Technische Universität München
Institut für Informatik

Construction and Stochastic Applications of Measure Spaces in Higher-Order Logic

Johannes Hölzl

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Thomas Huckle

Prüfer der Dissertation:

1. Univ.-Prof. Tobias Nipkow, Ph.D.
2. Univ.-Prof. Dr. Dr. h.c. Javier Esparza

Die Dissertation wurde am 10.10.2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 31.01.2013 angenommen.

Abstract

A rich formalization of measure and probability theory is a prerequisite to analyzing probabilistic program behaviour in interactive theorem provers.

When using probability theory it is important to have the necessary tools (definitions and theorems) to construct measure spaces with the desired properties. This thesis presents the formalization of a rich set of constructions. We start with discrete measures, distributions, densities and products. Then, we introduce the Lebesgue measure and products of probability spaces with an infinite index. The latter is used to construct the stochastic processes of discrete-time Markov chains on discrete state spaces.

For applications like randomized algorithms discrete probability spaces are enough. Here single elements have nonzero measures assigned. More advanced constructions are needed when we measure infinite traces, since sets of traces are not discrete. Important models for such trace spaces are discrete-time Markov chains. Here a trace is a sequence of states and the probability which state to choose next only depends on the previous state. We construct the trace measure of a Markov chain, based on the transition probabilities between states. With the formalization of Markov chains we verify probabilistic model checking, anonymity in the Crowds protocol, and the probability to allocate a free address in the ZeroConf protocol.

This development is done in the interactive theorem prover Isabelle/HOL.

Acknowledgment

I am deeply grateful to Tobias Nipkow for giving me the opportunity to work in his group. Already in my undergraduate studies, his "Perlen der Informatik" lectures introduced me to logic and interactive theorem proving. Later, he guided me toward this fruitful and interesting intersection between interactive theorem proving, mathematical analysis, and software verification. This happened with the help of Andrei Popescu, who suggested to formalize Markov chains, and of Marta Kwiatkowska, whose Marktoberdorf 2011 lecture inspired the formalization of pCTL model checking. I am also grateful to Javier Esparza for agreeing to act as a referee and examiner.

I want to thank all my (ex-)colleagues for the nice atmosphere in the Isabelle group: Stefan Berghofer, Jasmin Blanchette, Sascha Böhme, Lukas Bulwahn, Amine Chaieb, Brian Huffman, Dongchen Jiang, Cezary Kaliszyk, Alexander Krauss, Ondřej Kunčar, Peter Lammich, Lars Noschinski, Steven Obua, Andrei Popescu, Dmitriy Traytel, Thomas Türk, Christian Urban, and Makarius Wenzel. Thanks are also due to Armin Heller, Robert Himmelmann, and Fabian Immler, who helped to develop Isabelle's multivariate analysis and measure theory. Andrei, Ondřej, and Fabian deserve special thanks for reading drafts of this thesis and for giving me valuable suggestions.

This research was financially supported by the DFG RTG 1480 (PUMA).

Contents

Chapter 1

Introduction

This thesis describes the formalization of measure, probability and information theory in the interactive theorem prover Isabelle/HOL.

1.1 Motivation

Probability theory is an important tool used in computer science. Probabilities come either from outside of a computer system, such as physical characteristics, probabilistic behaviour of external components or human behaviour. Or they are internal to the computer system, such as the use of a random number generator for probabilistic choice for symmetry breaking or by employing randomized algorithms. Stochastic processes allow us to model such s system together with its timing behaviour. Discrete-time or continuous-time Markov chains are instances of stochastic processes often used to model the behaviour of probabilistic computer systems.

The modelling of probabilistic systems requires a toolbox of probability measures: beginning with discrete measures, products of infinitely many independent random variables, the Lebesgue measure (important to construct uniform, exponential, normal, etc. distributed random variables), and finally Markov chains. The probabilistic analysis requires the following concepts: the Lebesgue integral to handle expectation, independence of random variables, and mutual information and entropy to quantify the information stored in random variables. Since Kolmogorov's seminal work on probability theory [46, 47], we know that probability theory can be formally based on measure theory.

The goal of this thesis is to formalize these concepts and the necessary measure theory in the interactive theorem prover Isabelle/HOL. Here with formalization we mean to define a concept in terms of HOL, to prove its basic properties, to prove its relations to other concepts, and to apply it in concrete applications. The last two steps are very important: the relations to other concepts show the validity of the definition and the concrete applications show the value of the formalization.

An alternative method to verify probabilistic properties of computer systems is probabilistic model checking, as implemented by PRISM [51] or MRMC [45]. They interpret Markov chains and analyze quantitative properties, specified as probabilistic CTL (pCTL) [30] or CSL [7] formulas. While model checking works

	$\overline{\mathbb{R}}$	$\mathcal{B}_\mathcal{T}$	$\Omega \neq \mathcal{U}_\alpha$	Integral	$\lambda_{\mathbb{R}^n}$	Product	Dynkin
Hurd					$\{0 \mathinner{..} 1\}$		
Richter				✓			
Coble			✓	✓			
Mhamdi ITP '10		✓	✓	✓			
Mhamdi ITP '11	✓	✓	✓	✓			
Lester	✓	✓	✓	✓			
PVS	✓	✓	✓	✓	✓	✓	
Mizar	✓		✓	✓	✓		✓
HOL Light				✓	✓	$\mathbb{R}^{n+m}$	
Isabelle	✓	✓	✓	✓	✓	✓	✓

Table 1.1: Overview of the current formalizations of measure theory.

automatic, it is restricted to fixed finite models. Newer work in this area tries to mitigate these problems by applying a CEGAR-like approach [34], finding invariants [44], or introducing parametric Markov chains [29]. While these techniques move the boundary of what is possible in probabilistic model checking, there is still the point where automation fails. With the development in this thesis we hope to provide the basis for a method powerful enough to verify probabilistic models not fitted for model checking.

1.2 Related Work

Table 1.1 gives an overview of the current formalizations of measure theory we are aware of. The rows list first the work of Hurd [39], Richter [68], Coble [17], Mhamdi *et al.* [57, 58], and Lester [52]. The second part of the rows list the current state of theorem provers or libraries formalizing measure theory: beginning with the PVS-NASA library,[1] the Mizar Mathematical Library (MML), the multivariate analysis found in HOL Light and finally the work presented in this thesis. Mhamdi *et al.* [58] represents the current state of HOL4, hence HOL4 is not listed. The columns correspond to different measure theoretic concepts and features: using extended reals $\overline{\mathbb{R}}$ for measure values, using topological spaces to define the Borel sets $\mathcal{B}_\mathcal{T}$, measure spaces Ω are independent of the type universe $\mathcal{U}_\alpha$, the Lebesgue integral, the Lebesgue measure $\lambda_{\mathbb{R}^n}$, product measures, and formalization of Dynkin systems.

1.2.1 Measure Theory in the Mizar Mathematical Library

Unlike all other theorem provers mentioned in this section Mizar is based on set theory. The Mizar Mathematical Library (MML) is a rich mathematical library with the intention to formalize mathematics. There is already a big collection of formalized measure and probability theorems. Here a small excerpt of the available theories:

[1] http://shemesh.larc.nasa.gov/fm/ftp/larc/PVS-library/pvslib.html

- Nędzusiak [60, 61] has been the first one who formalized measures in an ITP. He defines probability measures on σ-algebras and the Borel sets generated by right-bounded intervals in $\mathbb{R}$.
- Białas extends these measures to the extended real numbers, introduces completion of measures [10], proves Caratheodory's extension theorem [11], and constructs the Lebesgue measure on real numbers [12].
- Endou *et al.* [24] introduce measurable functions and the Lebesgue integral. They close with proving its monotone convergence.
- Merkl [56] formalizes Dynkin systems and Dynkin's lemma.
- Doll [22] formalizes the independence of a family of events and uses it for Kolmogorov's 0-1-law.

All of these formalizations were published in *Formalized Mathematics (FM)*, a journal publishing annotated Mizar theories.

With these formalizations the MML contains already the foundations of measure and probability theory. Unfortunately, while the MML contains formalizations of mathematics, it does not contains applications onto problems in computer science. For example there is no formalization of Markov chains or trace spaces which could be used to formalize algorithms.

1.2.2 Probability Space on $\mathbb{N} \to \mathbb{B}$ by Hurd (hol98)

The formalization of probability theory in HOL starts with Hurd's thesis [39]. He introduces σ-algebras and measures, proves Caratheodory's extension theorem and uses it to introduce a probability space on infinite boolean sequences $\mathbb{N} \to \mathbb{B}$, isomorphic to the Lebesgue measure on the unit interval $\{0 \mathinner{..} 1\}$. The space of σ-algebras is type bound $\Omega = \mathcal{U}_\alpha$.

He models the execution with a random number generator as a monad operating on traces in the probability space $\mathbb{N} \to \mathbb{B}$. An atomic operation returns the first element, and passes the rest of the stream to further operations. This monad also provides a while-combinator with a rule for almost sure termination. Based on this he provides methods to generate discrete random variables with uniform, geometric, or Bernoulli distribution. Thus it allows him to model countably many independent random variables with these discrete distributions.

1.2.3 Lebesgue Integration by Richter (Isabelle/HOL)

Richter [68] formalizes the Lebesgue integral in Isabelle/HOL and uses it together with Hurd's probability space. Richter introduces the Borel sets, but only on right-bounded intervals in $\mathbb{R}$. He does not formalize extended real numbers, hence his positive integral is only defined if the result is a real number. Also the space of σ-algebras is fixed to the type universe $\Omega = \mathcal{U}_\alpha$. He does not formalize Caratheodory's theorem in Isabelle/HOL, instead he imports Hurd's probability space by using Obua's and Skalberg's HOL-import tool [64, 72].

1.2.4 Liveness Reasoning by Wang *et al.* (Isabelle/HOL)

Wang *et al.* [73] introduce measure spaces defined on arbitrary spaces $\Omega \neq \mathcal{U}_\alpha$. They prove Caratheodory's extension theorem and use it to construct a probability measure for execution traces of concurrent systems. They introduce parametric fairness on these concurrent systems and prove that the probability of all parametric fair traces is 1.

The formalization in this thesis was developed independent of Wang *et al.* [73]. First, while developing the theories presented in this thesis the author was not aware of their work. And second, the work described by Wang *et al.* [73] is unfortunately not publicly available.

1.2.5 Information Theory by Coble (HOL4)

Coble [17] ports Richter's formalization of the Lebesgue integral to HOL4 and generalizes the definition of σ-algebras, which are now defined on arbitrary spaces $\Omega \neq \mathcal{U}_\alpha$. For his goal of formalizing quantitative information flow analysis he requires product spaces and the Radon-Nikodým derivative to define mutual information. While his version of the Lebesgue integral works on continuous measure spaces, his other theorems about the binary product measure, the Radon-Nikodým theorem, and mutual information only work for discrete finite distributions.

1.2.6 Entropy Measures by Mhamdi *et al.* (HOL4)

Mhamdi *et al.* [57] extend Coble's [17] work. They define Borel sets as the σ-algebra generated by open sets comparable to the definition in this thesis. However, they do not formalize measure values as extended real numbers but only as plain reals. They define a more restricted version of the almost everywhere predicate, and do not give rules for the interaction with logical connectives. They prove Markov's inequality and the weak law of large numbers.

Later Mhamdi *et al.* [58] introduce extended real numbers, and use them for measures and the Lebesgue integral. Based on this they formalize the Radon-Nikodým theorem and relative entropy.

The work by Mhamdi *et al.* [57, 58] was done in parallel to the work presented in this thesis.

1.2.7 Measure and Probability Theory by Lester *et al.* (PVS)

There is also the PVS formalization of topology by Lester [52]. He gives a short overview of the measure theory based on his formalization of topology. This includes measures using extended real numbers, a definition of almost everywhere, Borel sets on topological spaces, and the Lebesgue integral. Daumas and Lester [20] and Daumas *et al.* [21] use this development to bound the probability that the rounding error of big sums exceeds a limit. For this Daumas and Lester [20] introduce martingales and apply Doob-Kolmogorov's inequality. Later Daumas *et al.* [21] prove Markov's and Levy's inequalities for this application. They need binary product spaces, finite families of independent random variables,

and martingales. In recent development the PVS-NASA library also contains the proof that the Lebesgue integral extends the Riemann integral. In PVS, abstract reasoning is performed using parametrized theories, similar to our usage of locales.

1.2.8 Multivariate Analysis by Harrison (HOL Light)

The gauge integral (an extension of the Lebesgue integral) in HOL Light is based on Harrison's work on Euclidean spaces [31]. It is used to define a subset of the Lebesgue measure, missing infinite measure values. Euclidean spaces are defined as functions $\alpha \to \mathbb{R}$, where α is a type with a finite type universe. The product of two Euclidean spaces $\alpha \to \mathbb{R}$ and $\beta \to \mathbb{R}$ is $\alpha \times \beta \to \mathbb{R}$. His theories are now ported to Isabelle/HOL and we use them to introduce the Lebesgue measure and to show that Lebesgue integrability implies gauge integrability and that in this case both integrals are equal.

1.2.9 Analysis of Random Variables by Hasan *et al.* (HOL4)

Hasan [32] formalizes continuous random variables on Hurd's probability space of boolean sequences $\mathbb{N} \to \mathbb{B}$. First, he constructs a uniformly distributed random variable $(\mathbb{N} \to \mathbb{B}) \to \{0 \mathinner{..} 1\}$. Then, by using the inverse transform method, he constructs random variables with an exponential, uniform, Rayleigh, and triangular distribution. Finally, the cumulative distribution function is verified for each random variable.

Hasan *et al.* [33] verify the expectations of these random variables. They use Coble's Lebesgue integral [17] on Hurd's probability space. Abbasi [1] builds on this and verifies the second moment and the variance of exponentially, uniformly, and triangularly distributed random variables.

The random variables are directly constructed on the probability space, and the distributions are not axiomatically introduced. As no product space is available in HOL4, this restricts the analysis to only one continuous random variable. While Abbasi [1] introduces independence of finite lists of random variables, he has no product spaces to *construct* a probability space with independent random variables.

1.2.10 Markov Chain Analysis by Liu *et al.* (HOL4)

Based on Hurd's and Hasan's work Liu *et al.* [54] formalize the concept of Markov chains. They do not construct a probability measure for the traces, and their Markov chain property

$$\Pr(X_{n+1} = s | X_n = t_n, \ldots, X_1 = t_1) = \Pr(X_{n+1} = s | X_n = t_n)$$

does not assume that $\Pr(X_n = t_n, \ldots, X_1 = t_1)$ is nonzero. For this, they only support Markov chains where each state is always reachable.

1.2.11 Formalizations of Discrete Probability Spaces

Continuous measure spaces are not always necessary in program verification or information theory. Usually, the cardinality of the result values is countable and in many cases even finite. This simplifies the necessary formalizations as measure theory is not necessary: measures are easily constructed, they are just discrete sums, each function with a finite range is measurable and hence also integrable.

- Hurd *et al.* [40] formalize weakest precondition semantics of the probabilistic guarded command language (pGCL) in HOL4. They introduce positive extended reals to formalize expectations on a discrete probability space. The programming language is deeply embedded allowing them to give an axiomatization of weakest precondition.
- Audebaud and Paulin-Mohring [5] formalize randomized algorithms in Coq. The programs are shallow embedded into Coq by representing them as expectations. For example, coin flip is represented as: $\mathit{flip}\ p = (\lambda f.\ p \cdot f\ \mathit{True} + (1-p) \cdot f\ \mathit{False})$. The probability that a program p returns a value v is simply expressed as $p\ (\chi\ \{v\})$. The formalization does not restrict the integrand to be measurable, so it only works for discrete measure spaces.
- Coble [17] formalizes the Lebesgue integral on arbitrary measure spaces, hence also on continuous measures. However, his formalization of product spaces, the Radon-Nikodým derivative, entropy, and mutual information is limited to discrete finite measure spaces. For definitions, he uses general measure theory, but for theorems he assumes a discrete finite measure space: $\mathit{finite}\ \Omega$ and all sets are measurable $\mathcal{A} = \mathcal{P}(\Omega)$.
- Affeld and Hagiwara [2] introduce discrete finite probability spaces in Coq to formalize Shannon's source and channel coding theorems. They formalize entropy only for discrete finite distributions, with this they avoid the formalization of Lebesgue integration.

1.3 Contributions

Our work started as an Isabelle/HOL port of the formalizations done by Hurd [39], Richter [68], and Coble [17]. Later, we reworked most of it to introduce a type for measures, use the extended reals as measure and integral values, and define the Borel sets to be generated by the open sets. Richter and Coble define the Lebesgue integral as the limit of simple integrals of simple functions converging to f. Also, their integral needs to be finite. Our definition of the Lebesgue integral is the one found in Schilling's textbook [69]. It defines the integral of f as the supremum of all simple integrals of simple functions bounded by f. With this definition the integrand is not required to be Borel-measurable, e.g. for monotony measurability is not required.

The first main contribution is the formalization of a generic version of measure, probability and information theory. As we saw in the previous section, most formalizations stick to one fixed measure space. They are restricted to sequences

of boolean values, the Lebesgue measure, or to discrete finite measure spaces. In contrast we construct multiple different measure spaces and use a generalized concept of random variables and their distributions. The range of random variables is not restricted to be discrete, finite, or real valued. We only assume that they map into a σ-algebra.

Compared to Coble [17] and Affeld and Hagiwara [2] we support concepts like entropy and mutual information also on continuous random variables. While Mhamdi *et al.* [58] introduce Radon-Nikodým on generic measure spaces, their information theoretic concepts are still limited to discrete random variables.

The second main contribution is the formalization of important probability spaces for stochastic processes. Often the analysis of probabilistic properties assumes probability spaces with random variables of a fixed distribution and with additional assumptions like independence, memorylessness, etc. As the probability spaces get more and more complicated, it gets more likely to introduce inconsistencies by just assuming their existence. The explicit construction of such probability spaces allows us to get rid of these assumptions. To support such constructions we provide the product of infinitely many independent random variables and the trace space for Markov chains. For Markov chains we also include important properties like state fairness and finite hitting time. These formalizations allow us to model and verify applications in computer science, like pCTL model checking or randomized network protocols.

1.4 Publications

Most parts of this thesis are based on the following publications. They are incorporated with the permissions of the coauthors.

1. Johannes Hölzl and Armin Heller. Three Chapters of Measure Theory in Isabelle/HOL. In M. C. J. D. van Eekelen, H. Geuvers, J. Schmaltz, and F. Wiedijk, editors, *Interactive Theorem Proving (ITP 2011)*, volume 6898 of LNCS, pages 135–151, 2011.

2. Johannes Hölzl and Tobias Nipkow. Verifying pCTL Model Checking. In C. Flanagan and B. König, editors, *Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2012)*, volume 7214 of LNCS, pages 347–361, 2012.

3. Johannes Hölzl and Tobias Nipkow. Interactive Verification of Markov Chains: Two Distributed Protocol Case Studies. In U. Fahrenberg, A. Legay, and C. Thrane: *Quantities in Formal Methods (QFM 2012)*, EPTCS, 2012.

The formalizations described in this thesis can be found in the Isabelle repository,[2] the formalization of Markov chains and its applications in the AFP entry `Markov_Models`[3] [36].

The following publications were written as part of the Ph.D. but do not fit thematically in this thesis.

[2] `http://isabelle.in.tum.de/repos/isabelle`
[3] `http://afp.sf.net/entries/Markov_Models.shtml`

4. Andrei Popescu, Johannes Hölzl and Tobias Nipkow. Proving Concurrent Noninterference. Accepted for *Certified Programs and Proofs (CPP 2012)*.

5. Fabian Immler and Johannes Hölzl Numerical Analysis of Ordinary Differential Equations in Isabelle/HOL. In L. Beringer and A. Felty, editors, *Interactive Theorem Proving (ITP 2012)*, volume 7406 of LNCS, pages 377–392, 2012.

6. Gilad Arnold, Johannes Hölzl, Ali Sinan Köksal, Rastislav Bodík, and Mooly Sagiv. Specifying and Verifying Sparse Matrix Codes. In *ACM SIGPLAN International Conference on Functional Progamming (ICFP 2010)*, pages 249–260, 2010.

7. Johannes Hölzl Proving Inequalities over Reals with Computation in Isabelle/HOL, In G. Dos Reis and L. Théry, editors, *ACM SIGSAM International Workshop on Programming Languages for Mechanized Mathematics Systems (PLMMS'09)*, pages 38–45, 2009.

1.5 Preliminaries

The formalizations presented in this thesis are done in the interactive theorem prover Isabelle/HOL. In this section we give an overview of our syntactic conventions.

The term syntax follows the λ-calculus, i.e. function application is juxtaposition as in $f\ t$. The notation $t :: \tau$ means that t has type τ. Types are built from the base types $\mathbb{B}$ (booleans), $\mathbb{N}$ (natural numbers), $\mathbb{R}$ (reals), type variables (α, β, etc), via the function type constructor $\alpha \to \beta$, via the set type constructor $\alpha\ \mathit{set}$, and via Cartesian products $\alpha \times \beta$. We also use extended real numbers $\overline{\mathbb{R}}$, see Appendix A.

When defining a new constant we first declare the type of the constant. If the constant has a special syntax, i.e. if it uses sub- or super-script for arguments, or if it has a mixfix syntax, then we use $\square$ to mark the positions in the constant for these special arguments.

Similar to functional programming, Isabelle/HOL provides an **if**-expression and a **let**-expression. The **if**-expression has the following mixfix syntax:

$$\textsf{if}\ \square\ \textsf{then}\ \square\ \textsf{else}\ \square \quad :: \quad \mathbb{B} \to \alpha \to \alpha \to \alpha$$

The expression $\textsf{let}\ x_1 = t_1; \cdots; x_n = t_n\ \textsf{in}\ t\ x_1\ \cdots\ x_n$ is translated into $t\ t_1\ \cdots\ t_n$. For further syntax conventions see Fig. 1.1.

Isabelle/HOL supports *type classes* allowing us to define constants and to state theorems about type variables with additional constraints. We use $\alpha :: \mathcal{T}$ to annotate that the type variable α is in the type class $\mathcal{T}$. Isabelle/HOL provides type classes for linear orders, complete lattices, monoids, groups, fields, etc. The type classes explicitly appearing in this thesis are topological spaces $\mathcal{T}$, types with a countable universe $\mathcal{C}$ and Euclidean spaces $\mathcal{E}$. For Euclidean spaces we also use a different annotation: we write $\mathbb{R}^n$, like a regular type. For example we write $t :: \mathbb{R}^n$ instead of writing $t :: \alpha :: \mathcal{E}$ and assuming that the dimensionality of the Euclidean space α is n.

For our development we heavily build on *locales* [28], a mechanism in Isabelle/HOL to introduce concepts combining variables with assumptions about these variables. For example we will introduce algebras as a space Ω and a set of sets $\mathcal{A}$ and the assumption that $\mathcal{A}$ is closed under the Ω-complement, union and intersection. The locale is then a context of constants (e.g. Ω and $\mathcal{A}$) and axioms (e.g. closure properties).

The **locale**-command introduces a new locale

$$\textbf{locale } \mathit{loc} = \mathit{par} + \textbf{fixes } x :: \alpha \textbf{ assumes } P_1\ x \textbf{ and } \ldots \textbf{ and } P_n\ x$$

This introduces the locale *loc* with a variable x and the assumptions $P_1\ x, \ldots, P_n\ x$. It inherits the context such as variables and assumptions, but also theorems, abbreviations, definitions, setup for the proof methods and more from its parent locale *par*. This command also introduces a predicate $\mathit{loc}\ x = P_1\ x \wedge \ldots P_n\ x$. We get the theorems for a specific instantiation $x = C$ by proving $\mathit{loc}\ C$. When we prove a theorem in the locale *loc*, we also have access to the theorems of *par*, i.e. a lemma in *algebra* is immediately available in the *σ-algebra* locale.

Of course we can use the locale predicate *loc* just as a regular predicate. In this thesis, when we work inside a locale *loc* with the variables x we write in italics *"In this section we assume x is a loc"*. This is the same as adding $\mathit{loc}\ x$ to all following theorems and x as parameter to all definitions.

$\mathit{SOME}\ x.\ P\ x$	Hilbert choice, chooses an arbitrary element x for which $P\ x$ holds: $(\exists x.\ P\ x) \longrightarrow P\ (\mathit{SOME}\ x.\ P\ x)$
$\mathit{LEAST}\ x.\ P\ x$	The least element x fulfilling $P\ x$: $P\ i \wedge \left(\forall k < i.\ \neg P\ k\right) \implies \left(\mathit{LEAST}\ x.\ P\ x\right) = i$
$\sup_{i\in I} f\ i$, $\inf_{i\in I} f\ i$	The supremum and infimum of $\{f\ i \mid i \in I\}$. We use it not only for complete lattices but also as minimum and maximum when I is finite.
x_i	The i-th component of the vector x.
$a < b$, $a \leq b$	The usual order relations, this is also defined for functions: $f \leq g \Leftrightarrow (\forall x.\ f\ x \leq g\ x)$ and for vectors $x \leq y \Leftrightarrow (\forall i.\ x_i \leq y_i)$
$\{a <..< b\}$	The open and closed interval ranging from a to b. For infinite intervals we just drop the infinite side before or after the dots, e.g. $\{a <..\} = \{x \mid a < x\}$. When a side is closed the <-symbol is omitted, e.g. $\{a\ ..\ b\} = \{x \mid a \leq x \wedge x \leq b\}$.
$\mathcal{P}(A)$	The power set $\mathcal{P}(A) = \{B \mid B \subseteq A\}$.
$\mathcal{U}_\alpha$	The universe for type α: $\mathcal{U}_\alpha :: \alpha\ \mathit{set} = \{x \mid \mathit{True}\}$
$f[A]$	The image of A under f: $f[A] = \{f\ x \mid x \in A\}$.
$f^{-1}[B]$	The inverse image of B under f: $f^{-1}[B] = \{x \mid f\ x \in B\}$
$A \times B$	The Cartesian set product: $A \times B = \{(a,b) \mid a \in A, b \in B\}$
$\mathit{undefined} :: \alpha$	An arbitrary element of type α.
$\bigtimes_{i\in I} A\ i$	The dependent function space (which is a set, not a type): $\bigtimes_{i\in I} A\ i = \{\omega \mid (\forall i \in I.\ \omega\ i \in A\ i) \wedge (\forall i \notin I.\ \omega\ i = \mathit{undefined})\}$ We require each function to have the *undefined* value outside of I, otherwise there is more than one function with the same values on I.
$I \to A$	The dependent function space when A is constant. This should not be confused with the function type annotation $\alpha \to \beta$.
$\omega{\restriction_I}$	The restriction of x to the domain J: $\omega{\restriction_I}\ \ i = x\ i$ if $i \in I$ otherwise $\omega{\restriction_I}\ \ I = \mathit{undefined}$.
$x{\cdot}\omega$	Prepending an element x to a sequence $\omega :: \mathbb{N} \to \alpha$ is written as $x{\cdot}\omega$, i.e. $(x{\cdot}\omega)\ 0 = x$ and $(x{\cdot}\omega)\ (n+1) = \omega\ n$.
$\chi\ A\ x$	The indicator function: $\chi\ A\ x = 1$ if $x \in A$ otherwise $\chi\ A\ x = 0$.
$\mathit{incseq}\ A$	The sequence A is increasing: $A\ 0 \leq A\ 1 \leq A\ 2 \leq \cdots$
$\mathit{decseq}\ A$	The sequence A is decreasing: $A\ 0 \geq A\ 1 \geq A\ 2 \geq \cdots$

Figure 1.1: Syntax conventions used in this thesis.

Chapter 2

Measure and Integration

Before we construct measures in the next chapter, we want to formalize them from an abstract point of view. This chapter starts with measurability, characterizing valid sets and functions we want to use on measure spaces. Then, we formalize uniqueness and Caratheodory's extension theorem to show that a unique measure exists when the measure values for easily characterizable sets are fixed. Finally, the Lebesgue integral is formalized.

The formalizations described in this chapter have quite some history in HOL theorem provers. It starts with Hurd's formalization [39] in `hol98`, which already covers an abstract description of measure spaces and Caratheodory's extension theorem. Parts of the measure space formalization was then ported to Isabelle/HOL and used by Richter [68] to formalize the Lebesgue integral. The Lebesgue integration was than ported back and generalized to HOL4 by Coble [17]. Finally, this theory was ported again to Isabelle/HOL, this time containing Caratheodory's extension theorem, started by L. C. Paulson and later continued by Hölzl and Heller.

These ported theories were then extended and generalized in multiple ways:

- The introduction of extended real numbers $\overline{\mathbb{R}}$ allows a generalization of measure values. This enables us to define measures with infinite measure values, like the Lebesgue measure. Lebesgue integration is defined for arbitrary nonnegative functions, also for functions with infinite integrals.

- We introduce Dynkin's lemma used to show that equality on measures can be reduced to equality on the generating sets. Dynkin's lemma is later also used to construct product measures and for the σ-closure of independent sets.

- Rings and semirings are introduced. Caratheodory's extension theorem is generalized to semirings. We formalize the proofs about premeasures on semirings found in Elstrodt [23].

- Isabelle/HOL specific concepts like type classes and locales are employed all throughout the development. Type classes allow us to define Borel sets on topological spaces. Locales help to describe the hierarchies of set systems and measure spaces.

- The concept of a measure space is not introduced as predicate, but cast into its own type. This improves the support for automation as there are no assumptions necessary when a σ-algebra or a measure space is needed.

The first three generalizations and extensions are based on the measure theory books from Bauer [9] and from Schilling [69].

The work described in this chapter is based on joint work with Armin Heller [35].

2.1 Measure Type and σ-Algebras

A central element in measure theory is the concept of measurable sets. A measure is not defined on all subsets of the space. The sets for which it is defined are the measurable sets. One obviously wants that as many sets as possible are measurable and that as many operations as possible result in measurable sets. Unfortunately, the downside is that the more sets are measurable the more difficult it gets to construct a measure with the desired properties.

A famous result in measure theory is Vitali's theorem[1], stating that for each translation invariant measure with $\mu\ \{a\ ..\ b\} = b - a$, there exists a non-measurable set. As we want a translation invariant Lebesgue measure λ we know due to Vitali's theorem that it cannot be defined for all subsets of $\mathbb{R}$. This does not only apply to the Lebesgue measure, but we can use it to construct similar counterexamples for infinite products and the trace measure of Markov chains. So we can even strengthen the last sentence in the previous paragraph: sometimes it is plainly impossible to construct a measure with the wanted properties where all sets are measurable.

The problem is solved in measure theory by introducing σ-algebras, which are families of sets closed under complement and countable intersection and union. For example, we can define the Lebesgue measure on Borel sets, the smallest σ-algebra containing all intervals $\{a\ ..\ b\}$. To now accommodate both sides the measurable sets are chosen to be the smallest σ-algebra containing sets for which the measure can be easily defined. The measurable sets are said to be generated by the aforementioned intervals. The goal of this section is now to provide tools to handle σ-algebras, and to reduce statements about σ-algebras to statements on their generating sets.

2.1.1 Families of Sets

We first introduce a couple of set systems we want to use as generators for σ-algebras: semirings, rings, and algebras. All these set systems can be organized in a hierarchy, each of them requires closure under a set of operations. We cast this hierarchy into a locale hierarchy: σ-algebra $\subseteq$ algebra $\subseteq$ ring $\subseteq$ semiring $\subseteq$ family of sets. A *family of sets* is a set of sets $\mathcal{A}$ where each set $A \in \mathcal{A}$ is a subset of

[1] A proof of Vitali's theorem is found in Appendix D in Schilling [69]. There are formalizations in interactive theorem provers, like Cowles and Gamboa [19] in ACL2, or in HOL Light (`http://hol-light.googlecode.com/svn/trunk/Examples/vitali.ml`).

the space Ω:

locale *family-of-sets* =
  fixes $\Omega :: \alpha\ set$ **and** $\mathcal{A} :: \alpha\ set\ set$
  assumes $\mathcal{A} \subseteq \mathcal{P}(\Omega)$

This is necessary as in many cases we are not interested in the type universe $\mathcal{U}_\alpha$, but only a subset of it, e.g. α can be the type of natural numbers $\mathbb{N}$ and we want to have a distribution on the finite subset $\Omega = \{0 .. N\}$.

Notation: For set comprehensions $\{x \in \Omega \mid P\ x\}$ where we can infer the space Ω we usually drop Ω and write just $\{x \mid P\ x\}$. This typically occurs when the set comprehension is used together with a family of sets, $\{x \mid P\ x\} \in \mathcal{A} = \{x \in \Omega \mid P\ x\} \in \mathcal{A}$, or with the measure type, $\mu_{\mathcal{M}}\ \{x \mid P\ x\} = \mu_{\mathcal{M}}\ \{x \in \Omega_{\mathcal{M}} \mid P\ x\}$.

A *semiring of sets*[2] contains the empty set, is closed under intersection, and the result of set difference is the union of finitely many, disjoint elements of the semiring.

$$\begin{array}{lcl} \textit{disjoint-family}_{\mathcal{I}}\ F & \Leftrightarrow & \forall i, j \in \mathcal{I}.\ \Big(i \neq j \implies F\ i \cap F\ j = \emptyset\Big) \end{array}$$

$$\begin{array}{lcl} \textit{disjoint} & :: & \alpha\ set\ set \to \mathbb{B} \\ \textit{disjoint}\ \mathcal{A} & \Leftrightarrow & \textit{disjoint-family}_{\mathcal{A}}\ id \end{array}$$

locale *semiring-of-sets* = *family-of-sets* +
  assumes $\emptyset \in \mathcal{A}$
  and $\forall A, B \in \mathcal{A}.\ A \cap B \in \mathcal{A}$
  and $\forall A, B \in \mathcal{A}.\ \exists \mathcal{D} \subseteq \mathcal{A}.\ finite\ \mathcal{D} \wedge disjoint\ \mathcal{D} \wedge A \setminus B = \bigcup \mathcal{D}$

An example for a semiring is the set of all intervals $\{a\ ..< b\}$ on a linear ordered type.

A *ring of sets*[3] is a family of sets containing the empty set and is closed under union, intersection and difference or, alternatively, a semiring closed under union. By introducing rings as a sub-locale of semirings we automatically enable all theorems about semirings for rings.

locale *ring-of-sets* = *semiring-of-sets* +
  assumes $\forall A, B \in \mathcal{A}.\ A \cup B \in \mathcal{A}$

But, we also want to show that a ring is defined by its closure properties:

lemma RING-OF-SETS-IFF:
  ring-of-sets $\Omega\ \mathcal{A} \Leftrightarrow$
    $\mathcal{A} \subseteq \mathcal{P}(\Omega) \wedge \emptyset \in \mathcal{A} \wedge (\forall A, B \in \mathcal{A}.\ A \cup B \in \mathcal{A} \wedge A \setminus B \in \mathcal{A})$

We will see for Caratheodory's extension theorem that it is important to have a ring on which a premeasure is defined. However, it is easier to define a semiring and then extend it to a ring. We generate a ring out of a semiring $\mathcal{A}$:

$$\begin{array}{lcl} \textit{generated-ring} & :: & \alpha\ set\ set \to \alpha\ set\ set \\ \textit{generated-ring}\ \mathcal{A} & = & \left\{ \bigcup G \mid G \subseteq \mathcal{A} \wedge finite\ G \wedge disjoint\ G \right\} \end{array}$$

[2]When not otherwise specified, *semiring* refers to a semiring of sets.
[3]When not otherwise specified, *ring* refers to a ring of sets.

This generates a ring:

lemma GENERATED-RING:
semiring-of-sets $\Omega\ \mathcal{A} \implies$ *ring-of-sets* Ω (*generated-ring* $\mathcal{A}$)

An *algebra* contains the empty set, the entire space and is closed under set difference, union and intersection. Again, we express this as an extension of a ring:

locale *algebra* = *ring-of-sets* +
assumes $\Omega \in \mathcal{A}$

It is enough to show that an algebra contains the empty space and is closed under union and complement:

lemma ALGEBRA-IFF:
algebra $\Omega\ \mathcal{A} \Leftrightarrow$
$\mathcal{A} \subseteq \mathcal{P}(\Omega) \wedge \emptyset \in \mathcal{A} \wedge (\forall A, B \in \mathcal{A}.\ A \cup B \in \mathcal{A} \wedge \Omega \setminus A \in \mathcal{A})$

Finally, our goal is to define σ-*algebras*: algebras closed under countable union. They are important for the definition of measures: the measurable sets form a σ-algebra. We can incorporate σ-algebras in our locale hierarchy:

locale σ-*algebra* = *algebra* +
assumes $\forall F \in \mathbb{N} \to \mathcal{A}.\ (\bigcup_i F\ i) \in \mathcal{A}$

By this definition σ-algebras are also algebras. To show that a family of sets is a σ-algebra it is not required to prove all the assumptions for a semiring, ring and algebra. It is enough to show that it contains the empty set and is closed under complement and countable union.

lemma σ-ALGEBRA-IFF:
σ-*algebra* $\Omega\ \mathcal{A} \Leftrightarrow \Big(\mathcal{A} \subseteq \mathcal{P}(\Omega)\ \wedge\ \emptyset \in \mathcal{A}\ \wedge\ (\forall A \in \mathcal{A}.\ \Omega \setminus A \in \mathcal{A}) \wedge$
$(\forall F \in \mathbb{N} \to \mathcal{A}.\ (\bigcup_i F\ i) \in \mathcal{A})\Big)$

From the definition of σ-*algebra* we easily derive that it also contains countable, finite and binary union and intersection, as well as set difference. *The following rules assume that we are in a* σ*-algebra* $\mathcal{A}$ *over a set* Ω*.*

lemmas
$A, B \in \mathcal{A} \implies A \cup B, A \cap B, A - B \in \mathcal{A}$
$A \subseteq \mathcal{A}\ \wedge$ *finite* $A \implies \bigcup A \in \mathcal{A}$
$A \subseteq \mathcal{A}\ \wedge$ *finite* $A\ \wedge\ A \neq \emptyset \implies \bigcap A \in \mathcal{A}$
$F \in \mathcal{U}_{\iota::\mathcal{C}} \to \mathcal{A} \implies (\bigcup_i F\ i), (\bigcap_i F\ i) \in \mathcal{A}$
$\emptyset, \Omega \in \mathcal{A}$

An alternative characterization of these rules uses set comprehension. As a reminder, the set comprehension syntax implicitly restricts the set on the space Ω,

i.e. $\{x \mid P\ x\} = \{x \in \Omega \mid P\ x\}$. This allows more generic rules for measurability: finite intersection is not restricted to an empty index set and we can match on negation. implication, and constant predicates.

lemmas

$\{x \mid P\ x\}, \{x \mid Q\ x\} \in \mathcal{A} \implies$
$\{x \mid P\ x \implies Q\ x\}, \{x \mid P\ x \Leftrightarrow Q\ x\}, \{x \mid P\ x \wedge Q\ x\}, \{x \mid P\ x \vee Q\ x\} \in \mathcal{A}$

$\mathit{finite}\ I \wedge (\forall i \in I.\ \{x \mid P\ i\ x\} \in \mathcal{A}) \implies$
$\{x \mid \forall i \in I.\ P\ i\ x\}, \{x \mid \exists i \in I.\ P\ i\ x\} \in \mathcal{A}$

$(\forall i :: \iota :: \mathcal{C}.\ \{x \mid P\ i\ x\} \in \mathcal{A}) \implies \{x \mid \forall i.\ P\ i\ x\}, \{x \mid \exists i.\ P\ i\ x\} \in \mathcal{A}$

$\{x \mid P\ x\} \in \mathcal{A} \implies \{x \mid \neg P\ x\} \in \mathcal{A}$

$\{x \mid k\} \in \mathcal{A}$

These alternative characterizations are handy if the measurable set can be expressed as set comprehension of a first-order logic formulae. For example, consider the proof of the fact that the *LEAST* operator on $\mathbb{N}$ is measurable:

lemma COLLECT-LEAST-$\mathbb{N}$:
$(\forall i :: \mathbb{N}.\ \{x \mid P\ i\ x\} \in \mathcal{A}) \implies \{x \mid Q\ (\mathit{LEAST}\ i.\ P\ i\ x)\} \in \mathcal{A}$

Proof. The least natural number i fulfilling a predicate $P\ i\ x$ is easy to represent: either an i exists such that $P\ i\ x$ holds, then the least element can be characterized, otherwise $P\ i\ x$ is always false:

$$Q\ (\mathit{LEAST}\ i.\ P\ i\ x) \Leftrightarrow \Big((\exists i.\ P\ i\ x) \implies \forall i.\ \big(P\ i\ x \implies (\forall j.\ P\ j\ x \implies i \leq j) \implies Q\ i\big)\Big) \wedge \Big(\neg(\exists i.\ P\ i\ x) \implies Q\ (\mathit{LEAST}\ i.\ \mathit{False})\Big)$$

By rewriting the set $\{x \mid Q\ (\mathit{LEAST}\ i.\ P\ i\ x)\}$ with this equation it is easy to prove measurability by using the set-comprehension lemmas. □

We define the *σ-algebra generated by $\mathcal{G}$* (also called *σ-closure of $\mathcal{G}$*) as an inductive set. This set contains the generator $\mathcal{G}$ and is closed under the defining properties of σ-algebras:

inductive $\sigma\text{-sets} :: \alpha\ \mathit{set} \to \alpha\ \mathit{set}\ \mathit{set} \to \alpha\ \mathit{set}\ \mathit{set}$
where $\mathcal{G} \subseteq \sigma\text{-sets}\ \Omega\ \mathcal{G}$
and $\emptyset \in \sigma\text{-sets}\ \Omega\ \mathcal{G}$
and $\forall A \in \sigma\text{-sets}\ \Omega\ \mathcal{G}.\ \Omega \setminus A \in \sigma\text{-sets}\ \Omega\ \mathcal{G}$
and $\forall F \in \mathbb{N} \to \sigma\text{-sets}\ \Omega\ \mathcal{G}.\ (\bigcup_i F\ i) \in \sigma\text{-sets}\ \Omega\ \mathcal{G}$

The **inductive** command provides us with an induction rule for sets in *σ-sets*:

lemma σ-SETS.INDUCT:
$A \in \sigma\text{-sets}\ \Omega\ \mathcal{G} \wedge$
$(\forall A \in \mathcal{G}.\ P\ A) \wedge P\ \emptyset \wedge (\forall A \in \sigma\text{-sets}\ \Omega\ \mathcal{G}.\ P\ A \implies P\ (\Omega \setminus A)) \wedge$
$(\forall F \in \mathbb{N} \to \sigma\text{-sets}\ \Omega\ \mathcal{G}.\ (\forall i.\ P\ (F\ i)) \implies P\ (\bigcup_i F\ i)) \implies$
$P\ A$

We show that for each family of sets $\mathcal{G}$, *σ-sets* generates a σ-algebra:

lemma σ-ALGEBRA-σ-SETS:
$\mathcal{G} \subseteq \mathcal{P}(\Omega) \implies \sigma\text{-algebra}\ \Omega\ (\sigma\text{-sets}\ \Omega\ \mathcal{G})$

To be usable as a generating operator, we also show that the σ-algebra defined by *σ-sets* is the smallest one:

lemma σ-ALGEBRA-σ-SETS-LEAST:
$\sigma\text{-algebra}\ \Omega\ \mathcal{A} \implies \forall \mathcal{G} \subseteq \mathcal{A}.\ \sigma\text{-sets}\ \Omega\ \mathcal{G} \subseteq \mathcal{A}$

Also, by the definition of *σ-sets* it follows that it is a superset of the generator $\mathcal{G}$. Generators need to be families of sets, no further restriction is required on generators when used with *σ-sets*. Later, when we show uniqueness of measures or when we construct measures, the generators will be $\cap$-stable families, semirings, rings, or even algebras.

The σ-closure operation is compatible with the closure operations for rings:

lemma σ-SETS-GENERATED-RING-EQ:
$\text{semiring-of-sets}\ \Omega\ \mathcal{G} \implies \sigma\text{-sets}\ \Omega\ (\text{generated-ring}\ \mathcal{G}) = \sigma\text{-sets}\ \Omega\ \mathcal{G}$

2.1.2 Dynkin Systems

The inductive definition of *σ-sets* gives us a nice induction rule to prove properties about sets from the σ-algebra generated by $\mathcal{G}$. However, this induction rule has one problem: the sets in the countable union case are not disjoint. This is problematic when we want to prove properties about measures, as it only commutes with sums and unions if the sets are disjoint. We will provide an induction rule for σ-algebras generated by $\cap$-stable sets where the countable union case is weakened to the countable union of disjoint sets.

First, we introduce *Dynkin systems* as a family of sets containing the space and closed under complement and countable unions of *disjoint sets*.

locale $\text{dynkin-system} = \text{family-of-sets}\ +$
assumes $\Omega \in \mathcal{A}$
and $\forall A \in \mathcal{A}.\ \Omega \setminus A \in \mathcal{A}$
and $\forall F \in \mathbb{N} \to \mathcal{A}.\ \text{disjoint-family}_{\mathbb{N}}\ F \implies (\bigcup_i F\ i) \in \mathcal{A}$

Second, we introduce the *smallest Dynkin system generated by* $\mathcal{G}$ (also called the *Dynkin closure of* $\mathcal{G}$). The **inductive** command would introduce the same definition, but we do not need an induction rule so we do not use that command.

