportions f_n differ from p by less than ϵ.

15.4 Random walk

Let $w_j = 1 - 2b_j$, so that $b_j = 0$ gives $w_j = 1$ and $b_j = 1$ gives $w_j = -1$. Then the sequence $x_n = w_1 + \cdots w_n$ is called *random walk* starting at zero. In the case when $p = q = 1/2$ this is called symmetric random walk.

Theorem 15.8 *Let* ρ_{01} *be the probability that the random walk starting at zero ever reaches 1. Then this is a solution of the equation*

$$q\rho^2 - \rho + p = (q\rho - p)(\rho - 1) = 0. \tag{15.23}$$

In particular, if $p = q = 1/2$*, then* $\rho_{01} = 1$.

Proof: Let $\rho = \rho_{01}$. The idea of the proof is to break up the computation of ρ into the case when the first step is positive and the case when the first step is negative. Then the equation

$$\rho = p + q\rho^2 \tag{15.24}$$

is intuitive. The probability of succeeding at once is p. Otherwise there must be a failure followed by getting from -1 to 0 and then from 0 to 1. However getting from -1 to 0 is of the same difficulty as getting from 0 to 1.

To make this intuition precise, lett τ_1 be the first time that the walk reaches one. Then

$$\rho = \mu(\tau_1 < +\infty) = \mu(w_1 = 1, \tau_1 < +\infty) + \mu(w_1 = -1, \tau_1 < +\infty). \tag{15.25}$$

The value of the first term is p.

The real problem is with the second term. Write it as

$$\mu(w_1 = -1, \tau_1 < +\infty) = \sum_{k=2}^{\infty} \mu(w_1 = -1, \tau_0 = k, \tau_1 < +\infty) = \sum_{k=2}^{\infty} q\mu(\tau_1 = k-1)\rho = q\rho^2. \tag{15.26}$$

This gives the conclusion. It may be shown that when $p < q$ the correct solution is $\rho = p/q$. □

Notice the dramatic fact that when $p = q = 1/2$ the probability that the random walk gets to the next higher point is one. It is not hard to extend this to show that the probability that the random walk gets to any other point is also one. So the symmetric random walk must do a lot of wandering.

Theorem 15.9 *Let m_{01} be the expected time until the random walk starting at zero reaches 1. Then m_{01} is a solution of*

$$m = 1 + 2qm. \tag{15.27}$$

In particular, when $p = q = 1/2$ the solution is $m = +\infty$.

Proof: Let $m = m_{01}$. The idea of the proof is to break up the computation of ρ into the case when the first step is positive and the case when the first step is negative. Then the equation

$$m = p + q(1 + 2m). \tag{15.28}$$

is intuitive. The probability of succeeding at once is p, and this takes time 1. Otherwise $\tau_1 = 1 + (\tau_0 - 1) + (\tau_1 - \tau_0)$. However the average of the time $\tau_0 - 1$ to get from -1 to 0 is the same as the average of the time $\tau_1 - \tau_0$ to get from 0 to 1.

A more detailed proof is to write

$$m = \mu(\tau_1) = \mu(\tau_1 1_{w_1=1}) + \mu(\tau_1 1_{w_1=-1}). \tag{15.29}$$

The value of first term is p.

The second term is

$$\mu(\tau_1 1_{w_1=-1}) = \mu((1+(\tau_0-1)+(\tau_1-\tau_0))1_{w_1=-1}) = q+q\mu(\tau_1)+q\mu(\tau_1) = q(1+2m). \tag{15.30}$$

It may be shown that when $p > q$ the correct solution is $m = 1/(p - q)$. □

When $p = q = 1/2$ the expected time for the random walk to get to the next higher point is infinite. This is because there is some chance that the symmetric random walk wanders for a very long time on the negative axis before getting to the points above zero.

Problems

1. Consider a random sample of size n from a very large population. The experimental question is to find what proportion p of people in the population have a certain opinion. The proportion in the random sample who have the opinion is f_n. How large must n be so that the standard deviation of f_n in this type of experiment is guaranteed to be no larger than one percent?

2. Recall that $f_n(x) \to f(x)$ as $n \to \infty$ means $\forall \epsilon > 0 \, \exists N \, \forall n \geq N \, |f_n(x) - f(x)| < \epsilon$. Show that $f_n \to f$ almost everywhere is equivalent to

$$\mu(\{x \mid \exists \epsilon > 0 \, \forall N \, \exists n \geq N \, |f_n(x) - f(x)| \geq \epsilon\}) = 0. \tag{15.31}$$

3. Show that $f_n \to f$ almost everywhere is equivalent to for all $\epsilon > 0$

$$\mu(\{x \mid \forall N \, \exists n \geq N \, |f_n(x) - f(x)| \geq \epsilon\}) = 0. \tag{15.32}$$

4. Suppose that the measure of the space is finite. Show that $f_n \to f$ almost everywhere is equivalent to for all $\epsilon > 0$

$$\lim_{N \to \infty} \mu(\{x \mid \exists n \geq N \, |f_n(x) - f(x)| \geq \epsilon\}) = 0. \tag{15.33}$$

   Show that this is not equivalent in the case when the measure of the space may be infinite. Note: Convergence almost everywhere occurs in the strong law of large numbers.

5. Say that $f_n \to f$ *in measure* if for all $\epsilon > 0$

$$\lim_{N \to \infty} \mu(\{x \mid |f_N(x) - f(x)| \geq \epsilon\}) = 0. \tag{15.34}$$

   Show that if the measure of the space is finite, then $f_n \to f$ almost everywhere implies $f_n \to f$ in measure. Note: Convergence in measure occurs in the weak law of large numbers.