Formalizing Integration Theory, with an Application to Probabilistic Algorithms

Stefan Richter
LuFG Theoretische Informatik
RWTH Aachen
AhornstraSSe 55
52056 Aachen
FRG
richter@informatik.rwth-aachen.de

March 17, 2025

Contents

Chapter 1

Prologue

> Verifying more examples of probabilistic algorithms will inevitably necessitate more formalization; in particular we already can see that a theory of expectation will be required to prove the correctness of probabilistic quicksort. If we can continue our policy of formalizing standard theorems of mathematics to aid verifications, then this will provide long-term benefits to many users of the HOL theorem prover.

This quote from the Future Work section of Joe Hurd's PhD thesis "Formal Verification of Probabilistic Algorithms" ([6] p. 131) served as a starting point for the following work. A theory of expectation is nothing but a theory of integration in its probability theoretic underpinnings. And though the proof of correctness for probabilistic quicksort might not need integration, an average runtime analysis certainly will.

As indicated in the very beginning, integration is needed in some way to talk about expectation in probability. The notion that is addressed here is a kind of average value of a random variable with respect to a (probability) measure. The concept of a *measure* lies at the heart of Lebesgue integration. A measure is simply a function satisfying a few sanity properties that maps sets to real numbers. Because the definition does not employ such concrete entities as intervals, it generalizes easily to functions that do not have the real numbers as their domain. In particular, the notion of measure is very natural in the field of probability theory, where a probability measure — nothing but a measure P with $P(\Omega) = 1$ — gives the probability of an event — a measurable subset of Ω.

This Ω might, for example, be the set of all infinite sequences of boolean values, as in Hurd's thesis[6]; our integral is then just a tool that extends this work in the sense depicted at the very beginning of this introduction.

We begin by declaring some preliminary notions, including elementary measure theory and monotone convergence. This leads into measurable real-valued functions, also known as random variables. A sufficient body of functions is shown to belong to this class. The central chapter is about integration proper. We build the integral for increasingly complex functions and prove essential properties, discovering the connection with measurability in the end.

Chapter 2

Measurable Functions

In this chapter, the focus is on the kind of functions to be integrated. As we will see later on, measurability is a good characterization for these functions. Moreover, the language of measure theory as well as the notion of monotone convergence is used frequently in the definition of the integral. So we begin by formalizing these necessary tools.

2.1 Preliminaries

2.1.1 Sigma algebras

theory *Sigma-Algebra* **imports** *Main* **begin**

The **theory** command commences a formal document and enumerates the theories it depends on. With the *Main* theory, a standard selection of useful HOL theories excluding the real numbers is loaded. This theory includes and builds upon a tiny theory of the same name by Markus Wenzel. This theory as well as *Measure* in 2.1.3 is heavily influenced by Joe Hurd's thesis [6] and has been designed to keep the terminology as consistent as possible with that work.

Sigma algebras are an elementary concept in measure theory. To measure — that is to integrate — functions, we first have to measure sets. Unfortunately, when dealing with a large universe, it is often not possible to consistently assign a measure to every subset. Therefore it is necessary to define the set of measurable subsets of the universe. A sigma algebra is such a set that has three very natural and desirable properties.

definition
sigma-algebra:: $'a$ *set set* $\Rightarrow$ *bool* **where**
sigma-algebra $A \longleftrightarrow$
$\{\} \in A \wedge (\forall a.\ a \in A \longrightarrow -a \in A) \wedge$
$(\forall a.\ (\forall\ i{::}nat.\ a\ i \in A) \longrightarrow (\bigcup i.\ a\ i) \in A)$

The **definition** command defines new constants, which are just named functions in HOL. Mind that the third condition expresses the fact that the union of countably many sets in A is again a set in A without explicitly defining the notion of countability.

Sigma algebras can naturally be created as the closure of any set of sets with regard to the properties just postulated. Markus Wenzel wrote the following inductive definition of the *sigma* operator.

inductive-set
$\mathit{sigma} :: 'a\ \mathit{set}\ \mathit{set} \Rightarrow 'a\ \mathit{set}\ \mathit{set}$
for $A :: 'a\ \mathit{set}\ \mathit{set}$
where
$\mathit{basic}\colon a \in A \Longrightarrow a \in \mathit{sigma}\ A$
$\mid \mathit{empty}\colon \{\} \in \mathit{sigma}\ A$
$\mid \mathit{complement}\colon a \in \mathit{sigma}\ A \Longrightarrow -a \in \mathit{sigma}\ A$
$\mid \mathit{Union}\colon (\bigwedge i{::}\mathit{nat}.\ a\ i \in \mathit{sigma}\ A) \Longrightarrow (\bigcup i.\ a\ i) \in \mathit{sigma}\ A$

He also proved the following basic facts. The easy proofs are omitted.

theorem $\mathit{sigma\text{-}UNIV}\colon \mathit{UNIV} \in \mathit{sigma}\ A\langle \mathit{proof}\rangle$

theorem $\mathit{sigma\text{-}Inter}\colon$
$(\bigwedge i{::}\mathit{nat}.\ a\ i \in \mathit{sigma}\ A) \Longrightarrow (\bigcap i.\ a\ i) \in \mathit{sigma}\ A\langle \mathit{proof}\rangle$

It is trivial to show the connection between our first definitions. We use the opportunity to introduce the proof syntax.

theorem assumes $\mathit{sa}\colon \mathit{sigma\text{-}algebra}\ A$
— Named premises are introduced like this.

shows $\mathit{sigma\text{-}sigma\text{-}algebra}\colon \mathit{sigma}\ A = A$
$\langle \mathit{proof}\rangle$

These two steps finish their respective proofs, checking that all subgoals have been proven.

To end this theory we prove a special case of the *sigma-Inter* theorem above. It seems trivial that the fact holds for two sets as well as for countably many. We get a first taste of the cost of formal reasoning here, however. The idea must be made precise by exhibiting a concrete sequence of sets.

primrec $\mathit{trivial\text{-}series} :: 'a\ \mathit{set} \Rightarrow 'a\ \mathit{set} \Rightarrow (\mathit{nat} \Rightarrow 'a\ \mathit{set})$
where
$\mathit{trivial\text{-}series}\ a\ b\ 0 = a$
$\mid \mathit{trivial\text{-}series}\ a\ b\ (\mathit{Suc}\ n) = b$

Using **primrec**, primitive recursive functions over inductively defined data types — the natural numbers in this case — may be constructed.

theorem assumes $s\colon \mathit{sigma\text{-}algebra}\ A$ **and** $a\colon a \in A$ **and** $b\colon b \in A$
shows $\mathit{sigma\text{-}algebra\text{-}inter}\colon a \cap b \in A$

⟨*proof*⟩

Of course, a like theorem holds for union instead of intersection. But as we will not need it in what follows, the theory is finished with the following easy properties instead. Note that the former is a kind of generalization of the last result and could be used to shorten its proof. Unfortunately, this one was needed — and therefore found — only late in the development.

theorem *sigma-INTER*:
  assumes $a{:}(\bigwedge i{::}nat.\ i \in S \Longrightarrow a\ i \in sigma\ A)$
  shows $(\bigcap i{\in}S.\ a\ i) \in sigma\ A$⟨*proof*⟩

lemma assumes *s*: *sigma-algebra a* **shows** *sigma-algebra-UNIV*: $UNIV \in a$⟨*proof*⟩

end

2.1.2 Monotone Convergence

theory *MonConv*
imports *Complex-Main*
begin

A sensible requirement for an integral operator is that it be "well-behaved" with respect to limit functions. To become just a little more precise, it is expected that the limit operator may be interchanged with the integral operator under conditions that are as weak as possible. To this end, the notion of monotone convergence is introduced and later applied in the definition of the integral.

In fact, we distinguish three types of monotone convergence here: There are converging sequences of real numbers, real functions and sets. Monotone convergence could even be defined more generally for any type in the axiomatic type class[1] *ord* of ordered types like this.

mon-conv $u\ f \equiv (\forall\, n.\ u\ n \le u\ (Suc\ n)) \wedge Sup\ (range\ u) = f$

However, this employs the general concept of a least upper bound. For the special types we have in mind, the more specific limit — respective union — operators are available, combined with many theorems about their properties. For the type of real- (or rather ordered-) valued functions, the less-or-equal relation is defined pointwise.