$\text{dynkin-sets} :: \alpha\ \text{set} \to \alpha\ \text{set set} \to \alpha\ \text{set set}$
$\text{dynkin-sets}\ \Omega\ \mathcal{G} = \bigcap \{\mathcal{A} \mid \mathcal{G} \subseteq \mathcal{A} \wedge \text{dynkin-system}\ \Omega\ \mathcal{A}\}$

The generated set is a Dynkin system:

lemma DYNKIN-SYSTEM-DYNKIN-SETS:
$\mathcal{G} \subseteq \mathcal{P}(\Omega) \implies \text{dynkin-system}\ \Omega\ (\text{dynkin-sets}\ \Omega\ \mathcal{G})$

Third, when the generator set $\mathcal{G}$ is $\cap$-*stable*, we know that the Dynkin closure is equal to the σ-closure:

$$\begin{array}{lll} \cap\text{-}\mathit{stable} & :: & \alpha\ \mathit{set}\ \mathit{set} \to \mathbb{B} \\ \cap\text{-}\mathit{stable}\ \mathcal{A} & \Leftrightarrow & \forall A, B \in \mathcal{A}.\ A \cap B \in \mathcal{A} \end{array}$$

theorem σ-SETS-EQ-DYNKIN-SETS:
$\mathcal{G} \subseteq \mathcal{P}(\Omega) \ \wedge\ \cap\text{-}\mathit{stable}\ \mathcal{G} \implies \mathit{dynkin\text{-}sets}\ \Omega\ \mathcal{G} = \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G}$

Finally, Dynkin systems are now used to prove Dynkin's lemma, which helps to generalize statements about all sets of a $\cap$-stable set to the σ-closure of that set.

lemma DYNKIN-LEMMA:
$\mathit{dynkin\text{-}system}\ \Omega\ \mathcal{A} \ \wedge\ \cap\text{-}\mathit{stable}\ \mathcal{G} \ \wedge\ \mathcal{G} \subseteq \mathcal{A} \implies \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G} \subseteq \mathcal{A}$

When we instantiate $\mathcal{A} = \{A \in \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G} \mid P\ A\}$ we gain a nice introduction rule for sets in the σ-closure of $\mathcal{G}$. Compared to Lemma σ-SETS.INDUCT, it assumes a $\cap$-stable generator $\mathcal{G}$ but the union case of this induction rule is now weakened to a *disjoint union*:

corollary σ-SETS-INDUCT-DISJOINT:
$$\begin{array}{l} \cap\text{-}\mathit{stable}\ \mathcal{G} \ \wedge\ \mathcal{G} \subseteq \mathcal{P}(\Omega) \ \wedge\ A \in \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G} \ \wedge \\ (\forall A \in \mathcal{G}.\ P\ A) \ \wedge\ P\ \emptyset \ \wedge \\ (\forall A \in \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G}.\ P\ A \implies P\ (\Omega \setminus A)) \ \wedge \\ (\forall F \in \mathbb{N} \to \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G}. \\ \quad \mathit{disjoint\text{-}family}_{\mathbb{N}}\ F \wedge (\forall i.\ P\ (F\ i)) \implies P\ (\bigcup_i F\ i)) \implies \\ P\ A \end{array}$$

The format of this rule allows us to apply Isabelle's **induct** proof method [74] resulting in nice structural proofs.

2.1.3 Measure Type

Before we continue with σ-algebras we define the concept of measure spaces. Up to now we used locales to work with σ-algebras, hence we would be inclined to use locales to introduce measure spaces. However this is a disadvantage if more than one measure space is used in a proof. Either we introduce a new locale having multiple measure space sublocales, or we instantiate multiple measure spaces in the proof itself. Both usages produce a lot of instantiated theorems and hence do not work very well with automation, especially the simplifier.

As an alternative we introduce a type representing measure spaces. First we introduce the concept of a *positive* and *countably additive* function μ on a set of sets $\mathcal{A}$:

$$\begin{array}{lll} \mathit{positive} & :: & \alpha\ \mathit{set}\ \mathit{set} \to \left(\alpha\ \mathit{set} \to \overline{\mathbb{R}}\right) \to \mathbb{B} \\ \mathit{positive}\ \mathcal{A}\ \mu & \Leftrightarrow & \mu\ \emptyset = 0 \wedge (\forall A \in \mathcal{A}.\ 0 \leq \mu\ A) \end{array}$$

$$\begin{array}{lll} \mathit{countably\text{-}additive} & :: & \alpha\ \mathit{set}\ \mathit{set} \to \left(\alpha\ \mathit{set} \to \overline{\mathbb{R}}\right) \to \mathbb{B} \\ \mathit{countably\text{-}additive}\ \mathcal{A}\ \mu & \Leftrightarrow & \end{array}$$
$$\forall F \in \mathbb{N} \to \mathcal{A}.\ \Big(\mathit{disjoint\text{-}family}_{\mathbb{N}}\ F \wedge (\bigcup_i F\ i) \in \mathcal{A} \implies \mu\ (\bigcup_i F\ i) = \sum_i \mu\ (F\ i)\Big)$$

Note that countable additivity implies commutativity of the measure w.r.t. sum and union only if the union of the sequence F is in $\mathcal{A}$. We do not yet require that $\mathcal{A}$ is a σ-algebra. This is important for Caratheodory's extension theorem in Section 2.2, where we extend premeasures to measures.

A *measure space* is a σ-algebra $\mathcal{A}$ with an associated measure, i.e. a positive, countably additive function μ:

$$\begin{aligned}&\mathit{measure\text{-}space} &&:: \quad \alpha\ \mathit{set} \to \alpha\ \mathit{set\,set} \to \left(\alpha\ \mathit{set} \to \overline{\mathbb{R}}\right) \to \mathbb{B}\\ &\mathit{measure\text{-}space}\ \Omega\ \mathcal{A}\ \mu \quad \Leftrightarrow\\ &\qquad \sigma\text{-}\mathit{algebra}\ \Omega\ \mathcal{A} \wedge \mathit{positive}\ \mathcal{A}\ \mu \wedge \mathit{countably\text{-}additive}\ \mathcal{A}\ \mu\end{aligned}$$

Now, we introduce the *measure type*. This puts together the space Ω, the measurable sets $\mathcal{A}$, the measure function μ, and the assumption that they form a measure space. The **typedef** command introduces the type α *measure* where the type universe is isomorphic to the set of all triples $(\Omega, \mathcal{A}, \mu)$ forming a measure space:

$$\begin{aligned}&\textbf{typedef}\ \alpha\ \mathit{measure} \quad =\\ &\quad \left\{(\Omega, \mathcal{A}, \mu) \mid (\forall A \notin \mathcal{A}.\ \mu\ A = 0) \wedge \mathit{measure\text{-}space}\ \Omega\ \mathcal{A}\ \mu\right\}\end{aligned}$$

We fix the measure $\mu\ A$ for non-measurable sets A to zero. As a result, we have for each measure space a unique element in α *measure*. We need to show that such a measure space exists, which is easy: we choose the power set $\mathcal{A} = \mathcal{P}(\Omega)$, and the constant zero function as measure. Then, the **typedef** command introduces the abstraction morphism

$$\mathit{Abs}_{\mathit{measure}} :: (\alpha\ \mathit{set} \times \alpha\ \mathit{set\,set} \times (\alpha\ \mathit{set} \to \overline{\mathbb{R}})) \to \alpha\ \mathit{measure}$$

and the representation morphism

$$\mathit{Rep}_{\mathit{measure}} :: \alpha\ \mathit{measure} \to (\alpha\ \mathit{set} \times \alpha\ \mathit{set\,set} \times (\alpha\ \mathit{set} \to \overline{\mathbb{R}}))\ .$$

The **typedef** command provides us also the following law: when $(\Omega, \mathcal{A}, \mu)$ forms a measure space then $\mathit{Rep}_{\mathit{measure}}\ (\mathit{Abs}_{\mathit{measure}}\ (\Omega, \mathcal{A}, \mu)) = (\Omega, \mathcal{A}, \mu)$.

First we introduce access functions to get the space, the measurable sets and the measure function of a measure:

$$\begin{aligned}\Omega_\square &:: \alpha\ \mathit{measure} \to \alpha\ \mathit{set}\\ \Omega_{\mathcal{M}} &= \textbf{let}\ (\Omega, \mathcal{A}, \mu) = \mathit{Rep}_{\mathit{measure}}\ \mathcal{M}\ \textbf{in}\ \Omega\\ \mathcal{A}_\square &:: \alpha\ \mathit{measure} \to \alpha\ \mathit{set\,set}\\ \mathcal{A}_{\mathcal{M}} &= \textbf{let}\ (\Omega, \mathcal{A}, \mu) = \mathit{Rep}_{\mathit{measure}}\ \mathcal{M}\ \textbf{in}\ \mathcal{A}\\ \mu_\square &:: \alpha\ \mathit{measure} \to \alpha\ \mathit{set} \to \overline{\mathbb{R}}\\ \mu_{\mathcal{M}} &= \textbf{let}\ (\Omega, \mathcal{A}, \mu) = \mathit{Rep}_{\mathit{measure}}\ \mathcal{M}\ \textbf{in}\ \mu\end{aligned}$$

The access function for measures $\mu_{\mathcal{M}}$ is called the extended measure, returning extended real values. Later we will introduce the finite measure $\mu^f_{\mathcal{M}}$, returning real values.

With the type definition of α *measure* two measures $\mathcal{M}$ and $\mathcal{N}$ are equal iff the measurable sets are equal and the measure functions are equal on these sets:

lemma MEASURE-EQI:
$$\mathcal{A}_{\mathcal{M}} = \mathcal{A}_{\mathcal{N}} \ \wedge\ (\forall A \in \mathcal{A}_{\mathcal{M}}.\ \mu_{\mathcal{M}}\ A = \mu_{\mathcal{N}}\ A) \implies \mathcal{M} = \mathcal{N}$$

In Section 2.3.2 we will see a variant of this lemma which assumes equality only on an $\cap$-stable generator.

We use the abstraction morphism to define measures by providing the space Ω, the generating sets $\mathcal{G}$ and the measure function μ:

$$\begin{aligned}&\mathit{measure\text{-}of}\ \Omega\ \mathcal{G}\ \mu\ =\\&\quad \textbf{let}\ \mathcal{A} = \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G}\\&\quad \textbf{in if}\ \mathit{measure\text{-}space}\ \Omega\ \mathcal{A}\ \mu\ \textbf{then}\ \mathit{Abs}_{\mathit{measure}}\ (\Omega, \mathcal{A}, \lambda A.\ \textbf{if}\ A \in \mathcal{A}\ \textbf{then}\ \mu\ A\ \textbf{else}\ 0)\\&\qquad\qquad\qquad\qquad\qquad\qquad \textbf{else}\ \mathit{Abs}_{\mathit{measure}}\ (\Omega, \mathcal{A}, \lambda A.\ 0)\end{aligned}$$

We force the measure function to be zero if we do not have a measure space. This allows us to reason separately about the σ-algebra and the associated measure function. With this trick we have the following rules:

lemma SPACE-MEASURE-OF:
$\mathcal{G} \subseteq \mathcal{P}(\Omega) \implies \Omega_{\mathit{measure\text{-}of}\ \Omega\ \mathcal{G}\ \mu} = \Omega$

lemma SETS-MEASURE-OF:
$\mathcal{G} \subseteq \mathcal{P}(\Omega) \implies \mathcal{A}_{\mathit{measure\text{-}of}\ \Omega\ \mathcal{G}\ \mu} = \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G}$

lemma μ-MEASURE-OF-σ:
$\sigma\text{-}\mathit{algebra}\ \Omega\ \mathcal{A} \ \wedge\ \mathit{positive}\ \mathcal{A}\ \mu \ \wedge\ \mathit{countably\text{-}additive}\ \mathcal{A}\ \mu \implies$
$\forall A \in \mathcal{A}.\ \mu_{\mathit{measure\text{-}of}\ \Omega\ \mathcal{A}\ \mu}\ A = \mu\ A$

Sometimes only the σ-algebra of a measure space is of interest. In this case we use the abbreviation $\sigma\text{-}\mathit{of}\ \Omega\ \mathcal{A}$ for $\mathit{measure\text{-}of}\ \Omega\ \mathcal{A}\ (\lambda x.\ 0)$.

The usage of the type $\alpha\ \mathit{measure}$ results in a different proof structure. If we used a predicate on the triple $(\Omega, \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G}, \mu)$ we would not need a proof to show the measure values for elements in $\mathcal{G}$. But we need to prove explicitly that it forms a measure space when applying theorems about measure spaces. In contrast, by using the measure type we know that the result of $\mathit{measure\text{-}of}\ \Omega\ \mathcal{G}\ \mu$ is a measure. When we apply a theorem about measure spaces this assumption is fulfilled by the type. Only when we want to show properties about the measure values for the elements in $\mathcal{G}$, we need to prove that that $(\Omega, \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G}, \mu)$ forms a measure space.

When constructing a measure with *measure-of* the measure function needs to be defined on all measurable sets. But often the measure function is more easily defined on the generating sets. Also, the generating sets are often the range of a function G over some index set I. It then is easier to define the measure value based on the index than on the set $G[I]$. For example, to define the Lebesgue measure the index set is $\{(a, b) \mid a \leq b\}$, the generating sets are all $\{a\ ..<\ b\}$, and the measure values are $b - a$. For this we introduce the helper function *extend-measure*:

$$\begin{aligned}&\mathit{extend\text{-}measure} :: \alpha\ \mathit{set} \to \iota\ \mathit{set} \to \Big(\iota \to \alpha\ \mathit{set}\Big) \to \Big(\iota \to \overline{\mathbb{R}}\Big) \to \alpha\ \mathit{measure}\\&\mathit{extend\text{-}measure}\ \Omega\ I\ G\ \mu =\\&\quad \textbf{let}\ P\ \mu' = (\forall i \in I.\ \mu'\ (G\ i) = \mu\ i) \wedge \mathit{measure\text{-}space}\ \Omega\ (\sigma\text{-}\mathit{sets}\ \Omega\ G[I])\ \mu'\\&\quad \textbf{in if}\ (\exists \mu'.\ P\ \mu') \wedge \neg(\forall i \in I.\ \mu\ i = 0)\ \textbf{then}\ \mathit{measure\text{-}of}\ \Omega\ G[I]\ (\mathit{SOME}\ \mu'.\ P\ \mu')\\&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textbf{else}\ \mathit{measure\text{-}of}\ \Omega\ G[I]\ (\lambda A.\ 0)\end{aligned}$$

With the left conjunct in the **if** expression we guarantee that the resulting measure is a zero measure even when the measure is not uniquely defined on $G[I]$ and μ is a zero premeasure. Similarly to *measure-of* we guarantee that the measurable sets are the sets generated by $G[I]$ even when μ does not generate a measure space:

lemma SPACE-EXTEND-MEASURE:
$G[I] \subseteq \mathcal{P}(\Omega) \implies \Omega_{\textit{extend-measure}\ \Omega\ I\ G\ \mu} = \Omega$

lemma SETS-EXTEND-MEASURE:
$G[I] \subseteq \mathcal{P}(\Omega) \implies \mathcal{A}_{\textit{extend-measure}\ \Omega\ I\ G\ \mu} = \textit{sigma-sets}\ \Omega\ G[I]$

To extend a premeasure μ to the measure *extend-measure* $\Omega\ I\ G\ \mu$ we need to give a witness measure function μ' extending μ:

lemma μ-EXTEND-MEASURE:
$\mathcal{M} = \textit{extend-measure}\ \Omega\ I\ G\ \mu \wedge$
$G[I] \subseteq \mathcal{P}(\Omega) \wedge \Big(\forall i \in I.\ \mu'\ (G\ i) = \mu\ i\Big) \wedge$
$\textit{positive}\ \mathcal{A}_{\mathcal{M}}\ \mu' \wedge \textit{countably-additive}\ \mathcal{A}_{\mathcal{M}}\ \mu' \wedge i \in I \implies$
$\mu_{\mathcal{M}}\ (G\ i) = \mu\ i$

In Section 2.2 we will see how to construct such a measure function. In Section 2.3.2 we will see that when the generating set is $\cap$-stable, then the extended measure is uniquely defined.

2.1.4 Measurable Functions

In measure theory a function f is called $\mathcal{N}$-*measurable* if for all $A \in \mathcal{A}_{\mathcal{N}}$ the inverse image $\{x \mid f\ x \in A\} = f^{-1}[A] \cap \Omega_{\mathcal{M}}$ is measurable in $\mathcal{M}$. The intersection of $f^{-1}[A]$ with $\Omega_{\mathcal{M}}$ is necessary in HOL: the function f is a total function and hence also defined outside of $\Omega_{\mathcal{M}}$. To ensure that the inverse image is still measurable we need to cut it with $\Omega_{\mathcal{M}}$. We also add the assumption $f \in \Omega_{\mathcal{M}} \to \Omega_{\mathcal{N}}$, forcing f to map elements from $\Omega_{\mathcal{M}}$ into $\Omega_{\mathcal{N}}$. These are two concession we need to make by not working type based, i.e. our measure spaces $\Omega_{\mathcal{M}}$ and $\Omega_{\mathcal{N}}$ are not restricted to the type universe.

$$\begin{aligned} &\textit{measurable} && :: && \alpha\ \textit{measure} \to \beta\ \textit{measure} \to \Big(\alpha \to \beta\Big)\ \textit{set} \\ &\textit{measurable}\ \mathcal{M}\ \mathcal{N} && = && \Big\{ f \in \Omega_{\mathcal{M}} \to \Omega_{\mathcal{N}} \;\Big|\; \forall A \in \mathcal{A}_{\mathcal{N}}.\ f^{-1}[A] \cap \Omega_{\mathcal{M}} \in \mathcal{A}_{\mathcal{M}} \Big\} \end{aligned}$$

This definition looks similar to continuity on topological spaces, just replace measurable sets by open sets. However, while it is easy to construct and analyze non-continuous functions this is not the case with measurability. Similar to measurable sets, the goal is to have measurable as many functions as possible. For example, most theorems about the Lebesgue integral require measurable functions.

Without knowing more about the concrete σ-algebras of $\mathcal{M}$ and $\mathcal{N}$ we only provide a couple of generic statements about the composition of measurable func-

tions:

lemma MEASURABLE-ID:
$(\lambda x.\ x) \in \mathit{measurable}\ \mathcal{M}\ \mathcal{M}$

lemma MEASURABLE-CONST:
$c \in \Omega_{\mathcal{N}} \implies (\lambda x.\ c) \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N}$

lemma MEASURABLE-COMP:
$f \in \mathit{measurable}\ \mathcal{M}_1\ \mathcal{M}_2 \wedge g \in \mathit{measurable}\ \mathcal{M}_2\ \mathcal{M}_3 \implies$
$(g \circ f) \in \mathit{measurable}\ \mathcal{M}_1\ \mathcal{M}_3$

lemma MEASURABLE-IF:
$f \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N} \wedge g \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N} \wedge A \in \mathcal{A}_{\mathcal{M}} \implies$
$(\lambda x.\ \textbf{if}\ x \in A\ \textbf{then}\ f\ x\ \textbf{else}\ g\ x) \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N}$

When the σ-algebra of $\mathcal{N}$ is generated by $\mathcal{G}$ we can reduce $\mathcal{N}$-measurability to measurability on the generator $\mathcal{G}$:

theorem MEASURABLE-σ:
$\mathcal{A}_{\mathcal{N}} = \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G} \wedge \mathcal{G} \subseteq \mathcal{P}(\Omega) \wedge f \in \Omega_{\mathcal{M}} \to \Omega \wedge$
$(\forall A \in \mathcal{G}.\ f^{-1}[A] \cap \Omega_{\mathcal{M}} \in \mathcal{A}_{\mathcal{M}}) \implies$
$f \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N}$

This simplifies the task of showing that a function is measurable, as the generating sets are often very easy to characterize.

2.1.5 Borel Sets

In general, the *Borel sets* form the canonical σ-algebra on the real numbers. They will be later crucially used for the Lebesgue measure and integral. For example, on the Lebesgue measure at least all bounded intervals $\{a\ ..<\ b\}$ should be measurable. For the Lebesgue integral all half open intervals $\{..<\ a\}$ should be measurable. It is easy to see that they generate the same σ-algebra: $\{a\ ..<\ b\} = \{..<\ b\} - \{..<\ a\}$ and $\{..<\ a\} = (\bigcup_n \{-n\ ..<\ a\})$. Actually, we can even go one step further: this σ-algebra contains also all open sets.

The *Borel sets* are the σ-algebra generated by the open sets of a topological space. In Isabelle/HOL topological spaces form the type class $\mathcal{T}$. For each type α in the type class $\mathcal{T}$, *open A* holds when $A :: \alpha\ \mathit{set}$ is open in the topology of $\alpha :: \mathcal{T}$. Types in $\mathcal{T}$ include Euclidean spaces (hence $\mathbb{R}$), pairs of topological spaces $\alpha :: \mathcal{T} \times \beta :: \mathcal{T}$ and $\overline{\mathbb{R}}$.

$\mathcal{B}_\alpha :: (\alpha :: \mathcal{T})\ \mathit{measure}$
$\mathcal{B}_\alpha = \sigma\text{-}\mathit{of}\ \mathcal{U}_\alpha\ \{G \mid \mathit{open}\ G\}$

In the introduction of Section 2.1 we learned about Vitali's theorem, stating the existence of a non Lebesgue-measurable set. As each Borel set is Lebesgue measurable there also exists a non-Borel set of real numbers, so the Borel σ-algebra is not the power set of the real numbers.

From the definition of $\mathcal{B}_\alpha$ it immediately follows that open, closed, and compact sets are $\mathcal{B}_\alpha$-measurable. On $\mathcal{B}_{\mathbb{R}^n}$ the intervals, like $\{a\ ..<\ b\}$ or $\{a\ <..\}$, are

also measurable including the singleton set. Is the other direction also true, is $\mathcal{B}_{\mathbb{R}^n}$ generated by intervals? As each open set can be covered with intervals with rational endpoints, the Borel sets are alternatively generated by intervals:

theorem BOREL-EQ-GREATERTHANLESSTHAN:
$\mathcal{B}_{\mathbb{R}^n} = \sigma\text{-of } \mathcal{U}_{\mathbb{R}^n} \{\{a <..< b\} \mid a, b\}$

corollary BOREL-EQ-LESSTHAN, -GREATERTHAN, -ATMOST, -ATLEAST:
$\mathcal{B}_{\mathbb{R}^n} = \sigma\text{-of } \mathcal{U}_{\mathbb{R}^n} \{\{..< a\} \mid a\}$,
$\mathcal{B}_{\mathbb{R}^n} = \sigma\text{-of } \mathcal{U}_{\mathbb{R}^n} \{\{a <..\} \mid a\}$,
$\mathcal{B}_{\mathbb{R}^n} = \sigma\text{-of } \mathcal{U}_{\mathbb{R}^n} \{\{.. a\} \mid a\}$,
$\mathcal{B}_{\mathbb{R}^n} = \sigma\text{-of } \mathcal{U}_{\mathbb{R}^n} \{\{a ..\} \mid a\}$

A continuous function maps open sets to open sets, hence from the definition of the Borel sets and Theorem MEASURABLE-σ it immediately follows that a continuous function is $\mathcal{B}_\alpha$-measurable. *continuous-on* A f holds when the function $f :: \alpha \to \beta$, for two topological spaces $\alpha :: \mathcal{T}$ and $\beta :: \mathcal{T}$, is continuous on $A :: \alpha$ *set*. The real analysis and the multivariate analysis in Isabelle/HOL already provide continuity results for arithmetic and trigonometric operations. By reducing Borel-measurability to continuity we can reuse these results.

lemma $\mathcal{B}_\alpha$-MEASURABLE-CONT:
continuous-on A $f \wedge$ *open* A $\Longrightarrow$
$(\lambda x.$ **if** $x \in A$ **then** $f\ x$ **else** $c) \in$ *measurable* $\mathcal{B}_\alpha$ $\mathcal{B}_\beta$

With this lemma we immediately show that unary minus (everywhere continuous: $A = \mathcal{U}_\alpha$), the inverse operation $1/x$ (continuous on $A = \mathcal{U}_\alpha - \{0\}$) and the logarithm $\log_b\ x$ (continuous on $A = \{0 <..\}$) are Borel-measurable. The logarithm is a little problematic, as it is unspecified for nonpositive values. However, while it is unspecified there from its definition follows that it is constant for nonpositive values, enough to prove measurability.

lemma $\mathcal{B}_\mathbb{R}$-MEASURABLE-UMINUS, -INVERSE, -LOG:
$f \in$ *measurable* $\mathcal{M}$ $\mathcal{B}_\mathbb{R}$ $\Longrightarrow$
$(\lambda x.\ -f\ x), (\lambda x.\ 1/f\ x), (\lambda x.\ \log_b(f\ x)) \in$ *measurable* $\mathcal{M}$ $\mathcal{B}_\mathbb{R}$

Lemma $\mathcal{B}_\alpha$-MEASURABLE-CONT is not enough to show the measurability for binary operators like addition and multiplication. For these we employ the product topology. It allows us to specify continuity of binary operators.

lemma $\mathcal{B}_\alpha$-MEASURABLE-CONT-PAIR:
$f \in$ *measurable* $\mathcal{M}$ $\mathcal{B}_{\mathbb{R}^n} \wedge g \in$ *measurable* $\mathcal{M}$ $\mathcal{B}_{\mathbb{R}^m} \wedge$
continuous-on $\mathcal{U}_{\mathbb{R}^n \times \mathbb{R}^m}$ $(\lambda x.\ H\ (\mathit{fst}\ x)\ (\mathit{snd}\ x)) \Longrightarrow$
$(\lambda x.\ H\ (f\ x)\ (g\ x)) \in$ *measurable* $\mathcal{M}$ $\mathcal{B}_\alpha$

This is the result of the Lemmas MEASURABLE-COMP and $\mathcal{B}_\alpha$-MEASURABLE-CONT and the fact that the pairing operation $(\lambda x.\ (f\ x, g\ x))$ is $\mathcal{B}_{\mathbb{R}^n \times \mathbb{R}^m}$-measurable. Finally, for addition, subtraction, and multiplication follows with this lemma and together

with their continuity that they are also $\mathcal{B}_{\mathbb{R}}$-measurable:

lemma $\mathcal{B}_{\mathbb{R}}$-MEASURABLE, -ADD, -MINUS, -TIMES:
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \ \wedge\ g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \implies$
$(\lambda x.\ f\ x + g\ x), (\lambda x.\ f\ x - g\ x), (\lambda x.\ f\ x \cdot g\ x) \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}}$

lemma $\mathcal{B}_{\mathbb{R}}$-MEASURABLE-SETSUM, -SETPROD:
$(\forall i \in I.\ f\ i \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}}) \implies$
$\left(\lambda x.\ \sum_{i\in I} f\ i\ x\right), \left(\lambda x.\ \prod_{i\in I} f\ i\ x\right) \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}}$

We know that $\mathcal{B}_{\mathbb{R}^n}$ is also generated by intervals. We show for each semi-open interval an alternative characterization for $\mathcal{B}_{\mathbb{R}^n}$-measurable functions, again using Theorem MEASURABLE-σ:

lemma $\mathcal{B}_{\mathbb{R}^n}$-MEASURABLE-IFF-LESS, -LE, -GREATER, -GE:
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}^n} \Leftrightarrow (\forall a.\ \{x \mid f\ x < a\} \in \mathcal{A}_{\mathcal{M}})$,
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}^n} \Leftrightarrow (\forall a.\ \{x \mid f\ x \le a\} \in \mathcal{A}_{\mathcal{M}})$,
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}^n} \Leftrightarrow (\forall a.\ \{x \mid a < f\ x\} \in \mathcal{A}_{\mathcal{M}})$,
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}^n} \Leftrightarrow (\forall a.\ \{x \mid a \le f\ x\} \in \mathcal{A}_{\mathcal{M}})$

While helpful to show measurability of functions, these equalities are not very helpful to prove the measurability of sets. Similar to the set comprehension lemmas in Section 2.1.1, we provide the following introduction rules:

lemma $\mathcal{B}_{\mathbb{R}}$-MEASURABLE-LESS, -LE, -EQ:
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \ \wedge\ g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \implies$
$\{x \mid f\ x < g\ x\}, \{x \mid f\ x \le g\ x\}, \{x \mid f\ x = g\ x\} \in \mathcal{A}_{\mathcal{M}}$

All these rules about $\mathcal{B}_{\mathbb{R}^n}$-measurable functions can be translated to rules about $\mathcal{B}_{\overline{\mathbb{R}}}$-measurable functions: the coercion functions $(\cdot)_{\overline{\mathbb{R}}}$ and $(\cdot)_{\mathbb{R}}$ are continuous and hence also measurable. The arithmetic operations can be represented as a case distinction on the input value, and using the coercion functions.

And finally we show that a function is measurable whenever it is the supremum, infimum, or limit of $\mathcal{B}_{\overline{\mathbb{R}}}$-measurable functions:

lemma $\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE-SUP, -INF, -LIMSUP, -LIMINF, -LIM:
$(\forall i :: \iota :: \mathcal{C} \in A.\ f\ i \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}) \implies$
$(\lambda x.\ \sup_{i\in A} f\ i\ x), (\lambda x.\ \inf_{i\in A} f\ i\ x) \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}$

$(\forall i :: \mathbb{N}.\ f\ i \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}) \implies$
$(\lambda x.\ \limsup_i f\ i\ x), (\lambda x.\ \liminf_i f\ i\ x) \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}$

$(\forall i :: \mathbb{N}.\ f\ i \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}) \ \wedge\ (\forall x \in \Omega_{\mathcal{M}}.\ \lim_i f\ i\ x = u\ x) \implies$
$u \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}$

With this we show the measurability of the limit of $\mathcal{B}_{\mathbb{R}}$-measurable functions:

lemma $\mathcal{B}_{\mathbb{R}}$-MEASURABLE-LIM:
$(\forall i :: \mathbb{N}.\ f\ i \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}}) \ \wedge\ (\forall x \in \Omega_{\mathcal{M}}.\ \lim_i f\ i\ x = u\ x) \implies$
$u \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}}$

2.2 Extending Premeasures

In this section we describe the formalization of Caratheodory's extension theorem, which extends a premeasure to a measure. A *premeasure* is a *countably-additive* and *positive* function on a family of sets.

The formalization of Caratheodory's theorem in Isabelle/HOL was originally ported from Hurd's thesis [39]. Later it was adapted to support measures on the extended reals and assumes the premeasure is defined on a ring of sets instead of an algebra. These extensions of the proofs are based on Bauer's [9] and Elstrodt's [23] book. The proof still uses the outer measure of μ, and shows that its λ-system contains the σ-algebra $\sigma\text{-sets}\ \Omega\ \mathcal{G}$.

theorem CARATHEODORY':
$$\textit{ring-of-sets}\ \Omega\ \mathcal{G}\ \wedge\ \textit{positive}\ \mathcal{G}\ \mu\ \wedge\ \textit{countably-additive}\ \mathcal{G}\ \mu\ \Longrightarrow$$
$$\exists \mu'.\ \textit{measure-space}\ \Omega\ (\sigma\text{-sets}\ \Omega\ \mathcal{G})\ \mu'\ \wedge\ (\forall G \in \mathcal{G}.\ \mu\ G = \mu'\ G)$$

This theorem greatly simplifies the construction of a measure space. But we still need to show that the premeasure is countably additive. Luckily, for this there are helpful alternative characterizations, if we know that the function is *additive* on the ring $\mathcal{A}$:

$$\textit{additive} \quad :: \quad \alpha\ \textit{set set} \to \left(\alpha\ \textit{set} \to \overline{\mathbb{R}}\right) \to \mathbb{B}$$
$$\textit{additive}\ \mathcal{A}\ \mu \quad \Leftrightarrow \quad \left(\forall A, B \in \mathcal{A}.\ A \cup B \in \mathcal{A} \Longrightarrow \mu\ A + \mu\ B = \mu\ (A \cup B)\right)$$

First, countable additivity can be replaced by commutation with the limit of increasing sequences. We say the premeasure is *continuous from below*.[4]

lemma COUNTABLY-ADDITIVE-IFF-CONTINUOUS-FROM-BELOW:
$$\textit{ring-of-sets}\ \Omega\ \mathcal{A}\ \wedge\ \textit{positive}\ \mathcal{A}\ \mu\ \wedge\ \textit{additive}\ \mathcal{A}\ \mu\ \Longrightarrow$$
$$\textit{countably-additive}\ \mathcal{A}\ \mu \Leftrightarrow$$
$$(\forall F \in \mathbb{N} \to \mathcal{A}.\ \textit{incseq}\ F \wedge (\bigcup_i F\ i) \in \mathcal{A} \wedge \mu\ (F\ i) \xrightarrow[i \to \infty]{} \mu\ (\bigcup_i F\ i)$$

corollary CARATHEODORY-CONTINUOUS-FROM-BELOW:
$$\textit{ring-of-sets}\ \Omega\ \mathcal{G}\ \wedge\ \textit{positive}\ \mathcal{G}\ \mu\ \wedge\ \textit{additive}\ \mathcal{G}\ \mu\ \wedge$$
$$(\forall F \in \mathbb{N} \to \mathcal{G}.\ \textit{incseq}\ F \wedge (\bigcup_i F\ i) \in \mathcal{A} \wedge \mu\ (F\ i) \xrightarrow[i \to \infty]{} \mu\ (\bigcup_i F\ i)) \Longrightarrow$$
$$\exists \mu'.\ \textit{measure-space}\ \Omega\ (\sigma\text{-sets}\ \Omega\ \mathcal{G})\ \mu'\ \wedge\ (\forall G \in \mathcal{G}.\ \mu\ G = \mu'\ G)$$

Second, when the measure values of the generating sets are finite, we can replace countable additivity by the commutativity with the limit of decreasing sequences converging to $\emptyset$. This enables a nice proof principle: we assume that the limit of the measure values $\mu\ (F\ i)$ is nonzero and show from this that $\bigcap_i F\ i$ is not empty. Then the premeasure is $\emptyset$-*continuous*:

lemma $\emptyset$-CONTINUOUS-IMP-COUNTABLY-ADDITIVE:
$$\textit{ring-of-sets}\ \Omega\ \mathcal{A}\ \wedge\ \textit{positive}\ \mathcal{A}\ \mu\ \wedge\ \textit{additive}\ \mathcal{A}\ \mu\ \wedge\ (\forall A \in \mathcal{A}.\ \mu\ A < \infty)\ \wedge$$
$$\left(\forall F \in \mathbb{N} \to \mathcal{A}.\ \textit{decseq}\ F \wedge (\bigcap_i F\ i) = \emptyset \wedge \mu\ (F\ i) \xrightarrow[i \to \infty]{} 0\right) \Longrightarrow$$
$$\textit{countably-additive}\ \mathcal{A}\ \mu$$

[4] This is similar to left continuity of a function f on metric spaces, stating that for each increasing sequence x_i converging to x, the sequence $f\ x_i$ converges to $f\ x$.

corollary CARATHEODORY-∅-CONTINUOUS:
ring-of-sets $\Omega\ \mathcal{G}\ \wedge$ *positive* $\mathcal{G}\ \mu\ \wedge$ *additive* $\mathcal{G}\ \mu\ \wedge\ (\forall G \in \mathcal{G}.\ \mu\ G < \infty) \wedge$
$(\forall F \in \mathbb{N} \to \mathcal{G}.\ \textit{decseq}\ F \wedge (\bigcap_i F\ i) = \emptyset \wedge \mu\ (F\ i) \xrightarrow[i\to\infty]{} 0)\ \Longrightarrow$
$\exists \mu'.\ \textit{measure-space}\ \Omega\ (\sigma\text{-}\textit{sets}\ \Omega\ \mathcal{G})\ \mu'\ \wedge\ (\forall G \in \mathcal{G}.\ \mu\ G = \mu'\ G)$

Sometimes, it is easier to define the premeasure on a semiring. In this case we need to extend the premeasure to a premeasure on the ring generated by the semiring. As the semiring is *not* closed under union, *additive* cannot be used to show finite additivity. For this we introduce *volumes*: positive and finitely additive functions.

$$\begin{aligned}
&\textit{volume} \quad :: \quad \alpha\ \textit{set set} \to \left(\alpha\ \textit{set} \to \overline{\mathbb{R}}\right) \to \mathbb{B}\\
&\textit{volume}\ \mathcal{A}\ \mu \quad \Leftrightarrow\\
&\quad \textit{positive}\ \mathcal{A}\ \mu\ \wedge\\
&\quad (\forall G \subseteq \mathcal{A}.\ \textit{finite}\ G \wedge \textit{disjoint}\ G \wedge \bigcup G \in \mathcal{A}\ \Longrightarrow\ (\textstyle\sum_{A\in G} \mu\ A) = \mu\ (\bigcup G))
\end{aligned}$$

We can extend a volume on a semiring to a volume on the generated ring:

lemma EXTEND-VOLUME:
semiring-of-sets $\Omega\ \mathcal{G}\ \wedge$ *volume* $\mathcal{G}\ \mu\ \Longrightarrow$
$\exists \mu'.\ \textit{volume}\ (\textit{generated-ring}\ \mathcal{A})\ \mu'\ \wedge\ (\forall G \in \mathcal{G}.\ \mu\ G = \mu'\ G)$

Each set in the generated ring is a union of disjoint sets in the semiring. With this it is obvious how to define the extended volume. However, we need to show that the sum is independent of the representation of the union.

Finally, we can generalize Caratheodory's extension theorem from rings to semirings:

theorem CARATHEODORY:
semiring-of-sets $\Omega\ \mathcal{G}\ \wedge$ *positive* $\mathcal{G}\ \mu\ \wedge$ *countably-additive* $\mathcal{G}\ \mu\ \Longrightarrow$
$\exists \mu'.\ \textit{measure-space}\ \Omega\ (\sigma\text{-}\textit{sets}\ \Omega\ \mathcal{G})\ \mu'\ \wedge\ (\forall G \in \mathcal{G}.\ \mu\ G = \mu'\ G)$

First we show that the positive, countably additive function μ is also a volume. The extended volume is also countably additive, as each sum of volumes of the ring can be "flattened" to a single sum of elements of the semiring. As the last step we apply Caratheodory's extension theorem and by Lemma σ-SETS-GENERATED-RING-EQ we know that the σ-closures of the generated ring and the semiring are equal.

2.3 Properties of Measure Spaces

In Section 2.1.3 the *measure* type was introduced. We now explore the behaviour of the measure function $\mu_{\mathcal{M}} :: \alpha\ \textit{set} \to \overline{\mathbb{R}}$ for a measure $\mathcal{M} :: \alpha\ \textit{measure}$. From the definition of the measure type we already know that $\mathcal{M}$ forms a measure space, hence *countably-additive* $\mathcal{A}_{\mathcal{M}}\ \mu_{\mathcal{M}}$ and *positive* $\mathcal{A}_{\mathcal{M}}\ \mu_{\mathcal{M}}$ hold. So we immediately

derive:

lemma μ-EMPTY: $\mu_{\mathcal{M}}\ \emptyset = 0$

lemma μ-POSITIVE: $0 \leq \mu_{\mathcal{M}}\ A$

lemma SUMINF-μ:
$\mathit{disjoint\text{-}family}_{\mathbb{N}}\ F\ \wedge\ F \in \mathbb{N} \to \mathcal{A}_{\mathcal{M}}\ \Longrightarrow\ (\sum_i \mu_{\mathcal{M}}\ (F\ i)) = \mu_{\mathcal{M}}\ (\bigcup_i F\ i)$

We also know that $\mu_{\mathcal{M}}\ A = 0$ for a non-measurable $A \notin \mathcal{A}_{\mathcal{M}}$, so no assumption about the measurability of A is necessary in Lemma μ-POSITIVE.

From these basic laws about measures, we easily derive that it is additive, monotone, finite additive and commutes with set difference. Again, it is monotone even if the smaller set is not measurable:[5]

lemma PLUS-μ:
$A, B \in \mathcal{A}_{\mathcal{M}}\ \wedge\ A \cap B = \emptyset\ \Longrightarrow\ \mu_{\mathcal{M}}\ A + \mu_{\mathcal{M}}\ B = \mu_{\mathcal{M}}\ (A \cup B)$

lemma μ-MONO:
$A \subseteq B\ \wedge\ B \in \mathcal{A}_{\mathcal{M}}\ \Longrightarrow\ \mu_{\mathcal{M}}\ A \leq \mu_{\mathcal{M}}\ B$

lemma SETSUM-μ:
$\mathit{disjoint\text{-}family}_I\ F\ \wedge\ F \in I \to \mathcal{A}_{\mathcal{M}}\ \wedge\ \mathit{finite}\ I\ \Longrightarrow$
$(\sum_{i \in I} \mu_{\mathcal{M}}\ (F\ i)) = \mu_{\mathcal{M}}\ (\bigcup_{i \in I} F\ i)$

lemma MINUS-μ:
$\mu_{\mathcal{M}}\ B < \infty\ \wedge\ A, B \in \mathcal{A}_{\mathcal{M}}\ \wedge\ B \subseteq A\ \Longrightarrow\ \mu_{\mathcal{M}}\ A - \mu_{\mathcal{M}}\ B = \mu_{\mathcal{M}}\ (A \setminus B)$

Based on countable additivity we show the limit of increasing or decreasing sequences of measure values. On the extended reals these limits are equal to the supremum respective infimum.

lemma LIM-μ-INCSEQ, SUP-μ-INCSEQ:
$F \in \mathbb{N} \to \mathcal{A}_{\mathcal{M}}\ \wedge\ \mathit{incseq}\ F\ \Longrightarrow$
$\mu_{\mathcal{M}}\ (F\ i) \xrightarrow[i \to \infty]{} \mu_{\mathcal{M}}\ (\bigcup_i F\ i),\ \ (\sup_i \mu_{\mathcal{M}}\ (F\ i)) = \mu_{\mathcal{M}}\ (\bigcup_i F\ i)$

lemma LIM-μ-DECSEQ, INF-μ-DECSEQ:
$F \in \mathbb{N} \to \mathcal{A}_{\mathcal{M}}\ \wedge\ \mathit{decseq}\ F\ \wedge\ (\forall i.\ \mu_{\mathcal{M}}\ (F\ i) < \infty)\ \Longrightarrow$
$\mu_{\mathcal{M}}\ (F\ i) \xrightarrow[i \to \infty]{} \mu_{\mathcal{M}}\ (\bigcup_i F\ i),\ \ (\inf_i \mu_{\mathcal{M}}\ (F\ i)) = \mu_{\mathcal{M}}\ (\bigcap_i F\ i)$

The decreasing case requires that the elements in the sequence have a finite measure. A famous counterexample is $F\ i = \{i <..\}$ on the counting measure on the natural numbers: $\bigcap_i F\ i = \emptyset$ but the limit of the measures is still infinity.

For estimations it is often helpful to have inequalities between the sums of measure values and the measure of their union. For these cases we do not assume disjoint sets.

lemma μ-SUBADDITIVE:
$A, B \in \mathcal{A}_{\mathcal{M}}\ \Longrightarrow\ \mu_{\mathcal{M}}\ (A \cup B) \leq \mu_{\mathcal{M}}\ A + \mu_{\mathcal{M}}\ B$

lemma μ-SUBADDITIVE-FINITE:
$F \in I \to \mathcal{A}_{\mathcal{M}}\ \wedge\ \mathit{finite}\ I\ \Longrightarrow\ \mu_{\mathcal{M}}\ (\bigcup_{i \in I} F\ i) \leq (\sum_{i \in I} \mu_{\mathcal{M}}\ (F\ i))$

lemma μ-SUBADDITIVE-COUNTABLY:
$F \in \mathbb{N} \to \mathcal{A}_{\mathcal{M}}\ \Longrightarrow\ \mu_{\mathcal{M}}\ (\bigcup_i F\ i) \leq (\sum_i \mu_{\mathcal{M}}\ (F\ i))$

[5] Often one needs to prove measurability anyway, but this avoids to show measurability twice.

All rules in this section were about the extended measure, i.e. $\mu_{\mathcal{M}} :: \alpha\ \mathit{set} \to \overline{\mathbb{R}}$. The extended real numbers have nice limit properties, as the infimum and supremum always exists. However, their arithmetic properties often require a case distinction on the finiteness of the involved extended real numbers. For this we introduce the finite measure $\mu^f_{\mathcal{M}}$:

$$\begin{array}{lcl} \mu^f_{\square} & :: & \alpha\ \mathit{measure} \to \alpha\ \mathit{set} \to \mathbb{R} \\ \mu^f_{\mathcal{M}}\ A & = & (\mu_{\mathcal{M}}\ A)_{\mathbb{R}} \end{array}$$

The conversion function $(\cdot)_{\mathbb{R}}$ is the inverse of $(\cdot)_{\overline{\mathbb{R}}}$ for finite values. So we can directly represent the extended measure as a finite measure:

lemma μ-EQ-EREAL-μ^f: $\mu_{\mathcal{M}}\ A < \infty \implies \mu_{\mathcal{M}}\ A = (\mu^f_{\mathcal{M}}\ A)_{\overline{\mathbb{R}}}$

With this property we easily transfer all the rules about $\mu_{\mathcal{M}}$ to rules about $\mu^f_{\mathcal{M}}$ assuming that all involved measures are finite.

2.3.1 Finite and σ-Finite Measures

The generic measure space does not restrict the measure of the entire space. We explicitly allow infinite measure. As mentioned, often properties can only be shown when the measure of $\Omega_{\mathcal{M}}$ is finite. But sometimes we can show that the property holds when the measure is *σ-finite*, i.e. has a countable cover of sets with finite measure. In this case the property is first shown on each of these finite sub-spaces, and then by a limit argument on the entire space.

locale *σ-finite-measure* = **fixes** $\mathcal{M} :: \alpha\ \mathit{measure}$
  assumes $\exists C \in \mathbb{N} \to \mathcal{A}_{\mathcal{M}}.\ (\bigcup_i C\ i) = \Omega_{\mathcal{M}} \wedge (\forall i.\ \mu_{\mathcal{M}}\ (C\ i) < \infty)$

Prominent σ-finite measures are the Lebesgue measure, the counting space on the natural numbers, and each finite measure.

We introduce finite measures as a specialization of σ-finite measures:

locale *finite-measure* = *σ-finite-measure* +
  assumes $\mu_{\mathcal{M}}\ \Omega_{\mathcal{M}} < \infty$

Here we know that each set has a finite measure:

lemma μ-FINITE: *finite-measure* $\mathcal{M} \implies \mu_{\mathcal{M}}\ A < \infty$

lemma μ-EQ-μ^f: *finite-measure* $\mathcal{M} \implies \mu_{\mathcal{M}}\ A = \left(\mu^f_{\mathcal{M}}\ A\right)_{\overline{\mathbb{R}}}$

With these lemmas we copy all lemmas about $\mu^f_{\mathcal{M}}$ with finiteness assumptions into the *finite-measure*-locale, removing the finiteness assumptions.

