15-816 Modal Logic, Project Report

An Intuitionistic Modal Logic for Probability Theory

William Lovas Chris Martens

May 11, 2010

Abstract

We present a natural deduction calculus for an explicit-worlds modal logic of probability. We treat random events as propositions, and the worlds that index them represent probabilistic "worlds" in which they occur. These worlds form an algebraic structure called a residuated lattice with divisibility, and they record enough information about where events occur to suffice for probabilistic reasoning within the logic without evaluating to probability measures until the end.

Our system differs from prior accounts of probability theory in that it uses intuitionistic, judgmental principles to motivate its structure: we check the definitions of connectives with local soundness and completeness and present progress toward the verification of global properties such as substitution. We consider the relationship of conditional probabilities to hypotheticals and argue that, while they are distinct, a calculus of probabilistic reasoning can be built around conditions in much the same way as a calculus of truth can be built around hypotheses.

Unfortunately, the principles we derive for a kind of "probability weakening" and logical weakening turn out to be incompatible – together, they actually allow us to *strengthen* a probability arbitrarily. We describe how this arises and sketch some potential solutions as future work.

1 Introduction

Probability theory is a mathematical formalism concerned with the analysis of random phenomena. Using it, one can calculate the likelihood that random events will occur and trace how uncertainties propagate through a sequence of random events. In particular, given some base probabilities for atomic events, one can perform analyses on combinations of events, such as *joint probability*, the likelihood that two events co-occur, and *conditional probability*, the likelihood that an event will occur given that another occurs. Probability theory is traditionally formulated in an axiomatic style based on set theory: an event is a set of outcomes, and a probability function $P(-)$ assigns to each event a real number between 0 and 1, subject to certain constraints.

We present a *logical* presentation of probability theory based on the principle of distinguishing judgments and propositions, pioneered by Martin-Löf [ML96] and popularized by Pfenning and Davies [PD01]. Events become propositions, and probability values associated with events are derived from the labels of an explicit-worlds modal logic in the style of Simpson [Sim94].

The motivations for such an undertaking are myriad. First, by building a logic for reasoning in the presence of uncertainty, we open the door to a whole host of practical applications. We can imagine for instance leveraging the power of automated theorem proving in the design of planning algorithms for autonomous systems, or combining our logic with knowledge-based logics to reason about multi-agent systems. Since we intend to build our logic on an intuitionistic, verificationist foundation that emphasizes the role of proofs as evidence, we can further imagine a computational interpretation that treats proofs as plans that can be executed to achieve a given goal. We find this motive particularly compelling due to our arrival at a logic that works within abstract probability values: in the design of a computational agent that acts probabilistically, one could use a type system based on such a logic to verify statically that its actions respect (in)equalities of probability measures without evaluating them to real numbers. In fact in many scenarios, the real numbers defining ground probabilities will not even be known at compile time—they may be learned by the agent dynamically—in which case a type system based on algebraic probabilities, comparable without these values, could offer enormous advantages for correctness checking.

Furthermore, a logical re-envisioning can increase the expressive power of the theory. For example, building on a logical foundation gives us an obvious way of internalizing the notion of conditional probability $P(A\,|\,B)$

via the proposition $B \supset A$, now a first-class event which can hold with a given probability just like any other. Such a notion can express higher-order propositions which would otherwise fall outside the purview of traditional probability theory.

Finally, by recasting any theory according to the judgmental methodology, we have at our disposal its tools for guiding the design of logical connectives at our disposal, namely local soundness and completeness. Further, by casting it in a logical light, we can find connections to familiar logical notions and exploit known properties of those notions toward a better understanding of the theory.

Our objective was to design a harmonious logic capturing standard accounts of probability theory. By "harmonious", we mean it should admit a principle allowing for chaining proofs together (substitution), and connective definitions should enjoy local soundness and completeness according to such a principle. By "captures standard accounts", we mean that the logic should also render admissible various modes of reasoning from traditional presentations of probability theory, such as Kolmogorov's axioms and Bayes's theorem[1].

Unfortunately, the logic design we present is unsuccessful with respect to our goals: two of the principles that arose during its development turned out to be incompatible, in the sense that together they allow us to derive anything with probability 1 as long as it is derivable with some non-zero probability. We describe this result and sketch some potential paths of recourse. Despite such a downfall, we hope that future work towards our goal can make some use of the ideas and general approach we present herein.

Outline In the remainder of the paper, we explore some of the related work that inspired us (Section 2), followed by an in-depth discussion of the philosophical considerations that motivated our logic design (Section 3). We then present the full logic and its algebra, along with some metatheoretic results (Section 4) and an example of how we imagine it being used (Section 5). Finally, we explain the incompatibility in our principles that led to the degeneracy of our logic (Section 6), sketching some ideas towards how we might fix things in future work.