$(f \le g) = (\forall\, x.\ f\, x \le g\ x)$

Now the foundations are laid for the definition of monotone convergence. To express the similarity of the different types of convergence, a single overloaded operator is used.

consts
  mon-conv:: $(nat \Rightarrow 'a) \Rightarrow 'a{::}ord \Rightarrow bool$ (‹-↑-› [60,61] 60)

[1] For the concept of axiomatic type classes, see [7, 9]

overloading
  mon-conv-real $\equiv$ *mon-conv* :: $- \Rightarrow real \Rightarrow bool$
  mon-conv-real-fun $\equiv$ *mon-conv* :: $- \Rightarrow ('a \Rightarrow real) \Rightarrow bool$
  mon-conv-set $\equiv$ *mon-conv* :: $- \Rightarrow 'a\ set \Rightarrow bool$
begin

definition $x{\uparrow}(y{::}real) \equiv (\forall n.\ x\ n \le x\ (Suc\ n)) \wedge x \longrightarrow y$
definition $u{\uparrow}(f{::}'a \Rightarrow real) \equiv (\forall n.\ u\ n \le u\ (Suc\ n)) \wedge\ (\forall w.\ (\lambda n.\ u\ n\ w) \longrightarrow f\ w)$
definition $A{\uparrow}(B{::}'a\ set) \equiv (\forall n.\ A\ n \le A\ (Suc\ n)) \wedge B = (\bigcup n.\ A\ n)$

end

theorem *realfun-mon-conv-iff*: $(u{\uparrow}f) = (\forall w.\ (\lambda n.\ u\ n\ w){\uparrow}((f\ w){::}real))$
  ⟨*proof*⟩

The long arrow signifies convergence of real sequences as defined in the theory *SEQ* [5]. Monotone convergence for real functions is simply pointwise monotone convergence.

Quite a few properties of these definitions will be necessary later, and they are listed now, giving only few select proofs.

lemma assumes *mon-conv*: $x{\uparrow}(y{::}real)$
  shows *mon-conv-mon*: $(x\ i) \le (x\ (m{+}i))$⟨*proof*⟩

lemma *limseq-shift-iff*: $(\lambda m.\ x\ (m{+}i)) \longrightarrow y = x \longrightarrow y$⟨*proof*⟩

theorem assumes *mon-conv*: $x{\uparrow}(y{::}real)$
  shows *real-mon-conv-le*: $x\ i \le y$
⟨*proof*⟩

theorem assumes *mon-conv*: $x{\uparrow}(y{::}('a \Rightarrow real))$
  shows *realfun-mon-conv-le*: $x\ i \le y$
⟨*proof*⟩

lemma assumes *mon-conv*: $x{\uparrow}(y{::}real)$
  and *less*: $z < y$
  shows *real-mon-conv-outgrow*: $\exists n.\ \forall m.\ n \le m \longrightarrow z < x\ m$
⟨*proof*⟩

theorem *real-mon-conv-times*:
  assumes *xy*: $x{\uparrow}(y{::}real)$ **and** *nn*: $0{\le}z$
  shows $(\lambda m.\ z{*}x\ m){\uparrow}(z{*}y)$⟨*proof*⟩

theorem *realfun-mon-conv-times*:
  assumes *xy*: $x{\uparrow}(y{::}'a{\Rightarrow}real)$ **and** *nn*: $0{\le}z$
  shows $(\lambda m\ w.\ z{*}x\ m\ w){\uparrow}(\lambda w.\ z{*}y\ w)$⟨*proof*⟩

theorem *real-mon-conv-add*:

assumes *xy*: $x\uparrow(y::real)$ **and** *ab*: $a\uparrow(b::real)$
shows $(\lambda m.\ x\ m + a\ m)\uparrow(y + b)\langle proof\rangle$
theorem *realfun-mon-conv-add*:
assumes *xy*: $x\uparrow(y::'a\Rightarrow real)$ **and** *ab*: $a\uparrow(b::'a \Rightarrow real)$
shows $(\lambda m\ w.\ x\ m\ w + a\ m\ w)\uparrow(\lambda w.\ y\ w + b\ w)\langle proof\rangle$

theorem *real-mon-conv-bound*:
assumes *mon*: $\bigwedge n.\ c\ n \le c\ (Suc\ n)$
and *bound*: $\bigwedge n.\ c\ n \le (x::real)$
shows $\exists\, l.\ c\uparrow l \wedge l{\le}x$
$\langle proof\rangle$

theorem *real-mon-conv-dom*:
assumes *xy*: $x\uparrow(y::real)$ **and** *mon*: $\bigwedge n.\ c\ n \le c\ (Suc\ n)$
and *dom*: $c \le x$
shows $\exists\, l.\ c\uparrow l \wedge l{\le}y$
$\langle proof\rangle$

theorem *realfun-mon-conv-bound*:
  assumes *mon*: $\bigwedge n.\ c\ n \le c\ (\mathit{Suc}\ n)$
  and *bound*: $\bigwedge n.\ c\ n \le (x{::}'a \Rightarrow \mathit{real})$
  shows $\exists\, l.\ c{\uparrow}l \wedge l{\le}x$⟨*proof*⟩

This brings the theory to an end. Notice how the definition of the limit of a real sequence is visible in the proof to *real-mon-conv-outgrow*, a lemma that will be used for a monotonicity proof of the integral of simple functions later on.

⟨*proof*⟩

end

2.1.3 Measure spaces

theory *Measure*
imports *Sigma-Algebra MonConv*
begin

Now we are already set for the central concept of measure. The following definitions are translated as faithfully as possible from those in Joe Hurd's thesis [6].

definition
  measurable:: $'a\ \mathit{set}\ \mathit{set} \Rightarrow\ 'b\ \mathit{set}\ \mathit{set} \Rightarrow (\,'a \Rightarrow\ 'b)\ \mathit{set}$ **where**
  $\mathit{measurable}\ F\ G = \{f.\ \forall\, g{\in}G.\ f\ -\text{`}\ g \in F\}$

So a function is called F-G-measurable if and only if the inverse image of any set in G is in F. F and G are usually the sets of measurable sets, the first component of a measure space[2].

definition
  measurable-sets:: $(\,'a\ \mathit{set}\ \mathit{set} * (\,'a\ \mathit{set} \Rightarrow \mathit{real})) \Rightarrow\ 'a\ \mathit{set}\ \mathit{set}$ **where**
  $\mathit{measurable\text{-}sets} = \mathit{fst}$

definition
  measure:: $(\,'a\ \mathit{set}\ \mathit{set} * (\,'a\ \mathit{set} \Rightarrow \mathit{real})) \Rightarrow (\,'a\ \mathit{set} \Rightarrow \mathit{real})$ **where**
  $\mathit{measure} = \mathit{snd}$

The other component is the measure itself. It is a function that assigns a nonnegative real number to every measurable set and has the property of being countably additive for disjoint sets.

definition
  positive:: $(\,'a\ \mathit{set}\ \mathit{set} * (\,'a\ \mathit{set} \Rightarrow \mathit{real})) \Rightarrow \mathit{bool}$ **where**
  $\mathit{positive}\ M \longleftrightarrow \mathit{measure}\ M\ \{\} = 0\ \wedge$
  $(\forall\, A.\ A{\in}\ \mathit{measurable\text{-}sets}\ M \longrightarrow 0 \le \mathit{measure}\ M\ A)$

[2] In standard mathematical notation, the universe is first in a measure space triple, but in our definitions, following Joe Hurd, it is always the whole type universe and therefore omitted.

definition
$countably\text{-}additive{::}\ ('a\ set\ set * ('a\ set => real)) => bool$ **where**
$countably\text{-}additive\ M \longleftrightarrow (\forall f{::}(nat => 'a\ set).\ range\ f \subseteq measurable\text{-}sets\ M$
$\wedge\ (\forall\, m\ n.\ m \neq n \longrightarrow f\ m \cap f\ n = \{\}) \wedge\ (\bigcup i.\ f\ i) \in measurable\text{-}sets\ M$
$\longrightarrow (\lambda n.\ measure\ M\ (f\ n))\ sums\ \ measure\ M\ (\bigcup i.\ f\ i))$

This last property deserves some comments. The conclusion is usually — also in the aforementioned source — phrased as
$measure\ M\ (\bigcup i.\ f\ i) = (\sum n.\ measure\ M\ (f\ n))$.
In our formal setting this is unsatisfactory, because the sum operator[3], like any HOL function, is total, although a series obviously need not converge. It is defined using the ε operator, and its behavior is unspecified in the diverging case. Hence, the above assertion would give no information about the convergence of the series.