2.3.2 Uniqueness of Measures

We want to reduce the equality of two measures to the equality on the generator of their σ-algebra. This is possible if the measure has an $\cap$-stable, σ-finite generator:

theorem MEASURE-EQI-GENERATOR-EQ:
  $\cap$-*stable* $\mathcal{G}\ \wedge\ \mathcal{G} \subseteq \mathcal{P}(\Omega)\ \wedge$
  $C \in \mathbb{N} \to \mathcal{G}\ \wedge\ (\bigcup_i C\ i) = \Omega\ \wedge\ (\forall i.\ \mu_{\mathcal{M}}\ (C\ i) < \infty)\ \wedge$
  $(\forall G \in \mathcal{G}.\ \mu_{\mathcal{M}}\ G = \mu_{\mathcal{N}}\ G)\ \wedge\ \mathcal{A}_{\mathcal{N}} = \mathcal{A}_{\mathcal{M}} = \sigma\text{-}\mathit{sets}\ \Omega\ \mathcal{G} \implies$
  $\mathcal{M} = \mathcal{N}$

We first show that for each i the measures $\mathcal{M}$ and $\mathcal{N}$ restricted to $C\ i$ are equal. We prove equality by induction with Corollary σ-SETS-INDUCT-DISJOINT. The generator is $\cap$-stable and equality of two finite measures is closed under disjoint union and complement. Each measurable set A is then expressed as the union of all $C\ i \cap A$, for which the two measure are equal.

2.3.3 Null Sets and AE-Quantifier

Null sets are the measurable sets in a measure space with zero measure:

$$\begin{array}{lcl} \mathit{null\text{-}sets}_{\square} & :: & \alpha\ \mathit{measure} \to \alpha\ \mathit{set\ set} \\ \mathit{null\text{-}sets}_{\mathcal{M}} & = & \{A \in \mathcal{A}_{\mathcal{M}} \mid \mu_{\mathcal{M}}\ A = 0\} \end{array}$$

We know that the empty set is a null set, and the union and intersection of null sets is again a null set. So, they form a ring:

lemma RING-OF-SETS-NULL-SETS: $\mathit{ring\text{-}of\text{-}sets}\ \Omega_{\mathcal{M}}\ \mathit{null\text{-}sets}_{\mathcal{M}}$

The null sets are closed under countable union and also closed under intersection or set difference with arbitrary measurable sets:

lemma NULL-SETS-UN:
$N \in \mathcal{U}_{i::\mathcal{C}} \to \mathit{null\text{-}sets}_{\mathcal{M}} \implies (\bigcup_i N\ i) \in \mathit{null\text{-}sets}_{\mathcal{M}}$

lemma NULL-SETS-INT1, -INT2, -DIFF:
$N \in \mathit{null\text{-}sets}_{\mathcal{M}} \wedge A \in \mathcal{A}_{\mathcal{M}} \implies N \cap A, A \cap N, N - A \in \mathit{null\text{-}sets}_{\mathcal{M}}$

Also, the measure $\mu_{\mathcal{M}}$ is invariant under adding or removing null sets:

lemma μ-DIFF-NULL-SET, -UN-NULL-SET:
$N \in \mathit{null\text{-}sets}_{\mathcal{M}} \wedge A \in \mathcal{A}_{\mathcal{M}} \implies$
$\mu_{\mathcal{M}}\ (A - N) = \mu_{\mathcal{M}}\ A,$
$\mu_{\mathcal{M}}\ (A \cup N) = \mu_{\mathcal{M}}\ A$

An important class of predicates in measure theory are predicates which are only false on a null set. The *almost everywhere* quantifier (AE-quantifier) is true on such a predicate. We use filters to introduce this quantifier.

Filters are a concept originally introduced to represent limits [14], like the limit of a sequence or of a function at an input value. However, they can also be used to introduced generalized quantifiers [42]. Usually, filters are represented as sets of sets, but in Isabelle/HOL they get their own type $\alpha\ \mathit{filter}$, similar to the type of measures. Here, $\mathit{Abs}_{\mathit{filter}}$ is the abstraction morphism, and *eventually* is the representation morphism, i.e. $\mathit{eventually}\ P\ (\mathit{Abs}_{\mathit{filter}}\ F) = F\ P$ holds for each filter F.

Now we define a filter for the AE-quantifier:

$$\begin{array}{lcl} \mathit{AE\text{-}filter}_{\square} & :: & \alpha\ \mathit{measure} \to \alpha\ \mathit{filter} \\ \mathit{AE\text{-}filter}_{\mathcal{M}} & = & \mathit{Abs}_{\mathit{filter}}\ (\lambda P.\ \exists N \in \mathit{null\text{-}sets}_{\mathcal{M}}.\ \{x \mid \neg P\ x\} \subseteq N) \end{array}$$

In textbooks the AE-quantifier is often written without an explicitly quantified variable but rather with an appended "a.e.". We use a syntax with an explicit binder:

$$\mathrm{AE}_{\mathcal{M}}\ x.\ P\ x \Leftrightarrow \mathit{eventually}\ (\lambda x.\ P\ x)\ \mathit{AE\text{-}filter}_{\mathcal{M}}$$

By using *eventually* we profit from the rules about filters, like modus ponens, and the commutativity of filters with logic connectives. More importantly, we obtain for free the automation setup and the proof method for filters. Before we can apply this AE-quantifier to measure theoretic properties, we need to show that it actually is a filter. This needs to be done when we equate the quantifier with its defining equation:

lemma AE-IFF:
$(\mathrm{AE}_{\mathcal{M}}\, x.\ P\ x) \Leftrightarrow (\exists N \in \mathit{null\text{-}sets}_{\mathcal{M}}.\ \{x \mid \neg P\ x\} \subseteq N)$

With $\mathit{AE\text{-}filter}_{\mathcal{M}}$ we actually defined a filter: the predicate which is always true is accepted, when we add values to an accepted predicate it is accepted, and the intersection of two accepted predicates is accepted. For this to work we need the completion of the null sets. Hence we accept all subsets of a null set. Our solution is different to the definition of almost everywhere in [39] and [57], where a predicate P only then holds almost everywhere when it is measurable.

As the AE-quantifier is a filter, we know that it commutes with finitely bounded universal quantification. Moreover, as null sets are closed under countable union, the AE-quantifier also commutes with countable universal quantification:

lemma AE-ALL-COUNTABLE:
$(\mathrm{AE}_{\mathcal{M}}\, x.\ \forall i :: \iota :: \mathcal{C}.\ P\ i\ x) \Leftrightarrow (\forall i.\ \mathrm{AE}_{\mathcal{M}}\, x.\ P\ i\ x)$

We can write more flexible congruence rules for the equality and monotony of measure functions in terms of the the AE-quantifier:

lemma AE-μ:
$(\mathrm{AE}_{\mathcal{M}}\, x.\ x \in A) \ \wedge\ A \in \mathcal{A}_{\mathcal{M}} \implies \mu_{\mathcal{M}}\, A = \mu_{\mathcal{M}}\, \Omega_{\mathcal{M}}$

lemma AE-μ-MONO:
$(\mathrm{AE}_{\mathcal{M}}\, x.\ x \in A \implies x \in B) \ \wedge\ B \in \mathcal{A}_{\mathcal{M}} \implies \mu_{\mathcal{M}}\, A \leq \mu_{\mathcal{M}}\, B$

lemma AE-μ-EQ:
$(\mathrm{AE}_{\mathcal{M}}\, x.\ x \in A \Leftrightarrow x \in B) \ \wedge\ A, B \in \mathcal{A}_{\mathcal{M}} \implies \mu_{\mathcal{M}}\, A = \mu_{\mathcal{M}}\, B$

As an example for the application of the AE-quantifier, we use that singletons are null sets in the Lebesgue measure, i.e. $\lambda_{\mathbb{R}}\ \{y\} = 0$ and hence $\mathrm{AE}_{\lambda_{\mathbb{R}}}\, x.\ x \neq y$. Now with these rules we show $\lambda_{\mathbb{R}}\ \{a\ ..\ b\} = \lambda_{\mathbb{R}}\ \{a <..< b\}$, simply applying Lemma AE-μ-EQ, since $\mathrm{AE}_{\lambda_{\mathbb{R}}}\, x.\ x \neq a \wedge x \neq b \implies \big(x \in \{a\ ..\ b\} \Leftrightarrow x \in \{a <..< b\}\big)$.

2.4 Lebesgue Integral

The Lebesgue integral is a generalization of finite and countable sums, yielding them as particular cases when the measure space is discrete. On the Lebesgue measure, it is a generalization of the Riemann integral. In comparison to the Riemann integral, it fulfills monotone convergence: for each rising sequence of integrable functions, the integral of their suprema exists and equals the supremum of the integrals. This makes the Lebesgue integral useful not only for calculus, but

also for measure theory itself: we will use it to add a density to a measure and to define products of measure spaces.

We will introduce the Lebesgue integral in the usual manner, by first introducing integration for step functions, then for positive functions and finally for arbitrary real-valued functions.

Notation: *Textbooks usually write $\int f\, x\, d\mu(x)$, where μ as a partial function also carries the σ-algebra of measurable sets. In HOL the measure function $\mu_{\mathcal{M}}$ is not enough. So we use the entire measure space $\mathcal{M}$ in our notation, and optionally bind the variable x following the integral symbol: $\int x.\, f\, x\, d\mathcal{M}$. If no variable is needed we write $\int f\, d\mathcal{M}$. The same holds for the simple integral $\int^S f\, d\mathcal{M}$ and the positive integral $\int^P f\, d\mathcal{M}$.*

2.4.1 Simple Functions

The definition of the Lebesgue integral requires the concept of *simple function*. A simple function is a Borel-measurable step function (i.e. its range is a finite set), or for $\overline{\mathbb{R}}$-functions equivalently: a step function where each inverse image is measurable. The second formulation has the advantage that the definition does not require the notion of Borel sets and is thus more general, as it allows arbitrary ranges. The predicate *simple-fn* is defined as follows:

$$\begin{array}{lcl} \textit{simple-fn} & :: & \alpha\ \textit{measure} \to (\alpha \to \beta) \to \mathbb{B} \\ \textit{simple-fn}\ \mathcal{M}\ f & \Leftrightarrow & \textit{finite}\ f[\Omega_{\mathcal{M}}] \wedge \forall x \in f[\Omega_{\mathcal{M}}].\ f^{-1}[\{x\}] \cap \Omega_{\mathcal{M}} \in \mathcal{A}_{\mathcal{M}} \end{array}$$

While we use this definition only for functions $f :: \alpha \to \overline{\mathbb{R}}$, this is a nice characterization for finite random variables in probability theory. The simple functions have also nice closure properties. Each composition where the input goes through a simple function, is again a simple function:

lemma SIMPLE-FN-COMPOSE1:
$\textit{simple-fn}\ \mathcal{M}\ f \implies \textit{simple-fn}\ \mathcal{M}\ (\lambda x.\ g\ (f\ x))$

lemma SIMPLE-FN-COMPOSE2:
$\textit{simple-fn}\ \mathcal{M}\ f \wedge \textit{simple-fn}\ \mathcal{M}\ g \implies \textit{simple-fn}\ \mathcal{M}\ (\lambda x.\ h\ (f\ x)\ (g\ x))$

Alternatively, we can express simple functions as Borel-measurable functions with a finite range. From this we immediately show that the constant function and the indicator function are simple functions:

lemma SIMPLE-FN-EQ-$\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE:
$\textit{simple-fn}\ \mathcal{M}\ f \Leftrightarrow \textit{finite}\ f[\Omega_{\mathcal{M}}] \wedge f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}$

lemma SIMPLE-FN-CONST: $\textit{simple-fn}\ \mathcal{M}\ (\lambda x.\ c)$

lemma SIMPLE-FN-χ: $A \in \mathcal{A}_{\mathcal{M}} \implies \textit{simple-fn}\ \mathcal{M}\ (\chi\ A)$

A simple function obviously is Borel-measurable, but can we express Borel-measurable functions in terms of simple functions? The answer is: yes, we can express each positive Borel-measurable function as the supremum of a sequence

of simple functions:

theorem $\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE-SIMPLE-FNS:
$u \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\exists F.\ \Big(\forall x.\ \sup_i F\ i\ x = \max\ 0\ (u\ x)\Big) \wedge$
$\Big(\forall i.\ \textit{simple-fn}\ \mathcal{M}\ (F\ i)\ \wedge\ \forall x.\ 0 \le F\ i\ x < \infty\Big)$

Instead of assuming a nonnegative function u, we return a sequence converging to the positive half of u. The sequence $F\ i$ is constructed by first stepping up to the integer component $\lfloor u\ x \rfloor$ and then, when this is reached, by approximating $\lfloor u\ x \cdot 2^i \rfloor \cdot 2^{-i}$. This theorem is helpful in proofs to replace a Borel-measurable function into a sequence of simple functions and then using monotone convergence.

A similar approach is used to define the Lebesgue integral. We begin by defining the integral on simple functions. We know that, as the range of f is finite, it is also representable as a sum:

lemma SIMPLE-FN-χ-REPRESENTATION:
$\textit{simple-fn}\ \mathcal{M}\ f \implies \forall x \in \Omega_{\mathcal{M}}.\ f\ x = \sum_{y \in f[\Omega_{\mathcal{M}}]} y \cdot \chi\ (f^{-1}[\{y\}] \cap \Omega_{\mathcal{M}})\ x$

This already suggests the definition of the *integral* $\int^S$ *of a simple function* f with respect to the measure space $\mathcal{M}$:

$$\begin{aligned}\int^S \square\, d\square &:: \alpha\ \textit{measure} \to \Big(\alpha \to \overline{\mathbb{R}}\Big) \to \overline{\mathbb{R}}\\ \int^S f\, d\mathcal{M} &= \sum_{y \in f[\Omega_{\mathcal{M}}]} y \cdot \mu_{\mathcal{M}}(f^{-1}[\{y\}] \cap \Omega_{\mathcal{M}})\end{aligned}$$

When a simple functions are a.e.-equal (a.e.-less than or equal) to another simple function then the two integrals are equal (less than or equal):

lemma $\int^S$-MONO-AE:
$\textit{simple-fn}\ \mathcal{M}\ f\ \wedge\ \textit{simple-fn}\ \mathcal{M}\ g\ \wedge\ \mathrm{AE}_{\mathcal{M}}\ x.\ f\ x \le g\ x \implies$
$\int^S f\, d\mathcal{M} \le \int^S g\, d\mathcal{M}$

lemma $\int^S$-CONG-AE:
$\textit{simple-fn}\ \mathcal{M}\ f\ \wedge\ \textit{simple-fn}\ \mathcal{M}\ g\ \wedge\ \mathrm{AE}_{\mathcal{M}}\ x.\ f\ x = g\ x \implies$
$\int^S f\, d\mathcal{M} = \int^S g\, d\mathcal{M}$

2.4.2 Integral of Positive $\overline{\mathbb{R}}$-Functions

To state the definition of the *positive integral* of nonnegative functions $f :: \alpha \to \overline{\mathbb{R}}$, simple functions have to be used as approximations of f from below. Then the integral is defined as the supremum of all the simple integrals of the approximations.

$$\begin{aligned}\int^P \square\, d\square &:: \alpha\ \textit{measure} \to \Big(\alpha \to \overline{\mathbb{R}}\Big) \to \overline{\mathbb{R}}\\ \int^P f\, d\mathcal{M} &= \sup\left\{\int^S g\, d\mathcal{M} \;\middle|\; \Big(\forall x.\ g\ x \le \max\ 0\ (f\ x)\Big) \wedge \textit{simple-fn}\ \mathcal{M}\ g\right\}\end{aligned}$$

The function $\lambda x.\ \max\ 0\ (f\ x)$ is the nonnegative part of f, i.e. it is zero when f is negative, otherwise it is equal to f. From the monotony of the simple integral follows the monotony of the positive integral, i.e. it is equal (less than or equal) when the integrands are a.e.-equal (a.e.-less than or equal).

lemma $\int^P$-MONO-AE:
$\mathrm{AE}_{\mathcal{M}}\ x.\ f\ x \leq g\ x \implies \int^P f\ d\mathcal{M} \leq \int^P g\ d\mathcal{M}$

lemma $\int^P$-CONG-AE:
$\mathrm{AE}_{\mathcal{M}}\ x.\ f\ x = g\ x \implies \int^P f\ d\mathcal{M} = \int^P g\ d\mathcal{M}$

From the definition of the integral it immediately follows that for simple functions the positive and simple integral are equal:

lemma $\int^P$-EQ-$\int^S$-AE:
simple-fn $\mathcal{M}\ f\ \wedge\ (\mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq f\ x) \implies \int^P f\ d\mathcal{M} = \int^S f\ d\mathcal{M}$

From the definition of the simple integral we know immediately that the indicator function over A maps to the measure of A:

lemma $\int^P$-χ:
$A \in \mathcal{A}_{\mathcal{M}} \implies \int^P \chi\ A\ d\mathcal{M} = \mu_{\mathcal{M}}\ A$

One way of constructing proofs about integrals of Borel-measurable functions $u :: \alpha \to \overline{\mathbb{R}}$ is: first prove the desired property about finite simple functions, then prove that the property is preserved under the pointwise monotone limit of functions. For this to work, we can use Theorem $\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE-SIMPLE-FNS. To use this with the Lebesgue integral, there is a compatibility theorem, called the monotone convergence theorem, which allows commuting the supremum operator and the positive integral:

theorem $\int^P$-MONOTONE-CONVERGENCE:
$(\forall i.\ f\ i \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}\ \wedge\ \mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq f\ i\ x \wedge f\ i\ x \leq f\ (i+1)\ x) \implies$
$\int^P (\sup_i f\ i)\ d\mathcal{M} = \sup_i \int^P f\ i\ d\mathcal{M}$

The monotone convergence theorem is now used to prove linearity of the positive integral. This is done in the aforementioned way: we obtain the sequences of simple functions to f, g, and $f+g$. By monotone convergence we replace the positive integrals with the suprema of the simple integrals of these sequences. Now it is simply a matter of linearity of the simple integral and of the suprema, that both sides are equal:

lemma $\int^P$-ADD:
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}\ \wedge\ g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge$
$(\mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq f\ x)\ \wedge\ (\mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq g\ x) \implies$
$\int^P x.\ f\ x + g\ x\ d\mathcal{M} = \int^P f\ d\mathcal{M} + \int^P g\ d\mathcal{M}$

lemma $\int^P$-CMULT:
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge 0 \leq c \implies \int^P x.\ c \cdot f\ x\ d\mathcal{M} = c \cdot \int^P f\ d\mathcal{M}$

corollary $\int^P$-SETSUM:
$(\forall i \in I.\ f\ i \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}) \wedge (\forall i \in I.\ \mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq f\ i\ x) \implies$
$\int^P x.\ \sum_{i \in I} f\ i\ x\ d\mathcal{M} = \sum_{i \in I} \int^P f\ i\ d\mathcal{M}$

corollary $\int^P$-SUMINF:
$(\forall i.\ f\ i \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}) \wedge (\forall i.\ \mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq f\ i\ x) \implies$
$\int^P x.\ \sum_{i} f\ i\ x\ d\mathcal{M} = \sum_{i} \int^P f\ i\ d\mathcal{M}$

The last two corollaries immediately follow from Lemma $\int^P$-ADD and monotone convergence. Note that the infinite sum in Corollary $\int^P$-SUMINF is the sum on extended reals, hence it is always defined for sums of positive values.

From the linearity of the positive integral, we derive the Markov inequality:

lemma $\int^P$-MARKOV-INEQUALITY:
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge \left(\mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq f\ x\right) \wedge A \in \mathcal{A}_{\mathcal{M}} \wedge 0 \leq c \implies$
$\mu_{\mathcal{M}} \left\{ x \in A \;\middle|\; 1 \leq c \cdot f\ x \right\} \leq c \cdot \int^P x.\ f\ x \cdot \chi\ A\ x\ d\mathcal{M}$

Utilizing this inequality we find an easy characterization to show that the positive integral of a function is zero:

lemma $\int^P$-0-IFF-AE:
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \implies \left(\int^P f\ d\mathcal{M} = 0\right) \Leftrightarrow \left(\mathrm{AE}_{\mathcal{M}}\ x.\ f\ x \leq 0\right)$

On the right side, we can only show that u is a.e.-nonpositive, as the positive integral maps negative values to zero.

2.4.3 Induction on Borel-Measurable Functions

We know that the Lebesgue integral is linear (Lemmas $\int^P$-ADD and $\int^P$-CMULT), admits the indicator function (Lemma $\int^P$-χ) and is monotone convergent (Lemma $\int^P$-MONOTONE-CONVERGENCE). Is the integral uniquely defined by these rules? Yes, at least when the integrand is a Borel-measurable function: each Borel-measurable function can be represented as a sequence of simple functions (Theorem $\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE-SIMPLE-FNS) and simple functions can be represented as sums of rectangle functions (Lemma SIMPLE-FN-χ-REPRESENTATION).

In Chapter 3 we will often show an alternative representation of the Lebesgue integral for a specific measure space. For this it is helpful to have an induction rule on positive Borel-measurable functions. We provide such an induction rule constructing Borel-measurable functions out of indicator functions on measurable sets, linear multiplication, addition, and the limit of an increasing sequence.

An alternative way to construct Borel-measurable functions would be just simple functions and the limit of an increasing sequence. However, this would require a second induction principle for simple functions. Also the proofs would require two steps, first prove the statement for simple functions and then for Borel-measurable functions.

We now give the induction rule:

corollary $\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE-INDUCT:

$$
\begin{aligned}
&f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge \Big(\forall x.\ 0 \le f\ x\Big) \wedge \\
&\Big(\forall f, g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}.\ (\forall x \in \Omega_{\mathcal{M}}.\ f\ x = g\ x) \wedge P\ f \implies P\ g\Big) \wedge \\
&\Big(\forall A \in \mathcal{A}_{\mathcal{M}}.\ P\ (\chi\ A)\Big) \wedge \\
&\Big(\forall f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}, c \ge 0.\ (\forall x.\ 0 \le f\ x) \wedge P\ f \implies P\ (\lambda x.\ c \cdot f\ x)\Big) \wedge \\
&\Big(\forall f, g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}. \\
&\quad (\forall x.\ 0 \le f\ x) \wedge P\ f \wedge (\forall x.\ 0 \le g\ x) \wedge P\ g \implies P\ (\lambda x.\ f\ x + g\ x)\Big) \wedge \\
&\Big(\forall F \in \mathbb{N} \to \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}. \\
&\quad \big(\forall i.\ (\forall x.\ 0 \le F\ i\ x \le (F\ (i+1)\ x)) \wedge P\ (F\ i)\big) \implies P\ (\sup_i F\ i)\Big) \implies \\
&P\ f
\end{aligned}
$$

The third conjunct of this induction rule is not used to construct the Borel-measurable function, but it shows that the predicate P does not care about function values outside the space $\Omega_{\mathcal{M}}$. Note also that the linearity is split into two cases for $f\ x + g\ x$ and $c \cdot f\ x$ instead of merging them into one $c \cdot f\ x + g\ x$. This is done deliberately as automation works better when only one operation is involved. This rule enables us to show equality between different representations of the Lebesgue integral, provided we have linearity and monotone convergence for both representations.

2.4.4 Integral of $\mathbb{R}$-Functions

The positive integral on $\overline{\mathbb{R}}$ has the ideal properties for an integral: linearity and monotone convergence. However, it cannot handle functions with negative values. For this we define integration for functions $f :: \alpha \to \mathbb{R}$ as the difference between the integral of the positive and the negative part of f. For this to be defined sensibly, we restrict integrable functions to Borel-measurable functions where the positive integrals of the positive and the negative parts are finite:

$$
\begin{aligned}
&\textit{integrable} && :: && \alpha\ \textit{measure} \to \Big(\alpha \to \mathbb{R}\Big) \to \mathbb{B} \\
&\textit{integrable}\ \mathcal{M}\ f && \Leftrightarrow && \Big(f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \wedge \\
& && && \quad \int^P x.\ f\ x\ d\mathcal{M} < \infty \wedge \int^P x.\ -f\ x\ d\mathcal{M} < \infty\Big) \\
&\int \Box d\Box && :: && \alpha\ \textit{measure} \to \Big(\alpha \to \mathbb{R}\Big) \to \mathbb{R} \\
&\int f\ d\mathcal{M} && = && \Big(\int^P x.\ f\ x\ d\mathcal{M}\Big)_{\mathbb{R}} - \Big(\int^P x.\ -f\ x\ d\mathcal{M}\Big)_{\mathbb{R}}
\end{aligned}
$$

(Note that explicit type conversions from $\mathbb{R}$ to $\overline{\mathbb{R}}$ have been omitted for the sake of readability.)

Many proofs of properties about the integral follow the scheme of the definitions and first establish the desired property for $\int^S$, then for $\int^P$, and eventually for $\int$. Congruence rules for a.e.-equality and a.e.-monotony follow this way:

lemma $\int$-CONG-AE:
$\mathrm{AE}_{\mathcal{M}}\, x.\ f\ x = g\ x \implies \int f\ d\mathcal{M} = \int g\ d\mathcal{M}$

lemma *integrable*-CONG-AE:
$f \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \wedge g \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \wedge \mathrm{AE}_{\mathcal{M}}\, x.\ f\ x = g\ x \implies$
$\mathit{integrable}\ \mathcal{M}\ f \Leftrightarrow \mathit{integrable}\ \mathcal{M}\ g$

lemma $\int$-MONO-AE:
$\mathit{integrable}\ \mathcal{M}\ f \wedge \mathit{integrable}\ \mathcal{M}\ g \wedge \mathrm{AE}_{\mathcal{M}}\, x.\ f\ x \le g\ x \implies$
$\int f\ d\mathcal{M} \le \int g\ d\mathcal{M}$

The integral equality follows directly from the equality of the positive integral, hence no integrability assumption is needed. For integrability the almost everywhere assumption does not imply measurability of the functions itself, hence there it is required. And for monotony of the integral we need to at least restrict g to a finite positive part and f to a finite negative part.

Linearity also follows directly from the properties of the positive integral. Borel-measurable is replaced by integrable:

lemma $\int$-ADD:
$\mathit{integrable}\ \mathcal{M}\ f \wedge \mathit{integrable}\ \mathcal{M}\ g \implies$
$\mathit{integrable}\ \mathcal{M}\ (\lambda x.\ f\ x + g\ x), \quad \int x.\ f\ x + g\ x\ d\mathcal{M} = \int f\ d\mathcal{M} + \int g\ d\mathcal{M}$

lemma $\int$-SETSUM:
$(\forall i \in I.\ \mathit{integrable}\ \mathcal{M}\ (f\ i)) \implies$
$\mathit{integrable}\ \mathcal{M}\ (\lambda x.\ \sum_{i \in I} f\ i\ x), \quad \int x.\ \sum_{i \in I} f\ i\ x\ d\mathcal{M} = \sum_{i \in I} \int f\ i\ d\mathcal{M}$

lemma $\int$-MINUS:
$\mathit{integrable}\ \mathcal{M}\ f \wedge \mathit{integrable}\ \mathcal{M}\ g \implies$
$\mathit{integrable}\ \mathcal{M}\ (\lambda x.\ f\ x - g\ x), \quad \int x.\ f\ x - g\ x\ d\mathcal{M} = \int f\ d\mathcal{M} - \int g\ d\mathcal{M}$

lemma $\int$-CMULT:
$\mathit{integrable}\ \mathcal{M}\ f \implies$
$\mathit{integrable}\ \mathcal{M}\ (\lambda x.\ a \cdot f\ x), \quad \int x.\ a \cdot f\ x\ d\mathcal{M} = a \cdot \int f\ d\mathcal{M}$

Monotone convergence follows also from the monotone convergence of the positive integral. Instead of using the supremum of a sequence we assume a sequence of functions converging from below. Also the integrals of these functions

need to converge:

theorem $\int$-MONOTONE-CONVERGENCE:
$(\forall i.\ \textit{integrable}\ \mathcal{M}\ (f\ i)) \wedge \left(\int x.\ f\ i\ x\right) \xrightarrow[i\to\infty]{} I \wedge u \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \wedge$
$(\mathrm{AE}_{\mathcal{M}}\ x.\ \textit{incseq}\ (\lambda i.\ f\ i\ x)) \wedge (\mathrm{AE}_{\mathcal{M}}\ x.\ f\ i\ x \xrightarrow[i\to\infty]{} u\ x) \Longrightarrow$
$\textit{integrable}\ \mathcal{M}\ u, \quad \int u\ d\mathcal{M} = I$

Instead of requiring an increasing sequence, we provide a relaxed version where the function sequence is absolutely bounded by an integrable function, i.e. dominated convergence. It can be used when the monotony of the function sequence does not hold.

theorem $\int$-DOMINATED-CONVERGENCE:
$(\forall i.\ \textit{integrable}\ \mathcal{M}\ (u\ i)) \wedge \textit{integrable}\ \mathcal{M}\ w \wedge u' \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \wedge$
$(\forall i.\ \mathrm{AE}_{\mathcal{M}}\ x.\ |u\ i\ x| \le w\ x) \wedge \mathrm{AE}_{\mathcal{M}}\ x.\ u\ i\ x \xrightarrow[i\to\infty]{} u'\ x \Longrightarrow$
$\textit{integrable}\ \mathcal{M}\ u', \quad \int u\ i\ d\mathcal{M} \xrightarrow[i\to\infty]{} \int u'\ d\mathcal{M}$

One advantage the Lebesgue integral gains by splitting integrability and measurability into two different concepts, is that we only need to find an upper bounding function to show integrability. This does not require dominated convergence, but we can easily show integrability by bounding a measurable function:

lemma *integrable*-BOUND:
$\textit{integrable}\ \mathcal{M}\ f \wedge g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \wedge (\mathrm{AE}_{\mathcal{M}}\ x.\ |g\ x| \le f\ x) \Longrightarrow$
$\textit{integrable}\ \mathcal{M}\ g$

It follows also directly that integrability is invariant under the absolute value function, as long as the inner integrand is measurable. From this we can then deduce the integrability of the minimum and maximum of functions.

lemma *integrable*-ABS-IFF:
$f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \Longrightarrow \textit{integrable}\ \mathcal{M}\ (\lambda x.\ |f\ x|) \Leftrightarrow \textit{integrable}\ \mathcal{M}\ f$

lemma *integrable*-MIN, -MAX:
$\textit{integrable}\ \mathcal{M}\ f \wedge \textit{integrable}\ \mathcal{M}\ g \Longrightarrow$
$\textit{integrable}\ \mathcal{M}\ (\lambda x.\ \min\ (f\ x)\ (g\ x)), \quad \textit{integrable}\ \mathcal{M}\ (\lambda x.\ \max\ (f\ x)\ (g\ x))$

With Lemma $\int$-MONO-AE we cannot show strict inequality. However, on a finite measure we know that the integral of a function is strictly monotone, if the functions are strictly monotone on a set with nonzero measure:

lemma $\int$-LESS-AE:
$\textit{finite-measure}\ \mathcal{M} \wedge \textit{integrable}\ \mathcal{M}\ f \wedge \textit{integrable}\ \mathcal{M}\ g \wedge A \in \mathcal{A}_{\mathcal{M}} \wedge$
$\mu_{\mathcal{M}}\ A \neq 0 \wedge (\mathrm{AE}_{\mathcal{M}}\ x \in A.\ f\ x \neq g\ x) \wedge (\mathrm{AE}_{\mathcal{M}}\ x.\ f\ x \le g\ x) \Longrightarrow$
$\int f\ d\mathcal{M} < \int g\ d\mathcal{M}$

Further formalized properties of Lebesgue integration described later on in this thesis, when we have developed the necessary mathematical machinery: it is used to construct weighted measures in Section 3.3, to introduce Fubini's theorem after product measures are introduced in Section 3.4.2, and, to relate it to the gauge integral after the Lebesgue measure is introduced in Section 3.5.

Chapter 3

Concrete Measures

The concept of measure spaces is generic and can be instantiated with a couple of different structures. Unfortunately, to show that a σ-algebra together with a premeasure raises a measure space can be complicated. We provide a couple of basic measure spaces which can be used to construct more complicated ones.

The counting measure ε assigns a set A its cardinality $\varepsilon\ A = |A|$ if A is finite and otherwise infinity: $\varepsilon\ A = \infty$. This measure maps concepts from measure theory to discrete concepts: measure equals cardinality and Lebesgue integration equals summation. This allows us to translate theorems about abstract measure spaces into theorems about discrete concepts.

The push-forward measure μ_X uses the $\mathcal{N}$-measurable function X to assign a measure to $\mathcal{N}$: $\mu_X\ A = \mu\ \{x \mid X\ x \in A\}$. For example X computes the time until a Markov chain τ terminates. Then the probability that the Markov chain terminates in n steps is $\mu_X\ \{..\ n\} = \tau\ \{\omega \mid X\ \omega \leq n\}$. So it defines a measure on the natural numbers by associating to each number n the probability of the traces terminating at time n.

The density measure $\mu_{\int f}$ weights the measure μ with the density f: $\mu_{\int f}\ A = \int_A f\ d\mu$. Here f is an extended real valued function on Ω. The intuitive understanding is that f weighs to each point in Ω. When we use the counting measure this is exactly what happens. Together with the push-forward measure density measures are often used to describe the distribution of a random variable. For example, X is an exponentially distributed random variable when

$$\mu_X\ A = \int_A x.\ \lambda \cdot \exp^{-\lambda \cdot x}\ d\lambda = \mu_{\int_x \lambda \exp^{-\lambda x}}\ A\ \ .$$

The binary product of measures $\pi = \mu_1 \otimes \mu_2$ maps Cartesian set products to the multiplication of their measure values: $\pi\ (A_1 \times A_2) = \mu_1\ A_1 \cdot \mu_2\ A_2$. With the product measure we show Fubini's theorem which allows us to commute integrals:

$$\int x.\ \left(\int y.\ f\ x\ y\ d\mu_2\right)\ d\mu_1 = \int y.\ \left(\int x.\ f\ x\ y\ d\mu_1\right)\ d\mu_2$$

The finite product of measure $\pi = \bigotimes_{i \in I} \mu_i$ iterates the product construction over a finite index I. The elements in this product space are functions from the index set I into the space Ω_i. The measure π maps dependent function spaces to multiplication: $\pi\ (\times_{i \in I} A_i) = \prod_{i \in I} \mu_i\ A_i$. This allows us to define the Lebesgue measure on Euclidean spaces $\mathbb{R}^n$.

The Lebesgue measure λ assigns a measure to subsets of the real line $\mathbb{R}$. The Lebesgue measure is uniquely defined by assigning each interval its length: $\lambda\ \{a\ ..\ b\} = b - a$. A famous result is that not all sets are Lebesgue measurable. We show that at least the Borel sets are measurable and that subsets of null sets are also measurable. The Lebesgue measure is important for real analysis: the Lebesgue integral on the Lebesgue measure is an extension of the Riemann integral.

In this chapter we will construct measures with the desired properties. But for a complete formalization we also look at the AE-quantifier and the Lebesgue integral on these measure spaces.

3.1 Counting Measure

The first concrete measure we introduce is the *counting measure*. It simply assigns the cardinality of a set as its measure. If the set is infinite the measure is ∞. The σ-algebra is discrete, i.e. all subsets of Ω are measurable.

$$\begin{array}{lcl} \mathit{count} & :: & \alpha\ \mathit{set} \to \alpha\ \mathit{measure} \\ \mathit{count}\ \Omega & = & \mathit{measure\text{-}of}\ \Omega\ \mathcal{P}(\Omega)\ (\lambda A.\ \textsf{if}\ \mathit{finite}\ A\ \textsf{then}\ \mathit{card}\ A\ \textsf{else}\ \infty) \end{array}$$

The count measure maps measure theory concepts like measurable sets and functions, measure, and the Lebesgue integral to their discrete counterparts: measurable sets are just subsets, each function into the space Ω is measurable, the measure is the cardinal of the set, and, as we will show later, the integral becomes sum. To show these results, we begin by defining the measure space. The power set is a σ-algebra and the count measure is countably additive, easy to show with the Lemma COUNTABLY-ADDITIVE-IFF-CONTINUOUS-FROM-BELOW. From this we derive:

lemma SPACE-COUNT: $\Omega_{\mathit{count}\ \Omega} = \Omega$

lemma SETS-COUNT: $\mathcal{A}_{\mathit{count}\ \Omega} = \mathcal{P}(\Omega)$

lemma MEASURABLE-COUNT: $f \in \mathit{measurable}\ (\mathit{count}\ \Omega)\ \mathcal{M} \Leftrightarrow f \in \Omega \to \Omega_{\mathcal{M}}$

lemma μ-COUNT-FINITE: $A \subseteq \Omega\ \wedge\ \mathit{finite}\ A\ \Longrightarrow\ \mu_{\mathit{count}\ \Omega}\ A = \mathit{card}\ A$

lemma μ-COUNT-INFINITE: $A \subseteq \Omega\ \wedge\ \neg \mathit{finite}\ A\ \Longrightarrow\ \mu_{\mathit{count}\ \Omega}\ A = \infty$

The counting measure assigns every element the measure value of 1: the only null set is the empty set. Hence the AE-quantifier is equal to the bounded universal quantifier:

lemma NULL-SETS-COUNT: $\mathit{null\text{-}sets}_{\mathit{count}\ \Omega} = \{\emptyset\}$

lemma AE-COUNT: $(\mathrm{AE}_{\mathit{count}\ \Omega}\ x.\ P\ x) \Leftrightarrow (\forall x \in \Omega.\ P\ x)$

The counting measure is only σ-finite if the space Ω is countable, otherwise we cannot find a σ-finite cover:

lemma σ-FINITE-COUNT: *σ-finite-measure* (*count* (Ω :: α :: $\mathcal{C}$ *set*))

lemma FINITE-MEASURE-COUNT: *finite* Ω $\Longrightarrow$ *finite-measure* (*count* Ω)

We could use a similar definition to introduce the point measure of p, where each element x has the weight $p\ x$. Instead of the cardinality of A, we would sum up over all $p\ x$: $\sum_{x \in A} p\ x$. However, it gets cumbersome to define the measure on infinite, countable sets, and it is even more cumbersome to prove that this is a measure space. In Section 3.3 we will use density measures to define the point measure.

3.1.1 Integration over a Count Measure

For the count measure the integral of f should be equal to the sum of f over all elements in Ω. First we notice that each function is Borel-measurable. Moreover, for a finite space each function is simple and integrable:

lemma SIMPLE-FN-COUNT: *simple-fn* (*count* Ω) $f \Leftrightarrow$ *finite* $f[\Omega]$

lemma $\mathcal{B}_\alpha$-MEASURABLE-COUNT: $f \in$ *measurable* (*count* Ω) $\mathcal{B}_\alpha$

lemma INTEGRABLE-COUNT: *finite* Ω $\Longrightarrow$ *integrable* (*count* Ω) f

The positive integral is well-defined also for an infinite space Ω. However, to equate it to sums we require that the function has a finite support, i.e. f is strictly positive only on a finite subset of Ω. The positive integral maps the negative values of f to zero, hence we are only interested in the positive support of f.

lemma $\int^P$-COUNT:

finite $\{x \in \Omega \mid 0 < f\ x\}$ $\Longrightarrow$ $\int^P f d(\mathit{count}\ \Omega) = \sum_{x \in \Omega\ \wedge\ 0 < f\ x} f\ x$

lemma $\int$-COUNT:

finite $\{x \in \Omega \mid f\ x \neq 0\}$ $\Longrightarrow$ $\int f d(\mathit{count}\ \Omega) = \sum_{x \in \Omega\ \wedge\ f\ x \neq 0} f\ x$

These equalities simply follow from the linearity of the Lebesgue integral: all functions are Borel-measurable on the count measure. As the support of f is finite we can represent it as a finite sum. These rules allow us to specialize theorems about integrals on measure spaces to finite sums.

3.2 Push-Forward Measure

A measurable function X from a measure $\mathcal{M}$ into a σ-algebra $\mathcal{N}$ induces a measure on $\mathcal{N}$, the so called *push-forward measure*. In measure theory books, the push-forward of the measure μ under the measurable function X is often written $X(\mu)$ or $X_*(\mu)$, implicitly assuming a σ-algebra $\mathcal{N}$. From the measure $\mathcal{N}$ we only need

the σ-algebra of its measurable sets. While the measure values are ignored, we still require to that it forms a σ-algebra. This is ensured by the type of $\mathcal{N} :: \beta$ *measure*.

$$\begin{array}{lcl} \mathit{distr} & :: & \alpha\ \mathit{measure} \to \beta\ \mathit{measure} \to (\alpha \to \beta) \to \beta\ \mathit{measure} \\ \mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X & = & \mathit{measure\text{-}of}\ \Omega_{\mathcal{N}}\ \mathcal{A}_{\mathcal{N}}\ (\lambda A.\ \mu_{\mathcal{M}}\ (X^{-1}[A] \cap \Omega_{\mathcal{M}})) \end{array}$$

We call the constant for the push-forward measure *distr*, as the push-forward measure of a random variable is its probability distribution. First, we show that *distr* is well-defined by deriving the following justifying theorems:

lemma SPACE-DISTR: $\Omega_{\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X} = \Omega_{\mathcal{N}}$

lemma SETS-DISTR: $\mathcal{A}_{\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X} = \mathcal{A}_{\mathcal{N}}$

lemma μ-DISTR:
$X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N} \ \wedge\ A \in \mathcal{A}_{\mathcal{N}} \implies \mu_{\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X}\ A = \mu_{\mathcal{M}}\ (X^{-1}[A] \cap \Omega_{\mathcal{M}})$

For the last theorem we show that $\mu_{\mathcal{M}}\ (X^{-1}[A] \cap \Omega_{\mathcal{M}})$ is countably additive in A. This holds as the inverse image of a measurable function maps countable, disjoint unions of measurable sets again to countable, disjoint unions of measurable sets.

We try to reduce operations that only talk about values mapped by X into operations on the push-forward measure under X. Obviously, we can transfer the measure itself $\mu_{\mathcal{M}}\ \{x \mid P\ (X\ x)\} = \mu_{\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X}\ \{x \mid P\ x\}$. This is also possible with the AE-quantifier and the Lebesgue integral.

lemma AE-DISTR:
$X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N} \implies (\mathrm{AE}_{\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X}\ x.\ P\ x) \Leftrightarrow (\mathrm{AE}_{\mathcal{M}}\ x.\ P\ (X\ x))$

theorem $\int^P$-DISTR:
$X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N} \ \wedge\ f \in \mathit{measurable}\ \mathcal{N}\ \mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\int^P f\ d(\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X) = \int^P x.\ f\ (X\ x)\ d\mathcal{M}$

corollary $\int$-DISTR:
$X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N} \ \wedge\ f \in \mathit{measurable}\ \mathcal{N}\ \mathcal{B}_{\mathbb{R}} \implies$
$\int f\ d(\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X) = \int x.\ f\ (X\ x)\ d\mathcal{M}$

corollary INTEGRABLE-DISTR-EQ:
$X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N} \ \wedge\ f \in \mathit{measurable}\ \mathcal{N}\ \mathcal{B}_{\mathbb{R}} \implies$
$\mathit{integrable}\ (\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X)\ f \Leftrightarrow \mathit{integrable}\ \mathcal{M}\ (\lambda x.\ f\ (X\ x))$

To prove Theorem $\int^P$-DISTR we use induction on the Borel-measurable function f. The integral of real functions and integrability is simply a matter of unfolding the definition, and then rewriting with Theorem $\int^P$-DISTR.

The composition of two push-forward measures is equal to the push-forward measure of the composition of the two measurable functions. This directly follows from the compositionality of inverse images: $(Y \circ X)^{-1}[A] = X^{-1}[Y^{-1}[A]]$.

lemma DISTR-DISTR-EQ:
$X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N} \ \wedge\ Y \in \mathit{measurable}\ \mathcal{N}\ \mathcal{L} \implies$
$\mathit{distr}\ (\mathit{distr}\ \mathcal{M}\ \mathcal{N}\ X)\ \mathcal{L}\ Y = \mathit{distr}\ \mathcal{M}\ \mathcal{L}\ (Y \circ X)$

These equalities are important for probability theory where a probability space with a couple of random variabels is assumed. The push-forward measure allows us to reduce statements about one of these random variables to statements about a probability space where the random variable does not occur anymore.

3.3 Density Measure

We define the *density measure* f[1] as follows: we use the positive integral to weigh each element x in the measure space $\mathcal{M}$ with $f\ x$. The function f is called the *density function*.

$$\begin{array}{lcl} \mathit{density} & :: & \alpha\ \mathit{measure} \to \left(\alpha \to \overline{\mathbb{R}}\right) \to \alpha\ \mathit{measure} \\ \mathit{density}\ \mathcal{M}\ f & = & \mathit{measure\text{-}of}\ \Omega_{\mathcal{M}}\ \mathcal{A}_{\mathcal{M}}\ \left(\lambda A.\ \int^P x.\ f\ x \cdot \chi\ A\ x\ d\mathcal{M}\right) \end{array}$$

For a Borel-measurable function f the integral behaves as a measure, it is positive and by monotone convergence it is also countably additive. So we defined a measure on the same σ-algebra as that of $\mathcal{M}$, expressible as integral of f:

lemma SPACE-DENSITY: $\Omega_{\mathit{density}\ \mathcal{M}\ f} = \Omega_{\mathcal{M}}$

lemma SETS-DENSITY: $\mathcal{A}_{\mathit{density}\ \mathcal{M}\ f} = \mathcal{A}_{\mathcal{M}}$

lemma μ-DENSITY:
$f \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge A \in \mathcal{A}_{\mathcal{M}} \implies$
$\mu_{\mathit{density}\ \mathcal{M}\ f}\ A = \int^P x.\ f\ x \cdot \chi\ A\ x\ d\mathcal{M}$

Intuitively, the null sets are extended with sets where f is a.e. nonpositive. This extends also nicely to the AE-quantifier:

lemma NULL-SETS-DENSITY:
$f \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\left(A \in \mathit{null\text{-}sets}_{\mathit{density}\ \mathcal{M}\ f}\right) \Leftrightarrow \left(A \in \mathcal{A}_{\mathcal{M}} \wedge \mathrm{AE}_{\mathcal{M}}\ x \in A.\ f\ x \leq 0\right)$

lemma AE-DENSITY:
$f \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\left(\mathrm{AE}_{\mathit{density}\ \mathcal{M}\ f}\ x.\ P\ x\right) \Leftrightarrow \left(\mathrm{AE}_{\mathcal{M}}\ x.\ 0 < f\ x \implies P\ x\right)$

Now, what happens when we integrate over a density measure? As expected, it results in an integral where the two functions are multiplied. This sticks with the intuition that we add a weight to each element.

theorem $\int^P$-DENSITY:
$f, g \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge \left(\mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq f\ x\right) \implies$
$\left(\int^P g\ d(\mathit{density}\ \mathcal{M}\ f)\right) = \left(\int^P x.\ f\ x \cdot g\ x\ d\mathcal{M}\right)$

We prove this rule by induction on the Borel-measurable function g.