[1] At least, as far as such modes of reasoning do not compromise the intuitionistic nature of our logic. For instance, we should not expect to capture $P(A \vee \neg A) = 1$

2 Related Work

Existing work on logics for probability chiefly focuses on defining an analogue of classical Kripke semantics for probabilistic reasoning [HM98, SA07]. Our project differs by providing a constructive, proof theoretic foundation for a logic based on probability theory. There has been some work on intuitionistic variants of Kolmogorov's axioms for probability theory [Wea03], which serves to motivate the philosophical basis on which to base our rules but whose goals differ from ours in that it does not aim to reexamine the logical structure of probabilistic reasoning, defaulting to Hilbert-style axiom schema.

Zhou [Zho09] proposes a deductive system for probability logic similar to ours in that it is based on Kripke-style modal logic, the modal operators are indexed by probability values, and their intended interpretation is an inequality ($L_r\phi$ means "ϕ holds with at least probability r"). It contains something similar to what we call a *narrowing* principle allowing a derivation of the truth of $L_r\phi$ to $L_q\phi$ if q is less than r. However, this system differs in that its definition uses rational numbers for probability values and relies on a *probability model*, which contains a *type function* that returns such a number given a proposition. Because they do not treat events as propositions, they cannot define probabilities based on the structure of events. Further, they employ a classical, axiomatic approach that cannot leverage judgmental principles.

After many iterations of design discussed in Section 3, we decided to explore a line of work pioneered by Girard [Gir87] and more recently Reed [RP09]. In Girard's phase semantics and Reed's constructive resource semantics, they use elements of an algebraic structure to index propositions in a similar vein as Kripke semantics. They use these structure elements to represent the *resources* used to prove a proposition; in this way, they can embed the substructural properties of *linear logic* (among other examples) into a first-order calculus. Our motivation is different in that we do not need to describe resources with our label algebra, but it originates from the same desire to have better "bookkeeping" on the particular circumstances in which a proposition can be thought of as true, threaded through each step of a derivation. Our particular choice of algebra was motivated primarily by work on the mathematical formalism of fuzzy logic [NPM99].

3 Design Philosophy

First, we endeavor to explain our design principles and the philosophical considerations that led to them. Although ultimately we will see that our principles are incompatible with each other (Section 6 below), there may yet be some value in elaborating how they came about in order to elucidate our thought process regarding this work.

The basic idea of our probabilistic logic is to move from a judgment like A *true* to an explicit-worlds judgment of the form $A\,[p]$, meaning A *holds with probability at least* p. The foundation of any logic is its notion of entailment or hypothetical judgment, which we derive from conditional probability: hypotheses represent conditions and therefore are unlabelled. We thus have a judgment of the form

$$A_1\ \mathit{true}, \ldots, A_n\ \mathit{true} \vdash C\,[p]$$

whose interpretation is "assuming events $A_1, \ldots, A_n$ happen, event C happens with probability at least p", or in traditional probability theoretic notation, $P(C \,|\, A_1 \wedge \ldots \wedge A_n) \geq p$. We have chosen a formulation based on approximation ($\geq$) rather than one based on on-the-nose equality ($=$) in order to be able to give elimination rules for conjunction (see below).

An important question to answer early on in the work is just what algebra the probability values p are to be taken from. An obvious answer is the unit interval $[0, 1]$ of the real line, but part of the motivation of this work is to be as abstract as possible with regards to p's algebraic structure. If we wish to compute real numbers at the end of the day, we can always instantiate the algebra appropriately, but we will develop our logic incrementally, only assuming what we need to about the world algebra.

Since our interpretation of probability judgments is approximate, we expect that any judgment giving a lower bound can be pushed down or "narrowed" to a less precise lower bound. The principle of narrowing captures this idea:

Principle 1 (Narrowing) *If* $\Gamma \vdash A\,[p]$ *and* $p \geq q$, *then* $\Gamma \vdash A\,[q]$.

We see from this that our world algebra must at least be a partially ordered set: the justification of the principle of narrowing depends on transitivity.

An aside is in order with regards to how low one may lower a probability value. Since we are designing a logic of positive information, we do not consider a probability value corresponding to zero probability; instead, we include an element ε corresponding to a very small but strictly positive

probability. This is more a matter of interpretation than pragmatics: ε will have many order-theoretic properties of a zero element, but this convention saves us from having to ascribe meaning to inference rules when they act on proofs of impossibility—anything provable with some probability *is* in fact possible.

Our rule for using conditions, an analogue of the hypothesis rule in standard logics, says that assuming A happens, then any probability at all may serve as a lower bound, including the probability 1:

$$\frac{}{\Gamma, A\ \mathit{true} \vdash A\,[p]}\ \text{cond}$$

One can see that narrowing is baked into the condition rule: for any lower bound we assign, we could have chosen one that is lower still.