Furthermore, the definition contains redundancy. Assuming that the countable union of sets is measurable is unnecessary when the measurable sets form a sigma algebra, which is postulated in the final definition[4].

definition
$measure\text{-}space{::}\ ('a\ set\ set * ('a\ set \Rightarrow real)) \Rightarrow bool$ **where**
$measure\text{-}space\ M \longleftrightarrow sigma\text{-}algebra\ (measurable\text{-}sets\ M) \wedge$
$positive\ M \wedge countably\text{-}additive\ M$

Note that our definition is restricted to finite measure spaces — that is, $measure\ M\ UNIV < \infty$ — since the measure must be a real number for any measurable set. In probability, this is naturally the case.

Two important theorems close this section. Both appear in Hurd's work as well, but are shown anyway, owing to their central role in measure theory. The first one is a mighty tool for proving measurability. It states that for a function mapping one sigma algebra into another, it is sufficient to be measurable regarding only a generator of the target sigma algebra. Formalizing the interesting proof out of Bauer's textbook [1] is relatively straightforward using rule induction.

theorem assumes sig: $sigma\text{-}algebra\ a$ **and** $meas$: $f \in measurable\ a\ b$ **shows**
$measurable\text{-}lift$: $f \in measurable\ a\ (sigma\ b)$
$\langle proof \rangle$

The case is different for the second theorem. It is only five lines in the book (ibid.), but almost 200 in formal text. Precision still pays here, gaining a detailed view of a technique that is often employed in measure theory — making a sequence of sets disjoint. Moreover, the necessity for the above-

[3] Which is merely syntactic sugar for the *suminf* functional from the *Series* theory [5].

[4] Joe Hurd inherited this practice from a very influential probability textbook [10]

mentioned change in the definition of countably additive was detected only in the formalization of this proof.

To enable application of the additivity of measures, the following construction yields disjoint sets. We skip the justification of the lemmata for brevity.

```
primrec mkdisjoint:: (nat ⇒ 'a set) ⇒ (nat ⇒ 'a set)
where
  mkdisjoint A 0 = A 0
| mkdisjoint A (Suc n) = A (Suc n) − A n

lemma mkdisjoint-un:
  assumes up: ⋀n. A n ⊆ A (Suc n)
  shows A n = (⋃ i∈{..n}. mkdisjoint A i)⟨proof⟩

lemma mkdisjoint-disj:
  assumes up: ⋀n. A n ⊆ A (Suc n) and ne: m ≠ n
  shows mkdisjoint A m ∩ mkdisjoint A n = {}⟨proof⟩

lemma mkdisjoint-mon-conv:
  assumes mc: A↑B
  shows (⋃ i. mkdisjoint A i) = B⟨proof⟩
```

Joe Hurd calls the following the Monotone Convergence Theorem, though in mathematical literature this name is often reserved for a similar fact about integrals that we will prove in 3.2.2, which depends on this one. The claim made here is that the measures of monotonically convergent sets approach the measure of their limit. A strengthened version would imply monotone convergence of the measures, but is not needed in the development.

```
theorem measure-mon-conv:
  assumes ms: measure-space M and
  Ams: ⋀n. A n ∈ measurable-sets M and AB: A↑B
  shows (λn. measure M (A n)) ⟶ measure M B
⟨proof⟩⟨proof⟩
```

2.2 Real-Valued random variables

theory *RealRandVar*
imports *Measure HOL−Library.Countable*
begin

While most of the above material was modeled after Hurd's work (but still proved independently), the original content basically starts here[5]. From now on, we will specialize in functions that map into the real numbers and are measurable with respect to the canonical sigma algebra on the reals, the Borel sigma algebra. These functions will be called real-valued random variables. The terminology is slightly imprecise, as random variables hint at a probability space, which usually requires *measure M UNIV = 1*. Notwithstanding, as we regard only finite measures (cf. 2.1.3), this condition can easily be achieved by normalization. After all, the other standard name, "measurable functions", is even less precise.

A lot of the theory in this and the preceding section has also been formalized within the Mizar project [3, 4]. The abstract of the second source hints that it was also planned as a stepping stone for Lebesgue integration, though further results in this line could not be found. The main difference lies in the use of extended real numbers — the reals together with ±∞ — in those documents. It is established practice in measure theory to allow infinite values, but "(...) we felt that the complications that this generated (...) more than canceled out the gain in uniformity (...), and that a simpler theory resulted from sticking to the standard real numbers." [6, p. 32f]. Hurd also advocates going directly to the hyper-reals, should the need for infinite measures arise. I agree, nevertheless sticking to his example for the reasons mentioned in the prologue.

definition
  Borelsets:: *real set set* (‹$\mathbb{B}$›) **where**
  $\mathbb{B}$ = *sigma* $\{S.\ \exists\, u.\ S{=}\{..u\}\}$

definition

  rv:: ($'a$ *set set* $*$ ($'a$ *set* $\Rightarrow$ *real*)) $\Rightarrow$ ($'a \Rightarrow$ *real*) *set* **where**
  rv M = $\{f.$ *measure-space M* $\wedge\ f \in$ *measurable* (*measurable-sets M*) $\mathbb{B}\}$

As explained in the first paragraph, the preceding definitions[6] determine the rest of this section. There are many ways to define the Borel sets. For example, taking into account only rationals for u would also have worked

[5] There are two main reasons why the above has not been imported using Sebastian Skalberg's import tool [8]. Firstly, there are inconveniences caused by different conventions in HOL, meaning predicates instead of sets foremost, that make the consistent use of such basic definitions impractical. What is more, the import tool simply was not available at the time these theories were written.

[6] The notation $\{..u\}$ signifies the interval from negative infinity to u included.

out above, but we can take the reals to simplify things. The smallest sigma algebra containing all the open (or closed) sets is another alternative; the multitude of possibilities testifies to the relevance of the concept.

The latter path leads the way to the fact that any continuous function is measurable. Generalization for $\mathbb{R}^n$ brings another unified way to prove all the measurability theorems in this theory plus, for instance, measurability of the trigonometric and exponential functions. This approach is detailed in another influential textbook by Billingsley [2]. It requires some concepts of topologic spaces, which made the following elementary course, based on Bauer's excellent book [1], seem more feasible.

Two more definitions go next. The image measure, law, or distribution — the last term being specific to probability — of a measure with respect to a measurable function is calculated as the measure of the inverse image of a set. Characteristic functions will be frequently needed in the rest of the development.

definition
distribution::
$('a\ set\ set * ('a\ set \Rightarrow real)) \Rightarrow ('a \Rightarrow real) \Rightarrow (real\ set \Rightarrow real)$ (‹*law*›) **where**
$f \in rv\ M \Longrightarrow law\ M\ f \equiv (measure\ M) \circ (vimage\ f)$

definition
characteristic-function:: $'a\ set \Rightarrow ('a \Rightarrow real)$ (‹χ -›) **where**
$\chi\ A\ x \equiv if\ x \in A\ then\ 1\ else\ 0$

lemma *char-empty*: $\chi\ \{\} = (\lambda t.\ 0)$
⟨*proof*⟩

Now that random variables are defined, we aim to show that a broad class of functions belongs to them. For a constant function this is easy, as there are only two possible preimages.

lemma assumes *sigma*: *sigma-algebra S*
shows *const-measurable*: $(\lambda x.\ (c::real)) \in measurable\ S\ X$
⟨*proof*⟩

theorem assumes *ms*: *measure-space M*
shows *const-rv*: $(\lambda x.\ c) \in rv\ M$ ⟨*proof*⟩

Characteristic functions produce four cases already, so the details are glossed over.