[1] In measure theory textbooks this is often called a measure having density f.

This can be easily extended to integration on real functions. Note that we now also assume that f is a real, a.e.-positive function, otherwise the types for the integrands would not match anymore.

corollary INTEGRABLE-, $\int$-DENSITY:
$f, g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}} \wedge (\mathrm{AE}_{\mathcal{M}}\, x.\ 0 \leq f\ x) \implies$
$\textit{integrable}\ (\textit{density}\ \mathcal{M}\ f)\ g \Leftrightarrow \textit{integrable}\ \mathcal{M}\ (\lambda x.\ f\ x \cdot g\ x)$,
$\left(\int g\ d(\textit{density}\ \mathcal{M}\ f)\right) = \left(\int x.\ f\ x \cdot g\ x\ d\mathcal{M}\right)$

From Theorem $\int^P$-DENSITY also immediately follows that sequentially applying two density functions to a measure is the same as concurrently applying the product of both density functions:

lemma DENSITY-DENSITY:
$f, g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge (\mathrm{AE}_{\mathcal{M}}\, x.\ 0 \leq f\ x) \implies$
$\textit{density}\ (\textit{density}\ \mathcal{M}\ f)\ g = \textit{density}\ \mathcal{M}\ (\lambda x.\ f\ x \cdot g\ x)$

The positive integral is equal if the integrands are a.e.-equal. From this it obviously follows that the density measures are equal when the density functions are a.e.-equal. Is the density function uniquely determined by its measure? Yes, for a σ-finite measure we show that density measures are equal iff their density functions are a.e.-equal:

theorem DENSITY-UNIQUE-IFF:
$\sigma\textit{-finite-measure}\ \mathcal{M} \wedge f, g \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge$
$(\mathrm{AE}_{\mathcal{M}}\, x.\ 0 \leq f\ x) \wedge (\mathrm{AE}_{\mathcal{M}}\, x.\ 0 \leq g\ x) \implies$
$(\textit{density}\ \mathcal{M}\ f = \textit{density}\ \mathcal{M}\ g) \Leftrightarrow (\mathrm{AE}_{\mathcal{M}}\, x.\ f\ x = g\ x)$

The proof for the left to right implication is quite involved, first we show that it holds when the density measures are finite, then when $\mathcal{M}$ is finite, and finally when $\mathcal{M}$ is σ-finite.

The density function is a.e.-finite iff the density measure is also σ-finite:

lemma σ-FINITE-IFF-FINITE-DENSITY:
$\sigma\textit{-finite-measure}\ \mathcal{M} \wedge f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\sigma\textit{-finite-measure}\ (\textit{density}\ \mathcal{M}\ f) \Leftrightarrow \mathrm{AE}_{\mathcal{M}}\, x.\ f\ x < \infty$

We use C, the σ-finite cover of $\mathcal{M}$, to construct the sets $F\ n\ m = \{x \mid f\ x < n\} \cap C\ m$. Each $F\ n\ m$ has finite measure, and their union covers the entire space.

3.3.1 Point Measure

By combining the density measure and the count measure we get the *point measure*, which assigns a measure value to each element in a discrete measure space:

$\textit{point} :: \alpha\ \textit{set} \to \left(\alpha \to \overline{\mathbb{R}}\right) \to \alpha\ \textit{measure}$
$\textit{point}\ A\ f = \textit{density}\ (\textit{count}\ A)\ f$

The measurable sets are obviously all subsets of A. To equate the measure value to the sum over f we require that f is nonnegative in X and that X is finite:

lemma μ-POINT:
$(\forall i \in X.\ 0 \leq f\ i) \wedge \mathit{finite}\ X \wedge X \subseteq A \implies \mu_{\mathit{point}\ A\ f}\ X = \sum_{i \in X} f\ i$

3.3.2 Radon-Nikodým Derivative

From the previous section we know that we can construct new measures by adding densities to a measure. Interestingly the other direction also holds: under some assumptions there exists a density between two measures on the same measurable sets. More precisely, the Radon-Nikodým theorem states that for each measure $\mathcal{N}$ that is *absolutely continuous* on $\mathcal{M}$ there exists an a.e.-unique density function f, s.t. $\mathit{density}\ \mathcal{M}\ f = \mathcal{N}$. This is used to define *conditional expectation* in probability theory and *mutual information* in information theory. The Radon-Nikodým theorem requires that $\mathcal{M}$ is σ-finite.

First we introduce *absolute continuity*. This is a minimal assumption for Radon-Nikodým, as we cannot weigh elements in a null set of $\mathcal{M}$ to get a non-null set in $\mathcal{N}$:

$$\begin{array}{lcl} \mathit{absolutely\text{-}continuous} & :: & \alpha\ \mathit{measure} \to \alpha\ \mathit{measure} \to \mathbb{B} \\ \mathit{absolutely\text{-}continuous}\ \mathcal{M}\ \mathcal{N} & \Leftrightarrow & \mathit{null\text{-}sets}_{\mathcal{M}} \subseteq \mathit{null\text{-}sets}_{\mathcal{N}} \end{array}$$

The implication between almost everywhere quantification immediately follows from the inclusion between null sets:

lemma ABSOLUTELY-CONTINUOUS-AE:
$\mathcal{A}_{\mathcal{M}} = \mathcal{A}_{\mathcal{N}} \wedge \mathit{absolutely\text{-}continuous}\ \mathcal{M}\ \mathcal{N} \implies$
$(\mathrm{AE}_{\mathcal{M}}\ x.\ P\ x) \implies (\mathrm{AE}_{\mathcal{N}}\ x.\ P\ x)$

With Lemma AE-DENSITY it directly follows that a density measure is absolutely continuous to its supporting measure:

lemma ABSOLUTELY-CONTINUOUS-DENSITY:
$f \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \implies \mathit{absolutely\text{-}continuous}\ \mathcal{M}\ (\mathit{density}\ \mathcal{M}\ f)$

As Radon-Nikodým states, the other direction is also true: if two measures are absolutely continuous, then there exists the exists the so-called *Radon-Nikodým derivative*, which is the density function between these two measures:

theorem RADON-NIKODÝM:
$\sigma\text{-}\mathit{finite\text{-}measure}\ \mathcal{M} \wedge \mathit{absolutely\text{-}continuous}\ \mathcal{M}\ \mathcal{N} \wedge \mathcal{A}_{\mathcal{M}} = \mathcal{A}_{\mathcal{N}} \implies$
$\exists f \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}.\ (\forall x.\ 0 \leq f\ x) \wedge \mathit{density}\ \mathcal{M}\ f = \mathcal{N}$

For the proof we first assume that $\mathcal{M}$ and $\mathcal{N}$ are finite measures. There we form $\mathcal{G} = \{g \mid \forall A.\ \mu_{\mathit{density}\ \mathcal{M}\ g}\ A \leq \mu_{\mathcal{N}}\ A\}$ and construct a sequence of $g_i \in \mathcal{G}$ with $\lim_i \int g_i\ d\mathcal{M} = \sup_{g \in \mathcal{G}} \int g\ d\mathcal{M}$ and finally choose $\lim_i g_i$ as Radon-Nikodým derivative. Then we assume that $\mathcal{N}$ is an arbitrary measure space, and we cover Ω with one set where $\mathcal{N}$ is a.e.-infinite and a sequence of sets with finite measure. Finally, we assume that $\mathcal{M}$ is σ-finite and weigh each set C_i in its σ-cover with

$\mu_{\mathcal{M}}(C_i) \cdot 2^{-i}$ then we have a finite measure, obtain f and revert the weighting. As this proof is quite involved we omitted further details from this presentation — our formalized proof follows Bauer [9], §17. The interested reader may look in the textbook, we do not give more details as all reusable theorems used by the proof are already presented.

We know from Theorem DENSITY-UNIQUE-IFF that the Radon-Nikodým derivative f is a.e.-uniquely defined by $\mathcal{M}$ and $\mathcal{N}$. So, together with the proof of its existence it is sensible to define the Radon-Nikodým derivative as a function of $\mathcal{M}$ and $\mathcal{N}$:

$$\begin{aligned}&\textit{RN-deriv} :: \alpha\ \textit{measure} \to \alpha\ \textit{measure} \to \left(\alpha \to \overline{\mathbb{R}}\right)\\ &\textit{RN-deriv}\ \mathcal{M}\ \mathcal{N} =\\ &\quad \textit{SOME}\ f \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}.\ (\forall x.\ 0 \le f\ x) \wedge \textit{density}\ \mathcal{M}\ f = \mathcal{N}\end{aligned}$$

Thanks to Radon-Nikodým, this definition is correct:

corollary RN-DERIV-$\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE, RN-DERIV-NONNEG, RN-DERIV-DENSITY:
σ-*finite-measure* $\mathcal{M} \wedge$ *absolutely-continuous* $\mathcal{M}\ \mathcal{N} \wedge \mathcal{A}_{\mathcal{M}} = \mathcal{A}_{\mathcal{N}} \implies$
RN-deriv $\mathcal{M}\ \mathcal{N} \in$ *measurable* $\mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}$,
$(\forall x.\ 0 \le$ *RN-deriv* $\mathcal{M}\ \mathcal{N}\ x)$,
density $\mathcal{M}$ (*RN-deriv* $\mathcal{M}\ \mathcal{N}$) $= \mathcal{N}$

Together with Theorem DENSITY-UNIQUE-IFF we show that the Radon-Nikodým derivative is a.e. the density function f when applied on the density measure of f:

lemma RN-DERIV-UNIQUE:
σ-*finite-measure* $\mathcal{M} \wedge f \in$ *measurable* $\mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge (\mathrm{AE}_{\mathcal{M}}\ x.\ 0 \le f\ x) \implies$
$\mathrm{AE}_{\mathcal{M}}\ x.$ *RN-deriv* $\mathcal{M}$ (*density* $\mathcal{M}\ f$) $x = f\ x$

The existence of the Radon-Nikodým derivative will justify the characterization of distributions in Section 4.3, and the definition of entropy and mutual information in Section 4.4.

3.4 Products of Measures

We first introduce the binary product of measure spaces, and then finite products of measure spaces. In Section 4.5 we will further extend them to infinite products.

3.4.1 Binary Product Measure

The *binary product measure* is the measure-theoretic counterpart of the Cartesian product of sets. The measurable sets are generated by the products of the measurable sets of its factors. The measure is defined using the iteration of the Lebesgue integral. With Fubini's theorems we later show that the result is independent of the order of iteration.

$$\begin{aligned}&\square \otimes \square :: \alpha\ \textit{measure} \to \beta\ \textit{measure} \to (\alpha \times \beta)\ \textit{measure}\\ &\mathcal{M} \otimes \mathcal{N} = \textit{measure-of}(\Omega_{\mathcal{M}} \times \Omega_{\mathcal{N}})\ \{A \times B | A \in \mathcal{A}_{\mathcal{M}}, B \in \mathcal{A}_{\mathcal{N}}\}\\ &\qquad \left(\lambda Q.\ \int^P x.\ \left(\int^P y.\ \chi\ Q\ (x,y)\ d\mathcal{N}\right) d\mathcal{M}\right)\end{aligned}$$

The space of the binary product measure is the Cartesian product of the spaces of its factors $\mathcal{M}$ and $\mathcal{N}$. And we also know that each Cartesian product of measurable sets is measurable in the binary product measure.

lemma SPACE-PAIR-MEASURE:
$\Omega_{\mathcal{M}\otimes\mathcal{N}} = \Omega_{\mathcal{M}} \times \Omega_{\mathcal{N}}$

lemma SETS-PAIR-MEASUREI:
$A \in \mathcal{A}_{\mathcal{M}} \ \wedge \ B \in \mathcal{A}_{\mathcal{N}} \implies A \times B \in \mathcal{A}_{\mathcal{M}\otimes\mathcal{N}}$

With these lemmas we verify that the projection functions *fst* and *snd* and the pair construction is measurable. All functions on the binary product type can be constructed using them. This also allows us to use characterize binary operators as measurable functions, e.g. $(\lambda(x,y).\ x+y) \in \mathit{measurable}\ (\mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}})\ \mathcal{B}_{\mathbb{R}}$.

lemma MEASURABLE-FST:
$\mathit{fst} \in \mathit{measurable}\ (\mathcal{M} \otimes \mathcal{N})\ \mathcal{M}$

lemma MEASURABLE-SND:
$\mathit{snd} \in \mathit{measurable}\ (\mathcal{M} \otimes \mathcal{N})\ \mathcal{N}$

lemma MEASURABLE-PAIR:
$f \in \mathit{measurable}\ \mathcal{L}\ \mathcal{M} \ \wedge \ g \in \mathit{measurable}\ \mathcal{L}\ \mathcal{N} \implies$
$(\lambda x.\ (f\ x, g\ x)) \in \mathit{measurable}\ \mathcal{L}\ (\mathcal{M} \otimes \mathcal{N})$

The last lemma requires Theorem MEASURABLE-σ, all others lemma follow by definition.

The next step is now to show that the iterated integral, we employ in the definition for $\mathcal{M} \otimes \mathcal{N}$, is really a measure. The idea is quite simple: the integral is positive and from monotone convergence follows countable additivity. However, to apply monotone convergence on the Lebesgue integral we need to show that the integrand $\lambda x.\ \int^P y.\ \chi\ Q\ (x,y)\ d\mathcal{N}$ is a Borel-measurable function for each measurable set Q. This function can be rewritten into the measure of the cut of Q: $\lambda x.\ \mu_{\mathcal{N}}\ \{y \mid (x,y) \in Q\}$. It is measurable when $\mathcal{N}$ is finite:

lemma MEASURABLE-FINITE-MEASURE-CUT:
$\mathit{finite\text{-}measure}\ \mathcal{N} \ \wedge \ Q \in \mathcal{A}_{\mathcal{M}\otimes\mathcal{N}} \implies$
$\Big(\lambda x.\ \mu_{\mathcal{N}}\ \{y \mid (x,y) \in Q\}\Big) \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}$

We prove this by induction over Q with Corollary σ-SETS-INDUCT-DISJOINT: the generator of $\mathcal{M} \otimes \mathcal{N}$ is $\cap$-stable, the cut is measurable for each generating set $A \times B$, and that property is closed under complement and countable union of disjoint sets.

Now we generalize the previous lemma to a σ-finite measure $\mathcal{N}$. With countable additivity of the measure $\mu_{\mathcal{M}}$ we represent the cut $\{x \mid (x,y) \in Q\}$ as the sum of all restrictions to the elements of the σ-cover. The restrictions being finite, with the Lemmas MEASURABLE-FINITE-MEASURE-CUT and $\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE-SUP it then follows:

lemma MEASURABLE-σ-FINITE-MEASURE-CUT:
$\sigma\text{-}\mathit{finite\text{-}measure}\ \mathcal{N} \ \wedge \ Q \in \mathcal{A}_{\mathcal{M}\otimes\mathcal{N}} \implies$
$\Big(\lambda x.\ \mu_{\mathcal{N}}\ \{y \mid (x,y) \in Q\}\Big) \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}}$

With the Borel-measurability of the integrand, we then infer monotone convergence for the iterated integral in the definition of the binary product measure. So we constructed a measure space where the measure is defined to be the iterated integral over χ Q:

theorem μ-PAIR-MEASURE:
σ-finite-measure $\mathcal{M} \wedge Q \in \mathcal{A}_{\mathcal{M}\otimes\mathcal{N}} \Longrightarrow$
$\mu_{\mathcal{M}\otimes\mathcal{N}}\ Q = \int^P x.\ \left(\int^P y.\ \chi\ Q\ (x,y)\ d\mathcal{N}\right) d\mathcal{M}$

The usual characterization of the binary product measure assumes that the measure of the Cartesian product equals multiplication of the measures. We verify this characterization by applying the equation $\chi\ (A \times B)\ (x,y) = \chi\ A\ x \cdot \chi\ B\ y$.

corollary μ-PAIR-MEASURE-TIMES:
σ-finite-measure $\mathcal{M} \wedge A \in \mathcal{A}_{\mathcal{M}} \wedge B \in \mathcal{A}_{\mathcal{N}} \Longrightarrow$
$\mu_{\mathcal{M}\otimes\mathcal{N}}\ (A \times B) = \mu_{\mathcal{M}}\ A \cdot \mu_{\mathcal{N}}\ B$

Also the product of σ-finite measure spaces is again a σ-finite measure space. This allows us to combine multiple product measure:

lemma σ-FINITE-MEASURE-PAIR-MEASURE:
σ-finite-measure $\mathcal{M} \wedge \sigma$-finite-measure $\mathcal{N} \Longrightarrow \sigma$-finite-measure $(\mathcal{M} \otimes \mathcal{N})$

The binary product measure is commutative in its factors. Formally, this translates to the measure preserving nature of $\big(\lambda(x,y).\ (y,x)\big)$:

lemma DISTR-PAIR-SWAP:
σ-finite-measure $\mathcal{M} \wedge \sigma$-finite-measure $\mathcal{N} \Longrightarrow$
$\mathcal{M} \otimes \mathcal{N} = \mathit{distr}\ (\mathcal{N} \otimes \mathcal{M})\ (\mathcal{M} \otimes \mathcal{N})\ \big(\lambda(x,y).\ (y,x)\big)$

This lemma is proved by Corollary μ-PAIR-MEASURE-TIMES and Theorem MEASURE-EQI-GENERATOR-EQ and commutativity of multiplication. From this immediately follows commutativity of the product measure, of the integral, and of the AE-quantifier.

For the AE-quantifier we also show iterativity of the product measure, i.e. the AE-quantifier over the measure is equal to an iteration of the AE-quantifier over both its factors:

lemma AE-PAIR-IFF:
σ-finite-measure $\mathcal{M} \wedge \sigma$-finite-measure $\mathcal{N} \wedge \{x \mid P\ x\} \in \mathcal{A}_{\mathcal{M}\otimes\mathcal{N}} \Longrightarrow$
$(\mathrm{AE}_{\mathcal{M}\otimes\mathcal{N}}\ x.\ P\ x) \Leftrightarrow (\mathrm{AE}_{\mathcal{M}}\ x.\ \mathrm{AE}_{\mathcal{N}}\ y.\ P\ (x,y))$

3.4.2 Fubini's Theorem

We show now that the Lebesgue integral over $\mathcal{M} \otimes \mathcal{N}$ equals the iterated integral of its factors $\mathcal{M}$ and $\mathcal{N}$ for all measurable functions. From this directly follows then Fubini's theorem, i.e. that integrals on σ-finite measure spaces are commutative. *In this section we assume that $\mathcal{M}$ and $\mathcal{N}$ are σ-finite measure spaces.*

For the iteration rule of the positive integral on the product measure it is necessary that the positive integral along one factor is measurable. This rule is not just an auxiliary lemma for Fubini's theorem, but it actually shows that the Lebesgue integral over a parametrized function is measurable.

lemma $\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE-$\int^P$-FST:
$f \in$ *measurable* $(\mathcal{M} \otimes \mathcal{N})$ $\mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\left(\lambda x.\ \int^P y.\ f\ (x,y)\ d\mathcal{N}\right) \in$ *measurable* $\mathcal{M}$ $\mathcal{B}_{\overline{\mathbb{R}}}$

We prove this by induction on the Borel-measurable function f. The base case is shown with Lemma MEASURABLE-σ-FINITE-MEASURE-CUT. All other cases are closure properties of the Lebesgue integral and of measurable functions.

Now we show the iteration rule for the Lebesgue integral:

theorem $\int^P$-FST:
$f \in$ *measurable* $(\mathcal{M} \otimes \mathcal{N})$ $\mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\int^P f\ d\,(\mathcal{M} \otimes \mathcal{N}) = \int^P x.\ \left(\int^P y.\ f\ (x,y)\ d\mathcal{N}\right) d\mathcal{M}$

For this we use again induction over the Borel-measurable function f. The base case is proved with Theorem μ-PAIR-MEASURE. All other cases are closure properties of the Lebesgue integral.

Instead of proving the symmetric variant of the last two lemmas we use the symmetry properties of the product space. With Lemma DISTR-PAIR-SWAP we know that the pair swap function $(\lambda(x,y).\ (y,x))$ is measure preserving between $\mathcal{M} \otimes \mathcal{N}$ and $\mathcal{N} \otimes \mathcal{M}$. This allows us to get symmetric variants of Lemma $\mathcal{B}_{\overline{\mathbb{R}}}$-MEASURABLE-$\int^P$-FST and Theorem $\int^P$-FST without writing two nearly identical proofs.

corollary $\int^P$-FUBINI:
$f \in$ *measurable* $(\mathcal{M} \otimes \mathcal{N})$ $\mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\int^P x.\ \left(\int^P y.\ f\ (x,y)\ d\mathcal{N}\right)\ d\mathcal{M} = \int^P y.\ \left(\int^P x.\ f\ (x,y)\ d\mathcal{M}\right)\ d\mathcal{N}$

To show Fubini's theorem also for real-valued functions we first need the integrability along one factor of the product measure. With Theorem $\int^P$-FST we show that from integrability on the product space follows a.e.-integrablity along one factor:

theorem $\int$-FST-INTEGRABLE:
integrable $(\mathcal{M} \otimes \mathcal{N})$ $f \implies \mathrm{AE}_{\mathcal{M}}\, x.$ *integrable* $\mathcal{N}$ $(\lambda y.\ f\ (x,y))$

Then, with this theorem we extend Theorem $\int^P$-FST to integration on functions into $\mathbb{R}$:

theorem $\int$-FST:
integrable $(\mathcal{M} \otimes \mathcal{N})$ $f \implies$
$\int x.\ \left(\int y.\ f\ (x,y)\ d\mathcal{M}\right) d\mathcal{N} = \int\ f\ d\,(\mathcal{M} \otimes \mathcal{N})$

Finally, we prove Fubini's theorem for integrable functions into $\mathbb{R}$:

corollary $\int$-FUBINI:
integrable $(\mathcal{M} \otimes \mathcal{N})$ $f \implies$
$$\int x.\ \left(\int y.\ f\ (x,y)\ d\mathcal{N}\right)\ d\mathcal{M} = \int y.\ \left(\int x.\ f\ (x,y)\ d\mathcal{M}\right)\ d\mathcal{N}$$

This corollary is proved with two instantiations of Theorem $\int$-FST, once with $\mathcal{M}$, $\mathcal{N}$ and f and once with the two measures swapped and the function $\lambda(y,x).\ f(x,y)$. With Lemma DISTR-PAIR-SWAP we show that $\lambda(y,x).\ f(x,y)$ is integrable on $\mathcal{N} \otimes \mathcal{M}$.

With Theorem $\int^P$-FST we can also show the commutation of the density measure with the pair measure.

lemma PAIR-MEASURE-DENSITY-FST:
$f \in$ *measurable* $\mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge (\mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq f\ x) \wedge$
σ-*finite-measure* (*density* $\mathcal{M}\ f$) $\implies$
(*density* $\mathcal{M}\ f$) $\otimes \mathcal{N}$ = *density* $(\mathcal{M} \otimes \mathcal{N})\ (\lambda(x,y).\ f\ x)$

Such a commutation law also holds for the push-forward measure.

lemma PAIR-MEASURE-DISTR-FST:
$T \in$ *measurable* $\mathcal{M}\ \mathcal{S} \wedge \sigma$-*finite-measure* (*distr* $\mathcal{M}\ \mathcal{S}\ T$) $\implies$
(*distr* $\mathcal{M}\ \mathcal{S}\ T$) $\otimes \mathcal{N}$ = *distr* $(\mathcal{M} \otimes \mathcal{N})\ (\mathcal{S} \otimes \mathcal{N})\ (\lambda(x,y).\ (T\ x,y))$

We only show the variants for the first element projection, with Lemma DISTR-PAIR-SWAP we can also show their symmetric variants. Both lemmas show how nice the density, the push-forward and the product measure fit into our measure-theoretic framework.

3.4.3 Product σ-Algebra on Dependent Function Space

In textbooks, product spaces with finitely many factors are usually defined as the iteration of the binary product. There the product space $\bigotimes_{i \in \{1,\ldots,n\}} \mathcal{M}_i$ is defined to be $\mathcal{M}_1 \otimes (\mathcal{M}_2 \otimes (\mathcal{M}_3 \otimes \cdots \otimes \mathcal{M}_n))$ and, since it is isomorphic under associative reordering: it can be written without parenthesis. The problem now is that the product measure type encodes the amount of factors and their order and types. This has several problems in Isabelle/HOL: (1) the types are not equal under associative reordering, forcing us to apply coercion functions between reordered product spaces and (2) this is problematic when we want to do induction on n, which is not possible as the type would depend on n.

To avoid these problems we model the *product measure space* $\bigotimes_{i \in I} \mathcal{M}\ i$ as dependent function space. The factors of the product measure are represented with a function $\mathcal{M}$ mapping from the index I into measure spaces. The product measure space provides the canonical σ-algebra for the function space: it is used for the finite product of σ-finite measures, infinite product of probability measures, and for the trace space of a stochastic process. The measure part of $\bigotimes_{i \in I} \mathcal{M}\ i$ is used as product measure in the finite and the infinite case.

Notation: *In this section we assume that $\mathcal{M}$ is a function from the index set I into measure spaces. We abbreviate $\Omega_{\mathcal{M}\ i}$ with Ω_i, $\mathcal{A}_{\mathcal{M}\ i}$ with $\mathcal{A}_i$, and $\mu_{\mathcal{M}\ i}$ with μ_i.*

The basic sets we want to measure are $\{\omega \in \times_{i\in I}\Omega_i \mid \omega\ i \in A\}$ for each i in I and A in $\mathcal{A}_i$. Yet, these basic sets are not suited to define a measure, as they are not closed under intersection. For the finite case we define the measure on the cubes $\times_{i\in I} A\ i$, where $A\ i$ is in $\mathcal{A}_i$ for all $i \in I$. This does not work for the infinite case, the index set may not be countable, and multiplication is hard to handle in the countable but infinite case. To express products that map to finite multiplication, we use the embedding of cubes with finite indices. For this we define the *embedding* of sets Q from $\times_{i\in J}\Omega_i$ in $\times_{i\in I}\Omega_i$, where $J \subseteq I$:

$$\begin{array}{lcl} \mathit{emb} & :: & \iota\ \mathit{set} \to (\iota \to \alpha)\ \mathit{set} \to (\iota \to \alpha)\ \mathit{set} \\ \mathit{emb}\ J\ Q & = & \{\omega \in \times_{i\in I}\Omega_i \mid \omega{\restriction_J} \in Q\} \end{array}$$

With *emb J Q* cubes on J are extended to cubes on I:

lemma EMB-$\times$:
$J \subseteq I \wedge (\forall i \in J.\ A\ i \subseteq \Omega_i) \implies$
$\mathit{emb}\ J\ (\times_{i\in J} A\ i) = (\times_{i\in I}\ \textbf{if}\ i \in J\ \textbf{then}\ A\ i\ \textbf{else}\ \Omega_i)$

The measurable sets of the product measure are generated by the embedding of all cubes of all finite index sets $J \subseteq I$. These generating sets are $\cap$-stable and we can easily assign measures to them.

$$\begin{array}{l} \bigotimes_{i\in\square}\square \quad :: \quad \iota\ \mathit{set} \to (\iota \to \alpha\ \mathit{measure}) \to (\iota \to \alpha)\ \mathit{measure} \\ \bigotimes_{i\in I}\mathcal{M}_i \quad = \\ \quad \mathit{extend\text{-}measure}\ (\times_{i\in I}\Omega_i) \\ \qquad \{(J,A) \mid (J \neq \emptyset \vee I \neq \emptyset) \wedge \mathit{finite}\ J \wedge J \subseteq I \wedge (\forall j \in J.\ A\ i \in \mathcal{A}_i)\} \\ \qquad \left(\lambda(J,A).\ \mathit{emb}\ J\ (\times_{j\in J} A\ j)\right) \\ \qquad \left(\lambda(J,A).\ \prod_{j\in J\cup\{i\in I.\ \mu_i\ \Omega_i\neq 1\}} \textbf{if}\ j \in J\ \textbf{then}\ \mu_j\ (A\ j)\ \textbf{else}\ \mu_j\ \Omega_j\right) \end{array}$$

The premeasure we use (the last parameter to *extend-measure*) multiplies over all indices whose measure is not 1. This is necessary when we define the infinite product measure in Section 4.5. In that case the index set equates J as $\mu_i\ \Omega_i$ will always be 1.

To simplify further proofs we restrict the generating sets to embeddings where the index set J is not empty if I is not empty. And for the measure we need to take care of the case where the factors are not probability spaces. When the index set is empty $I = \emptyset$, the generating sets contain only the entire space, i.e. the singleton set containing the everywhere undefined function. This gives us a probability measure for the case $I = \emptyset$.

Notation: *We abbreviate* $\Omega_{\bigotimes_{i\in I}\mathcal{M}_i}$ *with* $\Omega_\otimes$, $\mathcal{A}_{\bigotimes_{i\in I}\mathcal{M}_i}$ *with* $\mathcal{A}_\otimes$, *and* $\mu_{\bigotimes_{i\in I}\mathcal{M}_i}$ *with* $\mu_\otimes$.

Now we show that the measurable sets are generated by cubes for a finite I and alternatively by the projections of each index:

lemma SPACE-$\bigotimes$: $\Omega_\otimes = \times_{i\in I}\Omega_i$

lemma SETS-$\bigotimes$-FINITE:
$\mathit{finite}\ I \implies \mathcal{A}_\otimes = \sigma\text{-}\mathit{sets}\ \Omega_\otimes \left\{ \times_{i\in I} A\ i \;\middle|\; \forall i \in I.\ A\ i \in \mathcal{A}_i \right\}$

lemma SETS-⨂-SINGLE:
$\mathcal{A}_\otimes = \sigma\text{-sets}\ \Omega_\otimes \left\{ \{ f \in \Omega_\otimes \mid f\ i \in A \} \;\middle|\; i \in I, A \in \mathcal{A}_i \right\}$

This gives us a simple rule to show the measurability of functions into the product space.

lemma MEASURABLE-⨂-SINGLE:
$f \in \Omega_\mathcal{N} \to \Omega_\otimes \ \wedge\ (\forall i \in I, A \in \mathcal{A}_i.\ \{x \mid f\ x\ i \in A\} \in \mathcal{A}_\mathcal{N}) \implies$
$f \in \mathit{measurable}\ \mathcal{N}\ (\bigotimes_{i\in I} \mathcal{M}_i)$

So, f is measurable if the projection of each index is measurable, and the domain of the resulting function is restricted to I:

lemma MEASURABLE-⨂-RESTRICT:
$(\forall i \in I.\ f\ i \in \mathit{measurable}\ \mathcal{N}\ \mathcal{M}_i) \implies$
$(\lambda x.\ (\lambda i.\ f\ i\ x)\restriction_I) \in \mathit{measurable}\ \mathcal{N}\ (\bigotimes_{i\in I} \mathcal{M}_i)$

The opposite direction is also true, the projection of each index is measurable:

lemma MEASURABLE-⨂-COMPONENT:
$i \in I \implies (\lambda\omega.\ \omega\ i) \in \mathit{measurable}\ (\bigotimes_{i\in I} \mathcal{M}_i)\ \mathcal{M}_i$

The last two lemmas also imply that merging of two products and adding dimensions is also measurable. Together with the measurability of the projection at index $i \in I$, we relate the binary product measure $\mathcal{N} \otimes \mathcal{L}$ and the product measure $\bigotimes_{i\in I} \mathcal{M}_i$. Later we will show that this is measure preserving on product measures.

$$\begin{array}{lcl} \mathit{merge} & :: & \iota\ \mathit{set} \to \iota\ \mathit{set} \to \big((\iota \to \alpha) \times (\iota \to \alpha)\big) \to (\iota \to \alpha) \\ \mathit{merge}\ I_1\ I_2 & = & (\lambda(\omega_1, \omega_2)\ i.\ \textsf{if}\ i \in I_1\ \textsf{then}\ \omega_1\ i\ \textsf{else} \\ & & \qquad \textsf{if}\ i \in I_2\ \textsf{then}\ \omega_2\ i\ \textsf{else}\ \mathit{undefined}) \end{array}$$

lemma MEASURABLE-⨂-MERGE:
$\mathit{merge}\ I_1\ I_2 \in \mathit{measurable}\ \Big((\bigotimes_{i\in I_1} \mathcal{M}_i) \otimes (\bigotimes_{i\in I_2} \mathcal{M}_i)\Big)\ \Big(\bigotimes_{i\in I_1\cup I_2} \mathcal{M}_i\Big)$

lemma MEASURABLE-⨂-ADD-DIM:
$(\lambda(\omega, x).\ \omega(i := x)) \in \mathit{measurable}\ \Big((\bigotimes_{i\in I} \mathcal{M}_i) \otimes \mathcal{M}_i\Big)\ \Big(\bigotimes_{i\in\{i\}\cup I} \mathcal{M}_i\Big)$

The equation of Lemma SETS-⨂-SINGLE gives a powerful induction principle for measurable sets in $\mathcal{A}_\otimes$. We can strengthen it when each σ-algebra $\mathcal{A}_i$ is generated by $\mathcal{G}\ i$. Then we show that the σ-algebra $\mathcal{A}_\otimes$ is generated by projections of the generators $\mathcal{G}\ i$. A requirement is that each generator $\mathcal{G}\ i$ contains a σ-cover $C\ i$.

lemma PROD-GENERATOR:
$\Big(\forall i \in I.\ \mathcal{G}\ i \subseteq \mathcal{P}(\Omega_i)\ \wedge\ \mathcal{A}_i = \sigma\text{-sets}\ \Omega_i\ (\mathcal{G}\ i)\ \wedge$
$\qquad (\bigcup_j C\ i\ j) = \Omega_i\ \wedge\ C\ i \in \mathbb{N} \to \mathcal{G}\ i\Big) \implies$
$\mathcal{A}_\otimes = \sigma\text{-sets}\ \Omega_\otimes \left\{ \{x \in \Omega_\otimes \mid x\ i \in A\} \;\middle|\; i \in I, A \in \mathcal{G}\ i \right\}$

When the index set I is finite, the product σ-algebra $\mathcal{A}_\otimes$ is generated by products over $G\ i \in \mathcal{G}\ i$ for each element $i \in I$. While the description of these generating sets is more complicated, they have the advantage that the sets are smaller. In fact the sets may have finite measures.

lemma PROD-GENERATOR-FINITE:

$$\textit{finite}\ I \ \wedge\ \Big(\forall i \in I.\ \mathcal{G}\ i \subseteq \mathcal{P}(\Omega_i)\ \wedge\ \mathcal{A}_i = \sigma\textit{-sets}\ \Omega_i\ (\mathcal{G}\ i)\ \wedge \\ (\bigcup_j C\ i\ j) = \Omega_i\ \wedge\ C\ i \in \mathbb{N} \to \mathcal{G}\ i\Big) \implies \\ \mathcal{A}_\otimes = \sigma\textit{-sets}\ \Omega_\otimes \left\{ \times_{i\in I} A\ i \;\middle|\; \forall i \in I.\ A\ i \in \mathcal{G}\ i \right\}$$

This lemma will be used in Section 3.5 to prove that the product space of Borel sets is isomorphic to the Borel sets on the Euclidean space. This then allows us to prove that the Lebesgue measure on the Euclidean space equals the product of Lebesgue measures on the real line.

3.4.4 Finite Product Measures

In the previous section we proved lemmas about the measurable sets of the product space. Now we show that for a finite index set I the product *measure* exists, i.e. cubes are mapped to the multiplication of the measures of their factors. We only show the finite case in this section, deferring the infinite case to Section 4.5.

We assume an index set I, and that all measure spaces $\mathcal{M}_i$ are σ-finite.[2]

theorem μ-$\bigotimes$:

$$\textit{finite}\ I\ \wedge\ (\forall i \in I.\ A\ i \in \mathcal{A}_i) \implies \mu_\otimes\ (\times_{i\in I} A\ i) = \prod_{i\in I} \mu_i\ (A\ i)$$

We prove this equation by induction on I. In the induction case we use the binary product to construct a measure space with an additional factor. This works as the binary product measure requires σ-finiteness for only one of the measures.

For $I = \emptyset$, the product measure is equal to the count measure of the singleton set $\{\lambda\_.\ \textit{undefined}\}$. This is convenient as we get a probability measure in the case $\bigotimes_{i\in\emptyset} \mathcal{M}\ i$.

lemma $\bigotimes$-EMPTY: $\displaystyle\bigotimes_{i\in\emptyset} \mathcal{M}\ i = \textit{count}\ \{\lambda\_.\ \textit{undefined}\}$

The finite product measure is again σ-finite. The σ-cover uses the increasing σ-covers of the factors.

lemma σ-FINITE-$\bigotimes$: $\textit{finite}\ I \implies \sigma\textit{-finite-measure}\ (\bigotimes_{i\in I} \mathcal{M}_i)$

The merge function we defined in the previous section relates the binary measure $(\bigotimes_{i\in I} \mathcal{M}_i) \otimes (\bigotimes_{i\in J} \mathcal{M}_i)$ with the finite product measure $\bigotimes_{i\in I\cup J} \mathcal{M}_i$. With the previous lemmas we show that it is measure preserving between these measures:

lemma DISTR-MERGE:

$$\textit{finite}\ I_1\ \wedge\ \textit{finite}\ I_2\ \wedge\ I_1 \cap I_2 = \emptyset \implies \\ \textit{distr}\ ((\bigotimes_{i\in I_1} \mathcal{M}_i) \otimes (\bigotimes_{i\in I} \mathcal{M}_i))\ (\bigotimes_{i\in I_1\cup I_2} \mathcal{M}_i)\ (\textit{merge}\ I_1\ I_2) = \\ (\bigotimes_{i\in I_1\cup I_2} \mathcal{M}_i)$$

[2] This assumption is on all i, not only the ones in I. It is actually a locale assumption and allows us to show that the factors are in the σ-finite measure sublocale.

Similarly, we show that on a product with only one factor the component projection is measure preserving:

lemma DISTR-SINGLETON: $\mathit{distr}\,(\bigotimes_{i\in\{i\}}\mathcal{M}_i)\;\mathcal{M}_i\;(\lambda x.\;x\;i) = \mathcal{M}_i$

Sometimes the Lebesgue integral over a finite product can be solved by induction over the index set. For this we provide unfolding rules for the Lebesgue integral on nonnegative functions, or integrable functions.

theorem $\bigotimes$-$\int^P$-FOLD-INSERT:
$i \notin I \;\wedge\; \mathit{finite}\;I \;\wedge\; f \in \mathit{measurable}\,(\bigotimes_{i\in I\cup\{i\}}\mathcal{M}_i)\;\mathcal{B}_{\overline{\mathbb{R}}} \implies$
$$\int^P f\;d\bigotimes_{i\in I\cup\{i\}}\mathcal{M}_i = \int^P \omega.\;\left(\int^P x.\;f(\omega(i := x))\;d\mathcal{M}_i\right)\;d\bigotimes_{i\in I}\mathcal{M}_i$$

theorem $\bigotimes$-$\int$-FOLD-INSERT:
$i \notin I \;\wedge\; \mathit{finite}\;I \;\wedge\; \mathit{integrable}\,(\bigotimes_{i\in I\cup\{i\}}\mathcal{M}_i)\;f \implies$
$$\int f\;d\bigotimes_{i\in I\cup\{i\}}\mathcal{M}_i = \int\omega.\;\left(\int x.\;f(\omega(i := x))\;d\mathcal{M}_i\right)\;d\bigotimes_{i\in I}\mathcal{M}_i$$

We prove these lemmas by splitting the finite product into a binary product by Lemma DISTR-MERGE and Lemma DISTR-SINGLETON and finally applying Fubini on the binary product.

The last lemma proves the distributivity of multiplication and integration by induction on I. First we show the distributivity for nonnegative functions:

corollary $\bigotimes$-$\int^P$-SETPROD:
$\mathit{finite}\;I \;\wedge\; (\forall i \in I.\;f\;i \in \mathit{measurable}\;\mathcal{M}_i\;\mathcal{B}_{\overline{\mathbb{R}}}) \;\wedge\; (\forall i \in I, \omega.\;0 \le f\;i\;\omega) \implies$
$$\int^P \omega.\;\prod_{i\in I} f\;i\;(\omega\;i)\;d\bigotimes_{i\in I}\mathcal{M}_i = \prod_{i\in I}\int^P f\;i\;d\mathcal{M}_i$$

With this we show the integrability. Using again the unfolding lemmas and induction over the index I we show the distribution of the Lebesgue integral.

corollary $\bigotimes$-INTEGRABLE-SETPROD, $\bigotimes$-$\int$-SETPROD:
$\mathit{finite}\;I \;\wedge\; (\forall i \in I.\;\mathit{integrable}\;\mathcal{M}_i\;(f\;i)) \implies$
$\mathit{integrable}\,(\bigotimes_{i\in I}\mathcal{M}_i)\;\;(\lambda\omega.\;\prod_{i\in I} f\;i\;(\omega\;i))\,,$
$$\int\omega.\;\prod_{i\in I} f\;i\;(\omega\;i)\;d\bigotimes_{i\in I}\mathcal{M}_i = \prod_{i\in I}\int f\;i\;d\mathcal{M}_i$$

3.5 Lebesgue Measure

An important measure is the *Lebesgue measure* λ assigning each interval its length: $\lambda\;\{a\;..<\;b\} = b - a$. Usually the Lebesgue measure is constructed using Caratheodory's extension theorem (see Remark 3 in §8 of Bauer [9], or Chapter 1 of Ash [4]). The measurable sets are the Borel sets, hence this measure is called *Lebesgue-Borel measure*, its completion is then the Lebesgue measure. Another more direct way is to to use the outer Lebesgue measure λ^* and chose all λ^*-measurable sets as Lebesgue measurable sets (see Chapter 1 of Gordon [25]). Both definitions result in the same measure (see Problem 3 in Chapter 1 of Ash [4]).

Instead of following one of these constructions, we use the gauge integral (also called the Henstock-Kurzweil integral) available in the multivariate analysis in Isabelle/HOL.[3] The gauge integral is an extension of the Riemann integral and of the Lebesgue integral on Euclidean vector spaces (see Chapter 9 of Gordon [25]). We use this construction as it simplifies the existence proof of the Lebesgue measure and we can easily relate the gauge integral to the Lebesgue integral.

In Isabelle/HOL the predicate *HK-integrable* A f states that the function f is gauge integrable on the set A, in which case the gauge integral of f on the set A has the real value *HK-integral* A f. The gauge measure of a set A is the gauge integral of the constant 1 function on A. Since the gauge measure is restricted to sets with a finite measure, it cannot be directly used as Lebesgue measure since it is not a σ-algebra. However we can measure the indicator function χ A on the intervals $\{-n \mathinner{..} n\}$ for all natural numbers n. When χ A is measurable on all intervals, we deem it as Lebesgue measurable and the Lebesgue measure is the supremum of the gauge measures for all intervals $\{-n \mathinner{..} n\}$. To define the Lebesgue measure on multidimensional Euclidean spaces we use hypercubes $\{x \mid \forall i.\ |x_i| \leq n\}$. The σ-algebra of the Lebesgue measure on a Euclidean space $\mathbb{R}^n$ consists of all sets A gauge measurable set on all hypercubes.

$$
\begin{array}{lcl}
\mathit{cube} & :: & \mathbb{N} \to \mathbb{R}^n \ \mathit{set} \\
\mathit{cube}\ n & = & \{x \mid \forall i.\ |x_i| \leq n\} \\
\lambda^{\mathrm{HK}}_{\mathbb{R}^n} & :: & \mathbb{R}^n \ \mathit{measure} \\
\lambda^{\mathrm{HK}}_{\mathbb{R}^n} & = & \mathit{measure\text{-}of}\ \mathcal{U}_{\mathbb{R}^n}\ \{A \mid \forall n.\ \mathit{HK\text{-}integrable}\ (\mathit{cube}\ n)\ (\chi\ A)\} \\
 & & \qquad \Big(\lambda A.\ \sup_n \mathit{HK\text{-}integral}\ (\mathit{cube}\ n)\ (\chi\ A)\Big)
\end{array}
$$

The gauge integral is monotone convergent, hence the measurable sets form a σ-algebra:

lemma σ-ALGEBRA-LEBESGUE:
$\sigma\text{-}\mathit{algebra}\ \mathcal{U}_{\mathbb{R}^n}\ \{A \mid \forall n.\ \mathit{HK\text{-}integrable}\ (\mathit{cube}\ n)\ (\chi\ A)\}$

lemma SPACE-LEBESGUE:
$\Omega_{\lambda^{\mathrm{HK}}_{\mathbb{R}^n}} = \mathcal{U}_{\mathbb{R}^n}$

lemma SETS-LEBESGUE:
$\mathcal{A}_{\lambda^{\mathrm{HK}}_{\mathbb{R}^n}} = \{A \mid \forall n.\ \mathit{HK\text{-}integrable}\ (\mathit{cube}\ n)\ (\chi\ A)\}$

From the monotone convergence of the gauge integral it follows also that it forms a measure space, mapping cubes to the product of their edge lengths. Hence, the Lebesgue measure forms a σ-finite measure space.

theorem μ-LEBESGUE:
$A \in \mathcal{A}_{\lambda^{\mathrm{HK}}_{\mathbb{R}^n}} \implies \mu_{\lambda^{\mathrm{HK}}_{\mathbb{R}^n}} A = \sup_n \mathit{HK\text{-}integral}\ (\mathit{cube}\ n)\ (\chi\ A)$

corollary μ-LEBESGUE-ATLEASTATMOST:
$a \leq b \implies \mu_{\lambda^{\mathrm{HK}}_{\mathbb{R}^n}} \{a \mathinner{..} b\} = \prod_{i<n}(b_i - a_i)$

lemma σ-FINITE-LEBESGUE:
$\sigma\text{-}\mathit{finite\text{-}measure}\ \lambda^{\mathrm{HK}}_{\mathbb{R}^n}$

[3] The multivariate analysis in Isabelle/HOL is ported from a later version of [31].