The cornerstone of our logic is the joint event $A \wedge B$. Following the Bayesian tradition, the joint probability of two events cannot be computed directly from the probability of each event in isolation, but rather must involve the *conditional probability* of one of the two events assuming the other. Jaynes [Jay03] tells a story thusly: though it may be quite plausible that the next person you meet has one brown eye, and similarly plausible that the next person you meet has one blue eye, it is still extremely *implausible* that they will have one brown eye *and* one blue eye. So although conjunction is normally one of the simplest connectives in a logic, the joint event is one of the most interesting kinds of events in our probability logic.

The traditional Bayesian definition of joint probability comes from the following equation:

$$P(A \wedge B) = P(B \mid A) \cdot P(A)$$

We can use this to derive an introduction rule:

$$\frac{\Gamma \vdash A\,[p] \qquad \Gamma, A\ \mathit{true} \vdash B\,[q]}{\Gamma \vdash A \wedge B\,[p \otimes q]}\ \wedge I$$

Here, we introduce an operator $\otimes$ into our algebra to reflect "multiplication" of probability space values. Eventually, we will derive axioms it must obey, but for now, we note that we require at least that $\otimes$ be a commutative operator such that $p_1 \otimes p_2 \leq p_i$.

When you eliminate a conjunction that holds with a certain lower bound, you may learn that either of its components holds with that same lower bound.

$$\frac{\Gamma \vdash A \wedge B\,[p]}{\Gamma \vdash A\,[p]}\ \wedge E_1 \qquad \frac{\Gamma \vdash A \wedge B\,[p]}{\Gamma \vdash B\,[p]}\ \wedge E_2$$

These rules may be justified by the traditional axiom of additivity, which tells us that $P(A) \geq P(A \wedge B)$ and $P(B) \geq P(A \wedge B)$. The first rule is locally sound by the following reduction:

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{D}\\ \Gamma \vdash A\,[p]\end{array} \quad \begin{array}{c}\mathcal{E}\\ \Gamma, A\ \mathit{true} \vdash B\,[q]\end{array}}{\Gamma \vdash A \wedge B\,[p \otimes q]}\wedge I}{\Gamma \vdash A\,[p \otimes q]}\wedge E_1 \quad \Longrightarrow_R \quad \dfrac{\begin{array}{c}\mathcal{D}\\ \Gamma \vdash A\,[p]\end{array}}{\Gamma \vdash A\,[p \otimes q]}\text{ Narrow}$$

in which the double-line indicates a use of the narrowing principle, justified by our requirement from above that $p \otimes q \leq p$.

To show the second rule locally sound, we must formulate a principle relating conditional truth to absolute truth, i.e. a substitution principle. We take the following, combining the probability values of the two pieces using our multiplication $\otimes$:

Principle 2 (Substitution) *If* $\Gamma, A\ \mathit{true} \vdash B\,[p]$ *and* $\Gamma \vdash A\,[q]$, *then* $\Gamma \vdash B\,[p \otimes q]$.

Using this principle, we can demonstrate the following local reduction:

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{D}\\ \Gamma \vdash A\,[p]\end{array} \quad \begin{array}{c}\mathcal{E}\\ \Gamma, A\ \mathit{true} \vdash B\,[q]\end{array}}{\Gamma \vdash A \wedge B\,[p \otimes q]}\wedge I}{\Gamma \vdash B\,[p \otimes q]}\wedge E_2 \quad \Longrightarrow_R \quad \begin{array}{c}[\mathcal{D}]\mathcal{E}\\ \Gamma \vdash B\,[p \otimes q]\end{array}$$

in which the reduct $[\mathcal{D}]\mathcal{E}$ is derived from $\mathcal{D}$ and $\mathcal{E}$ using the substitution principle.

One runs into some difficulty in trying to demonstrate local completeness for these elimination rules: given an arbitrary deduction $\mathcal{D} :: \Gamma \vdash A \wedge B\,[p]$, we can extract proofs of $\Gamma \vdash A\,[p]$ and $\Gamma \vdash B\,[p]$, but these do not suffice to reintroduce the connective. The second premise of the $\wedge I$ rule requires a proof of B conditioned on a proof of A. In ordinary logic, we would simply employ weakening, but what weakening principle, if any, should hold in our probabilistic logic?

In order for the judgment $\Gamma \vdash A\,[p]$ to have any meaning, the conditions in Γ must be non-contradictory: the Borel-Kolmogorov paradox [Kol33] says that the conditional probability $P(A \,|\, \Gamma)$ is undefined if $P(\Gamma) = 0$. Therefore, throughout our development, we tacitly assume that all contexts are non-contradictory. But clearly, we cannot allow arbitrary weakening by any proposition: if we weakened by a proposition that contradicted other

conditions, our sequent would lose all meaning. Furthermore, even if a proposition is non-contradictory with our conditions, we must record the fact that a weakening occurred: $P(B \mid \Gamma \wedge A)$ may be arbitrarily different from $P(B \mid \Gamma)$.