lemma assumes *a*: $a \in S$ **and** *sigma*: *sigma-algebra S* **shows**
char-measurable : $\chi\ a \in measurable\ S\ x$⟨*proof*⟩

theorem assumes *ms*: *measure-space M* **and** *A*: $A \in measurable\text{-}sets\ M$
shows *char-rv*: $\chi\ A \in rv\ M$ ⟨*proof*⟩

For more intricate functions, the following application of the measurability lifting theorem from 2.1.3 gives a useful characterization.

theorem assumes *ms*: *measure-space* M **shows**
rv-le-iff: $(f \in rv\ M) = (\forall\, a.\ \{w.\ f\ w \leq a\} \in measurable\text{-}sets\ M)$
⟨*proof*⟩

The next four lemmata allow for a ring deduction that helps establish this fact for all of the signs $<$, $>$ and $\geq$ as well.

lemma assumes *sigma*: *sigma-algebra* A **and** *le*: $\forall\, a.\ \{w.\ f\ w \leq a\} \in A$
shows *le-less*: $\forall\, a.\ \{w.\ f\ w < (a::real)\} \in A$
⟨*proof*⟩

lemma assumes *sigma*: *sigma-algebra* A **and** *less*: $\forall\, a.\ \{w.\ f\ w < a\} \in A$
shows *less-ge*: $\forall\, a.\ \{w.\ (a::real) \leq f\ w\} \in A$
⟨*proof*⟩

lemma assumes *sigma*: *sigma-algebra* A **and** *ge*: $\forall\, a.\ \{w.\ a \leq f\ w\} \in A$
shows *ge-gr*: $\forall\, a.\ \{w.\ (a::real) < f\ w\} \in A$⟨*proof*⟩

lemma assumes *sigma*: *sigma-algebra* A **and** *gr*: $\forall\, a.\ \{w.\ a < f\ w\} \in A$
shows *gr-le*: $\forall\, a.\ \{w.\ f\ w \leq (a::real)\} \in A$⟨*proof*⟩
theorem assumes *ms*: *measure-space* M **shows**
rv-ge-iff: $(f \in rv\ M) = (\forall\, a.\ \{w.\ a \leq f\ w\} \in measurable\text{-}sets\ M)$
⟨*proof*⟩

theorem assumes *ms*: *measure-space* M **shows**
rv-gr-iff: $(f \in rv\ M) = (\forall\, a.\ \{w.\ a < f\ w\} \in measurable\text{-}sets\ M)$⟨*proof*⟩

theorem assumes *ms*: *measure-space* M **shows**
rv-less-iff: $(f \in rv\ M) = (\forall\, a.\ \{w.\ f\ w < a\} \in measurable\text{-}sets\ M)$⟨*proof*⟩

As a first application we show that addition and multiplication with constants preserve measurability. This is a precursor to the more general addition and multiplication theorems later on. You can see that quite a few properties of the real numbers are employed.

lemma assumes *g*: $g \in rv\ M$
shows *affine-rv*: $(\lambda x.\ (a::real) + (g\ x) * b) \in rv\ M$
⟨*proof*⟩

For the general case of addition, we need one more set to be measurable, namely $\{w.\ f\ w \leq g\ w\}$. This follows from a like statement for $<$. A dense and countable subset of the reals is needed to establish it.

Of course, the rationals come to mind. They were not available in Isabelle/HOL[7], so I built a theory with the necessary properties on my own. [Meanwhile Isabelle has proper rationals and SR's development of the rationals has been moved to and merged with Isabelle's rationals.

[7] At least not as a subset of the reals, to the definition of which a type of positive rational numbers contributed [5].

lemma assumes *f*: $f \in \mathit{rv}\ M$ **and** *g*: $g \in \mathit{rv}\ M$
  shows *rv-less-rv-measurable*: $\{w.\ f\ w < g\ w\} \in \mathit{measurable\text{-}sets}\ M$
$\langle \mathit{proof} \rangle$

lemma assumes *f*: $f \in \mathit{rv}\ M$ **and** *g*: $g \in \mathit{rv}\ M$
  shows *rv-le-rv-measurable*: $\{w.\ f\ w \leq g\ w\} \in \mathit{measurable\text{-}sets}\ M$ (**is** $?a \in\ ?M$)
$\langle \mathit{proof} \rangle$

lemma assumes *f*: $f \in \mathit{rv}\ M$ **and** *g*: $g \in \mathit{rv}\ M$
  shows *f-eq-g-measurable*: $\{w.\ f\ w = g\ w\} \in \mathit{measurable\text{-}sets}\ M$ $\langle \mathit{proof} \rangle$

lemma assumes *f*: $f \in \mathit{rv}\ M$ **and** *g*: $g \in \mathit{rv}\ M$
  shows *f-noteq-g-measurable*: $\{w.\ f\ w \neq g\ w\} \in \mathit{measurable\text{-}sets}\ M$ $\langle \mathit{proof} \rangle$

With these tools, a short proof for the addition theorem is possible.

theorem assumes *f*: $f \in \mathit{rv}\ M$ **and** *g*: $g \in \mathit{rv}\ M$
  shows *rv-plus-rv*: $(\lambda w.\ f\ w + g\ w) \in \mathit{rv}\ M$
$\langle \mathit{proof} \rangle$

To show preservation of measurability by multiplication, it is expressed by addition and squaring. This requires a few technical lemmata including the one stating measurability for squares, the proof of which is skipped.

lemma *pow2-le-abs*: $(a^2 \leq b^2) = (|a| \leq |b{::}\mathit{real}|)\langle \mathit{proof} \rangle$
lemma assumes *f*: $f \in \mathit{rv}\ M$
  shows *rv-square*: $(\lambda w.\ (f\ w)^2) \in \mathit{rv}\ M\langle \mathit{proof} \rangle$
lemma *realpow-two-binomial-iff*: $(f{+}g{::}\mathit{real})^2 = f^2 + 2{*}(f{*}g) + g^2$
  $\langle \mathit{proof} \rangle$
lemma *times-iff-sum-squares*: $f{*}g = (f{+}g)^2/4 - (f{-}g)^2/(4{::}\mathit{real})$
  $\langle \mathit{proof} \rangle$

theorem assumes *f*: $f \in \mathit{rv}\ M$ **and** *g*: $g \in \mathit{rv}\ M$
  shows *rv-times-rv*: $(\lambda w.\ f\ w * g\ w) \in \mathit{rv}\ M$
$\langle \mathit{proof} \rangle$

The case of substraction is an easy consequence of *rv-plus-rv* and *rv-times-rv*.

theorem *rv-minus-rv*:
  assumes *f*: $f \in \mathit{rv}\ M$ **and** *g*: $g \in \mathit{rv}\ M$
  shows $(\lambda t.\ f\ t - g\ t) \in \mathit{rv}\ M\langle \mathit{proof} \rangle$

Measurability for limit functions of monotone convergent series is also surprisingly straightforward.

theorem assumes *u*: $\bigwedge n.\ u\ n \in \mathit{rv}\ M$ **and** *mon-conv*: $u{\uparrow}f$
  shows *mon-conv-rv*: $f \in \mathit{rv}\ M$
$\langle \mathit{proof} \rangle$

Before we end this chapter to start the formalization of the integral proper, there is one more concept missing: The positive and negative part of a function. Their definition is quite intuitive, and some useful properties are given right away, including the fact that they are random variables, provided that their argument functions are measurable.