3.5.1 Lebesgue-Borel Measure

We know that $\lambda^{\text{HK}}_{\mathbb{R}^n}$ is a σ-algebra and since all intervals $\{a \mathinner{..} b\}$ are Lebesgue measurable all Borel sets are Lebesgue measurable:

lemma LEBESGUE-BOREL: $A \in \mathcal{A}_{\mathcal{B}_{\mathbb{R}^n}} \implies A \in \mathcal{A}_{\lambda^{\text{HK}}_{\mathbb{R}^n}}$

But the Lebesgue measure is not generated by $\{a \mathinner{..} b\}$. It contains not only all Borel sets, but is also complete, i.e. for each null set it also contains all subsets:

lemma LEBESGUE-COMPLETE:
$A \subseteq B \wedge B \in \mathit{null\text{-}sets}_{\lambda^{\text{HK}}_{\mathbb{R}^n}} \implies A \in \mathit{null\text{-}sets}_{\lambda^{\text{HK}}_{\mathbb{R}^n}}$

The Lebesgue measure is *the* completion of the Lebesgue-Borel measure (see Chapter 1 in Ash [4]). With these two lemmas we only proved that the Lebesgue measurable sets include the completion of the Borel sets. We do not prove the other direction, as it is complicated and we do not have an application for it.

While the Lebesgue measure admits more measurable sets than the Borel sets, its measurable sets are complicated to handle (see Remark 3 in §8 in Bauer [9]). It is easier to work on the Lebesgue-Borel measure, where the measurable sets are the Borel sets, with its nice generation property. For this, we introduce the Lebesgue-Borel measure by changing the measurable sets from the Lebesgue sets to the Borel sets.

$\lambda_{\mathbb{R}^n} :: \mathbb{R}^n \mathit{\ measure}$
$\lambda_{\mathbb{R}^n} = \mathit{measure\text{-}of}\ \mathcal{U}_{\mathbb{R}^n}\ \mathcal{A}_{\mathcal{B}_{\mathbb{R}^n}}\ \mu_{\lambda^{\text{HK}}_{\mathbb{R}^n}}$

Now the measure $\lambda_{\mathbb{R}^n}$ has the Borel sets as measurable sets, assigns to each cube the product of its edge lengths as measure, and hence is σ-finite:

lemma SPACE-$\lambda_{\mathbb{R}^n}$: $\Omega_{\lambda_{\mathbb{R}^n}} = \mathcal{U}_{\mathbb{R}^n}$

lemma SETS-$\lambda_{\mathbb{R}^n}$: $\mathcal{A}_{\lambda_{\mathbb{R}^n}} = \mathcal{A}_{\mathcal{B}_{\mathbb{R}^n}}$

lemma μ-$\lambda_{\mathbb{R}^n}$-ATMOSTATLEAST: $a \leq b \implies \mu_{\lambda_{\mathbb{R}^n}}\ \{a \mathinner{..} b\} = \prod_{i<n}(b_i - a_i)$

lemma σ-FINITE-$\lambda_{\mathbb{R}^n}$: $\sigma\mathit{\text{-}finite\text{-}measure}\ \lambda_{\mathbb{R}^n}$

As application of the fact that $\lambda_{\mathbb{R}^n}$ is generated by all intervals $\{a \mathinner{..} b\}$, we use Theorem MEASURE-EQI-GENERATOR-EQ to show that $\lambda_{\mathbb{R}^n}$ is equal to other measures introduced on the Borel sets and based on the volume of cubes.

theorem $\lambda_{\mathbb{R}^n}$-EQI:
$\Big(\forall a\ b.\ a \leq b \implies \mu_{\mathcal{M}}\ \{a \mathinner{..} b\} = \prod_{i<n}(b_i - a_i)\Big) \wedge \mathcal{A}_{\mathcal{M}} = \mathcal{A}_{\mathcal{B}_{\mathbb{R}^n}} \implies$
$\mathcal{M} = \lambda_{\mathbb{R}^n}$

We use this to show how to apply an affine transformation to the integral:

corollary $\lambda_{\mathbb{R}^n}$-REAL-AFFINE:
$c \neq 0 \implies \lambda_{\mathbb{R}} = \mathit{density}\ \big(\mathit{distr}\ \lambda_{\mathbb{R}}\ \mathcal{B}_{\mathbb{R}}\ (\lambda x.\ t + c \cdot x)\big)\ (\lambda\_.\ |c|)$

corollary $\lambda_{\mathbb{R}^n}$-$\int$-REAL-AFFINE:
$c \neq 0 \implies \int f\ d\lambda_{\mathbb{R}} = |c| \cdot \int x.\ f\ (t + c \cdot x)\ d\lambda_{\mathbb{R}}$

We will see another application in Section 3.5.3 when we equate the Lebesgue-Borel measure on multidimensional Euclidean spaces to the product of Lebesgue-Borel measures.

3.5.2 Lebesgue Integral and Gauge Integral

From the linearity of the gauge integral and from our definition of the Lebesgue measure it is easy to see that all Lebesgue measurable simple functions whose integral is finite are also gauge integrable. With the monotone convergence of the gauge integral we show that all nonnegative Lebesgue measurable functions with a finite integral are gauge integrable. And finally we show that all Lebesgue integrable functions are gauge integrable.

theorem HAS-INTEGRAL-IFF-$\int^P$-LEBESGUE:
$f \in \mathit{measurable}\ \lambda_{\mathbb{R}^n}^{\mathrm{HK}}\ \mathcal{B}_{\mathbb{R}}\ \wedge\ (\forall x.\ 0 \leq f\ x) \implies$
$\mathit{HK\text{-}has\text{-}integral}\ \mathcal{U}_{\mathbb{R}^n}\ f\ I \Leftrightarrow \left(\int^P f\ d\lambda_{\mathbb{R}^n}^{\mathrm{HK}} = (I)_{\overline{\mathbb{R}}} \right)$

corollary LEBESGUE-HAS-INTEGRAL:
$\mathit{integrable}\ \lambda_{\mathbb{R}^n}^{\mathrm{HK}}\ f \implies \mathit{HK\text{-}has\text{-}integral}\ \mathcal{U}_{\mathbb{R}^n}\ f \left(\int f\ d\lambda_{\mathbb{R}^n}^{\mathrm{HK}} \right)$

Please note that the right-side of Theorem HAS-INTEGRAL-IFF-$\int^P$-LEBESGUE equates two $\overline{\mathbb{R}}$ values, so it forces the integral to be finite.

The Lebesgue-Borel measure is defined as a sub-σ-algebra of the Lebesgue measure. Hence, for Borel-measurable functions, Lebesgue integrability equals Lebesgue-Borel integrability:

lemma $\lambda_{\mathbb{R}^n}^{\mathrm{HK}}$-$\int$-EQ-$\lambda_{\mathbb{R}^n}$:
$f \in \mathit{measurable}\ \mathcal{B}_{\mathbb{R}^n}\ \mathcal{B}_{\mathbb{R}} \implies$
$\mathit{integrable}\ \lambda_{\mathbb{R}^n}^{\mathrm{HK}}\ f \Leftrightarrow \mathit{integrable}\ \lambda_{\mathbb{R}^n}\ f,$
$\int f\ d\lambda_{\mathbb{R}^n}^{\mathrm{HK}} = \int f\ d\lambda_{\mathbb{R}^n}$

We lift now the fundamental theorem of calculus (FTC) for the gauge integral to the Lebesgue integral on the Lebesgue-Borel measure. First, we show that a function continuous on a closed interval is also integrable on this interval:

theorem INTEGRABLE-ATLEASTATMOST-ISCONT:
$a \leq b\ \wedge\ \mathit{continuous\text{-}on}\ \{a\ ..\ b\}\ f \implies \mathit{integrable}\ \lambda_{\mathbb{R}}\ (\lambda x.\ f\ x \cdot \chi\ \{a\ ..\ b\}\ x)$

In the next step we show FTC:

corollary $\int$-FTC:
$a \leq b\ \wedge\ \mathit{continuous\text{-}on}\ \{a\ ..\ b\}\ f\ \wedge\ \mathit{differentiable\text{-}on}\ \{a\ ..\ b\}\ F\ f \implies$
$\int x.\ f\ x \cdot \chi\ \{a\ ..\ b\}\ x\ d\lambda_{\mathbb{R}} = F\ b - F\ a$

Here $\mathit{differentiable\text{-}on}\ \{a\ ..\ b\}\ F\ f$ states that f is the differential of F on $\{a\ ..\ b\}$. The proof simply equates the Lebesgue-Borel integral to the gauge integral and then uses the FTC on the gauge integral.

3.5.3 Euclidean Spaces and Product Measures

We relate the Euclidean space $\mathbb{R}^n$ with the n-dimensional product of Lebesgue-Borel measures:[4]

$$\begin{aligned} \lambda^n &:: \left(\mathbb{N} \to \mathbb{R}\right) \textit{ measure} \\ \lambda^n &= \left(\bigotimes_{i \in \{1 \,..\, n\}} \lambda_{\mathbb{R}}\right) \end{aligned}$$

The function $p2e :: (\mathbb{N} \to \mathbb{R}) \to \mathbb{R}^n$ maps functions to vectors with $(p2e\ f)_i = f\ i$. We use Lemma PROD-GENERATOR-FINITE to show that $p2e$ is measurable:

lemma P2E-MEASURABLE: $p2e \in \textit{measurable}\ \lambda^n\ \mathcal{B}_{\mathbb{R}^n}$

With Theorem MEASURE-EQI-GENERATOR-EQ we show that it is measure preserving from λ^n to $\lambda_{\mathbb{R}^n}$.

theorem $\lambda_{\mathbb{R}^n}$-EQ-λ^n: $\lambda_{\mathbb{R}^n} = \textit{distr}\ \lambda^n\ \mathcal{B}_{\mathbb{R}^n}\ p2e$

With this, the Theorem $\int^P$-DISTR, and the Corollaries $\int$-DISTR, and INTEGRABLE-DISTR-EQ, it follows the equivalence of integrals:

corollary $\lambda_{\mathbb{R}^n}$-$\int^P$:
$f \in \textit{measurable}\ \mathcal{B}_{\mathbb{R}^n}\ \mathcal{B}_{\overline{\mathbb{R}}} \implies \int^P f\ d\lambda_{\mathbb{R}^n} = \int^P x.\ f\ (p2e\ x)\ d\lambda^n$

lemma $\lambda_{\mathbb{R}^n}$-INTEGRABLE, -$\int$:
$f \in \textit{measurable}\ \mathcal{B}_{\mathbb{R}^n}\ \mathcal{B}_{\mathbb{R}} \implies$
$\textit{integrable}\ \lambda_{\mathbb{R}^n}\ f \Leftrightarrow \textit{integrable}\ \lambda^n\ (f \circ p2e)$,
$\int f\ d\lambda_{\mathbb{R}^n} = \int x.\ f\ (p2e\ x)\ d\lambda^n$

These lemmas allow now proofs by induction over the dimension. While the Euclidean vector space formalization in Isabelle/HOL includes the dimensionality in the vector type and hence it is not possible to use induction over the dimensionality of the Euclidean space. With Lemmas $\lambda^{\text{HK}}_{\mathbb{R}^n}$-$\int$-EQ-$\lambda_{\mathbb{R}^n}$ and $\lambda_{\mathbb{R}^n}$-EQ-λ^n we equate the gauge integral to the Lebesgue integral over λ^n, we then use induction over n.

[4] n is the dimension of the Euclidean space $\mathbb{R}^n$.

Chapter 4

Probability

The concepts from measure theory, like measure, measurable sets and functions and Lebesgue integration map to probability, events, random variables and expectation. We introduce probability measures as measures with measure value 1 for the entire space. The other concepts map one-to-one onto the probabilistic concepts. Instead of introducing new constants we use the measure-theoretic ones.

For probability theory we formalize the following concepts:

Independence of events states that the occurrence of one of these events does not influence the occurrence of the others. While introductory textbooks often only introduce independence between *two* events or random variables, we introduce independence on *indexed families* of events and of random variables.

Distribution of a random variable is the push-forward measure of the random variable. When analysing the properties of random variables we often ignore the concrete result values and only look at the distribution. For this we equate the distribution of a random variable to a density measure.

Information theory quantifies the information stored in a random variable (entropy) or shared between two random variables (mutual information). We formalize entropy and mutual information and their conditional versions.

Infinite products $\pi = \prod_{i \in I} \mu_i$ extends the finite product of measures to an infinite index I, this works at least with probability measures. For distinct indices $i_1, i_2, \ldots, i_n$, and when μ_{i_l} is defined on the sets A_l for all $l \leq n$, then we have:

$$\pi \{\omega \mid \omega\ i_1 \in A_1 \wedge \omega\ i_2 \in A_2 \wedge \cdots \wedge \omega\ i_n \in A_n\} = \mu_{i_1}\ A_1 \cdot \mu_{i_2}\ A_2 \cdots \mu_{i_n}\ A_n$$

This is used in probability theory to construct a probability space with infinitely many independent random variables.

Trace measure τ is the stochastic process of a Markov chain defined by a transition matrix T. Where $T\ s\ t$ is the probability that the Markov chain transitions from state s into state t. The probability that a trace of a Markov chain starts with the states $s_1, s_2, \ldots, s_n$ is

$$\tau \{\omega \mid \forall i \leq n.\ \omega_i = s_i\} = \prod_{i<n} T\ s_i\ s_{i+1} \quad .$$

4.1 Probability Measures

A *probability measure* is a finite measure $\mathcal{M}$ where the measure value of the space is 1. The finite measure value $\mu^f_{\mathcal{M}}\ A$ for a measurable set A of a probability measure $\mathcal{M}$ is also called probability of A.

locale *prob-measure* = *finite-measure* $\mathcal{M}$ +
  assumes $\mu_{\mathcal{M}}\ \Omega_{\mathcal{M}} = 1$

lemma PROB-SPACEI: $\mu_{\mathcal{M}}\ \Omega_{\mathcal{M}} = 1 \implies$ *prob-measure* $\mathcal{M}$

Notation: *In this chapter we assume that $\mathcal{M}$ is a probability measure. We write* $\Pr_{\mathcal{M}}\ A = \mu^f_{\mathcal{M}}\ A$ *for the probability of a set A, and we write* $\Pr_{\mathcal{M}}(x.\ P\ x) = \mu^f_{\mathcal{M}}\ \{x \mid P\ x\}$ *for the probability of a measurable predicate P. We omit the probability measure $\mathcal{M}$ when writing the probability* $\Pr(x.\ P\ x) = \Pr_{\mathcal{M}}(x.\ P\ x)$ *and* $\Pr\ A = \Pr_{\mathcal{M}}\ A$.

4.1.1 Random Variables

A function f is a *random variable*[1] on the probability measure $\mathcal{M}$ into a measure space $\mathcal{N}$ if $f \in$ *measurable* $\mathcal{M}\ \mathcal{N}$. For many theorems in probability theory we assume a probability measure with a collection of random variables on it. The push-forward measure of a random variable f allows us to reason about the probability distribution of f. The push-forward measure of a random variable f is again a probability measure:

lemma PROB-SPACE-DISTR:
  $f \in$ *measurable* $\mathcal{M}\ \mathcal{N} \implies$ *prob-measure* (*distr* $\mathcal{M}\ \mathcal{N}\ f$)

4.1.2 Conditional Probability

If the probability of an event A depends on another event B, e.g. A is the outcome of a dice and B tells us if the dice was odd, we may want to ask what the probability is that A happens when we know that B happened. This is the *conditional probability of A given B*. Instead of events we define conditional probability on measurable predicates:

$$\begin{array}{lcl} \Pr_{\square}(x.\ \square \mid \square) & :: & \alpha\ \textit{measure} \to \Big(\alpha \to \mathbb{B}\Big) \to \Big(\alpha \to \mathbb{B}\Big) \to \mathbb{R} \\ \Pr_{\mathcal{M}}(x.\ P\ x \mid Q\ x) & = & \Pr(x.\ P\ x \wedge Q\ x) / \Pr(x.\ Q\ x) \end{array}$$

It is easy to see that conditional probability introduces a measure depending on the given event Q. We introduce it by using density and weigh each element in Q with the inverse of the probability of Q. Elements not in Q are weigh with 0.

$$\begin{array}{lcl} \textit{cond-measure} & :: & \alpha\ \textit{measure} \to \Big(\alpha \to \mathbb{B}\Big) \to \alpha\ \textit{measure} \\ \textit{cond-measure}\ \mathcal{M}\ Q & = & \textit{density}\ \mathcal{M}\ \Big(\lambda x.\ \chi\ \{x \mid Q\ x\}\ x / \Pr(x.\ Q\ x)\Big) \end{array}$$

[1] Some textbooks call it a random object or element [43], and only random variable when into $\mathbb{R}$.

When the probability of Q is not 0, then we have a probability measure:

lemma PROB-SPACE-COND-MEASURE:
$\Pr(x.\ Q\ x) \neq 0 \implies \textit{prob-measure}\ (\textit{cond-measure}\ \mathcal{M}\ Q)$

From $\Pr(x.\ Q\ x) \neq 0$ also follows that $\{x \mid Q\ x\}$ is measurable.

The conditional probability of a predicate P given a predicate Q equals the probability of P on the probability measure *cond-measure* $\mathcal{M}$ Q:

lemma PR-COND-MEASURE:
$\Pr(x.\ Q\ x) \neq 0 \ \wedge\ \{x \mid P\ x\} \in \mathcal{A}_{\mathcal{M}} \implies$
$\Pr_{\textit{cond-measure}\ \mathcal{M}\ Q}(x.\ P\ x) = \Pr(x.\ P\ x \mid Q\ x)$

4.1.3 Jensen's Inequality

The Lebesgue integral gains some nice properties when used on a probability measure. First we get some strict inequalities when the integrand is strictly less than or strictly greater than some boundary:

lemma $\int$-LESS: $\quad \textit{integrable}\ \mathcal{M}\ X \ \wedge\ \left(\mathrm{AE}_{\mathcal{M}}\ x.\ X\ x < b\right) \implies \int X\ d\mathcal{M} < b$

lemma $\int$-GREATER: $\quad \textit{integrable}\ \mathcal{M}\ X \ \wedge\ \left(\mathrm{AE}_{\mathcal{M}}\ x.\ a < X\ x\right) \implies a < \int X\ d\mathcal{M}$

Another important inequality is *Jensen's inequality*. It generalizes the fact that a convex function q at the middle of the interval $\{a\ ..\ b\}$ is less than or equal to the average of the values at the endpoints of the interval: $q\ ((a+b)/2) \leq (q\ a + q\ b)/2$. Jensen's inequality generalizes middle and average to integration. First, we defined the concept of convex functions on a carrier I:

$\textit{convex-on} \quad :: \quad \mathbb{R}\ \textit{set} \to (\mathbb{R} \to \mathbb{R}) \to \mathbb{B}$
$\textit{convex-on}\ I\ f \quad \Leftrightarrow$
$\left(\forall x, y \in I.\ \forall u, v \geq 0.\ u + v = 1 \implies f\ (u \cdot x + v \cdot y) \leq u \cdot f\ x + v \cdot f\ y\right)$

An example for a convex function is the vertically mirrored logarithm. We need its convexity later in information theory.

lemma MINUS-LOG-CONVEX: $1 < b \implies \textit{convex-on}\ \{0 <..\}\ (\lambda x.\ -\log_b x)$

Jensen's inequality is about a convex function q on an open and convex domain I. Hence I is an open interval allowing infinite endpoints.

theorem JENSENS-INEQUALITY:
$\left(I = \{a <..< b\} \ \vee\ I = \{a <..\} \ \vee\ I = \{..< b\} \ \vee\ I = \mathcal{U}_{\mathbb{R}}\right) \wedge$
$\textit{integrable}\ \mathcal{M}\ X \ \wedge\ \left(\mathrm{AE}_{\mathcal{M}}\ x.\ X\ x \in I\right) \wedge$
$\textit{integrable}\ \mathcal{M}\ (\lambda x.\ q\ (X\ x)) \ \wedge\ \textit{convex-on}\ I\ q \implies$
$q\ \left(\int X\ d\mathcal{M}\right) \leq \int x.\ q\ (X\ x)\ d\mathcal{M}$

For information theory Jensen's inequality is helpful to show that the Kullback-Leibler divergence and hence also mutual information is always nonnegative.

4.2 Families of Independent Sets and Functions

Two sets A and B are independent if $\Pr\ (A \cap B) = \Pr\ A \cdot \Pr\ B$. Independence is generalized to families of sets A_i indexed by $i \in I$: for each finite, nonempty subset J of the index set I the independence property $\Pr\ (\bigcap_{j\in J} A_j) = \prod_{j\in J} \Pr\ A_j$ holds. Note that this is stronger than the pairwise independence of each A_i and A_j. For example: assume we have the uniform probability measure on the space $\{0,1\} \times \{0,1\}$. The sets $\{(0,0),(0,1)\}$, $\{(0,0),(1,0)\}$, and $\{(0,0),(1,1)\}$ are pairwise independent, but all three together do not form an independent family.

4.2.1 Independent Sets of Sets

We introduce independence for an indexed family of sets of sets:

$$\begin{aligned}
&\mathit{indep\text{-}sets} \quad :: \quad \Big(\iota \to \alpha\ \mathit{set}\ \mathit{set}\Big) \to \iota\ \mathit{set} \to \mathbb{B} \\
&\mathit{indep\text{-}sets}\ F\ I \quad \Leftrightarrow \quad \Big(F \in I \to \mathcal{P}(\mathcal{A}_{\mathcal{M}}) \ \wedge\ \Big(\forall J \subseteq I.\ J \neq \emptyset \ \wedge\ \mathit{finite}\ J \implies \\
&\quad \forall A \in \times_{j\in J} F\ j.\ \Pr\ (\textstyle\bigcap_{j\in J} A\ j) = \prod_{j\in J} \Pr(A\ j)\Big)\Big)
\end{aligned}$$

This predicate *indep-sets* is monotone in the index set and in the sets of sets, i.e. we can restrict the index set and the range of F:

lemma INDEP-SETS-MONO:
$\mathit{indep\text{-}sets}\ F\ I \ \wedge\ J \subseteq I \ \wedge\ (\forall i \in J.\ G\ i \subseteq F\ i) \implies \mathit{indep\text{-}sets}\ G\ J$

So, we know that subsets are again independent, but is it possible to deduce independence of further sets? When A and B are independent then the complement of a A is also independent:

$$\Pr\ ((\Omega - A) \cap B) = \Pr\ (B - (A \cap B)) = \Pr\ B - \Pr(A \cap B) = (1 - \Pr\ A) \cdot \Pr\ B$$

This also works for the union of disjoint sets. Hence, for an independent family F the Dynkin closures of each element in F is independent:

theorem INDEP-SETS-DYNKIN-SETS:
$\mathit{indep\text{-}sets}\ F\ I \implies \mathit{indep\text{-}sets}\ (\lambda i.\ \mathit{dynkin\text{-}sets}\ \Omega_{\mathcal{M}}\ (F\ i))\ I$

When the family of independent sets F is $\cap$-stable, we know with Theorem σ-SETS-EQ-DYNKIN-SETS that the Dynkin closure is equal to the σ-closure. Hence the σ-closures of the sets of an independent family are again independent.

corollary INDEP-SETS-σ-SETS:
$\mathit{indep\text{-}sets}\ F\ I \ \wedge\ \big(\forall i \in I.\ \cap\text{-}\mathit{stable}\ (F\ i)\big) \implies$
$\mathit{indep\text{-}sets}\ (\lambda i.\ \sigma\text{-}\mathit{sets}\ \Omega_{\mathcal{M}}\ (F\ i))\ I$

We strengthen this rule by partitioning the index I of an independent family, and we then show that the σ-algebras generated by the union of each part are also

independent. The index set is now the union over all I, where J indexes the parts.

lemma INDEP-SETS-COLLECT-σ-SETS:
$\mathit{indep\text{-}sets}\ F\ (\bigcup_{j\in J} I\ j)\ \wedge\ (\forall j \in J, i \in I\ j.\ \cap\text{-}\mathit{stable}\ (F\ i))\ \wedge$
$\mathit{disjoint\text{-}family}_J\ I\ \Longrightarrow$
$\mathit{indep\text{-}sets}\ (\lambda j.\ \sigma\text{-}\mathit{sets}\ \Omega_{\mathcal{M}}\ (\bigcup_{i\in I\ j} F\ i))\ J$

4.2.2 Independent Random Variables

Based on independent sets of sets we describe an indexed family of independent random variables:

$$\mathit{indep\text{-}vars} :: \Big(\iota \to \beta\ \mathit{measure}\Big) \to \Big(\iota \to \alpha \to \beta\Big) \to \iota\ \mathit{set} \to \mathbb{B}$$
$$\mathit{indep\text{-}vars}\ \mathcal{N}\ X\ I \Leftrightarrow \Big(\Big(\forall i \in I.\ X\ i \in \mathit{measurable}\ \mathcal{M}\ (\mathcal{N}\ i)\Big) \wedge \mathit{indep\text{-}sets}\ \Big(\lambda i.\ \{\{x \mid X\ i\ x \in A\} \mid A \in \mathcal{A}_{\mathcal{N}\ i}\}\Big)\ I\Big)$$

Independent random variables obey a simple data flow rule: when a family of random variables is constructed such that each of these random variables is built out of different, independent random variables, then the result are again independent random variables. This rule follows from Lemma INDEP-SETS-COLLECT-σ-SETS. We prove a simpler variant, where each random variable is a composition with exactly one random variable from an independent family:

lemma INDEP-VARS-COMPOSE:
$\mathit{indep\text{-}vars}\ \mathcal{N}\ X\ I\ \wedge\ (\forall i \in I.\ Y\ i \in \mathit{measurable}\ (\mathcal{N}\ i)\ (\mathcal{L}\ i))\ \Longrightarrow$
$\mathit{indep\text{-}vars}\ \mathcal{L}\ (\lambda i.\ Y\ i \circ X\ i)\ I$

This lemma helps us to derive independence of random variables when we already know that the random variables they are based on are independent. But, how do we prove that random variables are independent in the first place? The definition of *indep-vars* assumes independence for each finite collection of measurable sets of $\mathcal{N}$, a σ-algebra. As usual we can simplify this by assuming that each $\mathcal{N}\ i$ is generated by a generator $\mathcal{G}\ i$. While it holds also for an infinite index set, we only prove the lemma for a finite index. This follows from Corollary INDEP-SETS-σ-SETS.

lemma INDEP-VARS-FINITE:
$I \neq \emptyset\ \wedge\ \mathit{finite}\ I\ \wedge\ (\forall i \in I.\ X\ i \in \mathit{measurable}\ \mathcal{M}\ (\mathcal{N}\ i))\ \wedge$
$(\forall i \in I.\ \mathcal{A}_{\mathcal{N}\ i} = \sigma\text{-}\mathit{sets}\ \Omega_{\mathcal{N}\ i}\ (\mathcal{G}\ i)\ \wedge\ \cap\text{-}\mathit{stable}\ (\mathcal{G}\ i)\ \wedge\ \Omega_{\mathcal{N}\ i} \in \mathcal{G}\ i\ \wedge$
$\quad \mathcal{G}\ i \subseteq \mathcal{P}(\Omega_{\mathcal{N}\ i}))\ \Longrightarrow$
$\mathit{indep\text{-}vars}\ \mathcal{N}\ X\ I \Leftrightarrow$
$\quad \big(\forall A \in (\times_{i\in I} \mathcal{G}\ i).\ \Pr(x.\ \forall i \in I.\ X\ i\ x \in A\ i) = \prod_{i\in I} \Pr(x.\ X\ i\ x \in A\ i)\big)$

The usual variant of independence is when we only look at two independent random variables:

$$\mathit{indep\text{-}var} :: \beta\ \mathit{measure}\beta\ \mathit{measure} \to\to (\alpha \to \beta) \to (\alpha \to \beta) \to \mathbb{B}$$
$$\mathit{indep\text{-}var}\ \mathcal{S}\ \mathcal{T}\ X\ Y \Leftrightarrow \mathit{indep\text{-}vars}\ (\lambda i.\ \textbf{if}\ i\ \textbf{then}\ \mathcal{S}\ \textbf{else}\ \mathcal{T})\ (\lambda i.\ \textbf{if}\ i\ \textbf{then}\ X\ \textbf{else}\ Y)\ \mathcal{U}_{\mathbb{B}}$$

For the case of two random variables, independence says that the joint distribution is equivalence to the product of both single distributions.

lemma INDEP-VAR-DISTRIBUTION-EQ:

$$\mathit{indep\text{-}var}\ \mathcal{S}\ \mathcal{T}\ X\ Y \Leftrightarrow \Big(X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{S}\ \wedge\ Y \in \mathit{measurable}\ \mathcal{M}\ \mathcal{T}\ \wedge \Big(\mathit{distr}\ \mathcal{M}\ \mathcal{S}\ X\Big) \otimes \Big(\mathit{distr}\ \mathcal{M}\ \mathcal{T}\ Y\Big) = \mathit{distr}\ \mathcal{M}\ \Big(\mathcal{S} \otimes \mathcal{T}\Big)\ \Big(\lambda x.\ (X\ x, Y\ x)\Big)\Big)$$

This also holds for family of independent random variables. However we need infinite products of probability spaces, hence we show it in Section 4.5.

4.2.3 Sequences of Independent Sets and 0-1-Laws

Often we want to have a more specialized version, where we only talk about a single set per index. For example we want to state that a sequence $F \in \mathbb{N} \to \alpha\ \mathit{set}$ is independent.

$$\begin{array}{lcl} \mathit{indep\text{-}events} & :: & \Big(\iota \to \alpha\ \mathit{set}\Big) \to \iota\ \mathit{set} \to \mathbb{B} \\ \mathit{indep\text{-}events}\ F\ I & \Leftrightarrow & \mathit{indep\text{-}sets}\ \big(\lambda i.\ \{F\ i\}\big)\ I \end{array}$$

As an application of independence we prove Kolmogorov's and Borel's 0-1-law. They are helpful to show that a limit of sets is either a.e.-true or a.e.-false. For Kolmogorov's 0-1-law, we start with introducing *tail events*. They are used to describe limiting properties of a sequence of σ-algebras A. These σ-algebras are often the σ-algebras generated by a sequence of random variables.

$$\begin{array}{lcl} \mathit{tail\text{-}events} & :: & \Big(\mathbb{N} \to \alpha\ \mathit{set}\ \mathit{set}\Big) \to \alpha\ \mathit{set}\ \mathit{set} \\ \mathit{tail\text{-}events}\ A & = & \bigcap_n \sigma\text{-}\mathit{sets}\ \Omega_{\mathcal{M}}\ \big(\bigcup_{i \geq n} A\ i\big) \end{array}$$

When A is a sequence of σ-algebras, then the tail events are also a σ-algebra:

lemma σ-ALGEBRA-TAIL-EVENTS:
$(\forall i.\ \sigma\text{-}\mathit{algebra}\ \Omega_{\mathcal{M}}\ (A\ i)) \implies \sigma\text{-}\mathit{algebra}\ \Omega_{\mathcal{M}}\ (\mathit{tail\text{-}events}\ A)$

lemma TAIL-EVENTS-SUBSET:
$(\forall i.\ A\ i \subseteq \mathcal{A}_{\mathcal{M}}) \implies \mathit{tail\text{-}events}\ A \subseteq \mathcal{A}_{\mathcal{M}}$

Kolmogorov's 0-1-law states that each tail event is either a.e. true or a.e. false.

theorem KOLMOGOROV-0-1-LAW:
$(\forall i.\ \sigma\text{-}\mathit{algebra}\ \Omega_{\mathcal{M}}\ (A\ i)) \wedge \mathit{indep\text{-}sets}\ A\ \mathcal{U}_{\mathbb{N}} \wedge X \in \mathit{tail\text{-}events}\ A \implies \Pr\ X = 0 \vee \Pr\ X = 1$

Using this we can easily show Borel's 0-1-law by instantiating A_i with the σ-algebra $\{\Omega, \Omega - F_i, F_i, \emptyset\}$, then $\bigcap_n \bigcup_{m \geq n} F_m$ is a tail event.

corollary BOREL-0-1-LAW:
$\mathit{indep\text{-}events}\ F\ \mathcal{U}_{\mathbb{N}} \implies \Pr\ (\bigcap_n \bigcup_{m \geq n} F\ m) = 0 \vee \Pr\ (\bigcap_n \bigcup_{m \geq n} F\ m) = 1$

4.3 Distributions of Random Variables

The *distribution of a random variable X into $\mathcal{S}$* is defined as its push-forward measure:[2] *distr* $\mathcal{M}$ $\mathcal{S}$ X, see Kallenberg [43]. When the measure space $\mathcal{S}$ is the Borel σ-algebra, the distribution is often expressed as *cumulative distribution function (cdf)* $F_X\ a = \Pr(x.\ X\ x \leq a)$. With Theorem MEASURE-EQI-GENERATOR-EQ it is easy to show that the cdf F_X uniquely determines the distribution of X. When the distribution can be expressed as the measure $\mathcal{S}$ with density f the function f is called *probability density function (pdf)* (for a discrete random variable X the function f is also called *probability mass function (pmf)*).

Distributions of random variables are mostly expressed as pdf on a counting space or on the Lebesgue measure. For example, the exponential distribution has the pdf $\lambda x.\ l \cdot \exp^{-l \cdot x}$, where l is a parameter for the exponential distribution. When the distribution of a random variable has a pdf it helps us to compute its probability, expectation, variance, and entropy by integration.

We introduce a predicate to easily express the distribution of a random variable. This predicate *distributed* $\mathcal{S}$ X P_X states that X is a random variable with the measure space $\mathcal{S}$ as range and that X has the pdf P_X.

$$\begin{array}{lcl} \mathit{distributed} & :: & \beta\ \mathit{measure} \to \Big(\alpha \to \beta\Big) \to \Big(\beta \to \overline{\mathbb{R}}\Big) \to \mathbb{B} \\ \mathit{distributed}\ \mathcal{S}\ X\ P_X & \Leftrightarrow & (X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{S} \wedge \\ & & \quad P_X \in \mathit{measurable}\ \mathcal{M}\ \mathcal{B}_{\overline{\mathbb{R}}} \wedge \mathrm{AE}_{\mathcal{S}}\ x.\ 0 \leq P_X\ x \wedge \\ & & \quad \mathit{distr}\ \mathcal{M}\ \mathcal{S}\ X = \mathit{density}\ \mathcal{S}\ P_X) \end{array}$$

We use the Lebesgue measure on *distributed* to represent the distribution of continuous random variables. For each Borel-measurable function f it is always possible to construct a random variable X_f with f as pdf. We simply use the random variable $X_f = \lambda x.\ x$ on the measure space *density* $\lambda_{\mathbb{R}}\ f$. It is not necessary to apply the the inverse transform method or the Box-Muller method to construct random variables out of binary sequences like Hasan [32] does.

From Theorem DENSITY-UNIQUE-IFF we know that *distributed* is a.e.-unique in the density function. Hence, for two different densities on the same random variable, we know that they are a.e.-equal:

lemma DISTRIBUTED-UNIQUE:
$\mathit{distributed}\ \mathcal{S}\ X\ P_1 \wedge \mathit{distributed}\ \mathcal{S}\ X\ P_2 \implies \mathrm{AE}_{\mathcal{S}}\ x.\ P_1\ x = P_2\ x$

When the random variable X induces the density P_X on the measure space $\mathcal{S}$, then the probability of X equals the integral over $\mathcal{S}$:

lemma DISTRIBUTED-PR:
$\mathit{distributed}\ \mathcal{S}\ X\ P_X \wedge A \in \mathcal{A}_{\mathcal{S}} \implies$
$\Pr\Big(x.\ X\ x \in A\Big) = \int^{P} x.\ P_X\ x \cdot \chi\ A\ x\ d\mathcal{S}$

Similarly we can replace the application of X in an integral by the multiplication

[2] That's why the constant for the push-forward measure is called *distr*.

with P_X on the measure space $\mathcal{S}$:

lemma DISTRIBUTED-$\int$:
distributed $\mathcal{S}\ X\ P_X\ \wedge\ g \in$ *measurable* $\mathcal{S}\ \mathcal{B}_{\mathbb{R}}\ \Longrightarrow$
$\left(\int x.\ g\ (X\ x)\ d\mathcal{M}\right) = \left(\int x.\ P_X\ x \cdot g\ x\ d\mathcal{S}\right)$

These are generic rules working with each random variables for which the pdf is known.

From Lemma DISTRIBUTED-PR it is easy to get the cdf of a random variable with a pdf on the Lebesgue measure. For example, for an exponentially distributed random variable X the equation

$$\Pr(x.\ X\ x \leq a) = \int_{0 \leq x \leq a} l \cdot \exp\ (-l \cdot x)\ d\lambda_{\mathbb{R}}$$

holds. But we also know that the other direction also holds, i.e. when the cdf of X is representable as an integral over P_X, then the random variable X is distributed with P_X as pdf. Fortunately, Theorem MEASURE-EQI-GENERATOR-EQ provides us with a technique to reduce the equality for all Borel sets to the equality for all intervals $\{..\ a\}$.

lemma DISTRIBUTEDI-BOREL-ATMOST:
$X \in$ *measurable* $\mathcal{M}\ \mathcal{B}_{\mathbb{R}}\ \wedge\ P_X \in$ *measurable* $\mathcal{B}_{\mathbb{R}}\ \mathcal{B}_{\mathbb{R}}\ \wedge\ \left(\mathrm{AE}_{\lambda_{\mathbb{R}}}\ x.\ 0 \leq P_X\ x\right) \wedge$
$\left(\forall a.\ \Pr\left(x.\ X\ x \leq a\right) = \left(\int^P x.\ P_X\ x \cdot \chi\ \{..\ a\}\ x\ d\lambda_{\mathbb{R}}\right)\right)\ \Longrightarrow$
distributed $\lambda_{\mathbb{R}}\ X\ P_X$

4.3.1 Joint Distribution

We model the *joint distribution* of two random variables X and Y as the distribution of $XY = \lambda x.\ (X\ x, Y\ x)$. When P_{XY} is the density of the joint distribution of X and Y we write *distributed* $\left(\mathcal{S} \otimes \mathcal{T}\right)\ XY\ P_{XY}$. When X is distributed with density P_X and Y is distributed with density P_Y then X and Y are the *marginal distributions* of XY, and P_X and P_Y are the *marginal pdfs*.

We know that the marginal pdfs P_X and P_Y always exist, and that they are the integral of P_{XY} along the measure space $\mathcal{T}$ and $\mathcal{S}$, respectively:

lemma DISTRIBUTED-MARGINAL1, -MARGINAL2:
σ-*finite-measure* $\mathcal{T}\ \wedge\ \sigma$-*finite-measure* $\mathcal{S}\ \wedge$
distributed $\left(\mathcal{T} \otimes \mathcal{S}\right)\ \left(\lambda x.\ (X\ x, Y\ x)\right)\ P_{XY}\ \Longrightarrow$
distributed $\mathcal{S}\ X\ \left(\lambda x.\ \int^P y.\ P_{XY}\ (x,y)\ d\mathcal{T}\right)$,
distributed $\mathcal{T}\ Y\ \left(\lambda y.\ \int^P x.\ P_{XY}\ (x,y)\ d\mathcal{S}\right)$,

We first prove Lemma DISTRIBUTED-MARGINAL1 and then use the symmetry of the

product space on distributed random variables:

lemma DISTRIBUTED-SWAP:
σ-finite-measure $\mathcal{T} \wedge$ σ-finite-measure $\mathcal{S} \wedge$
distributed $\Big(\mathcal{S} \otimes \mathcal{T}\Big)$ $\Big(\lambda x.\ (X\ x, Y\ x)\Big)$ $P_{XY} \implies$
distributed $\Big(\mathcal{T} \otimes \mathcal{S}\Big)$ $\Big(\lambda x.\ (Y\ x, X\ x)\Big)$ $(\lambda(y,x).\ P_{XY}\ (x,y))$

What do we know about the joint pdf of two random variables? A simple case is when the two random variables are independent, then the joint pdf is the product of their marginal pdfs:

lemma DISTRIBUTED-JOINT-INDEP:
σ-finite-measure $\mathcal{S} \wedge \sigma$-finite-measure $\mathcal{T} \wedge$
distributed $\mathcal{S}\ X\ P_X \wedge$ distributed $\mathcal{T}\ Y\ P_Y \wedge$ indep-var $\mathcal{S}\ \mathcal{T}\ X\ Y \implies$
distributed $\Big(\mathcal{S} \otimes \mathcal{T}\Big)$ $\Big(\lambda x.\ (X\ x, Y\ x)\Big)$ $\Big(\lambda(x,y).\ P_X\ x \cdot P_Y\ y\Big)$

There is no general description of the joint pdf when the variables are not independent. However, we can provide a simpler way to show that a function is the pdf for the joint distribution of X and Y. It is enough if P_{XY} is the pdf on each product $A \times B$:

lemma DISTRIBUTED-JOINTI:
σ-finite-measure $\mathcal{T} \wedge$ σ-finite-measure $\mathcal{S} \wedge$
$X \in$ measurable $\mathcal{M}\ \mathcal{S} \wedge Y \in$ measurable $\mathcal{M}\ \mathcal{T} \wedge$
$P_{XY} \in$ measurable $\Big(\mathcal{S} \otimes \mathcal{T}\Big)$ $\mathcal{B}_{\overline{\mathbb{R}}} \wedge \Big(\text{AE}_{\mathcal{S}\otimes\mathcal{T}}\ x.\ 0 \le P_{XY}\ x\Big) \wedge$
$\Bigg(\forall A \in \mathcal{A}_{\mathcal{S}}, B \in \mathcal{A}_{\mathcal{T}}.\ \Pr\Big(x.\ X\ x \in A \wedge Y\ x \in B\Big) =$
$\Big(\int^P x.\ \Big(\int^P y.\ P_{XY}\ (x,y) \cdot \chi\ B\ y\ d\mathcal{T}\Big) \cdot \chi\ A\ x\ d\mathcal{S}\Big)\Bigg) \implies$
distributed $\mathcal{M}$ $\Big(\mathcal{S} \otimes \mathcal{T}\Big)$ $\Big(\lambda x.\ (X\ x, Y\ x)\Big)$ P_{XY}

4.3.2 Uniform Distribution

A random variable X has a *uniform distribution* if the probability that it hits an element in its range is equal for all elements. The density function is $\lambda x.\ \chi\ A\ x / \mu^f_{\lambda_{\mathbb{R}}}\ A$, where A is the range of X.

With Lemma DISTRIBUTEDI-BOREL-ATMOST we prove that each random variable with the cumulative distribution function $(t-a)/(b-a)$ is uniformly distributed on the interval $\{a\ ..\ b\}$:

theorem UNIFOM-DISTRIBUTED-IFF:
distributed $\lambda_{\mathbb{R}}\ X\ (\lambda x.\ \chi\ \{a\ ..\ b\}\ x / \mu^f_{\lambda_{\mathbb{R}}}\ \{a\ ..\ b\}) \Leftrightarrow$
$\Big(X \in$ measurable $\mathcal{M}\ \mathcal{B}_{\mathbb{R}} \wedge a < b \wedge$
$(\forall a \le t \le b.\ \Pr(x.\ X\ x \le t) = (t-a)/(b-a))\Big)$

When a random variable is uniformly distributed on $\{a \mathinner{..} b\}$, then we also know that its expectation is $(a+b)/2$. This is easily proved with the fundamental theorem of calculus on the integrand $\lambda x.\ x \cdot \chi\ \{a \mathinner{..} b\}\ x/(b-a)$.

lemma UNIFORM-DISTRIBUTED-EXPECTATION:
$$\textit{distributed}\ \lambda_{\mathbb{R}}\ X\ \left(\lambda x.\ \chi\ \{a \mathinner{..} b\}\ x/\mu^{f}_{\lambda_{\mathbb{R}}}\ \{a \mathinner{..} b\}\right) \implies \int X\, d\mathcal{M} = (a+b)/2$$

4.3.3 Exponential Distribution

A random variable X is *exponentially distributed* when its distribution on the Lebesgue-Borel measure has the following density:

$$\begin{array}{lcl} \textit{exponential-density} & :: & \mathbb{R} \to \mathbb{R} \to \mathbb{R} \\ \textit{exponential-density}\ l\ x & = & \textbf{if}\ 0 \le x\ \textbf{then}\ l \cdot \exp\ (-x \cdot l)\ \textbf{else}\ 0 \end{array}$$

Exponentially distributed random variables appear as the transition times in continuous-time Markov chains. Here a random variable X describes the time between two transitions on a path. These random variables are memoryless, i.e. it forgets how much *time* has passed, so the probability that X is above $a+t$, under the condition that it is above a, does not depend on a, only on the distance t.

lemma DISTRIBUTED-EXPONENTIAL-MEMORYLESS:
$$\textit{distributed}\ \lambda_{\mathbb{R}}\ X\ (\textit{exponential-density}\ l)\ \wedge\ 0 \le a\ \wedge\ 0 \le t \implies$$
$$\Pr(x.\ a+t \le X\ x \mid a \le X\ x) = \Pr(x.\ t \le X\ x)$$

Similar to the uniform distribution, we show that each exponentially distributed random variable has $1 - \exp\ (-x \cdot l)$ as the cumulative distribution function.

theorem DISTRIBUTED-EXPONENTIAL-IFF:
$$\textit{distributed}\ \lambda_{\mathbb{R}}\ X\ (\textit{exponential-density}\ l) \Leftrightarrow$$
$$\Big(X \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}}\ \wedge\ 0 < l\ \wedge$$
$$\left(\forall a \ge 0.\ \Pr(x.\ X\ x \le a) = 1 - \exp(-a \cdot l)\right)\Big)$$

We also show that the expectation of such a random variable is $1/l$.

lemma EXPONENTIAL-DISTRIBUTED-EXPECTATION:
$$\textit{distributed}\ \lambda_{\mathbb{R}}\ X\ (\textit{exponential-density}\ l) \implies \int X\, d\mathcal{M} = 1/l$$

4.4 Information

Information theory is concerned with quantifying information represented by random variables. Shannon [70] introduced entropy. Examples where it is used in computer science are to reason about the average size of compressed data, quantitative information flow analysis (see Clark *et al.* [16]), or analysis of security properties of side channel attacks (see Köpf and Dürmuth [48], Section 5.3 and 5.4). A detailed introduction into information theory is given by Gray [26], as well as Cover and Thomas [18]. We used these books as basis for our formalization.