Given these considerations, we add another constructor to our algebra of probability values to record when weakening happens, written $p \to q$. The weakening principle then not only requires the new condition hold with non-zero probability—to avoid the Borel-Kolmogorov paradox—but it also records the structure of the probability value induced.

Principle 3 (Weakening) *If* $\Gamma \vdash A\,[p]$ *and* $\Gamma \vdash B\,[q]$, *then* $\Gamma, A \vdash B\,[p \to q]$

The implication value $p \to q$ forms a sort of inverse to multiplication $p \otimes q$, in a way that we will describe shortly.

Using the weakening principle, we can construct the following local expansion, demonstrating local completeness of our conjunction elimination rules.

$$\dfrac{\mathcal{D}}{\Gamma \vdash A \wedge B\,[p]} \;\Longrightarrow_E\; \dfrac{\dfrac{\dfrac{\mathcal{D}}{\Gamma \vdash A \wedge B\,[p]}}{\Gamma \vdash A\,[p]}\wedge E_1 \qquad \dfrac{\dfrac{\dfrac{\mathcal{D}}{\Gamma \vdash A \wedge B\,[p]}}{\Gamma \vdash A\,[p]}\wedge E_1 \quad \dfrac{\dfrac{\mathcal{D}}{\Gamma \vdash A \wedge B\,[p]}}{\Gamma \vdash B\,[p]}\wedge E_2}{\Gamma, A\;true \vdash B\,[p \to p]}\text{Weak}}{\Gamma \vdash A \wedge B\,[p \otimes (p \to p)]}\wedge I$$

in which the righthand subderivation of the expansion employs the weakening principle. It follows from the algebraic properties of $\otimes$ and $\to$ that $p \otimes (p \to p) = p$, since $\to$ is something like the inverse of $\otimes$, described in more detail below.

4 A Probability Calculus With Algebraic Labels

4.1 Residuated Lattices

The algebraic structure we use to index propositions is a *residuated lattice with divisibility*, a kind of algebra that plays a central role in the study of fuzzy logic [NPM99]. A residuated lattice[2] is an algebra $\mathcal{L} = \langle L, \wedge, \vee, \otimes, \to, \varepsilon, 1\rangle$ where:

[2] The definition here comes from Bělohlávek (2003) [Běl03], though variations exist in the literature, notably non-commutative residuated lattices.

1. $\langle L, \wedge, \vee, \varepsilon, 1\rangle$ is a lattice with least element ε,

2. $\langle L, \otimes, 1\rangle$ is a commutative monoid, and

3. $x \otimes y \leq z$ iff $x \leq y \to z$ (the "adjointness condition").

The "adjointness condition" is the sense in which the $\to$ operator is inverse to $\otimes$. The elements of a residuated lattice can serve as truth values in an algebraic semantics for fuzzy logic, which makes them seem like an attractive choice for a probability logic, a similar "logic of uncertainties".

In all residuated lattices, it is the case that $x \otimes (x \to y) \leq y$. A residuated lattice satisfies *divisibility* just in case $x \otimes (x \to y) = x \wedge y$ exactly. Divisibility is equivalent to the following statement [Běl03]: for any $x \leq y$, there is some z such that $x = y \otimes z$.

Although the specification of a residuated lattice is very simple, many non-trivial properties follow. We give two small examples here. First, the property above, that $x \otimes (x \to y) \leq y$:

$$\begin{array}{ll} x \to y \leq x \to y & \text{by reflexivity} \\ (x \to y) \otimes x \leq y & \text{by adjointness} \\ x \otimes (x \to y) \leq y & \text{by commutativity} \end{array}$$

Next, we show that $\otimes$ is monotonic in both arguments. Suppose $x \leq x'$ and $y \leq y'$:

$$\begin{array}{ll} x' \otimes y' \leq x' \otimes y' & \text{by reflexivity} \\ x' \leq y' \to x' \otimes y' & \text{by adjointness} \\ x \leq y' \to x' \otimes y' & \text{by transitivity} \\ x \otimes y' \leq x' \otimes y' & \text{by adjointness} \\ y' \leq x \to x' \otimes y' & \text{by commutativity/adjointness} \\ y \leq x \to x' \otimes y' & \text{by transitivity} \\ x \otimes y \leq x' \otimes y' & \text{by adjointness/commutativity} \end{array}$$

Most reasoning in residuated lattices follows this structure: move something out of the way using adjointness, employ lattice or monoid reasoning, move something back using adjointness.