definition
nonnegative:: $('a \Rightarrow ('b::\{ord,zero\})) \Rightarrow bool$ **where**
$nonnegative\ f \longleftrightarrow (\forall x.\ 0 \le f\ x)$

definition
positive-part:: $('a \Rightarrow ('b::\{ord,zero\})) \Rightarrow ('a \Rightarrow 'b)$ (‹*pp*›) **where**
$pp\ f\ x = (if\ 0{\le}f(x)\ then\ f\ x\ else\ 0)$

definition
negative-part:: $('a \Rightarrow ('b::\{ord,zero,uminus,minus\})) \Rightarrow ('a \Rightarrow 'b)$ (‹*np*›) **where**
$np\ f\ x = (if\ 0{\le}f(x)\ then\ 0\ else\ -f(x))$

lemma *f-plus-minus*: $((f\ x)::real) = pp\ f\ x - np\ f\ x$
⟨*proof*⟩

lemma *f-plus-minus2*: $(f::'a \Rightarrow real) = (\lambda t.\ pp\ f\ t - np\ f\ t)$
⟨*proof*⟩

lemma *f-abs-plus-minus*: $(|f\ x|::real) = pp\ f\ x + np\ f\ x$
⟨*proof*⟩

lemma *nn-pp-np*: **assumes** *nonnegative f*
shows $pp\ f = f$ **and** $np\ f = (\lambda t.\ 0)$ ⟨*proof*⟩

lemma *pos-pp-np-help*: $\bigwedge x.\ 0{\le}f\ x \Longrightarrow pp\ f\ x = f\ x \wedge np\ f\ x = 0$
⟨*proof*⟩

lemma *real-neg-pp-np-help*: $\bigwedge x.\ f\ x \le (0::real) \Longrightarrow np\ f\ x = -f\ x \wedge pp\ f\ x = 0$⟨*proof*⟩
lemma *real-neg-pp-np*: **assumes** $f \le (\lambda t.\ (0::real))$
shows $np\ f = (\lambda t.\ -f\ t)$ **and** $pp\ f = (\lambda t.\ 0)$ ⟨*proof*⟩

lemma assumes *a*: $0{\le}(a::real)$
shows *real-pp-np-pos-times*:
$pp\ (\lambda t.\ a{*}f\ t) = (\lambda t.\ a{*}pp\ f\ t) \wedge\ np\ (\lambda t.\ a{*}f\ t) = (\lambda t.\ a{*}np\ f\ t)$⟨*proof*⟩
lemma assumes *a*: $(a::real){\le}0$
shows *real-pp-np-neg-times*:
$pp\ (\lambda t.\ a{*}f\ t) = (\lambda t.\ -a{*}np\ f\ t) \wedge\ np\ (\lambda t.\ a{*}f\ t) = (\lambda t.\ -a{*}pp\ f\ t)$⟨*proof*⟩

lemma *pp-np-rv*:
assumes *f*: $f \in rv\ M$
shows $pp\ f \in rv\ M$ **and** $np\ f \in rv\ M$
⟨*proof*⟩

theorem *pp-np-rv-iff*: $(f::'a \Rightarrow real) \in rv\ M = (pp\ f \in rv\ M \wedge np\ f \in rv\ M)$⟨*proof*⟩

This completes the chapter about measurable functions. As we will see in the next one, measurability is the prime condition on Lebesgue integrable functions; and the theorems and lemmata established here suffice — at least in principle — to show it holds for any function that is to be integrated there.

end

Chapter 3

Integration

The chapter at hand assumes a central position in the present paper. The Lebesgue integral is defined and its characteristics are shown in 3.2. To illustrate the problems arising in doing so, we first look at implementation alternatives that did not work out.

3.1 Two approaches that failed

Defining Lebesgue integration can be quite involved, judging by the process in 3.2 that imitates Bauer's way [1]. So it is quite tempting to try cutting a corner. The following two alternative approaches back up my experience that this almost never pays in formalization. The theory that seems most complex at first sight is often the one that is closest to formal reasoning and deliberately avoids "hand-waving".

3.1.1 A closed expression

In contrast, Billingsley's definition [2, p. 172] is strikingly short. For non-negative measurable functions f:

> $\int f d\mu = sup \sum_i \left[inf_{\omega \in A_i} f(w) \right] \mu(A_i)$.
>
> The supremum here extends over all finite decompositions $\{A_i\}$ of Ω into $\mathcal{F}$-sets.[1]

Like the definition, the proofs of the essential properties are also rather short, about three pages in the textbook for almost all the theorems in 3.2; and a proof of uniqueness is obsolete for a closed expression like this. Therefore, I found this approach quite tempting. It turns out, however,

[1] The $\mathcal{F}$-sets are just the measurable sets of a measure space.

that it is unfortunately not well suited for formalization, at least with the background we use.

A complication shared by all possible styles of definition is the lack of infinite values in our theory, combined with the lack of partial functions in HOL. Like the sum operator in 2.1.3, the integral has to be defined indirectly. The classical way to do this employs predicates, invoking ε to choose the value that satisfies the condition:

$\int f\, dM \equiv (\varepsilon\ i.\ \textit{is-integral}\ M\ f\ i)$

To sensibly apply this principle, the predicate has to be ε-free to supply the information if the integral is defined or not. Now the above definition contains up to three additional ε when formalized naively in HOL, namely in the supremum, infimum and sum operators. The sum is over a finite set, so it can be replaced by a total function. For nonnegative functions, the infimum can also be shown to exist everywhere, but, like the supremum, must itself be replaced by a predicate.

Also note that predicates require a proof of uniqueness, thus losing the prime advantage of a closed formula anyway. In this case, uniqueness can be reduced to uniqueness of the supremum/infimum. The problem is that neither suprema nor infima come predefined in Isabelle/Isar as of yet. It is an easy task to make up for this — and I did — but a much harder one to establish all the properties needed for reasoning with the defined entities.

A lot of such reasoning is necessary to deduce from the above definition (or a formal version of it, as just outlined) the basic behavior of integration, which includes additivity, monotonicity and especially the integral of simple functions. It turns out that the brevity of the proofs in the textbook stems from a severely informal style that assumes ample background knowledge. Formalizing all this knowledge started to become overwhelming when the idea of a contrarian approach emerged.

3.1.2 A one-step inductive definition

This idea was sparked by the following note: “(. . .) the integral is uniquely determined by certain simple properties it is natural to require of it” [2, p. 175]. Billingsley goes on discussing exactly those properties that are so hard to derive from his definition. So why not simply define integration using these properties? That is the gist of an inductive set definition, like the one we have seen in 2.1.1. This time a functional operator is to be defined, but it can be represented as a set of pairs, where the first component is the function and the second its integral. To cut a long story short, here is the definition.

inductive-set
  integral-set:: $('a\ set\ set * ('a\ set \Rightarrow real)) \Rightarrow (('a \Rightarrow real) * real)\ set$
  for $M :: 'a\ set\ set * ('a\ set \Rightarrow real)$

```
 where
  char: ⟦f = χ A; A ∈ measurable-sets M⟧ ⟹ (f,measure M A) ∈ integral-set M
 | add: ⟦f = (λw. g w + h w); (g,x) ∈ integral-set M; (h,y) ∈ integral-set M⟧
  ⟹ (f,(x + y)) ∈ integral-set M
 | times: ⟦f = (λw. a∗g w); (g,x) ∈ integral-set M⟧ ⟹ (f,a∗x) ∈ integral-set M
 | mon-conv: ⟦u↑f; ⋀n. (u n, x n) ∈ integral-set M; x↑y⟧
  ⟹ (f,y) ∈ integral-set M
```

The technique is also encountered in the *Finite-Set* theory from the Isabelle library. It is used there to define the *sum* function, which calculates a sum indexed over a finite set and is employed in 3.2. The definition here is much more intricate though.

An obvious advantage of this approach is that almost all important properties are gained without effort. The introduction rule *mon-conv* corresponds to what is known as the Monotone Convergence Theorem in scientific literature; negative functions are also provided for via the *times* rule. To be precise, there is exactly one important theorem missing — uniqueness. That is, every function appears in at most one pair.

From uniqueness together with the introduction rules, all the other statements about integration, monotonicity for example, could be derived. On the other hand, monotonicity implies uniqueness. Much to my regret, none of these two could be proven. The proof would basically amount to a double induction to show that an integral gained via one rule is the same when derived by another. A lot of effort was spent trying to strengthen the induction hypothesis or reduce the goal to a simpler case. All of this was in vain though, and it seems that the hypothesis would have to be strengthened as far as to include the concept of integration in the first place, which in a way defeats the advantages of the approach.

3.2 The three-step approach

```
theory Integral
imports RealRandVar
begin
```

Having learnt from my failures, we take the safe and clean way of Heinz Bauer [1]. It proceeds as outlined in the introduction. In three steps, we fix the integral for elementary ("step-")functions, for limits of these, and finally for differences between such limits.

