Formalization of the Lebesgue Measure in MathComp-Analysis

Reynald Affeldt[1] and Cyril Cohen[2]

[1] National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan
[2] Université Côte d'Azur, Inria, Sophia Antipolis, France

Motivation and Contribution Our goal is to add integration theory to MathComp-Analysis [1, 2], an extension of MathComp dedicated to classical analysis (topology, real and complex analysis, etc.). In this abstract, we report on the construction of the Lebesgue measure, which is a necessary and difficult step in the formalization of Lebesgue integration and its variants. The originality of our approach is the use of the mathematical structure of algebras of sets as a ground for Carathéodory's extension theorem. This is how the construction is often taught in undergraduate classes (here, we follow [5]) and we believe that this approach improves the modularity of the formalization because it favors abstract lemmas. Beside this contribution, our experiment[1] leads us to extend MathComp-Analysis with reusable support for sequences, series, and extended real numbers. It also takes advantage of the Hierarchy-Builder [3] tool and MathComp's `interval.v`.

Construction of the Lebesgue Measure The main mathematical definitions we introduce are σ-algebra and measure. A *σ-algebra* is a set of sets that contains the empty set, and is stable by complement and by countably infinite union. Sets from a σ-algebra are said to be *measurable*. A *measure* μ is a non-negative function defined over a σ-algebra such that the measure of the empty set is 0 and such that $\mu(\cup_i F_i) = \sum_i^\infty \mu(F_i)$ for any sequence F of pairwise-disjoint measurable sets. σ-algebras are also algebras of sets. An *algebra of sets* is simpler than a σ-algebra: it is a set of sets stable by complement and union, formalized below using Hierarchy-Builder:

```
HB.factory Record isAlgebraOfSets T := {
  measurable : set (set T) ;
  measurable0 : measurable set0 ;
  measurableU : ∀ A B, measurable A → measurable B → measurable (A `|` B) ;
  measurableC : ∀ A, measurable A → measurable (~` A) }.
```

We formalize the construction of the Lebesgue measure as the composition of three main results.

The first result builds a σ-algebra and a measure over it given an *outer measure* μ (an outer measure is intuitively a "relaxed" definition of measure). The resulting σ-algebra is formed of *Caratheodory-measurable sets*, i.e., sets A such that $\forall X, \mu(X) = \mu(X \cap A) + \mu(X \cap \bar{A})$. In order to obtain the Lebesgue measure, it suffices to apply this result to the outer measure given by the next result.

The second result builds an outer measure given a measure *defined over an algebra of sets*. More precisely, the outer measure in question of a set X is defined as the infimum of the measures of covers $\inf \{\sum_i^\infty \mu(F_i) | (\forall i, \mathtt{measurable}(F_i)) \wedge X \subseteq \bigcup_i F_i\}$:

```
Definition measurable_cover X := [set F | (∀ i, measurable (F i)) ∧ X ⊆ \bigcup_i (F i)].
Definition mu_ext mu X := ereal_inf [set \sum_(i <oo) mu (F i) | F in measurable_cover X].
```

We show that this outer measure coincides with the input measure `mu`. We also show that the application of the first result above to this outer measure produces a σ-algebra that contains all the measurable sets of the smallest σ-algebra that contains the algebra of sets. To complete the construction of the Lebesgue measure, it now suffices to exhibit the Borel measure on the algebra of sets of finite unions of intervals.

The third result is the construction of this Borel measure. It starts with the formalization of the algebra of sets of *simple sets* (`sset`), i.e., sets that can be covered by a finite union of intervals. In proving that such sets indeed form an algebra of sets, the technical part is the stability by complement, since the complement of an interval is not an interval in general. For this purpose we provide a decomposition procedure `Decompose` that turns a sequence of (possibly overlapping) intervals into a sequence of non-overlapping intervals with the same cover. The length `slength` of a simple set is easily defined: call the decomposition procedure to produce an equivalent sequence of non-overlapping intervals and sum their

[1]The material corresponding to this abstract can be found online (`https://github.com/math-comp/analysis`). The first part of the experiment has already made its way to the files `ereal.v`, `sequences.v`, `csum.v`, `measure.v`, etc.; the rest appears as PR#371.

lengths. We then prove that `slength` is additive (this is not difficult since it does not involve infinite sums) and σ-additive, i.e., `(fun n ⇒ \sum_(i < n) slength (F i)) ⟶ slength (\bigcup_i F i)` (under appropriate measurability and pairwise-disjointness for the sequence `F`). This is the last technical step of the construction of the Lebesgue measure which we get by first proving that the length is actually σ-subadditive on the subclass of simple sets that corresponds to intervals. This is the main lemma and it is no surprise that it is indeed the focus point of many lecture notes we explored for the purpose of this work. It uses as an intermediate step the Borel-Lebesgue theorem that MathComp-Analysis conveniently already had.