4.2 Other Connectives

We present here locally sound and complete rules for implication and truth, and we speculate on rules for disjunction. We begin with implication:

$$\frac{\Gamma \vdash A\,[p] \quad \Gamma, A \vdash B\,[q]}{\Gamma \vdash A \supset B\,[q]}\supset I \qquad \frac{\Gamma \vdash A \supset B\,[p]}{\Gamma \vdash A\,[\varepsilon]}\supset E_1 \qquad \frac{\Gamma \vdash A \supset B\,[p] \quad \Gamma \vdash A\,[q]}{\Gamma \vdash B\,[p \otimes q]}\supset E_2$$

First, in order to avoid creating an invalid context, the introduction rule has to check that A occurs with some probability. Then, in order to ensure local completeness, we need to extract the information that A is possible. Finally, the usual implication elimination rule combines the probability values of its premises using $\otimes$ since the local soundness argument uses substitution. The local reductions and expansions follow:

$$\frac{\dfrac{\begin{array}{c}\mathcal{D}_1\\ \Gamma \vdash A\,[p]\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \Gamma, A \vdash B\,[q]\end{array}}{\Gamma \vdash A \supset B\,[q]}\supset I}{\Gamma \vdash A\,[\varepsilon]}\supset E_1 \quad \Longrightarrow_R \quad \frac{\begin{array}{c}\mathcal{D}_1\\ \Gamma \vdash A\,[p]\end{array}}{\Gamma \vdash A\,[\varepsilon]}\text{Narrow}$$

$$\frac{\dfrac{\begin{array}{c}\mathcal{D}_1\\ \Gamma \vdash A\,[\_]\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \Gamma, A \vdash B\,[p]\end{array}}{\Gamma \vdash A \supset B\,[p]}\supset I \quad \begin{array}{c}\mathcal{D}_3\\ \Gamma \vdash A\,[q]\end{array}}{\Gamma \vdash B\,[p \otimes q]}\supset E_2 \quad \Longrightarrow_R \quad \begin{array}{c}[\mathcal{D}_3]\mathcal{D}_2\\ \Gamma \vdash B\,[p \otimes q]\end{array}$$

$$\begin{array}{c}\mathcal{D}\\ \Gamma \vdash A \wedge B\,[p]\end{array} \Longrightarrow_E$$

$$\frac{\dfrac{\begin{array}{c}\mathcal{D}\\ \Gamma \vdash A \supset B\,[p]\end{array}}{\Gamma \vdash A\,[\varepsilon]}\supset E_1 \quad \dfrac{\dfrac{\dfrac{\begin{array}{c}\mathcal{D}\\ \Gamma \vdash A \supset B\,[p]\end{array}}{\Gamma \vdash A\,[\varepsilon]}\supset E_1 \quad \begin{array}{c}\mathcal{D}\\ \Gamma \vdash A \supset B\,[p]\end{array}}{\Gamma, A \vdash A \supset B\,[\varepsilon \to p]}\text{Weak} \quad \dfrac{}{\Gamma, A \vdash A\,[p]}\text{cond}}{\Gamma, A \vdash B\,[(\varepsilon \to p) \otimes p]}\supset E_2}{\Gamma \vdash A \supset B\,[p]}\supset I$$

In the local expansion, the final inference has the correct probability value because $\varepsilon \to p = 1$. The use of weakening with a condition of ε probability suggests a worry which is borne out in Section 6.

The connective $\top$ represents an event which always happens.

$$\frac{}{\Gamma \vdash \top\,[p]}\;\top I$$

There is no local reduction, since there are no elimination rules. The local expansion is essentially the standard one:

$$\begin{array}{c}\mathcal{D}\\ \Gamma \vdash \top\,[p]\end{array} \Longrightarrow_E \frac{}{\Gamma \vdash \top\,[p]}\;\top I$$

4.3 Metatheory

We sketch here a proof of the metatheoretic result of narrowing. Other results we defer to future work, when a suitable calculus has been properly defined.

Principle 1 (Narrowing) *If* $\Gamma \vdash A\,[p]$ *and* $p \geq q$, *then* $\Gamma \vdash A\,[q]$.

Proof. Straightforward induction on the inference rules of the logic, with an appeal in the $\wedge I$ case to the following lemma:

> *If* $r \leq p \otimes q$, *then* $r = p' \otimes q'$ *for some* p' *and* q' *such that* $p' \leq p$ *and* $q' \leq q$.

The lemma follows from divisibility: since $r \leq p \otimes q$, there is some s such that $r = p \otimes q \otimes s$. Let $p' = p$ and $q' = q \otimes s$. □

A Twelf formalization of this proof may be found in Appendix A; it uses `%trustme` to assume our unproven weakening principle and to obviate some simple algebraic relationships.