3.2.1 Simple functions

A simple function is a finite sum of characteristic functions, each multiplied with a nonnegative constant. These functions must be parametrized by measurable sets. Note that to check this condition, a tuple consisting of a set of measurable sets and a measure is required as the integral operator's second argument, whereas the measure only is given in informal notation. Usually the tuple will be a measure space, though it is not required so by the definition at this point.

It is most natural to declare the value of the integral in this elementary case by simply replacing the characteristic functions with the measures of their respective sets. Uniqueness remains to be shown, for a function may have infinitely many decompositions and these might give rise to more than one integral value. This is why we construct a *simple function integral set* for any function and measurable sets/measure pair by means of an inductive set definition containing but one introduction rule.

inductive-set
sfis:: $('a \Rightarrow real) \Rightarrow ('a\ set\ set * ('a\ set \Rightarrow real)) \Rightarrow real\ set$
for f :: $'a \Rightarrow real$ **and** M :: $'a\ set\ set * ('a\ set \Rightarrow real)$
where
base: $[\![f = (\lambda t.\ \sum i{\in}(S{::}nat\ set).\ x\ i * \chi\ (A\ i)\ t);$
$\forall\, i \in S.\ A\ i \in measurable\text{-}sets\ M;\ nonnegative\ x;\ finite\ S;$
$\forall\, i{\in}S.\ \forall\, j{\in}S.\ i \neq j \longrightarrow A\ i \cap A\ j = \{\};\ (\bigcup i{\in}S.\ A\ i) = UNIV]\!]$
$\Longrightarrow (\sum i{\in}S.\ x\ i * measure\ M\ (A\ i)) \in sfis\ f\ M$

As you can see we require two extra conditions, and they amount to the sets being a partition of the universe. We say that a function is in normal form if it is represented this way. Normal forms are only needed to show additivity and monotonicity of simple function integral sets. These theorems can then be used in turn to get rid of the normality condition.

More precisely, normal forms play a central role in the *sfis-present* lemma. For two simple functions with different underlying partitions it states the existence of a common finer-grained partition that can be used to represent the functions uniformly. The proof is remarkably lengthy, another case where informal reasoning is more intricate than it seems. The reason it is included anyway, with the exception of the two following lemmata, is that it gives insight into the arising complication and its formal solution.

The problem is in the use of informal sum notation, which easily permits for a change in index sets, allowing for a pair of indices. This change has to be rectified in formal reasoning. Luckily, the task is eased by an injective function from $\mathbb{N}^2$ into $\mathbb{N}$, which was developed for the rationals mentioned in 2.2. It might have been still easier if index sets were polymorphic in our integral definition, permitting pairs to be formed when necessary, but the logic doesn't allow for this.

lemma assumes *un*: $(\bigcup i{\in}R.\ B\ i) = UNIV$ **and** *fin*: *finite R*
and *dis*: $\forall\, j1{\in}R.\ \forall\, j2{\in}R.\ j1 \neq j2 \longrightarrow (B\ j1) \cap (B\ j2) = \{\}$
shows *char-split*: $\chi\ A\ t = (\sum j{\in}R.\ \chi\ (A \cap B\ j)\ t)$⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩⟨*proof*⟩**lemma**
assumes *measure-space M* **and** $a \in sfis\ f\ M$ **and** $b \in sfis\ g\ M$

shows *sfis-present*: $\exists\ z1\ z2\ C\ K.$
$f = (\lambda t.\ \sum i{\in}(K{::}nat\ set).\ z1\ i * \chi\ (C\ i)\ t) \land g = (\lambda t.\ \sum i{\in}K.\ z2\ i * \chi\ (C\ i)\ t)$
$\land\ a = (\sum i{\in}K.\ z1\ i * measure\ M\ (C\ i)) \land b = (\sum i{\in}K.\ z2\ i * measure\ M\ (C\ i))$
$\land\ finite\ K \land (\forall i{\in}K.\ \forall j{\in}K.\ i \neq j \longrightarrow C\ i \cap C\ j = \{\})$
$\land\ (\forall i \in K.\ C\ i \in measurable\text{-}sets\ M) \land (\bigcup i{\in}K.\ C\ i) = UNIV$
$\land\ nonnegative\ z1 \land nonnegative\ z2$
⟨*proof*⟩

Additivity and monotonicity are now almost obvious, the latter trivially implying uniqueness.

lemma assumes *ms*: *measure-space M* **and** *a*: $a \in sfis\ f\ M$ **and** *b*: $b \in sfis\ g\ M$
shows *sfis-add*: $a{+}b \in sfis\ (\lambda w.\ f\ w + g\ w)\ M$
⟨*proof*⟩

lemma assumes *ms*: *measure-space M* **and** *a*: $a \in sfis\ f\ M$
and *b*: $b \in sfis\ g\ M$ **and** *fg*: $f{\le}g$
shows *sfis-mono*: $a \le b$
⟨*proof*⟩

lemma *sfis-unique*:
assumes *ms*: *measure-space M* **and** *a*: $a \in sfis\ f\ M$ **and** *b*: $b \in sfis\ f\ M$
shows $a{=}b$
⟨*proof*⟩

The integral of characteristic functions, as well as the effect of multiplication with a constant, follows directly from the definition. Together with a generalization of the addition theorem to sums, a less restrictive introduction rule emerges, making normal forms obsolete. It is only valid in measure spaces though.

lemma *sfis-char*:
assumes *ms*: *measure-space M* **and** *mA*: $A \in measurable\text{-}sets\ M$
shows $measure\ M\ A \in sfis\ \chi\ A\ M$⟨*proof*⟩

lemma *sfis-times*:
assumes *a*: $a \in sfis\ f\ M$ **and** *z*: $0{\le}z$
shows $z{*}a \in sfis\ (\lambda w.\ z{*}f\ w)\ M$ ⟨*proof*⟩

lemma assumes *ms*: *measure-space M*
and *a*: $\forall i{\in}S.\ a\ i \in sfis\ (f\ i)\ M$ **and** *S*: *finite S*
shows *sfis-sum*: $(\sum i{\in}S.\ a\ i) \in sfis\ (\lambda t.\ \sum i{\in}S.\ f\ i\ t)\ M$ ⟨*proof*⟩

lemma *sfis-intro*:
assumes *ms*: *measure-space M* **and** *Ams*: $\forall i \in S.\ A\ i \in measurable\text{-}sets\ M$
and *nn*: *nonnegative x* **and** *S*: *finite S*
shows $(\sum i{\in}S.\ x\ i * measure\ M\ (A\ i)) \in$
$sfis\ (\lambda t.\ \sum i{\in}(S{::}nat\ set).\ x\ i * \chi\ (A\ i)\ t)\ M$
⟨*proof*⟩

That is nearly all there is to know about simple function integral sets. It will be useful anyway to have the next two facts available.

lemma *sfis-nn*:
assumes *f*: $a \in \mathit{sfis}\ f\ M$
shows $\mathit{nonnegative}\ f\ \langle\mathit{proof}\rangle$
lemma *sum-rv*:
assumes *rvs*: $\forall k{\in}K.\ (f\ k) \in \mathit{rv}\ M$ **and** *ms*: $\mathit{measure\text{-}space}\ M$
shows $(\lambda t.\ \sum k{\in}K.\ f\ k\ t) \in \mathit{rv}\ M\langle\mathit{proof}\rangle$
lemma *sfis-rv*:
assumes *ms*: $\mathit{measure\text{-}space}\ M$ **and** *f*: $a \in \mathit{sfis}\ f\ M$
shows $f \in \mathit{rv}\ M\ \langle\mathit{proof}\rangle$

3.2.2 Nonnegative Functions

There is one more important fact about *sfis*, easily the hardest one to see. It is about the relationship with monotone convergence and paves the way for a sensible definition of *nnfis*, the nonnegative function integral sets, enabling monotonicity and thus uniqueness. A reasonably concise formal proof could fortunately be achieved in spite of the nontrivial ideas involved — compared for instance to the intuitive but hard-to-formalize *sfis-present*. A small lemma is needed to ensure that the inequation, which depends on an arbitrary z strictly between 0 and 1, carries over to $z = 1$, thereby eliminating z in the end.