As the data represented in computer systems is discrete the first formalization of information theory in theorem provers was of discrete nature. Coble's formalization [17] reasons about anonymity. While his theorems are restricted to discrete finite random variables, his definitions are already suited for the continuous case.

Our definition of Kullback-Leibler divergence, conditional entropy, mutual information and conditional mutual information is similar to Coble's definitions [17]. Especially, the conditional mutual information is also represented using mutual information on joint distributions. This is necessary as we have no formalization of conditional distributions. With our formalizations of product measures, integrals and distributions we are able to prove information theory properties on continuous random variables.

The random variables are defined on a probability space $\mathcal{M}$, but we also want to abstract the *size* of a bit. For binary data this is usually 2, but in the continuous case this is sometimes also the Euler constant e. To abstract over this we introduce information spaces:

locale *information-space* = *prob-measure* $\mathcal{M}$ +
  fixes $b :: \mathbb{R}$ **assumes** $1 < b$

From now on we will assume that $\mathcal{M}$ and b form an information space.

Entropy, mutual information and their conditional extensions work on distributions of random variables. We name the random variables X, Y, and Z and their resulting measure spaces $\mathcal{S}$, $\mathcal{T}$, and $\mathcal{U}$, respectively.

Notation: *To save space, we abbreviate the (joint) distribution measures of these random variables:*

$$\begin{aligned}
\mathcal{D}_X &= \mathit{distr}\ \mathcal{M}\ \mathcal{S}\ X \\
\mathcal{D}_{XY} &= \mathit{distr}\ \mathcal{M}\ (\mathcal{S} \otimes \mathcal{T})\ (\lambda x.\ (X\ x, Y\ x)) \\
\mathcal{D}_{XYZ} &= \mathit{distr}\ \mathcal{M}\ (\mathcal{S} \otimes \mathcal{T} \otimes \mathcal{U})\ (\lambda x.\ (X\ x, Y\ x, Z\ x))
\end{aligned}$$

4.4.1 Entropy

The *entropy* $H(X)$ of a random variable X quantifies the uncertainty of X. In the discrete case it is defined to be $-\sum_x P_X\ x \cdot \log_b\ (P_X\ x)$, where P_X is the pdf of X. To extend this to the continuous case we use the Radon-Nikodým derivative of $\mathcal{D}_X$ as pdf of X, and the sum of the range of X is replaced by the integral over $\mathcal{D}_X$. We also introduce $H(X)$ for the discrete case.

$$\begin{aligned}
&\mathit{entropy} && :: && \beta\ \mathit{measure} \to \Big(\alpha \to \beta\Big) \to \mathbb{R} \\
&\mathit{entropy}\ \mathcal{S}\ X && = && -\int x.\ \log_b\ (\mathit{RN\text{-}deriv}\ \mathcal{S}\ \mathcal{D}_X\ x)\ d\mathcal{D}_X \\
&H(X) && = && \mathit{entropy}\ (\mathit{count}\ X[\Omega_{\mathcal{M}}])\ X
\end{aligned}$$

Note that $\mathcal{D}_X$ also depends on $\mathcal{S}$. When X is distributed with the pdf P_X, we express the entropy as an integral over $\mathcal{S}$.

lemma ENTROPY-DISTR:
  $\mathit{distributed}\ \mathcal{S}\ X\ P_X \implies \mathit{entropy}\ \mathcal{S}\ X = -\int x.\ P_X\ x \cdot \log_b(P_X\ x)\ d\mathcal{S}$

This equation tells us that to work with entropy we need to assume that the integral in this equation is finite, and that the pdf P_X exists. This is actually a generalization of discrete finite random variables: for them, the pdf always exists and the entropy is always defined (since the entropy density is always integrable). Since these two assumptions appear quite often, we introduce a predicate to characterize random variables with a finite entropy:

$$\begin{array}{lcl} \textit{finite-entropy} & :: & \beta\ \textit{measure} \to (\alpha \to \beta) \to (\beta \to \mathbb{R}) \to \mathbb{B} \\ \textit{finite-entropy}\ \mathcal{S}\ X\ P_X & \Leftrightarrow & \textit{distributed}\ \mathcal{S}\ X\ P_X \wedge \\ & & \textit{integrable}\ \mathcal{S}\ (\lambda x.\ P_X\ x \cdot \log_b\ (P_X\ x)) \end{array}$$

As mention, each discrete finite random variable has a finite entropy:

lemma FINITE-ENTROPY-SIMPLE-FN:
$$\textit{simple-fn}\ \mathcal{M}\ X \implies \textit{finite-entropy}\ (\textit{count}\ X[\Omega_{\mathcal{M}}])\ X\ \Big(\lambda a.\ \Pr\{x \mid X\ x = a\}\Big)$$

Now we analyze random variables with finite entropy and with a finite support, i.e. the set where P_X is nonzero has a finite measure. We show that the entropy has an upper bound in this case:

theorem ENTROPY-LE:
$$\textit{finite-entropy}\ \mathcal{S}\ X\ P_X \wedge \mu_{\mathcal{S}}\ \{x \mid P_X\ x \neq 0\} < \infty \implies \textit{entropy}\ \mathcal{S}\ X \leq \log_b\ (\mu^f_{\mathcal{S}}\ \{x \mid P_X\ x \neq 0\})$$

This theorem is proved using Theorem JENSENS-INEQUALITY with convexity of the logarithm $-\log_b\ (f\ x)$.

We know even for which distribution the maximum is reached: when we have a uniformly distributed random variable X, then the entropy *equals* the logarithm of the measure of the support of X.

corollary ENTROPY-UNIFORM:
$$\textit{distributed}\ \mathcal{S}\ X\ (\lambda x.\ \chi\ A\ x / \mu^f_{\mathcal{S}}\ A) \implies \textit{entropy}\ \mathcal{S}\ X = \log_b\ (\mu^f_{\mathcal{S}}\ A)$$

This tells us that the uniformly distribution is the maximum entropy distribution for random variables with a finite support. This lemma also tells us that the entropy can be negative: when $\mu^f_{\mathcal{S}}\ A < 1$.

4.4.2 Conditional Entropy

Conditional entropy $H(X|Y)$ quantifies the uncertainty of the random variable X, when the outcome of the random variable Y is already known. Coble [17] defines conditional entropy in terms of entropy: $H(X|Y) = H(X,Y) - H(Y)$. Gray [26] calls it conditional relative entropy and gives two definitions, the first one integrates over the pdf of X conditioned by Y, the second one uses the Kullback-Leibler divergence. Gray's Kullback-Leibler divergence is infinite if the two measures are not absolutely continuous, hence the second form is more general. However, in our setting we only allow the conditional entropy to assume finite values, hence we chose the integral version. We assume that the random variable Y has the pdf

P_Y, and P_{XY} is the joint pdf of X and Y. In this case the pdf of X conditioned by Y is $P_{X|Y}(x,y) = P_{XY}\ (x,y)/P_Y\ y$. In the discrete case, conditional entropy is

$$-\sum_{x,y} P_{XY}\ (x,y) \cdot \log_b\ (P_{X|Y}\ (x,y))\ .$$

Similar to our definition of entropy this equation maps nicely to the general case by using the Radon-Nikodým derivative and the Lebesgue integral. We use the abbreviation $H(X|Y)$ for the discrete case:

$$\begin{aligned}
&\mathit{conditional\text{-}entropy} :: \beta\ \mathit{measure} \to \gamma\ \mathit{measure} \to \Big(\alpha \to \beta\Big) \to \Big(\alpha \to \gamma\Big) \to \mathbb{R} \\
&\mathit{conditional\text{-}entropy}\ \mathcal{S}\ \mathcal{T}\ X\ Y = \\
&\quad -\int (x,y).\ \log_b \Big(\mathit{RN\text{-}deriv}\ \Big(\mathcal{S} \otimes \mathcal{T}\Big)\ \mathcal{D}_{XY}\ (x,y)\ /\ \mathit{RN\text{-}deriv}\ \mathcal{T}\ \mathcal{D}_Y\ y\Big)\ d\mathcal{D}_{XY} \\
&H(X|Y) = \mathit{conditional\text{-}entropy}\ (\mathit{count}\ X[\Omega_{\mathcal{M}}])\ (\mathit{count}\ Y[\Omega_{\mathcal{M}}])\ X\ Y
\end{aligned}$$

Now we show that the conditional entropy fulfills the so called *chain rule*: $H(X|Y) = H(X,Y) - H(Y)$. We need to assume that the random variable X and the joint random variables of X and Y have a finite entropy.

theorem CONDITIONAL-ENTROPY-CHAIN-RULE:
$$\begin{aligned}
&\sigma\text{-}\mathit{finite\text{-}measure}\ \mathcal{S}\ \wedge\ \sigma\text{-}\mathit{finite\text{-}measure}\ \mathcal{T}\ \wedge \\
&\mathit{finite\text{-}entropy}\ \mathcal{T}\ Y\ P_Y\ \wedge\ \mathit{finite\text{-}entropy}\ (\mathcal{S} \otimes \mathcal{T})\ (\lambda x.\ (X\ x, Y\ x))\ P_{XY}\ \Longrightarrow \\
&\mathit{conditional\text{-}entropy}\ \mathcal{S}\ \mathcal{T}\ X\ Y = \\
&\quad \mathit{entropy}\ (\mathcal{S} \otimes \mathcal{T})\ (\lambda x.\ (X\ x, Y\ x)) - \mathit{entropy}\ \mathcal{T}\ Y
\end{aligned}$$

4.4.3 Kullback-Leibler Divergence

To define mutual information we formalize *Kullback-Leibler divergence*. It quantifies the similarity of two probability measure $\mathcal{S}$ and $\mathcal{T}$ on the same σ-algebra $\mathcal{A}_{\mathcal{S}} = \mathcal{A}_{\mathcal{T}}$. It is often seen as some kind of distance between these two measures. This works if $\mathcal{T}$ is expressible as a density of $\mathcal{S}$. We integrate over the logarithm of the Radon-Nikodým derivative.

$$\begin{aligned}
&\mathit{KL\text{-}divergence} &&:: \mathbb{R} \to \alpha\ \mathit{measure} \to \beta\ \mathit{measure} \to \mathbb{R} \\
&\mathit{KL\text{-}divergence}\ b\ \mathcal{S}\ \mathcal{T} &&= \int x.\ \log_b\ (\mathit{RN\text{-}deriv}\ \mathcal{S}\ \mathcal{T}\ x)\ d\mathcal{T}
\end{aligned}$$

When both measure spaces are representable as density functions f and g, then we represent the Kullback-Leibler divergence as integral replacing the Radon-Nikodým derivative with $g\ x/f\ x$.

lemma KL-DIVERGENCE-DENSITY:
$$\begin{aligned}
&\sigma\text{-}\mathit{finite\text{-}measure}\ \mathcal{S}\ \wedge \\
&f \in \mathit{measurable}\ \mathcal{S}\ \mathcal{B}_{\mathbb{R}}\ \wedge\ \mathrm{AE}_{\mathcal{S}}\ x.\ 0 \le f\ x\ \wedge \\
&g \in \mathit{measurable}\ \mathcal{S}\ \mathcal{B}_{\mathbb{R}}\ \wedge\ \mathrm{AE}_{\mathcal{S}}\ x.\ 0 \le g\ x\ \wedge \\
&\mathrm{AE}_{\mathcal{S}}\ x.\ (f\ x = 0\ \Longrightarrow\ g\ x = 0)\ \Longrightarrow \\
&\mathit{KL\text{-}divergence}\ b\ (\mathit{density}\ \mathcal{S}\ f)\ (\mathit{density}\ \mathcal{S}\ g) = \int x.\ g\ x \cdot \log_b\ \frac{g\ x}{f\ x}\ d\mathcal{S}
\end{aligned}$$

The Kullback-Leibler divergence is often seen as some kind of metric or distance. This only holds for absolutely continuous probability measures. Only for

them the Radon-Nikodým derivative exists. This does not yet imply symmetry, but at least the Kullback-Leibler divergence is then nonnegative and only 0 iff the two measures are equal.

theorem KL-DIVERGENCE-EQ-0, -NONNEG:
$$\begin{aligned}&\textit{prob-measure}\ (\textit{density}\ \mathcal{M}\ D)\ \wedge\ \textit{integrable}\ \mathcal{M}\ (\lambda x.\ D\ x \cdot \log_b\ (D\ x))\ \wedge\\ &D \in \textit{measurable}\ \mathcal{M}\ \mathcal{B}_{\mathbb{R}}\ \wedge\ \Big(\mathrm{AE}_{\mathcal{M}}\ x.\ 0 \leq D\ x\Big)\ \Longrightarrow\\ &\Big(\textit{KL-divergence}\ b\ \mathcal{M}\ (\textit{density}\ \mathcal{M}\ D) = 0\Big) \Leftrightarrow \Big(\textit{density}\ \mathcal{M}\ D = \mathcal{M}\Big),\\ &0 \leq \textit{KL-divergence}\ b\ \mathcal{M}\ (\textit{density}\ \mathcal{M}\ D)\end{aligned}$$

In the following sections we will apply the Kullback-Leibler divergence only on distributions for which we know the pdf, hence these two lemmas will be applicable.

4.4.4 Mutual Information

Mutual information $I(X;Y)$ quantifies the information shared by the two random variables X and Y. For the discrete case information theory defines it as

$$\sum_{x,y} P_{XY}\ (x,y) \cdot \log_b\ \frac{P_{XY}\ (x,y)}{P_X\ x \cdot P_Y\ y}\ .$$

Gray [26] uses Kullback-Leibler divergence between the joint distribution of X and Y and the product of the distributions to define the mutual information also for continuous random variables. Coble [17] defines it in the same way, but only uses it for finite random variables. For finite random variables this definition is equal to the previous equation, hence we introduce the abbreviation $I(X;Y)$ for mutual information on discrete random variables.

$$\begin{aligned}&\textit{mutual-information} :: \beta\ \textit{measure} \to \gamma\ \textit{measure} \to \Big(\alpha \to \beta\Big) \to \Big(\alpha \to \gamma\Big) \to \mathbb{R}\\ &\textit{mutual-information}\ \mathcal{S}\ \mathcal{T}\ X\ Y = \textit{KL-divergence}\ b\ (\mathcal{D}_X \otimes \mathcal{D}_Y)\ \mathcal{D}_{XY}\\ &I(X;Y) \qquad\qquad = \textit{mutual-information}\ (\textit{count}\ X[\Omega_{\mathcal{M}}])\ (\textit{count}\ Y[\Omega_{\mathcal{M}}])\ X\ Y\end{aligned}$$

From Theorem KL-DIVERGENCE-NONNEG we derive that mutual information is nonnegative. When the pdf for X, Y and their joint distribution is given then we equate mutual information to the integral over $P_{XY}\ (x,y) \cdot \log_b\ (P_{XY}\ (x,y)/(P_X\ x \cdot P_Y\ y))$.

lemma MUTUAL-INFORMATION-DISTR, -NONNEG:
$$\begin{aligned}&\sigma\textit{-finite-measure}\ \mathcal{S}\ \wedge\ \sigma\textit{-finite-measure}\ \mathcal{T}\ \wedge\\ &\textit{finite-entropy}\ \mathcal{S}\ X\ P_X\ \wedge\ \textit{finite-entropy}\ \mathcal{T}\ Y\ P_Y\ \wedge\\ &\textit{finite-entropy}\ (\mathcal{T} \otimes \mathcal{S})\ (\lambda x.\ (X\ x, Y\ y))\ P_{XY}\ \Longrightarrow\\ &\textit{mutual-information}\ \mathcal{S}\ \mathcal{T}\ X\ Y =\\ &\quad \int (x,y).\ P_{XY}\ (x,y) \cdot \log_b\ \frac{P_{XY}\ (x,y)}{P_X\ x \cdot P_Y\ y}\ d\Big(\mathcal{S} \otimes \mathcal{T}\Big),\\ &0 \leq \textit{mutual-information}\ \mathcal{S}\ \mathcal{T}\ X\ Y\end{aligned}$$

This version assumes that all occurring random variables have a finite entropy.[3] But note that the mutual information may be defined even when the random variables do not have a finite entropy.

[3] There is also an alternative version of this theorem in Isabelle/HOL, assuming only integrability of $P_{XY}\ (x,y) \cdot \log_b\ (P_{XY}\ (x,y)/(P_X\ x \cdot P_Y\ y))$.

An example where this may not hold are independent random variables. For them the mutual information is always zero, even when both do not have a finite entropy. We proved that two random variables X and Y are independent if and only if the joint distribution equals the product distribution: $\mathcal{D}_{XY} = \mathcal{D}_X \otimes \mathcal{D}_Y$. With Theorem KL-DIVERGENCE-EQ-0 we also know that the Kullback-Leibler divergence is zero if and only if the two measures are equal. From this immediately follows that the mutual information is zero if and only if the random variables are independent:

theorem MUTUAL-INFORMATION-INDEP-VAR:
$$\textit{indep-var}\ \mathcal{S}\ \mathcal{T}\ X\ Y \Leftrightarrow \Big(X \in \textit{measurable}\ \mathcal{M}\ \mathcal{S} \wedge Y \in \textit{measurable}\ \mathcal{M}\ \mathcal{T} \wedge \textit{integrable}\ \mathcal{D}_{XY}\ (\textit{entropy-density}\ b\ (\mathcal{D}_X \otimes \mathcal{D}_Y)\ \mathcal{D}_{XY}) \wedge \textit{absolutely-continuous}\ (\mathcal{D}_X \otimes \mathcal{D}_Y)\ \mathcal{D}_{XY} \wedge \textit{mutual-information}\ \mathcal{S}\ \mathcal{T}\ X\ Y = 0\Big)$$

An alternative way to define mutual information is to use entropy: $I(X;Y) = H(X) - H(X|Y)$. However, for this equation it is again necessary that all occurring random variables have a finite entropy:

theorem MUTUAL-INFORMATION-EQ-ENTROPY-CONDITIONAL-ENTROPY:
$$\sigma\textit{-finite-measure}\ \mathcal{S} \wedge \sigma\textit{-finite-measure}\ \mathcal{T} \wedge \textit{finite-entropy}\ \mathcal{S}\ X\ P_X \wedge \textit{finite-entropy}\ \mathcal{T}\ Y\ P_Y \wedge \textit{finite-entropy}\ (\mathcal{S} \otimes \mathcal{T})\ (\lambda x.\ (X\ x, Y\ x))\ P_{XY} \implies \textit{mutual-information}\ \mathcal{S}\ \mathcal{T}\ X\ Y = \textit{entropy}\ \mathcal{S}\ X - \textit{conditional-entropy}\ \mathcal{S}\ \mathcal{T}\ X\ Y$$

This equation and Lemma MUTUAL-INFORMATION-NONNEG allow us to give entropy as an upper bound for conditional entropy.

corollary CONDITIONAL-ENTROPY-LE-ENTROPY:
$$\sigma\textit{-finite-measure}\ \mathcal{S} \wedge \sigma\textit{-finite-measure}\ \mathcal{T} \wedge \textit{finite-entropy}\ \mathcal{S}\ X\ P_X \wedge \textit{finite-entropy}\ \mathcal{T}\ Y\ P_Y \wedge \textit{finite-entropy}\ (\mathcal{S} \otimes \mathcal{T})\ (\lambda x.\ (X\ x, Y\ x))\ P_{XY} \implies \textit{conditional-entropy}\ \mathcal{S}\ \mathcal{T}\ X\ Y \leq \textit{entropy}\ \mathcal{S}\ X$$

4.4.5 Conditional Mutual Information

Conditional mutual information quantifies the information shared by two random variables, under the assumption that the result of a third random variable is already known. The discrete definition in information theory is similar to mutual information, it sums up for all possible outcomes z of Z the mutual information conditioned that $Z = z$.

$$\sum_z P_Z\ z \sum_{x,y} P_{XY|Z}(x,y,z) \cdot \log_b \frac{P_{XY|Z}(x,y,z)}{P_{X|Z}(x,z) \cdot P_{Y|Z}(y,z)}$$

As usual we could try to map this to integration over the Radon-Nikodým derivatives of the distributions of X, Y and Z. Gray [26] uses the Kullback-Leibler diver-

gence on conditional distributions of Z. For continuous random variables this requires in both cases the formalization of conditional distributions. To avoid this we use the same definition as Coble [17], the equality $I(X;Y|Z) = I(X;Y,Z) - I(X;Z)$:

conditional-mutual-information :: β *measure* $\to$ γ *measure* $\to$ δ *measure* $\to$
$\Big(\alpha \to \beta\Big) \to \Big(\alpha \to \gamma\Big) \to \Big(\alpha \to \delta\Big) \to \mathbb{R}$
conditional-mutual-information $\mathcal{S}$ $\mathcal{T}$ $\mathcal{U}$ X Y Z =
mutual-information $\mathcal{S}$ $(\mathcal{T} \otimes \mathcal{U})$ X $(\lambda x.\ (Y\ x, Z\ x))$ −
mutual-information $\mathcal{S}$ $\mathcal{U}$ X Z

$I(X;Y|Z)$ = *conditional-mutual-information*
(*count* $X[\Omega_{\mathcal{M}}]$) (*count* $Y[\Omega_{\mathcal{M}}]$) (*count* $Z[\Omega_{\mathcal{M}}]$) X Y Z

Similar to mutual information we can show that conditional mutual information is nonnegative and is an integral over the pdf of the joint random variable of X, Y and Z. We require that for the random variables X, Y and Z the pdfs P_X, P_Z, P_{YZ}, P_{XZ}, and P_{XYZ} are defined and have a finite entropy.[4]

lemma CONDITIONAL-MUTUAL-INFORMATION-EQ, -NONNEG:
σ-*finite-measure* $\mathcal{S}$ $\wedge$ σ-*finite-measure* $\mathcal{T}$ $\wedge$ σ-*finite-measure* $\mathcal{U}$ $\wedge$
finite-entropy $\mathcal{S}$ X P_X $\wedge$ *finite-entropy* $\mathcal{U}$ Z P_Z $\wedge$
finite-entropy $(\mathcal{T} \otimes \mathcal{U})$ $(\lambda x.\ (Y\ x, Z\ x))$ P_{YZ} $\wedge$
finite-entropy $(\mathcal{S} \otimes \mathcal{U})$ $(\lambda x.\ (X\ x, Z\ x))$ P_{XZ} $\wedge$
finite-entropy $(\mathcal{S} \otimes \mathcal{T} \otimes \mathcal{U})$ $(\lambda x.\ (X\ x, Y\ x, Z\ x))$ P_{XYZ} $\Longrightarrow$
conditional-mutual-information $\mathcal{S}$ $\mathcal{T}$ $\mathcal{U}$ X Y Z =
$$\int (x,y,z).\ P_{XYZ}\ (x,y,z) \cdot \log_b \frac{P_{XYZ}\ (x,y,z) \cdot P_Z\ z}{P_{XZ}\ (x,z) \cdot P_{YZ}\ (y,z)}\ d\Big(\mathcal{S} \otimes \mathcal{T} \otimes \mathcal{U}\Big),$$
$0 \leq$ *conditional-mutual-information* $\mathcal{S}$ $\mathcal{T}$ $\mathcal{U}$ X Y Z

4.5 Infinite Product of Probability Spaces

The finite product measure allows us to construct a measure space with a finite set of independent random variables. In probability theory it is often necessary to have an infinite set of independent random variables. For this, we introduce now $\bigotimes_{i \in I} \mathcal{M}_i$, the product of infinitely many probability measures $\mathcal{M}_i$. The measure $\bigotimes_{i \in I} \mathcal{M}_i$ defined in Section 3.4.3 is already usable for a finite index I. In this section we prove that its elementary property holds for an infinite index I: each set X embedded from a finite product measure $\bigotimes_{i \in J} \mathcal{M}_i$ (with $J \subseteq I$) has the same measure value $\mu_{\bigotimes_{i \in I} \mathcal{M}_i}$ (*emb* J X) $= \mu_{\bigotimes_{i \in J} \mathcal{M}_i}$ X. We used the proof in [8] as the basis of our formalization of infinite products.

We start by introducing the locale *product-prob-measures*. It assumes that $\mathcal{M}_i$ is a probability measure and I an index set:

locale *product-prob-measures* =
fixes $\mathcal{M}$ :: $\iota \to \alpha$ *measure* **and** I :: ι *set*
assumes $\forall i$. *prob-measure* $\mathcal{M}_i$

[4] Again there is a version of this theorem in Isabelle/HOL which relaxes these assumptions to only require the integrals in both mutual informations are defined.

Notation: *Note that we write $\mathcal{M}_i$ instead of $\mathcal{M}$ i. In this section we will assume this locale, i.e. all $\mathcal{M}_i$ are probability measures. For technical purposes we do not restrict that assumption to indices from I.*

We will prove the existence of the infinite product space with Caratheodory's extension theorem in the form of Corollary CARATHEODORY-∅-CONTINUOUS. This is done in three steps: (1) define a generating algebra $\mathcal{G}$, (2) define a volume μG on $\mathcal{G}$, and (3) show that μG is ∅-continuous hence extensible to a measure.

As generating sets we need a family of sets which is at least a ring. This rules out the projections at a single index $\{\omega \mid \omega\ i \in A\}$ as this is neither stable under intersection nor under union. It also rules out projections of finite Cartesian products *emb J* $(\times_{j\in J} A\ j)$ as this is not stable under union. What we can use are the projections of the σ-algebra of finite products: *emb J X* for $X \in \mathcal{A}_{\bigotimes_{j\in J} \mathcal{M}_i}$. We introduce the set $\mathcal{G}$ as all these projections:

$$\mathcal{G} :: \left(\iota \to \alpha\right) \textit{ set set}$$
$$\mathcal{G} = \{\textit{emb } J\ A \mid \textit{finite } J \wedge J \subseteq I \wedge J \neq \emptyset \wedge A \in \mathcal{A}_{\bigotimes_{i\in J} \mathcal{M}_j}\}$$

This generator is more complicated than the generator used to define $\bigotimes_{i\in I} \mathcal{M}_i$. But still, its generated σ-algebra equals the one from the product measure $\bigotimes_{i\in I} \mathcal{M}_i$:

lemma SETS-⊗-$\mathcal{G}$: $\mathcal{A}_{\bigotimes_{i\in I} \mathcal{M}_i} = \sigma\textit{-sets } (\times_{i\in I} \Omega_i)\ \mathcal{G}$

This generator $\mathcal{G}$ forms a ring, and even an algebra (we require that the index set is nonempty, otherwise it does not contain the space $\times_{i\in I} \Omega_i$):

lemma ALGEBRA-$\mathcal{G}$: $I \neq \emptyset \implies \textit{algebra } (\times_{i\in I} \Omega_i)\ \mathcal{G}$

This will simplify the proof of countable additivity, since it now is enough to show that there exists a ∅-continuous volume μG. This volume μG on the generator $\mathcal{G}$ is now defined using the finite product measure:

$$\mu G\ :: \left(\iota \to \alpha\right) \textit{ set} \to \overline{\mathbb{R}}$$
$$\mu G\ X = \Big(\textit{SOME } m.\ \forall J \subseteq I.\ \textit{finite } J \wedge J \neq \emptyset \implies \Big(\forall A \in \mathcal{A}_{\bigotimes_{i\in J} \mathcal{M}_j}.\ X = \textit{emb } J\ A \implies m = \mu_{\bigotimes_{j\in J} \mathcal{M}_j}\ A\Big)\Big)$$

The Hilbert choice in the definition forces us to show that the resulting value is uniquely defined. This requires that we can embed two different representations of X with index sets J_1 and J_2 into the same finite product measure $J_1 \cup J_2$. For this we prove the following fact: each finite product measure can be expressed as an embedding into a finite product measure with a bigger index set. In other words, the restriction to a smaller index set is measure preserving:

lemma DISTR-RESTRICT:
$J \neq \emptyset \ \wedge\ J \subseteq K \ \wedge\ \textit{finite } K \implies$
$(\bigotimes_{i\in J} \mathcal{M}_i) = \textit{distr } (\bigotimes_{i\in K} \mathcal{M}_i)\ (\bigotimes_{i\in J} \mathcal{M}_i)\ (\lambda\omega.\ \lambda i \in J.\ \omega\ i)$

This is easily shown by Theorem MEASURE-EQI-GENERATOR-EQ, the generating sets are the products over all J, as shown by Lemma SETS-⊗-FINITE.

With Lemma DISTR-RESTRICT we show then the defining equation for μG.

lemma μG-EQ:
$\mathit{finite}\ J \wedge J \subseteq I \wedge J \neq \emptyset \wedge A \in \mathcal{A}_{\bigotimes_{i\in J} \mathcal{M}_j} \implies \mu G\ (\mathit{emb}\ J\ A) = \mu_{\bigotimes_{i\in J} \mathcal{M}_j}\ A$

That μG is positive follows directly from the specification and for additivity we have an argument similar to the proof of Lemma μG-EQ.

lemma POSITIVE-μG: $I \neq \emptyset \implies \mathit{positive}\ \mathcal{G}\ \mu G$

lemma ADDITIVE-μG: $I \neq \emptyset \implies \mathit{additive}\ \mathcal{G}\ \mu G$

With these two lemmas we showed that μG is a volume, and then Caratheodory's extension theorem tells us that a probability measure exists mapping $\mathit{emb}\ J\ (\times_{i\in J} A\ i)$ to $\prod_{i\in J} \mu\ i\ (A\ i)$. From the existence of such a probability measure follows the property for the infinite product measure $\bigotimes_{i\in I} \mathcal{M}_i$:

theorem μ-$\bigotimes$-INF:
$J \subseteq I \wedge \mathit{finite}\ J \wedge \left(\forall i \in J.\ A\ i \in \mathcal{A}_i\right) \implies$
$\mu_{\bigotimes_{i\in I} \mathcal{M}_i}\ \left(\mathit{emb}\ J\ (\times_{i\in J} A\ i)\right) = \prod_{j\in J} \mu_i\ (A\ i)$

Since we use Caratheodory's extension theorem CARATHEODORY-$\emptyset$-CONTINUOUS, we assume a decreasing sequence of sets $A_i \in \mathcal{G}$ whose measures converge to a nonzero limit.. Each element A_i is an embedding of some finite product measure with dimension J_i. We construct inductively an element ω_0 contained in all A_i. Hence ω_0 is in the limit of the sequence A_i and our measure is countably additive. For more details see Bauer [8], §9.

With this we have a probability measure usable to construct arbitrary many, independent random variables with arbitrary distributions. Comparing this to the formalization in [39] which only provides a probability measure on sequences $\mathbb{N} \to \mathbb{B}$, and hence induces random variables with only a discrete distribution.

The infinite product measure does not only provide us with a probability space providing an infinite amount of independent random variables, but it also allows us to show that the joint distribution of a family of independent random variables is equal to the product of the single distributions:

theorem INDEP-VARS-IFF-DISTR-EQ-$\bigotimes$:
$I \neq \emptyset \ \wedge\ (\forall i.\ X_i \in \mathit{measurable}\ \mathcal{M}\ \mathcal{N}_i) \implies$
$\mathit{indep\text{-}vars}\ \mathcal{N}\ X\ I \Leftrightarrow$
$\left(\mathit{distr}\ \mathcal{M}\ (\bigotimes_{i\in I} \mathcal{N}_i)\ (\lambda\omega.\ \lambda i \in I.\ X_i\ \omega) = (\bigotimes_{i\in I} \mathit{distr}\ \mathcal{M}\ \mathcal{N}_i\ X_i)\right)$

4.6 Markov Chains

We introduce Markov chains as probabilistic automata, i.e. as discrete-time time-homogeneous finite-space Markov processes. A Markov chain is defined by its state space S and a transition matrix τ. We assume no initial distribution or starting state, however when measuring paths we provide the starting state to the

probability function. A path (or trace) on a Markov chain is a function $\mathbb{N} \to S$, i.e. an infinite sequence of states visited in the Markov chain.

locale *markov-chain* =
  fixes $S :: \alpha$ *set* **and** $\tau :: \alpha \to \alpha \to \mathbb{R}$
  assumes *finite* S **and** $S \neq \emptyset$
  assumes $\forall s, s' \in S.\ 0 \leq \tau\ s\ s'$ **and** $\forall s \in S.\ \left(\sum_{s' \in S} \tau\ s\ s'\right) = 1$

In this section we will assume this locale, i.e. a discrete-time Markov chain with the finite state space S and transition matrix τ.

We write $E\ s$ for the set of all *successor states*, i.e. all $s' \in S$ with $\tau\ s\ s' \neq 0$. Note that a path ω does not require that $\omega\ (i+1)$ is a successor of $\omega\ i$.

$$\begin{array}{lcl} E & :: & \alpha \to \alpha\ \textit{set} \\ E\ s & = & \{s' \in S \mid \tau\ s\ s' \neq 0\} \end{array}$$

4.6.1 Construction

The transition matrix τ defines for each state the distribution of the next state. We introduce $\mathcal{D}_s$ as this transition distribution:[5]

$$\begin{array}{lcl} \mathcal{D}_\square & :: & \alpha \to \alpha\ \textit{measure} \\ \mathcal{D}_s & = & \textit{point}\ S\ (\tau\ (\textbf{if}\ s \in S\ \textbf{then}\ s\ \textbf{else}\ s_0)) \end{array}$$

lemma SPACE-$\mathcal{D}$: $\Omega_{\mathcal{D}_s} = S$

lemma SETS-$\mathcal{D}$: $\mathcal{A}_{\mathcal{D}_s} = \mathcal{P}(S)$

lemma PR-$\mathcal{D}$: $s \in S \ \wedge\ A \subseteq S \implies \Pr_{\mathcal{D}_s}\ A = \left(\sum_{s' \in A} \tau\ s\ s'\right)$

lemma PROB-SPACE-$\mathcal{D}$: *prob-measure* $\mathcal{D}_s$

Here s_0 is an arbitrary element from S. Using this in the definition, the distribution $\mathcal{D}_s$ is always a probability measure.

Our first goal is to define a probability measure $\mathcal{T}_s$ on the space of all paths $\mathbb{N} \to S$.[6] Where s is the starting state. The path space $\mathcal{T}_s$ should assign the probability $\tau\ s\ \omega_0 \cdot \tau\ \omega_0\ \omega_1 \cdot \cdots \tau\ \omega_{n-1}\ \omega_n$ to the set of all paths starting with $\omega_0, \omega_1, \ldots, \omega_n$. The measurable sets $\mathcal{A}_{\mathcal{T}_s}$ should allow the projection on each time point, so that the sets $\{\omega \mid \omega\ n = s\}$ are measurable. For this we use the σ-algebra from the product space $\bigotimes_{i \in \mathbb{N}}$ *count* S.

How can we proof the existence of such a measure? Our first option is to use the method by Hurd [39] and define the algebra of finite unions of cylinders, i.e. sets starting with the same prefix. We have Caratheodory's extension theorem available, but to operate on finite unions of cylinders is very cumbersome. Introductions into Markov chains, like in Kwiatkowska *et al.* [50], use the semiring of the cylinders. This simplifies the proof, and Caratheodory's extension theorem is available on semirings too, however we still need to show the countable additivity of the premeasure.

[5] The function $\mathcal{D}_s$ is also called a Markov kernel.
[6] We call it $\mathcal{T}$ for trace space

An alternative is to cast the path measure $\mathcal{T}_s$ out of the infinite product measure with $\mathbb{N} \times S$ as index set. This product measure $\bigotimes_{(n,s)\in\mathbb{N}\times S} \mathcal{D}_s$ is indexed by time n and the current state s, as product factors we use the distribution $\mathcal{D}_s$ where s is the current state from the index. We can simply map an element from this product space into a path in $\mathcal{T}_s$, by starting with $(0, s)$ and then following the states selected at each time point. For this we define *path* recursively:

$$\begin{array}{lcl} \mathit{path} & :: & \alpha \to (\mathbb{N} \times \alpha \to \alpha) \to (\mathbb{N} \to \alpha) \\ \mathit{path}\ s\ \omega\ 0 & = & \omega\ (0, \textbf{if}\ s \notin S\ \textbf{then}\ s_0\ \textbf{else}\ s) \\ \mathit{path}\ s\ \omega\ (n+1) & = & \omega\ (n+1, \mathit{path}\ s\ \omega\ n) \end{array}$$

Using s_0, *path* s is even measurable when s is not in S.

lemma PATH-MEASURABLE:
$\mathit{path}\ s \in \mathit{measurable}\ (\bigotimes_{(n,s)\in\mathbb{N}\times S} \mathcal{D}_s)\ (\bigotimes_{n\in\mathbb{N}} \mathit{count}\ S)$

This guarantees that *path* s defines a probability measure on $\bigotimes_{n\in\mathbb{N}} \mathit{count}\ S$. We call this probability measure $\mathcal{T}_s$:

$$\begin{array}{lcl} \mathcal{T}_\square & :: & \alpha \to \Big(\mathbb{N} \to \alpha\Big)\ \mathit{measure} \\ \mathcal{T}_s & = & \mathit{distr}\ (\bigotimes_{(n,s)\in\mathbb{N}\times S} \mathcal{D}_s)\ (\bigotimes_{n\in\mathbb{N}} \mathit{count}\ S)\ (\mathit{path}\ s) \end{array}$$

$\mathcal{T}_s$ is now the trace space for our Markov chain. With Theorem μ-$\bigotimes$-INF and Lemma μ-DISTR we derive the equation for cylinder sets, i.e. the set of all traces starting with the same prefix:

theorem PR-$\mathcal{T}$:
$(\forall i < n.\ \omega\ i \in S) \ \wedge\ s \in S \implies$
$\Pr_{\mathcal{T}_s}\ \{\omega' \mid \forall i < n.\ \omega'\ i = \omega\ i\} = \prod_{i<n} \tau\ ((s'{\cdot}\omega)\ i)\ (\omega\ i)$

Notation: *In the rest of this thesis we write the AE-quantifier on the path measure $\mathcal{T}_s$ as $AE_s\ \omega.\ P\ \omega$ instead of $AE_{\mathcal{T}_s}\ \omega.\ P\ \omega$ and the probability $\Pr_{\mathcal{T}_s}$ is written as $\Pr_s$.*

From this we derive the *Markov property*, i.e. the probability that the Markov chain transitions from $t\ n$ to $t\ (n+1)$ is independent of the Markov chain's previous states. This property is also called *memoryless* as the Markov chain does not remember the states before its current state. While we use the conditional probability on two different events, we only need to show that $\forall i \leq n.\ \omega\ i = t\ i$ has a nonzero probability, as the probability for $\omega\ n = t\ n$ is then also nonzero.

theorem MARKOV-PROPERTY:
$s \in S\ \wedge\ \Pr_s(\omega.\ \forall i \leq n.\ \omega\ i = t\ i) \neq 0 \implies$
$\Pr_s(\omega.\ \omega\ (n+1) = t\ (n+1) \mid \forall i \leq n.\ \omega\ i = t\ i) =$
$\Pr_s(\omega.\ \omega\ (n+1) = t\ (n+1) \mid \omega\ n = t\ n)$

The Markov chain we construct is also *time homogeneous*, i.e. the transition probability is not time dependent.

theorem TIME-HOMOGENEOUS:
$s \in S \wedge \Pr_s(\omega.\ \omega\ i = a) \neq 0 \wedge \Pr_s(\omega.\ \omega\ j = a) \neq 0 \implies$
$\Pr_s(\omega.\ \omega\ (i+1) = b \mid \omega\ i = a) = \Pr_s(\omega.\ \omega\ (j+1) = b \mid \omega\ j = a)$

The statements proved in the Theorems MARKOV-PROPERTY and TIME-HOMOGENEOUS are the defining properties of a discrete-time time-homogeneous Markov chain. Hence we know that we constructed the correct probability measure for the Markov chain given by τ.

4.6.2 Iterative Equations

The Markov chain induces *iterative equations* on the measure $\mathcal{T}_s$, the Lebesgue integral and the AE-quantifier, relating properties about s to properties of $E\ s$, states that are not successors of s are ignored. These equations are often useful in inductive proofs.

theorem PR-EQ-SUM:
$s \in S \ \wedge\ A \in \mathcal{A}_{\mathcal{T}_s} \implies$
$\Pr_s\ A = (\sum_{s' \in E\ s} \tau\ s\ s' \cdot \Pr_{s'}(\omega.\ s' \cdot \omega \in A))$

theorem $\int^P$-EQ-SUM:
$s \in S \ \wedge\ f \in \mathit{measurable}\ \mathcal{T}_s\ \mathcal{B}_{\overline{\mathbb{R}}} \implies$
$\int^P f\ d\mathcal{T}_s = \sum_{s' \in E\ s} \tau\ s\ s' \cdot \int^P \omega.\ f\ (s' \cdot \omega)\ d\mathcal{T}_{s'}$

theorem AE-EQ-SUM:
$s \in S \ \wedge\ \{\omega \mid P\ \omega\} \in \mathcal{A}_{\mathcal{T}_s} \implies$
$\Big(\mathrm{AE}_s\ \omega.\ P\ \omega\Big) \Leftrightarrow \Big(\forall s' \in E\ s.\ \mathrm{AE}_{s'}\ \omega.\ P\ (s' \cdot \omega)\Big)$

We prove the iterative equation for $\Pr_s$ by using Theorem MEASURE-EQI-GENERATOR-EQ. We show the equality of $\mathcal{T}_s$ to an iterative measure whose measure equals the right side of Theorem PR-EQ-SUM. As generator we use the cylinder sets $\{\omega \mid \forall i < n.\ \omega\ i = \omega'\ i\}$. Based on this the integral equation is shown by induction on the Borel-measurable function f.

4.6.3 Reachability

A state s' is *reachable in* Φ *starting in* s iff there is a nonzero probability to reach s' by only going through the specific set of states Φ. The starting state s and the final state s' are not necessary in Φ.

$\mathit{reachable} :: \alpha\ \mathit{set} \to \alpha \to \alpha\ \mathit{set}$
$\mathit{reachable}\ \Phi\ s \ \Leftrightarrow\ \{s' \in S \mid \exists \omega \in \mathbb{N} \to S, n.\ (\forall i \leq n.\ \omega\ i \in E\ ((s \cdot \omega)\ i) \wedge \omega\ n = s' \ \wedge\ (\forall i < n.\ \omega\ i \in \Phi)\}$

Reachability is a purely qualitative property, as it is defined on the graph of nonzero transitions. Hence an upper bound R of $\mathit{reachable}\ \Phi\ s$ is given when all successor states of $R \cap \Phi$ are in R again.

lemma REACHABLE-CLOSED:
$s \in R \cap \Phi \ \wedge\ (\forall t \in R \cap \Phi.\ E\ t \subseteq R) \ \wedge\ R \subseteq S \ \wedge\ \Phi \subseteq S \implies$
$\mathit{reachable}\ \Phi\ s \subseteq R$

The *until-operator* introduces a similar concept on paths. Its definition does not assume that a state is a successor state of the previous one, as this is already ensured by the probability measure $\mathcal{T}_s$.

$$\begin{array}{lcl} \mathit{until} & :: & \alpha\ \mathit{set} \to \alpha\ \mathit{set} \to \left(\mathbb{N} \to \alpha\right) \mathit{set} \\ \mathit{until}\ \Phi\ \Psi & = & \{\omega \mid \exists n.\ (\forall i < n.\ \omega\ i \in \Phi) \wedge \omega\ n \in \Psi\} \end{array}$$

Can we compute the probability of $\Pr_s(\mathit{until}\ \Phi\ \Psi)$ by only using *reachable*? It is easy to show that $\Pr_s(\mathit{until}\ \Phi\ \Psi) = 0$ iff $(\mathit{reachable}\ \Phi\ s) \cap \Psi = \emptyset$. But is there also a method to characterize $\Pr_s(\mathit{until}\ \Phi\ \Psi) = 1$ in terms of *reachable*? For this we need to introduce state fairness. A path ω is *state fair* w.r.t. s and t if t appears infinitely often as the successor of s in ω, provided that s appears infinitely often. The definition and proofs about state fairness are based on Baier [6].

$$\begin{array}{lcl} \mathit{fair} & :: & \alpha \to \alpha \to \left(\mathbb{N} \to \alpha\right) \mathit{set} \\ \mathit{fair}\ s\ t & = & \left\{\omega \;\middle|\; \mathit{finite}\left\{i \mid \omega\ i = s \wedge \omega\ (i+1) = t\right\} \implies \mathit{finite}\left\{i \mid \omega\ i = s\right\}\right\} \end{array}$$

Baier [6] defines state fairness and a more general version called p-fairness, but we only need state fairness. We show that almost every path is state fair for each state and its successors.

theorem AE-FAIR:
$s \in S \ \wedge\ s' \in S \ \wedge\ t' \in E\ s' \implies \mathrm{AE}_s\ \omega.\ s{\cdot}\omega \in \mathit{fair}\ s'\ t'$

Using this we prove that starting in a state s almost every path fulfills $\mathit{until}\ \Phi\ \Psi$ if (1) all states reachable by Φ are in Φ or Ψ and (2) each state reachable from s has again the possibility to reach Ψ. This theorem allows us to prove that $\mathit{until}\ \Phi\ \Psi$ a.e.-holds by a reachability analysis on the graph, and hence $\Pr_s(\mathit{until}\ \Phi\ \Psi) = 1$.

corollary AE-UNTIL:
$s \in \Phi \wedge \Phi \subseteq S \wedge \mathit{reachable}\ (\Phi \setminus \Psi)\ s \subseteq \Phi \cup \Psi \wedge$
$\left(\forall t \in (\mathit{reachable}\ (\Phi \setminus \Psi)\ s \cup \{s\}) \setminus \Psi.\ \mathit{reachable}\ (\Phi \setminus \Psi)\ t \cap \Psi \neq \emptyset\right) \implies$
$\mathrm{AE}_s\ \omega.\ s{\cdot}\omega \in \mathit{until}\ \Phi\ \Psi$

4.6.4 Hitting Time

The *hitting time* on a path ω is the first index at which a state from a set Φ occurs.

$$\begin{array}{lcl} \mathit{hitting} & :: & \alpha\ \mathit{set} \to \left(\mathbb{N} \to \alpha\right) \to \mathbb{N} \\ \mathit{hitting}\ \Phi\ \omega & = & \mathit{LEAST}\ i.\ \omega\ i \in \Phi \end{array}$$

For the computation of rewards it is important to know if the expected hitting time is finite. Standard textbook proofs assume an irreducible chain. We took such a proof from [53], and adapted it to our setting. Instead of an irreducible chain we assume Φ is always reached from s. We show that the expected hitting time of Φ for paths starting in s is finite if almost every path starting in s reaches Φ.

theorem $\int^P$-HITTING-TIME-FINITE:
$s \in S \ \wedge\ \Phi \subseteq S \ \wedge\ (\mathrm{AE}_s\ \omega.\ s{\cdot}\omega \in \mathit{until}\ S\ \Phi) \implies$
$\left(\int^P \omega.\ \mathit{hitting}\ \Phi\ (s{\cdot}\omega)\ d\mathcal{T}_s\right) < \infty$

Chapter 5

Applications

In this chapter we apply the probability theory developed in the previous chapters to the following applications.

pCTL model checking: Probabilistic model checkers, like PRISM [51] or MRMC [45], interpret Markov chains and analyze quantitative properties, specified as probabilistic CTL (pCTL) formulas [30]. We formalize and verify the algorithm used by these probabilistic model checkers.