4.4 Alternate Rules

Since our algebra has meets $p \wedge q$, one might speculate the existence of a rule for conjunction such as the following:

$$\frac{\Gamma \vdash A\,[p] \quad \Gamma \vdash B\,[q]}{\Gamma \vdash A \wedge B\,[p \wedge q]}\;\wedge I'$$

In fact, as long as our residuated lattice is divisible, this rule and our rule are inter-admissible. First, we derive $\wedge I'$ from our rule:

$$\dfrac{\Gamma \vdash A\,[p] \qquad \dfrac{\Gamma \vdash A\,[p] \quad \Gamma \vdash B\,[q]}{\Gamma, A \vdash B\,[p \to q]}\text{ Weak}}{\Gamma \vdash A \wedge B\,[p \otimes (p \to q)]}\ \wedge I$$

Recall that the axiom of divisibility says that $p \otimes (p \to q) = p \wedge q$. Next, we can go in the reverse direction:

$$\dfrac{\Gamma \vdash A\,[p] \qquad \dfrac{\Gamma \vdash A\,[p] \quad \Gamma \vdash B\,[q]}{\Gamma \vdash B\,[p \otimes q]}\text{ Subst}}{\Gamma \vdash A \wedge B\,[p \wedge (p \otimes q)]}\ \wedge I'$$

The final line has the correct label: since $p \otimes q \le p$, we have $p \wedge (p \otimes q) = p \otimes q$.

Note that both admissibilities would hold even in a system based on on-the-nose probability values rather than lower bounds.

5 Example of Use

To give the reader an idea of how we imagine someone using such a system, consider a simple example: suppose one rolls a single, fair, six-sided die. We would like to assign atomic events to each possible roll (one for each of the six sides), and we would like to deduce some conditional probabilities, e.g. the probability of a 2 given that we rolled an even number and how that compares to the probability of a 2 given that we rolled a prime, or the probability of an even given that we rolled a prime.

One thing we might like to do is actually derive real probability values from the interval $(0, 1]$. After deriving the algebraic probability of an event in the logic, we could imagine instantiating the lattice with sets of disjoint atomic events where meets are their intersections and joins are their unions. For the event space Ω, define a function $p : \Omega \to \mathbb{R}$ to give probabilities to atomic events—in the case of the dice roll, $p(-) = 1/6$. Then define a measurement function $\mathcal{M} : \mathcal{L} \to \mathbb{R}$ as follows:

$$\begin{aligned}
\mathcal{M}(S) &= \sum_{\omega \in S} p(\omega) && \text{where } S \text{ is a set} \\
\mathcal{M}(p \otimes q) &= \mathcal{M}(p) \times \mathcal{M}(q) \\
\mathcal{M}(p \to q) &= \mathcal{M}(p \cap q)/\mathcal{M}(p)
\end{aligned}$$

As long as an algebraic probability value represents a statement that can be made in the language of axiomatic probability theory—e.g. there are no intersections of residuals like $(p \to q) \cap (r \to s)$—the measurement function should be defined on it.

Now suppose we associate the event *Even* with a probability value *ev* and the event *Prime* with a probability value *pr*. To answer a probability theory query like $P(\mathit{Even} \mid \mathit{Prime}) = ?$, we first derive the judgment $\mathit{Even} \vdash \mathit{Prime}\,[ev \to pr]$ by weakening. Then we map *ev* and *pr* down to sets and measure:

$$
\begin{aligned}
\mathcal{M}(ev \to pr) &= \mathcal{M}(\{2,4,6\} \to \{2,3,5\}) \\
&= \mathcal{M}(\{2,4,6\} \cap \{2,3,5\}) / \mathcal{M}(\{2,4,6\}) \\
&= \mathcal{M}(\{2\}) / \mathcal{M}(\{2,4,6\}) \\
&= \frac{1}{6} / \frac{1}{2} \\
&= \frac{1}{3}
\end{aligned}
$$

The answer is one third, as expected.

Although it is somewhat abbreviated, we hope this example will give the reader an idea how our system relates to the traditional account of probability theory. We leave it to future work to find interesting examples of what can be done with "higher-order" probability values that do not fit into the language of traditional probability theory, like $(p \to q) \wedge (r \to s)$ or $(p \to q) \to r$.

6 Incompatibility of Weakening and Narrowing

Using the stated principles of narrowing and weakening, we can produce a derivation of $\Gamma \vdash A\,[1]$ from any derivation of $\Gamma \vdash A\,[p]$ as follows:

$$
\begin{prooftree}
\AxiomC{$\mathcal{D}$}
\noLine
\UnaryInfC{$\Gamma \vdash A\,[p]$}
\AxiomC{}
\RightLabel{$\top I$}
\UnaryInfC{$\Gamma \vdash \top\,[1]$}
\doubleLine
\RightLabel{Narrow}
\UnaryInfC{$\Gamma \vdash \top\,[p]$}
\doubleLine
\RightLabel{Weak}
\BinaryInfC{$\Gamma, \top \vdash A\,[p \to p]$}
\AxiomC{}
\RightLabel{$\top I$}
\UnaryInfC{$\Gamma \vdash \top\,[1]$}
\doubleLine
\RightLabel{Subst}
\BinaryInfC{$\Gamma \vdash A\,[(p \to p) \otimes 1]$}
\doubleLine
\RightLabel{Narrow}
\UnaryInfC{$\Gamma \vdash A\,[1]$}
\end{prooftree}
$$

The critical problem is that the weakening principle allows us to weaken with a narrowed condition, thereby strengthening the "weakened" proof of the conditioned event – in this case, widening its probability from $1 \to p$ (which is equal to p) to $p \to p$ (equal to 1).