Useful extensions to MathComp-Analysis Since a measure is potentially infinite, it is represented by extended real numbers. A prerequisite for the construction of the Lebesgue measure we have explained above was therefore the development of the theory of extended real numbers and of their sequences. This actually called for a substantial extension of the MathComp-Analysis library. The latter provides several mathematical structures (topological, uniform, pseudometric spaces, normed modules, etc.) together with the expected lemmas. It was therefore important to first equip extended real numbers with the structures available in MathComp-Analysis, so as to enjoy already proved lemmas. On the topological side, extended real numbers form a pseudometric space. On the algebraic side, $\infty - \infty$ is undefined in the mathematical practice, but defining it to be $-\infty$ is crucial because it makes the extended real numbers a commutative monoid, which enables the use of the `bigop` library. There is no hope to get a better structure on the full type though. Establishing these facts was the first step of our formalization of sequences of extended real numbers, which goes on by proving the expected lemmas such as the fact that the limit of a sum is the sum of limits (assuming the quantities involved are properly defined), etc.

The first two results of the construction of the Lebesgue measure form Carathéodory's extension theorem. Given our newly developed theory of sequences of extended real numbers, the proofs of the first result are almost a direct translation of pencil-and-paper proofs. However, the proof of the rest of Carathéodory's extension theorem is more technical in that it requires new developments about sequences of extended real numbers, namely the definitions of sums (of non-negative terms) over general sets, i.e., the notation $\sum_{i\in S} a_i$. The latter is defined by $\sup\left\{\sum_{i\in F} a_i \,|\, F \text{ non-empty finite subset of } S\right\}$, and generalizes the notation for the limit of sequences of extended real numbers.

An important aspect of the third result (the construction of a measure over the algebra of sets of simple sets) is the procedure that decomposes a simple sets into non-overlapping intervals. This procedure uses an ordering of intervals which takes advantage of `path.v`, `order.v`, and the recently updated `interval.v` of MathComp. As already emphasized above, lecture notes about the construction of the Lebesgue measure seem to emphasize the lemma showing the σ-additivity for intervals. We found the generalization of σ-additivity from intervals to simple sets technically interesting. Indeed, it exhibits for example technical reindexings due to the combined used of infinite sums and infinite unions (that come from MathComp-Analysis) and finite sums and finite unions from MathComp's `bigop` library.

Related Work Isabelle/HOL [4] and Lean (https://github.com/leanprover-community/mathlib) have different constructions of the Lebesgue measure respectively based on the gauge integral and on the Lebesgue outer measure. Neither of which formally applies the Carathéodory's extension theorem to an algebra of sets.

References

[1] R. Affeldt, C. Cohen, M. Kerjean, A. Mahboubi, D. Rouhling, and K. Sakaguchi. Competing inheritance paths in dependent type theory: a case study in functional analysis. In *IJCAR 2020*, volume 12167(2) of *LNAI*, pages 3–20. Springer, Jul 2020.

[2] R. Affeldt, C. Cohen, and D. Rouhling. Formalization techniques for asymptotic reasoning in classical analysis. *Journal of Formalized Reasoning*, 11(1):43–76, 2018.

[3] C. Cohen, K. Sakaguchi, and E. Tassi. Hierarchy Builder: Algebraic hierarchies made easy in Coq with Elpi (system description). In *FSCD 2020*, volume 167 of *LIPIcs*, pages 34:1–34:21, 2020.

[4] J. Hölzl and A. Heller. Three chapters of measure theory in Isabelle/HOL. In *ITP 2011*, volume 6898 of *LNCS*, pages 135–151. Springer, 2011.

[5] D. Li. Construction de la mesure de Lebesgue. Université d'Artois, Faculté des Sciences Jean Perrin, 2008.