This work is published in Hölzl and Nipkow [38].

ZeroConf protocol: Network protocols without central services need randomization for symmetry breaking. We analyze the probability for double allocation and the expected runtime for the ZeroConf protocol [15]. We model its address allocation run as a Markov chain, based on Bohnenkamp *et al.* [13].

This work is published in Hölzl and Nipkow [37].

Crowds protocol: Anonymizing services deploy random choice to conceal the original sender when connecting to a server. We formalize the Crowds protocol by Reiter and Rubin [67]. We analyze the probability that the initiating sender of a message contacts an attacker and the information the attacker gains when this happens. To analyze this the path establishment is modelled as a Markov chain.

This work is published in Hölzl and Nipkow [37].

Köpf-Dürmuth countermeasure: Köpf and Dürmuth [48] analyze a countermeasure they developed against side channel attacks. We formalized this analysis using the information theory developed in this thesis.

The Isabelle theories for pCTL model checking, the ZeroConf protocol and the Crowds protocol can be found in the AFP [36]. The Köpf-Dürmuth countermeasure is found in the Isabelle repository.

5.1 pCTL Model Checking

5.1.1 pCTL Formulas

We do not introduce a labeled Markov chain as [50] does, instead we define labels to be subsets of S. We introduce a Markov chain with rewards as a Markov chain with ρ, the rewards associated per state, and ι, the rewards associated per transitions. These rewards are nonnegative, real numbers.

locale *markov-chain-with-reward* = *markov-chain*+
  fixes $\rho :: \alpha \to \mathbb{R}$ **and** $\iota :: \alpha \to \alpha \to \mathbb{R}$
  assumes $\forall s \in S.\ 0 \leq \rho\ s$ **and** $\forall s, s' \in S.\ 0 \leq \iota\ s\ s'$

For the rest of this section we assume a Markov chain with rewards, with the state space S, the transition matrix τ, and the reward functions ρ and ι.

The pCTL syntax is introduced as an inductive data type.

$$\begin{array}{lrcl}
\textbf{datatype} & \mathit{sform} & = & \mathit{label}\ \mathcal{P}(S) \mid \neg \mathit{sform} \mid \mathit{sform} \wedge \mathit{sform} \mid \\
 & & & P^{\bowtie \mathbb{R}}\ \mathit{pform} \mid E^{\bowtie \mathbb{R}}\ \mathit{eform} \\
\textbf{and} & \mathit{pform} & = & X\ \mathit{sform} \mid \mathit{sform}\ U^{\leq \mathbb{N}}\ \mathit{sform} \mid \mathit{sform}\ U^{\infty}\ \mathit{sform} \\
\textbf{and} & \mathit{eform} & = & C^{<\mathbb{N}} \mid I^{=\mathbb{N}} \mid F^{\infty}\ \mathit{sform} \\
\textbf{and} & \bowtie & = & \leq \mid < \mid = \mid > \mid \geq
\end{array}$$

Informally, a state s fulfills $P^{\bowtie r}\ \Phi$ (or $E^{\bowtie r}\ \Phi$) if the probability (expected reward) of the paths starting in s and fulfilling Φ is related with $\bowtie r$. A path fulfills $X\ \Phi$ if its second state fulfills Φ. A path fulfills $\Phi\ U^{\leq k}\ \Psi$ (or $\Phi\ U^{\infty}\ \Psi$, the unbounded until) if it stays in Φ, until it reaches Ψ in at least k steps (at some step). The reward $C^{<k}$ sums all state and transitions rewards for the first k steps, $I^{=k}$ is the state reward at step k, and the unbounded cumulated reward $F^{\infty}\ \Phi$ sums rewards until Φ is reached, if it is never reached it is infinity.

We define now semantics to assign a formal meaning to the pCTL syntax, cf. [30, 50].

$$\begin{array}{lcl}
[\![\Box]\!]_S & :: & \mathit{sform} \to \alpha\ \mathit{set} \\
[\![\mathit{label}\ S']\!]_S & = & \{s \in S \mid s \in S'\} \\
[\![\neg\ \Phi]\!]_S & = & S \setminus [\![\Phi]\!]_S \\
[\![\Phi\ \wedge\ \Psi]\!]_S & = & [\![\Phi]\!]_S \cap [\![\Psi]\!]_S \\
[\![P^{\bowtie r}\ \Phi]\!]_S & = & \left\{ s \in S \;\middle|\; \Pr_s\left(\omega.\ [\![\Phi, s{\cdot}\omega]\!]_P\right) \bowtie r \right\} \\
[\![E^{\bowtie r}\ \Phi]\!]_S & = & \left\{ s \in S \;\middle|\; \int \omega.\ [\![\Phi, s{\cdot}\omega]\!]_E\, d\mathcal{T}_s \bowtie r \right\}
\end{array}$$

$$\begin{array}{lcl}
[\![\Box, \Box]\!]_P & :: & \mathit{pform} \to (\mathbb{N} \to \alpha) \to \mathbb{B} \\
[\![X\ \Phi, \omega]\!]_P & = & \omega\ 1 \in [\![\Phi]\!]_S \\
[\![\Phi\ U^{\leq k}\ \Psi, \omega]\!]_P & = & \exists n \leq k.\ \omega\ n \in [\![\Psi]\!]_S \wedge \left(\forall i < n.\ \omega\ i \in [\![\Phi]\!]_S\right) \\
[\![\Phi\ U^{\infty}\ \Psi, \omega]\!]_P & = & \exists n.\ \omega\ n \in [\![\Psi]\!]_S \wedge \left(\forall i < n.\ \omega\ i \in [\![\Phi]\!]_S\right)
\end{array}$$

$$\begin{array}{lcl}
[\![\Box, \Box]\!]_E & :: & \mathit{eform} \to (\mathbb{N} \to \alpha) \to \mathbb{R} \\
[\![C^{<k}, \omega]\!]_E & = & \sum_{i<k} \rho\ (\omega\ i) + \iota\ (\omega\ i)\ (\omega\ (i+1)) \\
[\![I^{=k}, \omega]\!]_E & = & \rho\ (\omega\ k) \\
[\![F^{\infty}\ \Phi, \omega]\!]_E & = & \textsf{if}\ \ \exists i.\ \omega\ i \in [\![\Phi]\!]_S\ \ \textsf{then}\ \left[\!\!\left[C^{<\mathit{hitting}\ [\![\Phi]\!]_S\ \omega}, \omega\right]\!\!\right]_E \\
 & & \textsf{else}\ \infty
\end{array}$$

We see that $[\![\Phi]\!]_S$ is a subset of S and hence also finite. The set $\{\omega \mid [\![\Phi, \omega]\!]_P\}$ is measurable in $\mathcal{T}_s$, and $\lambda\omega.\ [\![\Phi, \omega]\!]_E$ is Borel-measurable on $\mathcal{T}_s$. So the probability for $[\![P^{\bowtie r}\ \Phi]\!]_S$, and the integral for $[\![E^{\bowtie r}\ \Phi]\!]_S$ are well-defined.

5.1.2 Computable HOL Fragment

The pCTL model checking algorithm solves linear equation systems. This may (in general) fail. To cater for this possibility we use the option values in our computation and formulate our algorithm with the help of the **do**-syntax.

To represent such non-total functions in HOL we use the option data type

$$\textsf{datatype}\ \alpha\ \mathit{option} = \mathit{Some}\ \alpha \mid \mathit{None}$$

whose values are *Some* x for $x :: \alpha$ and *None*. We introduce the option-monad to combine non-total functions to new non-total functions. The infix bind-operator $\gg\!=$ is defined by the equations $((\mathit{Some}\ x) \gg\!= f) = f\ x$ and $(\mathit{None} \gg\!= f) = \mathit{None}$. Notation **return** is equal to *Some*. Similar to Haskell's monad-syntax we use the **do**-syntax to represent chains of bind-operators, for example:

$$\begin{array}{lll}
\textsf{do} & x \leftarrow f & \qquad\qquad f \gg\!= (\lambda x. \\
 & y \leftarrow g\ x & \Longrightarrow \quad g\ x \gg\!= (\lambda y. \\
 & \textsf{let}\ z = h\ x\ y & \qquad \mathit{Some}\ (x + y + h\ x\ y))) \\
 & \textsf{return}\ (x + y + z) &
\end{array}$$

We use the option-monad not only to represent non-total functions, but also to write the algorithm in a more imperative style. The only non-total function in the pCTL model checking algorithm is Gauss-Jordan elimination.

The while-combinator **while** satisfies the standard recursion equation:

$$\begin{array}{lcl}
\textsf{while} & :: & (\alpha \to \mathbb{B}) \to (\alpha \to \alpha) \to \alpha \to \alpha \\
\textsf{while}\ P\ f\ x & = & \textsf{if}\ P\ x\ \textsf{then while}\ P\ f\ (f\ x)\ \textsf{else}\ x
\end{array}$$

5.1.3 Verifying the Algorithm

The model checking algorithm *Sat* for pCTL formulas is based on three methods:

- Iterative methods to compute the probability of bounded until and the expectation of bounded rewards
- Reachability analysis on the graph of nonzero transitions to compute the sets $[\![P^{=0}(\Phi\ U^{\infty}\ \Psi)]\!]_S$ and $[\![P^{=1}(\Phi\ U^{\infty}\ \Psi)]\!]_S$.

- Solving systems of linear equations for the unbounded until operator and unbounded rewards. This requires the previous methods to construct a system of linear equations with a unique solution.

The definition and the correctness proof of the algorithm *Sat* is by induction over the syntax of pCTL formulas. For a better overview of the formalization we split the definition of *Sat* into multiple parts interleaved with the necessary auxiliary definitions. The final soundness theorem states that *Sat* Φ returns a result and computes the set of states s for which $s \in [\![\Phi]\!]_S$ holds, i.e. *Sat* Φ = *Some* $[\![\Phi]\!]_S$.

The definition of *Sat* on *label* S', $\neg\Phi$, $\Phi \wedge \Psi$, and $P^{\bowtie r}(X\Phi)$ is easy. The soundness proof of the first three is done automatically, the last one needs Theorem Pr-eq-sum.

$$
\begin{array}{lcl}
\mathit{Sat} & :: & \mathit{sform} \to \alpha\ \mathit{set\ option} \\
\mathit{Sat}\ (\mathit{label}\ S') & = & \textbf{return}\ \{s \in S \mid s \in S'\} \\
\mathit{Sat}\ (\neg\ \Phi) & = & \textbf{do} \\
 & & F \leftarrow \mathit{Sat}\ \Phi \\
 & & \textbf{return}\ (S \setminus F) \\
\mathit{Sat}\ (\Phi\ \wedge\ \Psi) & = & \textbf{do} \\
 & & F_1 \leftarrow \mathit{Sat}\ \Phi \\
 & & F_2 \leftarrow \mathit{Sat}\ \Psi \\
 & & \textbf{return}\ (F_1 \cap F_2) \\
\mathit{Sat}\ (P^{\bowtie r}\ (X\ \Phi)) & = & \textbf{do} \\
 & & F \leftarrow \mathit{Sat}\ \Phi \\
 & & \textbf{return}\ \{s \in S \mid (\sum_{s' \in F} \tau\ s\ s') \bowtie r\}
\end{array}
$$

The iterative methods to compute bounded until (*ProbUB* $k\ s\ S_1\ S_2$), cumulative expectation (*ExpC* $k\ s$) and state expectation (*ExpI* $k\ s$) are simply defined by recursion on the bounding value k. Soundness is proved by induction on the bounding value k and using the iterative equations given by Theorems Pr-eq-sum and $\int^P$-eq-sum.

$$
\begin{array}{lcl}
\mathit{ProbUB} & :: & \mathbb{N} \to \alpha \to \alpha\ \mathit{set} \to \alpha\ \mathit{set} \to \mathbb{R} \\
\mathit{ProbUB}\ 0\ s\ S_1\ S_2 & = & \textbf{if}\ s \in S_2\ \textbf{then}\ 1\ \textbf{else}\ 0 \\
\mathit{ProbUB}\ (k+1)\ s\ S_1\ S_2 & = & \textbf{if}\ s \in S_1 \setminus S_2 \\
 & & \textbf{then}\ \sum_{s' \in S} \tau\ s\ s' \cdot \mathit{ProbUB}\ k\ s'\ S_1\ S_2 \\
 & & \textbf{else}\ (\textbf{if}\ s \in S_2\ \textbf{then}\ 1\ \textbf{else}\ 0)
\end{array}
$$

$$
\begin{array}{lcl}
\mathit{ExpC} & :: & \mathbb{N} \to \alpha \to \mathbb{R} \\
\mathit{ExpC}\ 0\ s & = & 0 \\
\mathit{ExpC}\ (k+1)\ s & = & \rho\ s + \sum_{s' \in S} \tau\ s\ s' \cdot (\iota\ s\ s' + \mathit{ExpC}\ k\ s')
\end{array}
$$

$$
\begin{array}{lcl}
\mathit{ExpI} & :: & \mathbb{N} \to \alpha \to \mathbb{R} \\
\mathit{ExpI}\ 0\ s & = & \rho\ s \\
\mathit{ExpI}\ (k+1)\ s & = & \sum_{s' \in S} \tau\ s\ s' \cdot \mathit{ExpI}\ k\ s'
\end{array}
$$

$$
\begin{array}{lcl}
\mathit{Sat}\ (P^{\bowtie r}\ (\Phi\ U^{\leq k}\ \Psi)) & = & \textbf{do} \\
 & & F_1 \leftarrow \mathit{Sat}\ \Phi \\
 & & F_2 \leftarrow \mathit{Sat}\ \Psi \\
 & & \textbf{return}\ \left\{s \in S \mid \mathit{ProbUB}\ k\ s\ F_1\ F_2 \bowtie r\right\} \\
\mathit{Sat}\ (E^{\bowtie r}\ (C^{<k})) & = & \textbf{return}\ \left\{s \in S \mid \mathit{ExpC}\ k\ s \bowtie r\right\} \\
\mathit{Sat}\ (E^{\bowtie r}\ (I^{=k})) & = & \textbf{return}\ \left\{s \in S \mid \mathit{ExpI}\ k\ s \bowtie r\right\}
\end{array}
$$

Our next step is to check the unbounded until operator. Here we compute the probability $P_{\Phi,\Psi}(s) = \Pr_s\left(\omega.\ [\![\Phi\ U^{\infty}\ \Psi, s{\cdot}\omega]\!]_P\right)$ for each state s by setting up a system of linear equations. From Theorem PR-EQ-SUM and the behavior of the unbounded until operator we derive a system of linear equations for $P_{\Phi,\Psi}(s)$.

$$
P_{\Phi,\Psi}(s) = \begin{cases} \sum_{s' \in E(s)} \tau\ s\ s' \cdot P_{\Phi,\Psi}(s') & \text{if} \quad s \in \Phi \setminus \Psi \\ 1 & \text{if} \quad s \in \Psi \\ 0 & \text{otherwise} \end{cases}
$$

We show that such a linear equation system has a unique solution, with two conditions: (1) the solutions are equal on Ψ and (2) the solutions are equal in all states which never reach Ψ, i.e. $P_{\Phi,\Psi}(s) = 0$. We proved this lemma following the uniqueness proof in [30].

lemma UNIQUE-SOLUTION:

$$
\begin{array}{l}
\Phi \subseteq S\ \wedge\ \Psi \subseteq N \subseteq S\ \wedge \\
\left(\forall s \in S.\ P_{\Phi,\Psi}(s) = 0\ \Longrightarrow\ s \in N\right)\ \wedge \\
\left(\forall s \in S \setminus N.\ l_1\ s - c\ s = \textstyle\sum_{s' \in S} \tau\ s\ s' \cdot l_1\ s'\right)\ \wedge \\
\left(\forall s \in S \setminus N.\ l_2\ s - c\ s = \textstyle\sum_{s' \in S} \tau\ s\ s' \cdot l_2\ s'\right)\ \wedge \\
\left(\forall s \in N.\ l_1\ s = l_2\ s\right)\ \Longrightarrow \\
\forall s \in S.\ l_1\ s = l_2\ s
\end{array}
$$

To find a solution of such a system of linear equations, we formalized Gauss-Jordan elimination on matrices represented as functions [59]. Then we adapted this to use states as indices instead of natural numbers. Correctness says that if *gauss M a* returns *Some x*, then x is a solution to the equation system $M \cdot x = a$.

lemma GAUSS-JORDAN-ELIMINATION:

$$
\mathit{gauss}\ M\ a = \mathit{Some}\ x\ \Longrightarrow\ \forall s \in S.\ (\textstyle\sum_{s' \in S} M\ s\ s' \cdot x\ s') = a\ s
$$

Before we use the uniqueness of our system of linear equations, Lemma UNIQUE-SOLUTION requires us to compute the states with $P_{\Phi,\Psi}(s) = 0$ before the algorithm builds the system of linear equations. *ProbZ* computes the set of all states with $P_{\Phi,\Psi}(s) > 0$ and returns the complement. The set of all s with $P_{\Phi,\Psi}(s) > 0$ is computed by starting with $R = \Psi$ and adding states to R which are in Φ and are predecessors of a state in R. With Lemma REACHABLE-CLOSED we know that R contains all reachable states, hence $P_{\Phi,\Psi}(s) > 0$ for all $s \in R$. The termination measure for the **while**-combinator is the difference $S \setminus R$, with each step either states are

added, or the loop terminates.

$$\begin{array}{lcl} \mathit{pred} & :: & \alpha\ \mathit{set} \to \alpha\ \mathit{set} \to \alpha\ \mathit{set} \\ \mathit{pred}\ \Phi\ R & = & \{s \in \Phi \mid R \cap E(s) \neq \emptyset\} \\ \\ \mathit{ProbZ} & :: & \alpha\ \mathit{set} \to \alpha\ \mathit{set} \to \alpha\ \mathit{set} \\ \mathit{ProbZ}\ \Phi\ \Psi & = & S \setminus \textbf{while}\ (\lambda R.\ \neg \mathit{pred}\ \Phi\ R \subseteq R)\ (\lambda R.\ R \cup \mathit{pred}\ \Phi\ R)\ \Psi \end{array}$$

The system of linear equations solved by *gauss M a* needs to be in the right form, i.e. the matrix M contains all variable coefficients and a all constants. We introduce *LES F* to define the matrix of the linear equation system $l\ s = (\sum_{s \in S} \tau\ s\ s' \cdot l\ s') + a\ s$ for $s \notin F$, and $l\ s = a\ s$ if $s \in F$.

$$\begin{array}{lcl} \mathit{LES} & :: & \alpha\ \mathit{set} \to \alpha \to \alpha \to \mathbb{R} \\ \mathit{LES}\ F\ r\ c & = & \textbf{if}\ r \in F\ \ \textbf{then}\ (\textbf{if}\ c = r\ \textbf{then}\ 1\ \textbf{else}\ 0) \\ & & \quad \textbf{else}\ (\textbf{if}\ c = r\ \textbf{then}\ \tau\ r\ c - 1\ \textbf{else}\ \tau\ r\ c) \end{array}$$

Combining all these functions we can finally compute the probability of an unbounded until formula. We prove its soundness using Lemmas GAUSS-JORDAN-ELIMINATION and UNIQUE-SOLUTION, and Theorem PR-EQ-SUM.

$$\begin{array}{lcl} \mathit{Sat}\ (P^{\bowtie r}\ (\Phi\ U^{\infty}\ \Psi)) & = & \textbf{do} \\ & & \quad F_1 \leftarrow \mathit{Sat}\ \Phi \\ & & \quad F_2 \leftarrow \mathit{Sat}\ \Psi \\ & & \quad p \leftarrow \mathit{gauss}\ (\mathit{LES}\ (F_2 \cup \mathit{ProbZ}\ F_1\ F_2)) \\ & & \qquad\qquad (\lambda s.\ \textbf{if}\ s \in F_2\ \textbf{then}\ 1\ \textbf{else}\ 0) \\ & & \quad \textbf{return}\ \{s \in S \mid p\ s \bowtie r\} \end{array}$$

The last equation of *Sat* computes the unbounded reward $E^{\bowtie r}(F^{\infty}\ \Phi)$. Similar to the unbounded until operator, we introduce a system of linear equations for $R_\Phi(s) = \int_\omega [\![F^{\infty}\ \Phi, s{\cdot}\omega]\!]_E d\Pr_s$. With Theorem $\int^P$-HITTING-TIME-FINITE we know that $R_\Phi(s)$ is finite if $P_{S,\Phi}(s) = 1$. If $P_{S,\Phi}(s) < 1$ there is a nonzero probability that Φ is never reached, and hence $R_\Phi(s) = \infty$.

$$R_\Phi(s) = \begin{cases} \sum_{s' \in E(s)} \tau\ s\ s' \cdot (\rho\ s + \iota\ s\ s' + R_\Phi(s')) & \text{if} \quad P_{S,\Phi}(s) = 1 \wedge s \notin \Phi \\ 0 & \text{if} \quad s \in \Phi \\ \infty & \text{otherwise} \end{cases}$$

To be usable with *LES*, we rewrite the first equation into:

$$R_\Phi(s) - \left(\rho\ s + \sum_{s' \in E(s)} \tau\ s\ s' \cdot \iota\ s\ s'\right) = \sum_{s' \in E(s)} \tau\ s\ s' \cdot R_\Phi(s')\ .$$

The Gauss-Jordan elimination we use works only on real numbers, luckily we can replace ∞ by 0 and replace it again after we solved the equation system. This is sound since for each s and $s' \in E(s)$ with $R_\Phi(s') = \infty$ either $s \in \Phi$ or $R_\Phi(s) = \infty$ hold. The states s with $P_{S,\Phi}(s) = 1$ are computed by *ProbOne*, building on *ProbZ*.

$$\begin{array}{lcl} \mathit{ProbOne} & :: & \alpha\ \mathit{set} \to \alpha\ \mathit{set} \to \mathbb{R} \\ \mathit{ProbOne}\ \Phi\ \Psi & = & \mathit{ProbZ}\ (\Phi \setminus \Psi)\ (\mathit{ProbZ}\ \Phi\ \Psi) \end{array}$$

We know that the resulting states only reach states which again reach Ψ, hence the assumptions of Corollary AE-UNTIL are fulfilled, and we know that *ProbOne S* Φ is the set of all states s with $P_{S,\Phi}(s) = 1$. With all this, we can formalize the last equation for *Sat*.

$$\begin{aligned}
\mathit{Sat}\ (E^{\bowtie r}\ (F^{\infty}\ \Phi)) \quad = \quad & \textbf{do} \\
& F \leftarrow \mathit{Sat}\ \Phi \\
& \textbf{let}\ Y = \mathit{ProbOne}\ S\ F \\
& l \leftarrow \mathit{gauss}\ (\mathit{LES}\ (S \setminus (Y \setminus F))) \\
& \quad (\lambda s.\ \textbf{if}\ i \in Y \setminus F\ \textbf{then}\ -(\rho\ s + (\textstyle\sum_{s' \in S}.\ \tau\ s\ s' \cdot \iota\ s\ s')) \\
& \qquad\qquad \textbf{else}\ 0) \\
& \textbf{let}\ e = (\lambda s.\ \textbf{if}\ s \in Y\ \textbf{then}\ l\ s\ \textbf{else}\ \infty) \\
& \textbf{return}\ \{s \in S \mid e\ s \bowtie r\}
\end{aligned}$$

Finally we show the soundness of *Sat* by induction on the structure of Φ. If we assume that *Sat* terminates with a result F, then F is the same set as defined by the semantic.

theorem SOUND-SAT: $\mathit{Sat}\ \Phi = \mathit{Some}\ F \implies [\![\Phi]\!]_S = F$

Now we turn to completeness. The only case in which *Sat* returns *None* is when the Gauss-Jordan elimination does not find a unique solution. Hence we need the property that if a unique solution exists, then *gauss* returns this solution.

If there is a unique solution x for $M \cdot x = a$ then *gauss* returns a result:

lemma COMPLETE-GAUSS:
$\Big(\forall s \in S.\ \textstyle\sum_{s' \in S} M\ s\ s' \cdot x\ s' = a\ s\Big) \wedge$
$\Big(\forall y.\ (\forall s \in S.\ \textstyle\sum_{s' \in S} M\ s\ s' \cdot y\ s' = a\ s) \implies \forall s \in S.\ x\ s = y\ s\Big) \implies$
$\exists x'.\ \mathit{gauss}\ M\ a = \mathit{Some}\ x'$

With this and Lemma UNIQUE-SOLUTION we prove that *Sat* always returns a result:

theorem COMPLETE-SAT: $\exists F.\ \mathit{Sat}\ \Phi = \mathit{Some}\ F$

Using Theorem SOUND-SAT we finally show

corollary SOUND-AND-COMPLETE-SAT: $\mathit{Sat}\ \Phi = \mathit{Some}\ [\![\Phi]\!]_S$

5.1.4 Discussion

We used the tutorial [50] as a guideline to formalize the pCTL model checking algorithm. Most parts of the soundness proof are straightforward. Three parts, however, required a more substantial formalization of the background theory:

- The correctness of *ProbOne* is based on Theorem AE-FAIR, which required us to formalize state fairness as found in [6].
- For the unbounded until and the unbounded rewards we solve a linear equation system. We needed to show that the solution of this equation system is unique, for which we followed the original proof from [30].

- The unbounded reward for a state can only be characterized as a linear equation if the reward is finite. We needed Theorem $\int^P$-HITTING-TIME-FINITE to show that the reward is finite, if the final states are almost always reached.

Technically, the largest difference between our work and Kwiatkowska *et al.* [50] is the construction of the probability space of paths: we use infinite products of probability spaces, whereas they use Caratheodory on semirings. We do not need to show that the probability of cylinders is countably additive, this is generically done for infinite products. We want to reuse the infinite products for continuous-time Markov chains and Markov decision processes.

The equations we give for the algorithm are not directly executable by the code generator in Isabelle [27]. We use sets in our equations, and the adaption of Gauss-Jordan elimination uses an arbitrary mapping from $\{0, \ldots, |S| - 1\}$ to S. One method to obtain an executable version is to create a copy Sat_L of *Sat* operating on lists instead of subsets of S. We assume as input a list of states $xs = [s_0, s_1, \ldots s_n]$, and define the Markov chains on $S =$ *set-of* xs. It should be straightforward to show that *Sat* $\Phi =$ *Some* F implies *set-of* $(Sat_L\ \Phi) = F$. The biggest hurdle is the while-combinator in *ProbZ* and the adaption of Gauss-Jordan elimination.

5.2 ZeroConf Protocol

Ad-hoc networks usually do not have a central address authority assigning addresses to new nodes in the network. An example are consumer networks where users want to connect their laptops to exchange data or attach a network capable printer. When connecting with WiFi these devices use IPv4 and hence need IPv4 addresses to communicate with each other.

The *ZeroConf* protocol [15] is a distributed network protocol which allows new hosts in the network to allocate an unused link-local IPv4 address. A link-local address is only valid in the local network, e.g. a WiFi network. We assume point-to-point communication in our local network, and hence directly communicate with each host identified by a valid address. The problem with IPv4 addresses is that they are limited, i.e. they are represented by 32-bit numbers, and for the local network the addresses from 169.254.1.0 to 169.254.254.255 are available, hence we can chose from 65024 distinct addresses. ZeroConf works by randomly selecting an address from this pool and then probing if the address is already in use.

Bohnenkamp *et al.* [13] give a formal analysis of the probability that an address collision happens, i.e. two hosts end up with the same address. They also analyze the expected run time until a (not necessarily valid) address is chosen. As a case study we formalize their analysis in Isabelle/HOL.

Andova *et al.* [3] present a model-checking approach for discrete-time Markov reward chains and apply it to the ZeroConf protocol as a case study. They support multiple reward structures and can compute the probability based on multiple constraints on these reward structures. Kwiatkowska *et al.* [49] have modelled this protocol as a probabilistic timed automata in PRISM. Both models include more

features of the actual protocol than the model by Bohnenkamp *et al.* [13] that we follow.

5.2.1 Description of Address Allocation

We give a short description of the model used in Bohnenkamp *et al.* [13]. The address allocation in ZeroConf uses ARP (address resolution protocol) to detect if an address is in use or not. An ARP request is sent to detect if a specific IPv4 address is already in use. When a host has the requested IPv4 address it answers with an ARP response. ZeroConf allocates a new address as follows:

1. Uniformly select a random address in the range 169.254.1.0 to 169.254.254.255.
2. Send an ARP request to detect if the address is already in use.
3. When a host responds to the ARP request, the address is already taken and we need to start again (go back to 1).
4. When no response arrives before a time limit r, we again send an ARP request. This is repeated N times.
5. When no response arrived for N requests we assume our address is not in use and are finished.

This probabilistic process depends on two parameters: (1) The probability q that the random chosen address is already taken; this probability depends on the number of hosts in the network and the number of available addresses. (2) The probability p that either the ARP request or response is lost.

The Markov chain shown in Fig. 5.1 describes the address allocation from a global viewpoint. At *Start* a new host is added to the network, it chooses an address and sends the first ARP request. There are two alternatives.

- With probability $1 - q$ the host chooses an unused address, the allocation is finished, the Markov chain directly goes to *Ok*. Of course, the host does not know this, and still sends out $N + 1$ ARP probes. Hence we associate the time cost $r \cdot (N + 1)$ with this transition.
- With probability q the host chooses a used address and goes to the probing phase: in the *Prb n* state it sends an ARP request and waits until r time units have passed, or until it receives an ARP response from the address owner. With probability $1 - p$ the host receives an ARP response and needs to choose a new address—we go back to *Start*. With probability p this exchange fails and we go to the next probe phase. After $N + 1$ probes, the host assumes the chosen address is free. As two hosts in the network end up with the same address we reached the *Error* state. The time cost E models the cost to repair the double allocation. This might involve restarting a laptop.

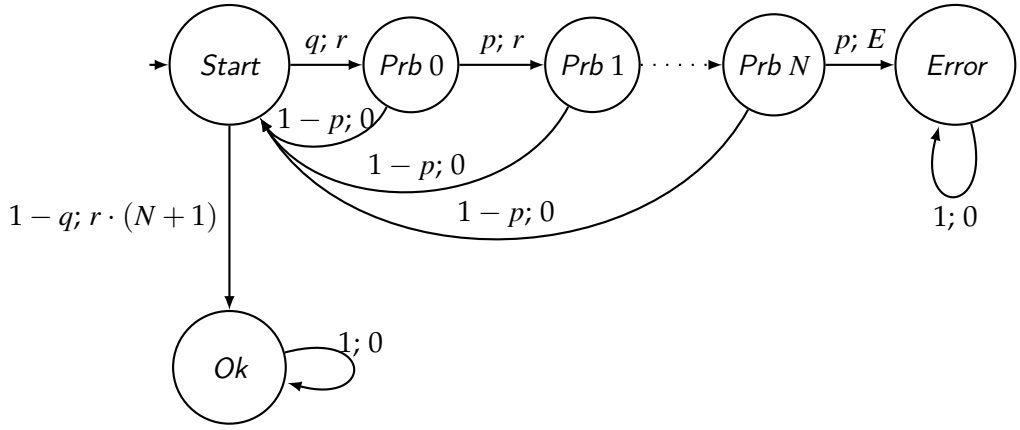

Figure 5.1: Markov chain of the ZeroConf protocol. The labels are annotated with $P;T$: the probability P to take this edge and the elapsed time T.

5.2.2 Formal Model of ZeroConf Address Allocation

The Isabelle/HOL model of the ZeroConf protocol describes the Markov chain in Fig. 5.1. We set up a context containing the probe numbers (starting with 0), the probabilities p and q, and the costs r and E:

locale *zeroconf* =
  fixes $N :: \mathbb{N}$ **and** $p\ q\ r\ E :: \mathbb{R}$
  assumes $0 < p$ **and** $p < 1$ **and** $0 < q$ **and** $q < 1$
  assumes $0 \leq E$ **and** $0 \leq r$

In the following sections we assume that these fixed variables N, p, q, r, and E fulfill the above assumptions of the ZeroConf protocol.

To represent the states in the Markov chain we introduce a new datatype:

datatype *zc-state* = *Start* | *Prb* $\mathbb{N}$ | *Ok* | *Error*

We have the type *zc-state* with the distinct objects *Start*, *Ok*, *Error*, and *Prb n* for all $n :: \mathbb{N}$. The valid states $S ::$ *zc-state set* are a restriction of this to only valid probe numbers. This also gives us a finite number of states.

S :: *zc-state set*
S = $\{Start, Ok, Error\} \cup \{Prb\ n \mid n \leq N\}$

The final modeling step is to define the transition matrix $\tau ::$ *zc-state* $\to$ *zc-state* $\to$ $\mathbb{R}$ and the cost function $\rho ::$ *zc-state* $\to$ *zc-state* $\to \mathbb{R}$. Both are defined by a case distinction on the current state and return the zero function $\mathbf{0}$ updated at the states with nonzero transition probability or cost.

τ :: *zc-state* $\to$ *zc-state* $\to \mathbb{R}$
$\tau\ s$ = **case** s **of** *Start* $\Rightarrow \mathbf{0}(Prb\ 0 := q, Ok := 1 - q)$
  | *Prb* $n \Rightarrow$ **if** $n < N$ **then** $\mathbf{0}(Prb\ (n+1) := p, Start := 1 - p)$
    else $\mathbf{0}(Error := p, Start := 1 - p)$
  | *Ok* $\Rightarrow \mathbf{0}(Ok := 1)$
  | *Error* $\Rightarrow \mathbf{0}(Error := 1)$

$$\begin{array}{lcl}
\rho & :: & \mathit{zc\text{-}state} \to \mathit{zc\text{-}state} \to \mathbb{R} \\
\rho\ s & = & \textbf{case}\ s\ \textbf{of}\ \mathit{Start} \Rightarrow \mathbf{0}(\mathit{Prb}\ 0 := r, \mathit{Ok} := r \cdot (N+1)) \\
 & & \quad \mid \mathit{Prb}\ n \Rightarrow \textbf{if}\ n < N\ \textbf{then}\ \mathbf{0}(\mathit{Prb}\ (n+1) := r) \\
 & & \qquad\qquad\qquad\qquad \textbf{else}\ \mathbf{0}(\mathit{Error} := E) \\
 & & \quad \mid \mathit{Ok} \Rightarrow \mathbf{0} \\
 & & \quad \mid \mathit{Error} \Rightarrow \mathbf{0}
\end{array}$$

Here $f(x := v)$ is the function f updated at x to the new value v, i.e. $f(x := v)\ x = v$.

We need to prove that we actually defined a Markov chain: as a consequence, Isabelle/HOL is able to provide the probabilities $\Pr_s A$ for each state s and path set A. For this we show that τ is a valid transition matrix for a Markov chain on S, and ρ is a valid cost function:

lemma τ-DTMC: *markov-chain-with-reward S τ ρ*

To prove this we need to show that τ and ρ are nonnegative for all states in S. And finally we need to show that $\tau\ s$ is a distribution for all s in S, which is easy to show by using the helper Lemma S-SPLIT:

lemma S-SPLIT: $\sum_{s \in S} f\ s = f\ \mathit{Start} + f\ \mathit{Ok} + f\ \mathit{Error} + \sum_{n \le N} f\ (\mathit{Prb}\ n)$

5.2.3 Probability of an Erroneous Allocation

The correctness property we want to verify is that no collision happens, i.e. we want to compute the probability that a protocol run ends in the *Error* state. The goal of this section is not only to show *what* we proved, but to show *how* we proved it. Most of the proofs are automatic by rewriting and we do not show the details. But we want to show the necessary lemmas and theorems needed to convince Isabelle/HOL.

We define $P_{err} :: \mathit{zc\text{-}state} \to \mathbb{R}$ to reason about the probability that a trace ω ends in the *Error* state when we started in a state s:

$$\begin{array}{lcl}
P_{err} & :: & \mathit{zc\text{-}state} \to \mathbb{R} \\
P_{err}\ s & = & \Pr_s(\omega.\ s{\cdot}\omega \in \mathit{until}\ S\ \{\mathit{Error}\})
\end{array}$$

Our final theorem will be to characterize $P_{err}\ \mathit{Start}$ only in terms of the system parameters p, q and N.

The first obvious result is that when we are already in *Error*, we will stay in *Error*, and when we are in *Ok* we will never reach *Error*:

lemma P_{err}-ERROR: $P_{err}\ \mathit{Error} = 1$
lemma P_{err}-OK: $P_{err}\ \mathit{Ok} = 0$

P_{err}-error is proved by rewriting: $\mathit{Error} \cdot \omega \in \mathit{until}\ S\ \{\mathit{Error}\}$ is always true. The *Ok* case is proved by $\mathit{reachable}\ (S \setminus \{\mathit{Error}\})\ \mathit{Ok} \subseteq \{\mathit{Ok}\}$. Together with lemma S-split and these two lemmas we provide an iterative lemma for P_{err}:

lemma P_{err}-ITER:
$s \in S \implies$
$P_{err}\ s = \tau\ s\ \mathit{Start} \cdot P_{err}\ \mathit{Start} + \tau\ s\ \mathit{Error} + \sum_{n \le N} \tau\ s\ (\mathit{Prb}\ n) \cdot P_{err}\ (\mathit{Prb}\ n)$

However this is a bad rewrite theorem, using it would result in non-termination of the rewrite engine. To avoid this we derive rules for specific states:

lemma P_{err}-LAST-PROBE: $P_{err}\ (\mathit{Prb}\ N) = p + (1-p) \cdot P_{err}\ \mathit{Start}$
lemma P_{err}-START-ITER: $P_{err}\ \mathit{Start} = q \cdot P_{err}\ (\mathit{Prb}\ 0)$

Our next step is to compute the probability to reach *Error* when we are in $\mathit{Prb}\ n$. This is the only proof which is not done by a simple rewrite step, but it requires induction and two separate rewrite steps. The induction is done over the number n of steps until we are in *Error*. To give the reader a better feeling for what these proofs look like, here is the skeleton of the Isabelle proof:

lemma P_{err}-PROBE-ITER:
  $n \leq N \implies P_{err}\ (\mathit{Prb}\ (N-n)) = p^{n+1} + (1-p^{n+1}) \cdot P_{err}\ \mathit{Start}$
proof $(\mathit{induct}\ n)$
  case $(n+1)$
  have $P_{err}\ (\mathit{Prb}\ (N-(n+1))) =$
    $p * (p^{n+1} + (1-p^{n+1}) * P_{err}\ \mathit{Start}) + (1-p) * P_{err}\ \mathit{Start}$
    <*proof*>
  also have $\cdots = p^{(n+1)+1} + (1-p^{(n+1)+1}) \cdot P_{err}\ \mathit{Start}$
    <*proof*>
  finally show $P_{err}\ (\mathit{Prb}\ (N-(n+1))) =$
    $p^{(n+1)+1} + (1-p^{(n+1)+1}) \cdot P_{err}\ \mathit{Start}$.
qed *simp* – The 0-case is a simple rewriting step with P_{err}-last-probe.

Together with P_{err}-start-iter we prove our final theorem:

theorem P_{err}-START: $P_{err}\ \mathit{Start} = (q \cdot p^{N+1})/(1-q \cdot (1-p^{N+1}))$

With typical parameters for the ZeroConf protocol (16 hosts ($q = 16/65024$), 3 probe runs ($N = 2$) and a probability of $p = 0.01$ to lose ARP packets) we compute (by rewriting) in Isabelle/HOL that the probability to reach *Error* is below $1/10^{13}$:

corollary P_{err}-START': $P_{err}\ \mathit{Start} \leq 1/10^{13}$

5.2.4 Expected Running Time of an Allocation Run

Users are not only interested in a very low error probability but also in fast allocation time for network address. Obviously there are runs which may take very long, but the probability for these runs are near zero. So we want to verify that the average running time of an allocation run is in the time range of milliseconds.

The running time of an allocation run $C_{fin} :: S \to \overline{\mathbb{R}}$ is modelled as the integral over the sum of all costs ρ for each step in each run. The sum of all steps until either *Ok* or *Error* is reached is simply *cost-until*:

$$\begin{aligned} C_{fin} &:: \mathit{zc\text{-}state} \to \overline{\mathbb{R}} \\ C_{fin}\ s &= \int_{\omega} \mathit{cost\text{-}until}\ \{\mathit{Error}, \mathit{Ok}\}\ (s \cdot \omega)\ \mathrm{dPr}_s \end{aligned}$$

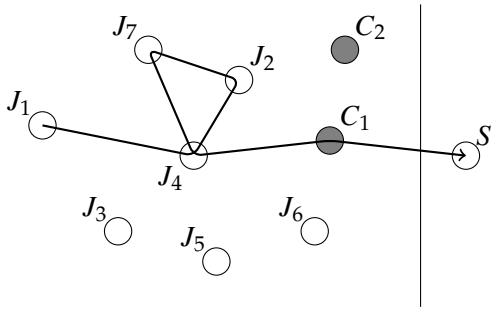

Figure 5.2: The established route $J_1 - J_4 - J_2 - J_7 - J_4 - C_1 - S$

In order to evaluate the integral we first show that it is finite. This is the case if *cost-until* $\{Error, Ok\}$ is a.e.-finite. So we first show that almost every path reaches $\{Error, Ok\}$:

lemma AE-TERM: $s \in S \implies \mathrm{AE}_s\, \omega.\ s \cdot \omega \in \mathit{until}\ S\ \{Error, Ok\}$

Using this we show an elementary form of C_{fin} in a similar way to P_{err}:

theorem C_{fin}-START:
$$C_{fin}\ Start = \frac{q \cdot (r + p^{N+1} \cdot E + r \cdot p \cdot (1 - p^N)/(1 - p)) + (1 - q) \cdot (r \cdot N + 1)}{1 - q + q \cdot p^{N+1}}$$

With typical values (16 hosts, 3 probe runs, a probability of $p = 0.01$ to lose ARP packets, 2 *ms* for an ARP round-trip ($r = 0.002$) and an error penalty of one hour ($E = 3600$)) we compute in Isabelle/HOL that the average time to terminate is less or equal 0.007 *s*:

corollary C_{fin}-START': $C_{fin}\ Start \leq 0.007$

5.3 Crowds Protocol

The *Crowds* protocol described by Reiter and Rubin [67] is an anonymizing protocol. The goal is to allow users to connect to servers anonymously. Neither the final server should know which user connects to it, nor attackers collaborating in the network. The Crowds protocol establishes an anonymizing route through a so called mix network: each user (Reiter and Rubin name them *jondo* pronounced "John Doe") is itself participating in the mix network. When a jondo establishes a route, it first connects to another random jondo which then decides based on a coin flip weighted with p_f if it should connect to the final server, or go through a further jondo, and so on. Figure 5.2 shows an established route through the jondos $J_1 - J_4 - J_2 - J_7 - J_4 - C_1 - S$. There is no global information about a route available to the participating jondos. For each connection a jondo only knows its immediate neighbours, but no other previous or following jondo, so it may happen that a route is going through a loop, as seen in Fig. 5.2.

First, Reiter and Rubin [67] show that the server has no chance to guess the original sender. In a second step they assume that some jondos collaborate to

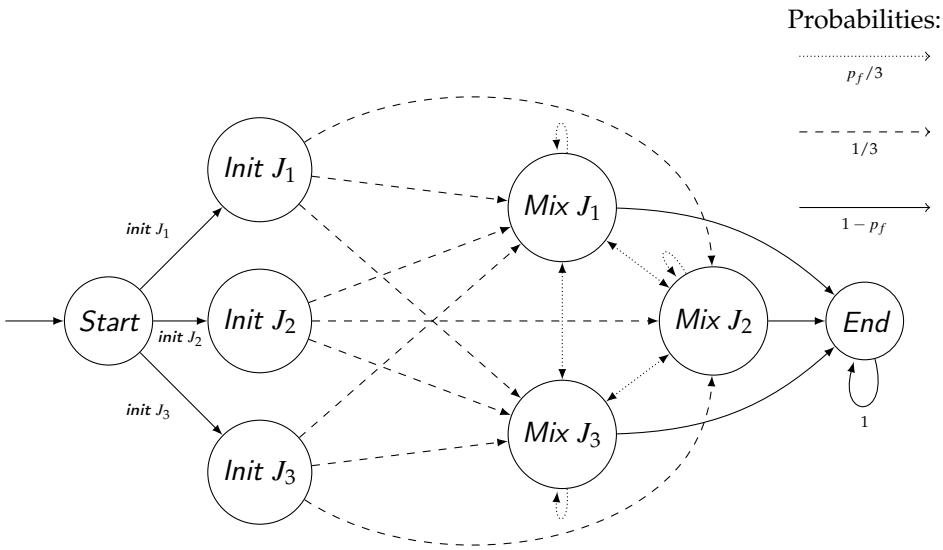

Figure 5.3: Example Markov chain of the small Crowds network $\{J_1, J_2, J_3\}$

guess the jondo initiating the route. They analyze the probability that a collaborating node is the successor of the initiating jondo. This analysis is affected by the fact that the route may go through the initiating jondo multiple times. An analysis of the Crowds protocol in PRISM, for specific sizes, has been conducted by Shmatikov [71].