This problem suggests that perhaps weakening should not permit events judged with an *inexact* probability into the condition context, or perhaps that it should require a separate judgment on them as having *less than* or equal to the given probability.

Although we realized this deficiency very late in the project lifetime, we began thinking through some solutions. Our first instinct was to return to trying for *exact* probabilities in light of what we learned since we initially strayed from them.

If we want an exact representation, the $\wedge$-elimination rules will no longer work, because each conjunct alone should have a greater probability than their conjunction. Here is a closed-scope elimination rule that may work instead:

$$\frac{\Gamma \vdash A \wedge B\,[p] \qquad \Gamma, A, B \vdash C\,[q]}{\Gamma \vdash C\,[p \otimes q]}\ \wedge E'$$

Keeping the introduction rule as previously stated, we can check local soundness and completeness for this rule:

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{D}_1\\ \Gamma \vdash A\,[p]\end{array} \qquad \begin{array}{c}\mathcal{D}_2\\ \Gamma, A \vdash B\,[q]\end{array}}{\Gamma \vdash A \wedge B\,[p \otimes q]}\ \wedge I \qquad \begin{array}{c}\mathcal{E}\\ \Gamma, A, B \vdash C\,[r]\end{array}}{\Gamma \vdash C\,[p \otimes q \otimes r]}\ \wedge E' \implies_R \begin{array}{c}[\mathcal{D}_1][\mathcal{D}_2]\mathcal{E}\\ \Gamma \vdash C\,[p \otimes q \otimes r]\end{array}$$

$$\begin{array}{c}\mathcal{D}\\ \Gamma \vdash A \wedge B\,[p]\end{array} \implies_E \dfrac{\begin{array}{c}\mathcal{D}\\ \Gamma \vdash A \wedge B\,[p]\end{array} \qquad \dfrac{\dfrac{}{\Gamma, A, B \vdash A\,[1]}\ \text{cond} \qquad \dfrac{}{\Gamma, A, B, A \vdash B\,[1]}\ \text{cond}}{\Gamma, A, B \vdash A \wedge B\,[1]}\ \wedge I}{\Gamma \vdash A \wedge B\,[p \otimes 1]}\ \wedge E'$$

This is what we want because $p \otimes 1 = p$.

These rules successfully buy us on-the-nose equality for conjunction, at least according to the local sanity checks. Continuing this line of thought may therefore be fruitful, and we have not had time to fully explore it.

However, we would need to reinvent the implication connective—now that we are dealing with exact equality, we cannot use the trick of encapsulating a proof of "A true with some non-zero probability" within a proof of $A \supset B$. It may be possible to recover the same behavior by using a *bona fide* possibility judgment.

Also, some speculation suggests that doing things this way would collapse the conjunctive algebraic operators, making it into something more like a Heyting algebra than a residuated lattice.

7 Conclusion

We have defined a calculus for probabilistic reasoning based on explicit worlds modal logic where worlds are elements of a residuated lattice with divisibility. We argue that this presentation serves as a novel way of thinking about probability theory with the important distinction that it abstracts from an events-as-sets interpretation and instead concerns itself with fundamental probabilistic reasoning principles.

The logic we present fails as a sound probability calculus. However, we consider this fact a result, if an incompletely-explored one, that probabilistic reasoning fails to agree with many existing intuitions from judgmental reasoning for modal logics. Many of our related difficulties have further clarified that the relationship of probability theory to hypothetical reasoning in general is complicated, and probabilistic conditionals are not at all straightforwardly related to logical hypotheticals.

There are many directions to take this work in the future! We have sketched several ideas for how we might recover our ideas, but only time will tell whether they can be made into a complete, consistent calculus for reasoning about probabilities. Along the way, we would do well to search for examples of the expressivity we would like to gain over traditional presentations of probability theory. And of course, we would strive to always keep in mind the lesson we have learned: although local sanity checks can help guide a logic's design, one must step back every so often and see whether the principles you have supposed capture the reasoning you intend.

References

[Běl03] Radim Bělohlávek. Some properties of residuated lattices. *Czechoslovak Mathematical Journal*, 53(128):161–171, 2003.

[Gir87] Jean-Yves Girard. Linear logic. *Theoretical Computer Science*, 50(1):1–102, 1987.