lemma *real-le-mult-sustain*:
assumes *zr*: $\bigwedge z.\ [\![0<z;\ z<1]\!] \Longrightarrow z * r \le y$
shows $r \le (y{::}\mathit{real})\langle\mathit{proof}\rangle$
lemma *sfis-mon-conv-mono*:
assumes *uf*: $u{\uparrow}f$ **and** *xu*: $\bigwedge n.\ x\ n \in \mathit{sfis}\ (u\ n)\ M$ **and** *xy*: $x{\uparrow}y$
and *sr*: $r \in \mathit{sfis}\ s\ M$ **and** *sf*: $s \le f$ **and** *ms*: $\mathit{measure\text{-}space}\ M$
shows $r \le y\ \ \langle\mathit{proof}\rangle$

Now we are ready for the second step. The integral of a monotone limit of functions is the limit of their integrals. Note that this last limit has to exist in the first place, since we decided not to use infinite values. Backed by the last theorem and the preexisting knowledge about limits, the usual basic properties are straightforward.

inductive-set
$\mathit{nnfis}{::}\ ('a \Rightarrow \mathit{real}) \Rightarrow ('a\ \mathit{set}\ \mathit{set} * ('a\ \mathit{set} \Rightarrow \mathit{real})) \Rightarrow \mathit{real}\ \mathit{set}$
for $f :: 'a \Rightarrow \mathit{real}$ **and** $M :: 'a\ \mathit{set}\ \mathit{set} * ('a\ \mathit{set} \Rightarrow \mathit{real})$
where
base: $[\![u{\uparrow}f;\ \bigwedge n.\ x\ n \in \mathit{sfis}\ (u\ n)\ M;\ x{\uparrow}y]\!] \Longrightarrow y \in \mathit{nnfis}\ f\ M$

lemma *sfis-nnfis*:
assumes *s*: $a \in \mathit{sfis}\ f\ M$
shows $a \in \mathit{nnfis}\ f\ M\langle\mathit{proof}\rangle$

lemma *nnfis-times*:
assumes *ms*: *measure-space M* **and** *a*: *a* ∈ *nnfis f M* **and** *nn*: *0≤z*
shows *z∗a* ∈ *nnfis* (λ*w*. *z∗f w*) *M* ⟨*proof*⟩

lemma *nnfis-add*:
assumes *ms*: *measure-space M* **and** *a*: *a* ∈ *nnfis f M* **and** *b*: *b* ∈ *nnfis g M*
shows *a+b* ∈ *nnfis* (λ*w*. *f w* + *g w*) *M* ⟨*proof*⟩

lemma assumes *ms*: *measure-space M* **and** *a*: *a* ∈ *nnfis f M*
and *b*: *b* ∈ *nnfis g M* **and** *fg*: *f≤g*
shows *nnfis-mono*: *a* ≤ *b* ⟨*proof*⟩

corollary *nnfis-unique*:
assumes *ms*: *measure-space M* **and** *a*: *a* ∈ *nnfis f M* **and** *b*: *b* ∈ *nnfis f M*
shows *a=b* ⟨*proof*⟩

There is much more to prove about nonnegative integration. Next up is a classic theorem by Beppo Levi, the monotone convergence theorem. In essence, it says that the introduction rule for *nnfis* holds not only for sequences of simple functions, but for any sequence of nonnegative integrable functions. It should be mentioned that this theorem cannot be formulated for the Riemann integral. We prove it by exhibiting a sequence of simple functions that converges to the same limit as the original one and then applying the introduction rule.

The construction and properties of the sequence are slightly intricate. By definition, for any f_n in the original sequence, there is a sequence $(u_{mn})_{m\in\mathbb{N}}$ of simple functions converging to it. The nth element of the new sequence is the upper closure of the nth elements of the first n sequences.

definition
upclose:: (*'a* ⇒ *real*) ⇒ (*'a* ⇒ *real*) ⇒ (*'a* ⇒ *real*) **where**
upclose f g = (λ*t*. *max* (*f t*) (*g t*))

primrec
mon-upclose-help :: *nat* ⇒ (*nat* ⇒ *nat* ⇒ *'a* ⇒ *real*) ⇒ *nat* ⇒ (*'a* ⇒ *real*) (‹*muh*›) **where**
muh 0 u m = *u m 0*
| *muh* (*Suc n*) *u m* = *upclose* (*u m* (*Suc n*)) (*muh n u m*)

definition
mon-upclose :: (*nat* ⇒ *nat* ⇒ *'a* ⇒ *real*) ⇒ *nat* ⇒ (*'a* ⇒ *real*) (‹*mu*›) **where**
mu u m = *muh m u m*

lemma *sf-norm-help*:
assumes *fin*: *finite K* **and** *jK*: *j* ∈ *K* **and** *tj*: *t* ∈ *C j* **and** *iK*: ∀ *i*∈*K*−{*j*}. *t* ∉ *C i*
shows (∑ *i*∈*K*. (*z i*) ∗ χ (*C i*) *t*) = *z j*⟨*proof*⟩
lemma *upclose-sfis*:
assumes *ms*: *measure-space M* **and** *f*: *a* ∈ *sfis f M* **and** *g*: *b* ∈ *sfis g M*

shows $\exists c.\ c \in \mathit{sfis}\ (\mathit{upclose}\ f\ g)\ M\ \langle \mathit{proof} \rangle$
lemma *mu-sfis*:
assumes u: $\bigwedge m\ n.\ \exists a.\ a \in \mathit{sfis}\ (u\ m\ n)\ M$ **and** ms: $\mathit{measure\text{-}space}\ M$
shows $\exists c.\ \forall m.\ c\ m \in \mathit{sfis}\ (\mathit{mu}\ u\ m)\ M\langle \mathit{proof} \rangle$

lemma *mu-help*:
assumes uf: $\bigwedge n.\ (\lambda m.\ u\ m\ n)\uparrow(f\ n)$ **and** fh: $f\uparrow h$
shows $(\mathit{mu}\ u)\uparrow h$ **and** $\bigwedge n.\ \mathit{mu}\ u\ n \leq f\ n$
$\langle \mathit{proof} \rangle$

theorem *nnfis-mon-conv*:
assumes fh: $f\uparrow h$ **and** xf: $\bigwedge n.\ x\ n \in \mathit{nnfis}\ (f\ n)\ M$ **and** xy: $x\uparrow y$
and ms: $\mathit{measure\text{-}space}\ M$
shows $y \in \mathit{nnfis}\ h\ M$
$\langle \mathit{proof} \rangle$

Establishing that only nonnegative functions may arise this way is a triviality.

lemma *nnfis-nn*: **assumes** $a \in \mathit{nnfis}\ f\ M$
shows $\mathit{nonnegative}\ f\ \langle \mathit{proof} \rangle$

3.2.3 Integrable Functions

Before we take the final step of defining integrability and the integral operator, we should first clarify what kind of functions we are able to integrate up to now. It is easy to see that all nonnegative integrable functions are random variables.

lemma assumes $\mathit{measure\text{-}space}\ M$ **and** $a \in \mathit{nnfis}\ f\ M$
shows *nnfis-rv*: $f \in \mathit{rv}\ M\ \langle \mathit{proof} \rangle$

The converse does not hold of course, since there are measurable functions whose integral is infinite. Regardless, it is possible to approximate any measurable function using simple step-functions. This means that all nonnegative random variables are quasi integrable, as the property is sometimes called, and brings forth the fundamental insight that a nonnegative function is integrable if and only if it is measurable and the integrals of the simple functions that approximate it converge monotonically. Technically, the proof is rather complex, involving many properties of real numbers.

lemma assumes $\mathit{measure\text{-}space}\ M$ **and** : $f \in \mathit{rv}\ M$ **and** $\mathit{nonnegative}\ f$
shows *rv-mon-conv-sfis*: $\exists u\ x.\ u\uparrow f \wedge (\forall n.\ x\ n \in \mathit{sfis}\ (u\ n)\ M)\langle \mathit{proof} \rangle$

The following dominated convergence theorem is an easy corollary. It can be effectively applied to show integrability.