Similar to the ZeroConf case, we only analyze the Markov chain having a global view on the protocol. We could model the individual behaviour of jondos in Isabelle/HOL and show that this induces our Markov chain model, but this is not in the scope of this thesis.

5.3.1 Formal Model of Route Establishment

We concentrate on the probabilistic aspects of route establishment in the Crowds protocol. We assume a set *jondos* of an arbitrary type α (which is just used to identify jondos), and a strict subset *colls*, the collaborating attackers. A jondo decides with probability p_f if it chooses another jondo as next step, or if it directly connects to the server. The distribution of the initiating jondos is given by *init*. Naturally the initiating jondo is not a collaborating jondo. In Isabelle this is expressed as the following context:

locale *crowds* =
  fixes *jondos colls* :: α *set* **and** p_f :: $\mathbb{R}$ **and** *init* :: $\alpha \to \mathbb{R}$
  assumes $0 < p_f$ **and** $p_f < 1$
  assumes *jondos* $\neq \emptyset$ **and** *colls* $\neq \emptyset$ **and** *finite jondos* **and** *colls* $\subset$ *jondos*
  assumes $\forall j \in$ *jondos*. $0 \leq$ *init* j **and** $\forall j \in$ *colls*. *init* $j = 0$
    and $\sum_{j \in jondos}$ *init* $j = 1$

The Markov chain has four different phases: start, the initial node, and the mixing phase, and finally the end phase where the server is contacted. See Fig. 5.3

for a small example. Our formalization of Markov chains requires a single start node, otherwise we could choose *init* as initial distribution. The type of the state α *c-state* depends on the type of the jondos α.

datatype α *c-state* = *Start* | *Init* α | *Mix* α | *End*

Similar to the ZeroConf protocol not all possible values of *c-state* are necessary. We restrict them further by only allowing non-collaborating jondos as initial jondos, and only elements from *jondos* participate in the mixing phase. With this definition it is easy to show that the set of states $S :: \alpha$ *c-state set* is finite.

$$\begin{array}{lcl} S & :: & \Big(\alpha \textit{ c-state}\Big) \textit{ set} \\ S & = & \{\textit{Start}\} \cup \{\textit{Init } j \mid j \in \textit{jondos} \setminus \textit{colls}\} \cup \{\textit{Mix } j \mid j \in \textit{jondos}\} \cup \{\textit{End}\} \end{array}$$

Often we are interested in the jondo referenced by the current state. We introduce *jondo-of* :: α *c-state* $\to \alpha$ returning the jondo if we are in an initial or mixing state:

$$\begin{array}{lcl} \textit{jondo-of} & :: & \alpha \textit{ c-state} \to \alpha \\ \textit{jondo-of } s & = & \textbf{case } s \textbf{ of } \textit{Init } j \Rightarrow j \mid \textit{Mix } j \Rightarrow j \end{array}$$

The transition matrix $\tau :: \alpha$ *c-state* $\to \alpha$ *c-state* $\to \mathbb{R}$ is defined by a case distinction on all possible transitions. The probability for steps from *Start* are given by the distribution of the initiating jondos *init*. The first routing jondo is arbitrarily chosen, and the probability of going from a mixing state to a mixing state is the product of p_f to stay in the mixing phase and the probability $1/J$ for the next jondo. With probability $1 - p_f$ the mixing state is finished and than the Markov chain stays in *End*. Figure 5.4 shows an example path through the different phases.

$$\begin{array}{lcl} J, H & :: & \mathbb{R} \\ J & = & |\textit{jondos}| \\ H & = & |\textit{jondos} \setminus \textit{colls}| \\ \\ \tau & :: & \alpha \textit{ c-state} \to \alpha \textit{ c-state} \to \mathbb{R} \\ \tau\ s\ t & = & \textbf{case } (s,t) \textbf{ of } (\textit{Start}, \quad \textit{Init } j) \quad \Rightarrow \textit{init } j \\ & & \quad \mid (\textit{Init } j, \quad \textit{Mix } j') \Rightarrow 1/J \\ & & \quad \mid (\textit{Mix } j, \quad \textit{Mix } j') \Rightarrow p_f/J \\ & & \quad \mid (\textit{Mix } j, \quad \textit{End}) \quad \Rightarrow 1 - p_f \\ & & \quad \mid (\textit{End}, \quad \textit{End}) \quad \Rightarrow 1 \\ & & \quad \mid \_ \qquad\qquad\quad \Rightarrow 0 \end{array}$$

Here $f(x := v)$ is the function f updated at x to the new value v, i.e. $f(x := v)\ x = v$.

This completes the definition of the Markov chain describing the route establishment in the Crowds protocol. Finally we show that S and τ describe a discrete-time Markov chain:

lemma MARKOV-CHAIN-CROWDS: *markov-chain S* τ

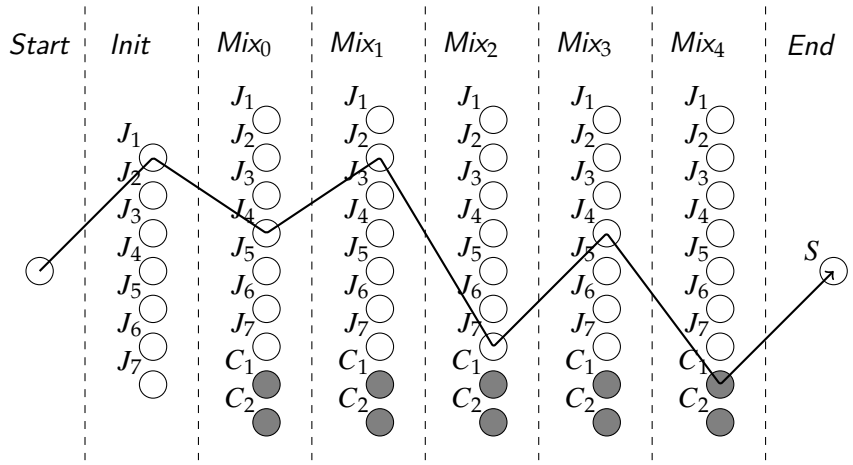

Figure 5.4: The established route $J_1 - J_4 - J_3 - J_5 - J_4 - C_1 - S$

5.3.2 Independence of Initiating Jondo and Contacting Jondo

We define a number of path properties of our Markov chain. The functions *len* :: $(\mathbb{N} \to \alpha$ *c-state*$) \to \mathbb{N}$, *first-jondo* :: $(\mathbb{N} \to \alpha$ *c-state*$) \to \alpha$ and *last-jondo* :: $(\mathbb{N} \to \alpha$ *c-state*$) \to \alpha$ operate on paths not containing the *Start* element. *len* returns the length of the mixing phase, i.e. how many *Mix* states are in the path until *End* is reached, *first-jondo* is the initiating jondo, and *last-jondo* is the jondo contacting the server.

$$\mathit{len}\ \omega = (\mathit{LEAST}\ n.\ \omega\ n = \mathit{End}) - 2$$
$$\mathit{first\text{-}jondo}\ \omega = \mathit{jondo\text{-}of}\ (\omega\ 0)$$
$$\mathit{last\text{-}jondo}\ \omega = \mathit{jondo\text{-}of}\ (\omega\ (\mathit{len}\ \omega + 1))$$

The path functions *len*, *first-jondo* and *last-jondo* are well-defined on almost every path. The paths in our Markov chain do not contain the *Start* element, so the paths start with an *Init* state. Hence for almost every path we know that the first element is an initiating state, then for the next *len* elements we have mixing states, and finally a tail of *End* states:

lemmas

$$AE_{\mathit{Start}}\ \omega.\ \omega \in \mathbb{N} \to S$$
$$AE_{\mathit{Start}}\ \omega.\ \exists j \in \mathit{jondos} \setminus \mathit{colls}.\ \omega\ 0 = \mathit{Init}\ j$$
$$AE_{\mathit{Start}}\ \omega.\ \forall i \leq \mathit{len}\ \omega.\ \exists j \in \mathit{jondos}.\ \omega\ (i+1) = \mathit{Mix}\ j$$
$$AE_{\mathit{Start}}\ \omega.\ \forall i > \mathit{len}\ \omega.\ \omega\ (i+1) = \mathit{End}$$

With this we can easily show that the jondo contacting the server is independent from the initiating jondo:

theorem $\mathit{indep\text{-}var}_{\mathit{Start}}\ (\mathit{count}\ (\mathit{jondos} \setminus \mathit{colls}))\ (\mathit{count}\ \mathit{jondos})\ \mathit{first\text{-}jondo}\ \mathit{last\text{-}ncoll}$

5.3.3 Probability that Initiating Jondo Contacts a Collaborator

The attacker model assumes that the collaborators want to detect the initiator of a route. This is obviously only possible if one of the collaborators is chosen as one of the mixing jondos. We have two goals: (1) If the numbers of collaborators is small, the probability to contact a collaborator should be near zero. (2) We want to analyze the probability that the initiating jondo directly contacts a collaborator. When we know the ratio of collaborators to jondos, how can we adjust p_f, so that this probability is less or equal to $1/2$?

The random variable *hit-colls* is true if a collaborator participates in the mixing phase, *first-coll* is the mixing phase in which the collaborator is hit, and *last-ncoll* is the last non-collaborating jondo, i.e. the jondo contacting a collaborator.

$$\begin{array}{lcl} \mathit{hit\text{-}colls} & :: & (\mathbb{N} \to \alpha\ \mathit{c\text{-}state}) \to \mathbb{B} \\ \mathit{hit\text{-}colls}\ \omega & = & \exists n, j \in \mathit{colls}.\ \omega\ n = \mathit{Mix}\ j \\ \\ \mathit{first\text{-}coll} & :: & (\mathbb{N} \to \alpha\ \mathit{c\text{-}state}) \to \mathbb{N} \\ \mathit{first\text{-}coll}\ \omega & = & (\mathit{LEAST}\ n.\ \exists j \in \mathit{colls}.\ \omega\ n = \mathit{Mix}\ j) - 1 \\ \\ \mathit{last\text{-}ncoll} & :: & (\mathbb{N} \to \alpha\ \mathit{c\text{-}state}) \to \alpha \\ \mathit{last\text{-}ncoll}\ \omega & = & \mathit{jondo\text{-}of}\ (\omega\ (\mathit{first\text{-}coll}\ \omega)) \end{array}$$

The property we want to check only makes sense if a collaborator participates in the mixing phase. So we first prove the probability to hit a collaborator:

lemma $\Pr_{\mathit{Start}}(\omega.\ \mathit{hit\text{-}colls}\ \omega) = (1 - H/J)/(1 - H/J \cdot p_f)$

We already see that the probability to hit a collaborator goes to 0 if the number of collaborators and p_f stay constant and $J \longrightarrow \infty$. Then $H/J \longrightarrow 1$ and hence $\Pr_{\mathit{Start}}(\omega.\ \mathit{hit\text{-}colls}\ \omega) \longrightarrow 0$. Thus our first goal is satisfied.

Additionally, we want to control the probability that the initiating jondo hits a collaborator. For this, we compute the probability to have a fixed first and last non-collaborating jondo before we hit a collaborator:

lemma P-FIRST-JONDO-LAST-NCOLL:
$l \in \mathit{jondos} \setminus \mathit{colls}$ **and** $i \in \mathit{jondos} \setminus \mathit{colls} \implies$
$\Pr_{\mathit{Start}}(\omega.\ \mathit{first\text{-}jondo}\ \omega = i \wedge \mathit{last\text{-}ncoll}\ \omega = l \mid \mathit{hit\text{-}colls}\ \omega) =$
$\quad \mathit{init}\ i \cdot (p_f/J + (\textbf{if}\ i = l\ \textbf{then}\ 1 - H/J \cdot p_f\ \textbf{else}\ 0))$

Note that the conditional probability does not divide by 0 because, by the previous lemma, we know that $\Pr_{\mathit{Start}}(\omega.\ \mathit{hit\text{-}colls}\ \omega) \neq 0$. By summing up over all possible non-collaborating jondos we show the probability that the last non-collaborating jondo is the initiating jondo:

theorem $\Pr_{\mathit{Start}}(\omega.\ \mathit{first\text{-}jondo}\ \omega = \mathit{last\text{-}ncoll}\ \omega \mid \mathit{hit\text{-}colls}\ \omega) = 1 - (H-1)/J \cdot p_f$

With this we can now enforce that the probability that the initiating jondo hits a collaborator is less or equal to $\frac{1}{2}$:

corollary
$H > 1 \ \wedge\ J/(2 \cdot (H-1)) \leq p_f \implies$
$\Pr_{\mathit{Start}}(\omega.\ \mathit{first\text{-}jondo}\ \omega = \mathit{last\text{-}ncoll}\ \omega \mid \mathit{hit\text{-}colls}\ \omega) \leq \frac{1}{2}$

Reiter and Rubin [67] call this probably innocent. Because $p_f < 1$ this is only possible if $1/2 < (H-1)/J$, i.e. more than half of the jondos are non-collaborating. This meets our second goal.

5.3.4 Information Gained by Collaborators

Obviously, in Isabelle/HOL we are not only restricted to state probabilities or expectations. For example, for quantitative information flow analysis, similar to the analysis by Malacaria [55], we are interested in the mutual information $\mathcal{I}_s(X;Y)$ between two random variables X and Y (c.f. Section 4.4). We know that if X and Y are simple functions, i.e. functions with a finite range, then $\mathcal{I}_s(X;Y)$ can be computed in the known discrete way:

lemma
$$\textit{simple-function}_s\ X \implies \textit{simple-function}_s\ Y \implies$$
$$\mathcal{I}_s(X;Y) = \textstyle\sum_{(x,y)\in\{(Xx,Yx)|x.x\in\Omega\}} . \Pr_s(\omega.\ X\ \omega = x \wedge Y\ \omega = y)\cdot$$
$$\log_2\left(\Pr_s(\omega.\ X\ \omega = x \wedge Y\ \omega = y)/(\Pr_s(\omega.\ X\ \omega = x)\cdot\Pr_s(\omega.\ Y\ \omega = y))\right)$$

We are only interested in runs which hit a collaborator. To use mutual information with this restriction we introduce the conditional probability $\Pr_{\textit{hit-colls}}$, with the condition that each run hits a collaborator. Its characteristic property (we omit the technical definition) is

lemma
$$\{x \mid P\ x\} \in \textit{measurable}_s \implies \Pr_{\textit{hit-colls}}(\omega.\ P\ \omega) = \Pr_{\textit{Start}}(\omega.\ P\ \omega \mid \textit{hit-colls}\ \omega)$$

With this property and Lemma P-FIRST-JONDO-LAST-NCOLL we can now show an upper bound for the information flow:

theorem $\mathcal{I}_{\textit{hit-colls}}(\textit{first-jondo}; \textit{last-ncoll}) \leq (1-(H-1)/J\cdot p_f)\cdot\log_2 H$

This supports the intuitive understanding that the information the attackers can gain is restricted by the probability that the initiating jondo is the jondo directly contacting a collaborator.

5.4 Köpf-Dürmuth Countermeasure

Köpf and Dürmuth [48] give a countermeasure against timing attacks. For this, they analyze the amount of information a deterministic side-channel attack can gain. They show that $|O|\cdot log_2(n+1)$ bits is an upper bound, where O is the set of possible observations and n is the number of attacks. In this section we formalize their analysis using the information theory developed in Section 4.4.

Before we start with the formalization, we introduce the locale *finite-information* to define a finite and discrete probability space with a *size* for bits b. A common instantiation is $b = 2$, i.e. when a unit of information is a binary digit.

locale *finite-information* =
fixes $\Omega :: \alpha\ set$ **and** $p :: \alpha \to \mathbb{R}$ **and** $b :: \mathbb{R}$
assumes *finite* Ω **and** $(\sum_{x\in\Omega} p\ x) = 1$ **and** $\forall x.\ 0 \leq p\ x$ **and** $1 < b$

The measure *point* Ω p (introduced in Section 3.3.1) is then an information space:

lemma *information-space* (*point* Ω p) b

Our model assumes an arbitrary distribution K on the finite set *keys* and an arbitrary distribution M on the finite set *messages*. The observation *observe* k m represents the side-channel information an attacker gains about a key k when exchanging the message m. In our analysis we assume that n messages are exchanged.

locale *koepf-duermuth* =
  finite-information keys K b + *finite-information messages M b*
  for *keys* :: k *set* **and** $K :: k \to \mathbb{R}$ **and**
    messages :: m *set* **and** $M :: m \to \mathbb{R}$ **and** $b :: \mathbb{R}+$
  fixes *observe* :: $k \to m \to o$ **and** $n :: \mathbb{N}$

The model assumes a probability space where the key and each of the n messages are independently distributed. The sequence of n messages is modelled as a list. We write *length xs* for the length of the list xs, *set xs* for the set of elements in the list xs, and *xs*!i for the i-th element of the list xs.

$$\begin{array}{lcl} \mathit{msgs} & :: & (k \times m\ \mathit{list})\ \mathit{set} \\ \mathit{msgs} & = & \mathit{keys} \times \{\mathit{ms} \mid \mathit{set}\ \mathit{ms} \subseteq \mathit{messages} \wedge \mathit{length}\ \mathit{ms} = n\} \end{array}$$

$$\begin{array}{lcl} P & :: & (k \times m\ \mathit{list}) \to \mathbb{R} \\ P\ (k, ms) & = & K\ k \cdot (\prod_{i<n} M\ (ms\ !\ i)) \end{array}$$

We show that P is a discrete distribution on the space *msgs*. We will use it as implicit probability space for the conditional entropy $H(X|Y)$ and for mutual information $I(X;Y)$.

lemma *finite-information msgs P b*

Our final theorem will be about the mutual information between the key and all observations on the sent messages. For this the random variable $\mathcal{O}$ is introduced representing all observations by an attacker in one run. The goal is then to give a bound on $I(\mathcal{K};\mathcal{O})$, where $\mathcal{K} = \mathit{fst}$ is the random variable returning the key of the run and $\mathcal{O}$ returns all observations by an attacker.

$$\begin{array}{lcl} \mathcal{O} & :: & k \times m\ \mathit{list} \to o\ \mathit{list} \\ \mathcal{O}\ (k, ms) & = & \mathit{map}\ (\mathit{observe}\ k)\ \mathit{ms} \end{array}$$

Here *map* is the map function on lists, the result of *map* f $[x_1, x_2, \ldots, x_n]$ is the list $[f\ x_1, f\ x_2, \ldots, f\ x_n]$.

The bound we provide depends on the set of all possible observations:

$$\begin{array}{lcl} \mathit{observations} & :: & o\ \mathit{set} \\ \mathit{observations} & = & \{\mathit{observe}\ k\ m \mid k \in \mathit{keys} \wedge m \in \mathit{messages}\} \end{array}$$

Now, we will show that $I(\mathcal{K};\mathcal{O}) \leq |\mathit{observations}| \cdot \log_b\,(n+1)$. First, we show that the order of the messages is irrelevant. We introduce $t\ \mathit{os}\ o$ returning how often each single observation o occurs in the observations os:

$$\begin{array}{lcl} t & :: & o\ \mathit{list} \to o \to \mathbb{N} \\ t\ \mathit{os} & = & \lambda o.\ \mathit{card}\ \{i < n.\ \mathit{os}\ !\ i = o\} \end{array}$$

The information an attacker gains stays equal no matter if he gets the observations as list or if he only gets the count for each observation:

theorem $I(\mathcal{K};\mathcal{O}) = I(\mathcal{K};t \circ \mathcal{O})$

We can also give an upper bound for the cardinality of all possible outcomes of t:

theorem $\mathit{card}\left(t \circ \mathcal{O}[\mathit{msgs}]\right) \leq (n+1)^{\mathit{card\ observations}}$

With Theorem ENTROPY-LE we know that $H(t \circ \mathcal{O}) \leq \mathit{card\ observations} \cdot \log_b\,(n+1)$. With the chain rule for entropy and the equality of mutual information and entropy we deduce $I(\mathcal{K};t \circ \mathcal{O}) \leq H(t \circ \mathcal{O})$. And finally, with the equation $I(\mathcal{K};\mathcal{O}) = I(\mathcal{K};t \circ \mathcal{O})$, we prove the upper bound for a side-channel attack:

corollary $I(\mathcal{K};\mathcal{O}) \leq \mathit{card\ observations} \cdot \log_b\,(n+1)$

Our proofs closely follow Köpf and Dürmuth [48]. They continue with introducing guessing entropy to quantify the expected effort for guessing the correct key. We stop at this point as we do not yet have guessing entropy in our information theory.

Chapter 6

Conclusion

6.1 Summary

In this thesis we developed measure, probability, and information theory in Isabelle/HOL.

One central goal of this development was the construction of measure spaces, starting with the explicit introduction of the push-forward measure and density measures. While this is usual in textbooks it is new in formalizations. The introduction of constants and verifying their properties not only allowed us to show that these measures exist, but they also encouraged us to introduce algebraic rules about combinations of these measures with products.

The next major construction is the product space. Bauer [9] introduces the finite product spaces as iteration of binary product spaces, which in HOL is not possible due to type constraints. Instead, we managed to introduce one measure space to represent both finite product spaces and infinite probability spaces. While the definition is more complicated, we gain a common ground (and Isabelle constant) for both concepts. This allows us to reuse the product σ-algebra in both cases so we only need half of the measurability proofs. Also the product σ-algebra is reused for stochastic processes as we see in the construction of the discrete-time Markov chain.

To analyze probability spaces we formalized tools to analyze random variables. Independence of a family of random variables relates their joint distribution to the product of their single distributions. For information theory mutual information of two random variables quantifies their shared information. This mutual information is exactly then zero if the two random variables are independent. To show this we introduced Kullback-Leibler divergence. For the analysis of random variables we expressed their distribution as a density measure and finally introduced uniform and exponential distributions.

The formalization of discrete-time Markov chains allowed us to model and verify the pCTL model checking algorithm. This required a couple of important theorems: the iterative equations, a.e.-state fairness and the finiteness of the hitting time (when the final states are a.e.-reached). Furthermore, these theorems helped us to verify the ZeroConf and the Crowds protocol. The analysis of the Crowds protocol used in addition independence and mutual information. This shows that

	Lines of theory (counted by `wc -l`)
Extended reals	3,900
Measure theory	13,300
Probability theory	7,400
Examples	3,100
Total	**27,700**
Multivariate analysis	31,700
Isabelle/HOL base image	68,300

Figure 6.1: Line count of the probability and measure theory compared to multivariate analysis and the Isabelle/HOL image.

the developed theories are already quite helpful to formalize probabilistic models.

Isabelle's probability theory is already used in further projects:

- Andrei Popescu, Johannes Hölzl and Tobias Nipkow [66, 65] formalize a framework for concurrent noninterference. There is an unpublished version of this framework, which uses a Markov decision process to describe the probabilistic behaviour of schedulers. The infinite product measures and the Markov chains were originally developed for this framework.

- Lars Noschinski proves the Girth-Chromatic number theorem [63, 62]. He uses a probabilistic proof, where the proof itself uses results from probability theory but no probabilistic concepts occur in the final theorem.

- Fabian Immler formalizes [41] the Daniell-Kolmogorov theorem allowing us to construct stochastic processes as the limit of their finite-dimensional distributions.

- Jeremy Avigad works together with his student on the formalization of characteristic functions and the central limit theorem in Isabelle/HOL.

Figure 6.1 gives an overview of the sizes of the files formalizing probability and measure theory and compares the cumulative size to the size of the theories it is based on. It only lists the cumulative line count of all `thy`-files, `ML`-files are not included. The base image is obviously a huge development containing already real analysis which we heavily use. The multivariate analysis is comparable with our theory, as it is also a big development of mathematical analysis. Integrating some of the work listed above will likely outgrow the multivariate analysis.

6.2 Future Work

The developed theories are a good foundation for further formalization of mathematical proofs. But they would also gain from extending the automation and additional support for common probability concepts.

- In interactive theorem proving, automation is very important. A particular optimization would be to add special support to show measurability. In Section 2.1 we declare quite a few introduction rules to show measurability for most logic combinators, and for each measure space we try to find the generic introduction rules. Unfortunately, the simplifier does not support long chains of introduction rules, or backtracking search. For this a special simplifier mechanism to solve measurability assumptions would be a valuable addition.

- When we model input values for an algorithm or a protocol, it is easy to assume a set of independent random variables that represent independent inputs. In HOL these random variables often have different types. Unfortunately, the current way to specify independent random variables with *indep-vars* and *indep-var* restricts them to have the same type. It is also often required to combine some random variables and then show that the result is still independent from the other random variables. This is only partly possible with Lemma INDEP-VARS-COMPOSE.

  A solution would be to use the σ-algebras generated by each random variables and represent the independence directly with *indep-sets*. However, it would be nice to have special combinators to manage the index sets and composition of random variables.

- Currently, we only allow finite real values for mutual information and entropy. This is due to the problem that (1) the entropy may also be negative and (2) the logarithm in the integrand in these definitions assumes also negative values. A solution would be to define entropy and mutual information to be infinite if the integral is not defined, i.e. the integral is not finite or the random variables do not have pdfs.

- For mathematical analysis an important tool are characteristic functions. It is an alternative way to characterize distributions and is helpful when we need to analyze sums of independent random variables. An application of this is the proof of the central limit theorem. As mentioned in the previous section, Jeremy Avigad is currently working on formalizing the central limit theorem and thus also formalizes characteristic functions.

  This would also introduce the normal distribution, a distribution central to stochastics. The central limit theorem tells us that the normal distribution is the limit of a sequence of sums of independent, identically distributed random variables.

  In general a further expansion on distributions and characteristics of distributions would be interesting. Besides the normal distribution there is also the Poisson distribution, which is often used in computer science and queueing theory. Maximum entropy statements like Theorem ENTROPY-LE are also interesting for the exponential and normal distribution.

- For computer science interesting probabilistic models include Markov chains and Markov decision processes not only with discrete-time but also with

continuous-time. For continuous-time Markov decision processes we need to construct continuous-space Markov chains, requiring the formalization of the Daniell-Kolmogorov theorem (already done by Immler [41]).

What is missing is the iteration of Markov kernels (the generalization of transition matrices into functions from states to probability distributions of states) to generate the finite-dimensional distributions of the Markov chain. This should enable us to construct the trace spaces of Markov chains and Markov decision processes.

Similar to our verification of pCTL model checking in Section 5.1 we can then start to verify probabilistic model checking for these models.

Appendix A

Extended Real Numbers

Extended reals are used in measure theory to represent measure values. For example the Lebesgue measure $\lambda_{\mathbb{R}}$ takes infinite values, as there is no real number we can reasonably assign to $\lambda_{\mathbb{R}}$. So we need a type containing the real numbers and a distinct value for infinity. We introduce the type $\overline{\mathbb{R}}$ as the reals extended with a positive and a negative infinite element.

$$\textbf{datatype } \overline{\mathbb{R}} = \infty \mid (\mathbb{R})_{\overline{\mathbb{R}}} \mid -\infty \qquad \mathit{real} :: \overline{\mathbb{R}} \to \mathbb{R}$$
$$\mathit{real}\,(r)_{\overline{\mathbb{R}}} = r \qquad \mathit{real}\,\infty = 0 \qquad \mathit{real}\,(-\infty) = 0$$

The conversion function *real* restricts the extended reals to the real numbers and maps $\pm\infty$ to 0. For the sake of readability we hide this conversion function.

$$\begin{array}{rcll} (r)_{\overline{\mathbb{R}}} \le (p)_{\overline{\mathbb{R}}} & \Leftrightarrow & r \le p & x \le \infty \qquad -\infty \le x \\ -(r)_{\overline{\mathbb{R}}} & = & (-r)_{\overline{\mathbb{R}}} & -(-\infty) = \infty \\ (r)_{\overline{\mathbb{R}}} + (p)_{\overline{\mathbb{R}}} & = & (r+p)_{\overline{\mathbb{R}}} & \infty + x = \infty \qquad x + \infty = \infty \\ (r)_{\overline{\mathbb{R}}} \cdot (p)_{\overline{\mathbb{R}}} & = & (r \cdot p)_{\overline{\mathbb{R}}} & x \cdot \pm\infty = \pm\infty \cdot x = \begin{cases} 0 & \text{if } x = 0 \\ \operatorname{sgn} x \cdot \pm\infty & \text{otherwise} \end{cases} \end{array}$$

For measure theory it is suitable to define $\infty \cdot 0 = 0$. Using *min* and *max* as meet and join, we get that $\overline{\mathbb{R}}$ is a complete lattice where *bot* is $-\infty$ and *top* is ∞.

Our next step is to define the topological structure on $\overline{\mathbb{R}}$. This is an extension of the topological structure on real numbers. However we need to take care of what happens when $\pm\infty$ is in the set.

$$\begin{aligned} &\mathit{open}\ A \Leftrightarrow \mathit{open}\ \{r \mid (r)_{\overline{\mathbb{R}}} \in A\} \wedge \\ &\quad (\infty \in A \implies \exists x.\ \forall y > x.\ (y)_{\overline{\mathbb{R}}} \in A) \wedge (-\infty \in A \implies \exists x.\ \forall y < x.\ (y)_{\overline{\mathbb{R}}} \in A) \end{aligned}$$

From this definition the continuity of $(\cdot)_{\overline{\mathbb{R}}}$ follows directly. The definition of limits of sequences in Isabelle/HOL is based on topological spaces. This allows us to reuse these definitions and also some of the proofs such as uniqueness of limits. We also verify that the limits and infinite sums on real numbers are the same as the limits and sums on extended reals:

$$(\lambda n.\ (f\ n)_{\overline{\mathbb{R}}}) \xrightarrow[n\to\infty]{} (r)_{\overline{\mathbb{R}}} \qquad \Leftrightarrow \qquad (\lambda n.\ f\ n) \xrightarrow[n\to\infty]{} r$$

If f is summable, then $\sum_n (f\ n)_{\overline{\mathbb{R}}} = (\sum_n f\ n)_{\overline{\mathbb{R}}}$.

Hurd [40] formalizes similar positive extended reals and also defines a complete lattice on them. Our $\overline{\mathbb{R}}$ includes negative numbers and we not only show that it forms a complete lattice but also that it forms a topological space. The complete lattice is used for monotone convergence and the topological space is used to define the Borel sets on $\overline{\mathbb{R}}$.

Bibliography

[1] Naeem Ahmad Abbasi. *Formal Reliability Analysis using Higher-Order Logic Theorem Proving*. PhD thesis, The Department of Electrical and Computer Engineering, Concordia University, Montréal, Québec, Canada, 2012.

[2] Reynald Affeldt and Manabu Hagiwara. Formalization of Shannon's theorems in SSReflect-Coq. 2012.

[3] Suzana Andova, Holger Hermanns, and Joost-Pieter Katoen. Discrete-time rewards model-checked. In Kim Guldstrand Larsen and Peter Niebert, editors, *Formal Modeling and Analysis of Timed Systems*, volume 2791 of *LNCS*, pages 88–104, 2003.

[4] Robert B. Ash. *Real Analysis and Probability*, volume 11 of *Probability and Mathematical Statistics*. Academic Press, 1972.

[5] Philippe Audebaud and Christine Paulin-Mohring. Proofs of randomized algorithms in Coq. *Science of Computer Programming*, 74(8):568–589, 2009. ISSN 0167-6423. doi: 10.1016/j.scico.2007.09.002. Special Issue on Mathematics of Program Construction (MPC 2006).

[6] Christel Baier. *On the Algorithmic Verification of Probabilistic Systems*. Habilitation, Universität Mannheim, 1998.

[7] Christel Baier, Boudewijn R. Haverkort, Holger Hermanns, and Joost-Pieter Katoen. Model-checking algorithms for continuous-time markov chains. *IEEE Trans. Software Eng.*, 29(6):524–541, 2003. doi: 10.1109/TSE.2003.1205180.

[8] Heinz Bauer. *Probability Theory*. de Gruyter, 1995. ISBN 3-11-013925-9.

[9] Heinz Bauer. *Measure and Integration theory*. de Gruyter, 2001. ISBN 3-11-016719-0.

[10] Józef Białas. The σ-additive measure theory. *Formalized Mathematics*, 2(2): 263–270, 1991.

[11] Józef Białas. Properties of Caratheodor's measure. *Formalized Mathematics*, 3 (1):67–70, 1992.

[12] Józef Białas. The one-dimensional Lebesgue measure. *Formalized Mathematics*, 5(2):253–258, 1996.

[13] Henrik Bohnenkamp, Peter van der Stok, Holger Hermanns, and Frits Vaandrager. Cost-optimisation of the IPv4 Zeroconf protocol. In *Dependable Systems and Networks (DSN'03)*, pages 531–540. IEEE CS Press, 2003.

[14] N. Bourbaki. *General Topology (Part I)*. Addison-Wesley, 1966.

[15] S. Cheshire, B. Aboba, and E. Guttman. Dynamic configuration of IPv4 link-local addresses. RFC 3927 (Proposed Standard), may 2005. URL http://www.ietf.org/rfc/rfc3927.txt.

[16] David Clark, Sebastian Hunt, and Pasquale Malacaria. A static analysis for quantifying information flow in a simple imperative language. *Journal of Computer Security*, 15(3):321–371, 2007. ISSN 0926-227X.

[17] Aaron R. Coble. *Anonymity, Information, and Machine-Assisted Proof*. PhD thesis, King's College, University of Cambridge, 2009.

[18] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991. ISBN 0-471-06259-6.

[19] John R. Cowles and Ruben Gamboa. Using a first order logic to verify that some set of reals has no Lebesgue measure. In Matt Kaufmann and Lawrence C. Paulson, editors, *Interactive Theorem Proving (ITP 2010)*, volume 6172 of *LNCS*, pages 25 – 34, 2010.

[20] Marc Daumas and David R. Lester. Stochastic formal methods: An application to accuracy of numeric software. In *Hawaii International International Conference on Systems Science (HICSS 2007)*, pages 262–269. IEEE Computer Society, 2007.

[21] Marc Daumas, David R. Lester, Érik Martin-Dorel, and Annick Truffert. Improved bound for stochastic formal correctness of numerical algorithms. *Innovations in Systems and Software Engineering*, 6(3):173–179, 2010. doi: 10.1007/s11334-010-0128-x.

[22] Agnes Doll. Kolmogorov's zero-one law. *Formalized Mathematics*, 17(1–4): 73–77, 2009.

[23] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Springer, 1996.

[24] Noboru Endou, Keiko Narita, and Yasunari Shidama. The Lebesgue monotone convergence theorem. *Formalized Mathematics*, 16(2):167–175, 2008. doi: 10.2478/v10037-008-0023-1.

[25] Russel A. Gordon. *The Integrals of Lebesgue, Denjoy, Perron, and Henstock*, volume 4 of *Graduate Studies in Mathematics*. American Mathematical Society, 1994.

[26] Robert M. Gray. *Entropy and information theory*. Springer-Verlag, 1990.

[27] Florian Haftmann and Tobias Nipkow. Code generation via higher-order rewrite systems. In M. Blume, N. Kobayashi, and G. Vidal, editors, *Functional and Logic Programming (FLOPS 2010)*, volume 6009 of *LNCS*, pages 103–117, 2010.

[28] Florian Haftmann and Makarius Wenzel. Local theory specifications in Isabelle/Isar. In Stefano Berardi, Ferruccio Damiani, and Ugo de'Liguoro, editors, *TYPES 2008*, volume 5497 of *LNCS*, pages 153–168. Springer, 2009. ISBN 978-3-642-02443-6. doi: 10.1007/978-3-642-02444-3_10.

[29] Ernst Moritz Hahn, Holger Hermanns, and Lijun Zhang. Probabilistic reachability for parametric markov models. *International Journal on Software Tools for Technology Transfer (STTT)*, 13(1):3–19, 2011. doi: 10.1007/s10009-010-0146-x.

[30] Hans Hansson and Bengt Jonsson. A logic for reasoning about time and reliability. Technical Report SICS/R90013, Swedish Institute of Computer Science, Dec 1994.

[31] John Harrison. A HOL theory of Euclidean space. In Joe Hurd and Tom Melham, editors, *Theorem Proving in Higher Order Logics, 18th International Conference, TPHOLs 2005*, volume 3603 of *Lecture Notes in Computer Science*, pages 114–129, 2005.

[32] Osman Hasan. *Formal Probabilistic Analysis using Theorem Proving*. PhD thesis, The Department of Electrical and Computer Engineering, Concordia University, Montréal, Québec, Canada, 2008.

[33] Osman Hasan, Naeem Abbasi, Behzad Akbarpour, Sofiène Tahar, and Reza Akbarpour. Formal reasoning about expectation properties for continuous random variables. In Ana Cavalcanti and Dennis Dams, editors, *FM 2009: Formal Methods*, volume 5850 of *LNCS*, pages 435–450. 2009. ISBN 978-3-642-05088-6.

[34] Holger Hermanns, Björn Wachter, and Lijun Zhang. Probabilistic CEGAR. In Aarti Gupta and Sharad Malik, editors, *Computer Aided Verification, 20th International Conference, CAV 2008, Princeton, NJ, USA, July 7-14, 2008, Proceedings*, volume 5123 of *LNCS*, pages 162–175, 2008.

[35] Johannes Hölzl and Armin Heller. Three chapters of measure theory in Isabelle/HOL. In Marko C. J. D. van Eekelen, Herman Geuvers, Julien Schmaltz, and Freek Wiedijk, editors, *Interactive Theorem Proving (ITP 2011)*, volume 6898 of *LNCS*, pages 135–151, 2011.

[36] Johannes Hölzl and Tobias Nipkow. Markov models. *The Archive of Formal Proofs*, Jan 2012. ISSN 2150-914x. http://afp.sf.net/entries/Markov_Models.shtml, Formal proof development.

[37] Johannes Hölzl and Tobias Nipkow. Interactive verification of Markov chains: Two distributed protocol case studies. In U. Fahrenberg, A. Legay, and C. Thrane, editors, *Quantities in Formal Methods (QFM 2012)*, EPTCS, 2012.

[38] Johannes Hölzl and Tobias Nipkow. Verifying pCTL model checking. In C. Flanagan and B. König, editors, *Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2012)*, volume 7214 of *LNCS*, pages 347–361, 2012.

[39] Joe Hurd. *Formal Verification of Probabilistic Algorithms*. PhD thesis, University of Cambridge, 2002.

[40] Joe Hurd, Annabelle McIver, and Carroll Morgan. Probabilistic guarded commands mechanized in HOL. *Theoretical Computer Science*, 346(1):96–112, Nov 2005.

[41] Fabian Immler. Generic construction of probability spaces for paths of stochastic processes in isabelle/hol. Master's thesis, Technische Universität München, Oct 2012.

[42] A.G. ter Meulen J.F.A.K. van Benthem. *Generalized quantifiers in natural language*. de Gruyter, 1985.

[43] Olav Kallenberg. *Foundations of Modern Probability Theory*. Probability and its Application. Springer, 1997.

[44] J.-P. Katoen, A. McIver, L. Meinicke, and C. C. Morgan. Linear-invariant generation for probabilistic programs: Automated support for proof-based methods. In R. Cousot and M. Martel, editors, *Static Analysis (SAS 2010)*, volume 6337 of *LNCS*, pages 390–406, 2010. doi: 10.1007/978-3-642-15769-1_24.

[45] Joost-Pieter Katoen, Ivan S. Zapreev, Ernst Moritz Hahn, Holger Hermanns, and David N. Jansen. The ins and outs of the probabilistic model checker MRMC. *Performance Evaluation*, 68:90–104, 2011.

[46] Andrei Kolmogorow. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933.

[47] Andrei Kolmogorow. *Foundations of the theory of probability*. Chelsea Publishing Company, New York, 1950.

[48] Boris Köpf and Markus Dürmuth. A Provably Secure and Efficient Countermeasure against Timing Attacks. In *Proc. 22nd IEEE Computer Security Foundations Symposium (CSF '09)*, pages 324–335, 2009.

[49] Marta Kwiatkowska, Gethin Norman, David Parker, and J. Sproston. Performance analysis of probabilistic timed automata using digital clocks. *Formal Methods in System Design*, 29:33–78, 2006.

[50] Marta Kwiatkowska, Gethin Norman, and David Parker. Stochastic model checking. In M. Bernardo and J. Hillston, editors, *Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation (SFM 2007)*, volume 4486 of *LNCS*, pages 220–270, 2007.

[51] Marta Kwiatkowska, Gethin Norman, and David Parker. PRISM 4.0: Verification of probabilistic real-time systems. In G. Gopalakrishnan and S. Qadeer, editors, *Computer Aided Verification (CAV 2011)*, volume 6806 of *LNCS*, pages 585–591, 2011.

[52] David R Lester. Topology in PVS: continuous mathematics with applications. In *Proceedings of the second workshop on Automated formal methods*, AFM '07, pages 11–20, 2007. doi: 10.1145/1345169.1345171.

[53] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

[54] Liya Liu, Osman Hasan, and Sofiene Tahar. Formalization of finite-state discrete-time markov chains in HOL. In T. Bultan and P.-A. Hsiung, editors, *Automated Technology for Verification and Analysis (ATVA 2011)*, volume 6996 of *LNCS*, pages 90–104, 2011.

[55] Pasquale Malacaria. Assessing security threats of looping constructs. In *Proceedings of the 34th Annual ACM SIGPLAN-SIGACT symposium on Principles of Programming Languages (POPL'07)*, pages 225–235, 2007. doi: 10.1145/1190215.1190251.

[56] Franz Merkl. Dynkin's lemma in measure theory. *Formalized Mathematics*, 9 (3):591–595, 2001.

[57] Tarek Mhamdi, Osman Hasan, and Sofiène Tahar. On the formalization of the Lebesgue integration theory in HOL. In Matt Kaufmann and Lawrence C. Paulson, editors, *Proceedings of ITP 2010*, volume 6172 of *LNCS*, pages 387–402, 2010.

[58] Tarek Mhamdi, Osman Hasan, and Sofiène Tahar. Formalization of entropy measures in hol. In Marko C. J. D. van Eekelen, Herman Geuvers, Julien Schmaltz, and Freek Wiedijk, editors, *Interactive Theorem Proving (ITP 2011)*, volume 6898 of *LNCS*, pages 233–248, 2011.

[59] Tobias Nipkow. Gauss-Jordan elimination for matrices represented as functions. In Gerwin Klein, Tobias Nipkow, and Lawrence Paulson, editors, *The Archive of Formal Proofs*. http://afp.sf.net/entries/Gauss-Jordan-Elim-Fun.shtml, Aug 2011. Formal proof development.

[60] Andrzej Nędzusiak. σ-fields and probability. *Formalized Mathematics*, 1(2): 401–407, 1990.

[61] Andrzej Nędzusiak. Probability. *Formalized Mathematics*, 1(**4**):745–749, 1990.

[62] Lars Noschinski. A probabilistic proof of the girth-chromatic number theorem. *The Archive of Formal Proofs*, Feb 2012. ISSN 2150-914x. http://afp.sf.net/entries/Girth_Chromatic.shtml, Formal proof development.

[63] Lars Noschinski. Proof pearl: A probabilistic proof for the Girth-Chromatic number theorem. In Lennart Beringer and Amy Felty, editors, *Interactive Theorem Proving (ITP 2012)*, volume 7406 of *LNCS*, 2012.

[64] Steven Obua and Sebastian Skalberg. Importing hol into isabelle/hol. In Ulrich Furbach and Natarajan Shankar, editors, *Automated Reasoning*, volume 4130 of *LNCS*, pages 298–302. 2006. doi: 10.1007/11814771_27.

[65] Andrei Popescu and Johannes Hölzl. Possibilistic noninterference. *The Archive of Formal Proofs*, Sep 2012. ISSN 2150-914x. http://afp.sf.net/entries/Possibilistic_Noninterference.shtml, Formal proof development.

[66] Andrei Popescu, Johannes Hölzl, and Tobias Nipkow. Proving concurrent noninterference. In *Certified Programs and Proofs (CPP 2012)*, LNCS. Springer, 2012.

[67] M. Reiter and A. Rubin. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security (TISSEC)*, 1(1):66–92, 1998.

[68] Stefan Richter. Formalizing integration theory with an application to probabilistic algorithms. In Konrad Slind, Annette Bunker, and Ganesh Gopalakrishnan, editors, *Proceedings of TPHOLs 2004*, volume 3223 of *LNCS*, pages 271–286, 2004.

[69] René L. Schilling. *Measures, Integrals and Martingales*. Cambridge University Press, 2005. ISBN 978-0-521-61525-9.

[70] Claude Elwood Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423 and 623–656, Jul 1948. URL http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html.

[71] Vitaly Shmatikov. Probabilistic analysis of an anonymity system. *Journal of Computer Security*, 12:355–377, 2004.

[72] Sebastian Skalberg. Import tool. URL http://www.mangust.dk/skalberg/isabelle.php.

[73] Jinshuang Wang, Huabing Yang, and Xingyuan Zhang. Liveness reasoning with Isabelle/HOL. In Stefan Berghofer, Tobias Nipkow, Christian Urban, and Makarius Wenzel, editors, *Theorem Proving in Higher Order Logics*, volume 5674 of *LNCS*, pages 485–499. 2009. doi: 10.1007/978-3-642-03359-9_33.

[74] Makarius Wenzel. Structured induction proofs in isabelle/isar. In Jonathan M. Borwein and William M. Farmer, editors, *Mathematical Knowledge Management (MKM 2006)*, volume 4108 of *LNCS*, pages 17–30. Springer, 2006.