corollary assumes ms: $\mathit{measure\text{-}space}\ M$ **and** f: $f \in \mathit{rv}\ M$
and b: $b \in \mathit{nnfis}\ g\ M$ **and** fg: $f{\leq}g$ **and** nn: $\mathit{nonnegative}\ f$
shows *nnfis-dom-conv*: $\exists a.\ a \in \mathit{nnfis}\ f\ M \wedge a \leq b\ \langle \mathit{proof} \rangle$

Speaking all the time about integrability, it is time to define it at last.

definition
integrable:: $('a \Rightarrow real) \Rightarrow ('a\ set\ set * ('a\ set \Rightarrow real)) \Rightarrow bool$ **where**

integrable f M $\longleftrightarrow$ *measure-space M* $\wedge$
$(\exists x.\ x \in nnfis\ (pp\ f)\ M) \wedge (\exists y.\ y \in nnfis\ (np\ f)\ M)$

definition
integral:: $('a \Rightarrow real) \Rightarrow ('a\ set\ set * ('a\ set \Rightarrow real)) \Rightarrow real$ (‹$\int$ - ∂-›) **where**
integrable f M $\Longrightarrow \int\ f\ \partial M = (THE\ i.\ i \in nnfis\ (pp\ f)\ M) -$
$(THE\ j.\ j \in nnfis\ (np\ f)\ M)$

So the final step is done, the integral defined. The theorems we are already used to prove from the earlier stages are still missing. Only now there are always two properties to be shown: integrability and the value of the integral. Isabelle makes it possible two have both goals in a single theorem, so that the user may derive the statement he desires. Two useful lemmata follow. They help lifting nonnegative function integral sets to integrals proper. Notice how the dominated convergence theorem from above is employed in the latter.

lemma *nnfis-integral*:
assumes *nn*: $a \in nnfis\ f\ M$ **and** *ms*: *measure-space M*
shows *integrable f M* **and** $\int\ f\ \partial\ M = a$
⟨*proof*⟩

lemma *nnfis-minus-nnfis-integral*:
assumes *a*: $a \in nnfis\ f\ M$ **and** *b*: $b \in nnfis\ g\ M$
and *ms*: *measure-space M*
shows *integrable* $(\lambda t.\ f\ t - g\ t)\ M$ **and** $\int\ (\lambda t.\ f\ t - g\ t)\ \partial\ M = a - b$
⟨*proof*⟩

Armed with these, the standard integral behavior should not be hard to derive. Mind that integrability always implies a measure space, just like random variables did in 2.2.

theorem assumes *integrable f M*
shows *integrable-rv*: $f \in rv\ M$⟨*proof*⟩
theorem *integral-char*:
assumes *ms*: *measure-space M* **and** *mA*: $A \in measurable\text{-}sets\ M$
shows $\int\ \chi\ A\ \partial\ M = measure\ M\ A$ **and** *integrable* $\chi\ A\ M$⟨*proof*⟩

theorem *integral-add*:
assumes *f*: *integrable f M* **and** *g*: *integrable g M*
shows *integrable* $(\lambda t.\ f\ t + g\ t)\ M$
and $\int\ (\lambda t.\ f\ t + g\ t)\ \partial M = \int\ f\ \partial M + \int\ g\ \partial M$
⟨*proof*⟩

theorem *integral-mono*:

```
  assumes f: integrable f M
  and g: integrable g M  and fg: f≤g
  shows ∫ f ∂M ≤ ∫ g ∂M
⟨proof⟩

theorem integral-times:
  assumes int: integrable f M
  shows integrable (λt. a∗f t) M and ∫ (λt. a∗f t) ∂M = a∗∫ f ∂M⟨proof⟩
```

To try out our definitions in an application, only one more theorem is missing. The famous Markov–Chebyshev inequation is not difficult to arrive at using the basic integral properties.

```
theorem assumes int: integrable f M and a: 0<a and intp: integrable (λx. |f x|
^ n) M
  shows markov-ineq: law M f {a..} ≤ ∫ (λx. |f x| ^ n) ∂M / (a^n)
⟨proof⟩

end
```

Chapter 4

Epilogue

To come to a conclusion, a few words shall subsume the work done and point out opportunities for future research at the same time.

What has been achieved? After opening with some introductory notes, we began translating the language of measure theory into machine checkable text. For the material in section 2.1, this had been done before. Besides laying the foundation for the development, the style of presentation should make it noteworthy.

It is a particularity of the present work that its theories are written in the Isar language, a declarative proof language that aims to be “intelligible”. This is not a novelty, nor is it the author’s merit. Still, giving full formal proofs in a text intended to be read by people is in a way experimental. Clearly, it is bound to put some strain on the reader. Nevertheless, I hope that we have made a little step towards formalizing mathematical knowledge in a way that is equally suitable for computation and understanding. One aim of the research done has been to demonstrate the viability of this approach. Unquestionably, there is plenty room for improvement regarding the quality of presentation. The language itself has, in my opinion, proven to be fit for a wide range of applications, including the classical mathematics we used it for.

Returning to a more content-centered viewpoint, we discussed the measurability of real-valued functions in section 2.2. As explained there, earlier scholarship has resulted in related theories for the MIZAR environment though the development seems to have stopped. Anyway, the mathematics covered should be new to HOL-based systems.

More functions could obviously be demonstrated to be random variables. We shortly commented on an alternative approach in the section just mentioned. It is applicable to continuous functions, proving these measurable all at once. Efforts on topological spaces would be required, but they constitute an interesting field themselves, so it is probably worth the while.

In the third chapter, integration in the Lebesgue style has been looked at in depth. To my knowledge, no similar theory had been developed in a theorem prover up to this point. We managed to systematically establish the integral of increasingly complex functions. Simple or nonnegative functions ought to be treated in sufficient detail by now. Of course, the repository of potential supplementary facts is vast. Convergence theorems, as well as the interrelationship with differentiation or concurrent integral concepts, are but a few examples. They leave ample space for subsequent work.

A shortcoming of the present development lies in the lack of user assistance. Greater care could be taken to ensure automatic application of appropriate simplification rules — or to design such rules in the first place. Likewise, the principal requirement of integrability might hinder easy usage of the integral. Fixing a default value for undefined integrals could possibly make some case distinctions obsolete. Facets like these have not been addressed in their due extent.

Bibliography

[1] Heinz Bauer. *MaSS- und Integrationstheorie.* de Gruyter, 1990.

[2] Patrick Billingsley. *Probability and Measure.* John Wiley, second edition, 1986.

[3] Noboru Endou, Katsumi Wasaki, and Yasunari Shidama. Definitions and basic properties of measurable functions. *Journal of Formalized Mathematics*, 12, 2000. Available on the web as http://mizar.uwb.edu.pl/JFM/Vol12/mesfunc1.html.

[4] Noboru Endou, Katsumi Wasaki, and Yasunari Shidama. The measurability of extended real valued functions. *Journal of Formalized Mathematics*, 12, 2000. Available on the web as http://mizar.uwb.edu.pl/JFM/Vol12/mesfunc2.html.

[5] Jacques D. Fleuriot and Lawrence C. Paulson. Mechanizing nonstandard real analysis. *LMS Journal of Computation and Mathematics*, 3:140–190, 2000. Available on the web as http://www.lms.ac.uk/jcm/3/lms1999-027/.

[6] Joe Hurd. *Formal Verification of Probabilistic Algorithms.* PhD thesis, University of Cambridge, 2002. Available on the web as http://www.cl.cam.ac.uk/~jeh1004/research/papers/thesis.html.

[7] Tobias Nipkow. Order-sorted polymorphism in isabelle. In Gérard Huet and Gordon Plotkin, editors, *Logical Environments*, pages 164–188. Cambridge University Press, 1993. Available on the web as http://www4.informatik.tu-muenchen.de/~nipkow/pubs/lf91.html.

[8] Sebastian Skalberg. Import tool. Available on the web as http://www.mangust.dk/skalberg/isabelle.php.

[9] Markus Wenzel. Using axiomatic type classes in Isabelle, 2002. Unpublished. Available on the web as http://isabelle.in.tum.de/dist/Isabelle2002/doc/axclass.pdf.

[10] David Williams. *Probability with Martingales.* Cambridge University Press, 1991.