[HM98] Aviad Heifetz and Philippe Mongin. The modal logic of probability. In *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'98)*, pages 175–185. Morgan Kaufmann Publishers, Inc., 1998.

[Jay03] E. T. Jaynes. *Probability Theory: the Logic of Science*. Cambridge University Press, 2003.

[Kol33] Andrey Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. Reprinted in English translation as [Kol56].

[Kol56] Andrey Kolmogorov. *Foundations of the Theory of Probability*. Chelsea, New York, 1956. Available on the web from http://www.mathematik.com/Kolmogorov/index.html.

[ML96] Per Martin-Löf. On the meanings of the logical constants and the justifications of the logical laws. *Nordic Journal of Philosophical Logic*, 1(1):11–60, 1996.

[NPM99] Vilém Novák, Irina Perfilieva, and Jiří Močkoř. *Mathematical Principles of Fuzzy Logic*. Kluwer Academic Publishers, Norwell, Massachusetts, 1999.

[PD01] Frank Pfenning and Rowan Davies. A judgmental reconstruction of modal logic. *Mathematical Structures in Computer Science*, 11:511–540, 2001. Notes to an invited talk at the *Workshop on Intuitionistic Modal Logics and Applications* (IMLA'99), Trento, Italy, July 1999.

[RP09] Jason Reed and Frank Pfenning. A constructive approach to the resource semantics of substructural logics. Submitted, April 2009.

[SA07] Afsaneh Shirazi and Eyal Amir. Probabilistic modal logic. In *Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI'07)*, pages 489–495. AAAI Press, 2007.

[Sim94] Alex K. Simpson. *The Proof Theory and Semantics of Intuitionistic Modal Logic*. PhD thesis, University of Edinburgh, 1994.

[Wea03] Brian Weatherson. From classical to intuitionistic probability. *Notre Dame Journal of Formal Logic*, 44(2):111–123, 2003.

[Zho09] Chunlai Zhou. A complete deductive system for probability logic. *Journal of Logic and Computation*, 19(6):1427–1454, 2009.

A Twelf Proof of Narrowing

```
label : type.           %name label P.
meet : label -> label -> label.
join : label -> label -> label.
tensor : label -> label -> label.
arrow : label -> label -> label.
one : label.

leq : label -> label -> type.

leq/refl : leq P P.
leq/trans : leq P R
             <- leq P Q
             <- leq Q R.

leq/one   : leq P one.

leq/meet1 : leq (meet P Q) P.
leq/meet2 : leq (meet P Q) Q.

leq/join1 : leq P (join P Q).
leq/join2 : leq Q (join P Q).

% monoid laws
eq : label -> label -> type.
leq/eq1 : leq P Q
          <- eq P Q.
leq/e12 : leq Q P
           <- eq P Q.
eq/leq : eq P Q
          <- leq P Q
          <- leq Q P.

%abbrev eq/refl : eq P P = (eq/leq leq/refl leq/refl).

eq/tensor-unit  : eq (tensor one P) P.
eq/tensor-assoc : eq (tensor (tensor P Q) R) (tensor P (tensor Q R)).
eq/tensor-comm  : eq (tensor P Q) (tensor Q P).

leq/adj-arr-tens  : leq (tensor P Q) R
```

```
                     <- leq P (arrow Q R).
leq/adj-tens-arr  : leq P (arrow Q R)
                     <- leq (tensor P Q) R.

prop : type. %name prop A.
and : prop -> prop -> prop.
or  : prop -> prop -> prop.
imp : prop -> prop -> prop.
top : prop.
bot : prop.

true : prop -> label -> type.
hyp : prop -> type.

top-i : true top P.

% bot-e : true A P
%         <- true bot P.

and-i : true (and A B) (tensor P Q)
         <- true A P
         <- (hyp A -> true B Q).

and-e1 : true A P
          <- true (and A B) P.

and-e2  : true B P
           <- true (and A B) P.

var     : true A P
           <- hyp A.

%block assm : some {A:prop} block {d:hyp A}.

weakening : true B Q -> true A P
             -> (hyp A -> true B (arrow P Q)) -> type.
%mode weakening +X1 +X2 -X3.

%worlds () (weakening _ _ _).
%trustme %total {} (weakening _ _ _).

%% label theorems %%

eq-trans  : eq P Q -> eq Q R -> eq P R -> type.
%mode eq-trans +X1 +X2 -X3.

%worlds () (eq-trans _ _ _).
%trustme %total {} (eq-trans _ _ _).
```

```
tensor-compat-eq : eq P P' -> eq Q Q'
                    -> eq (tensor P Q) (tensor P' Q') -> type.
%mode tensor-compat-eq +X1 +X2 -X3.

%worlds () (tensor-compat-eq _ _ _).
%trustme %total {} (tensor-compat-eq _ _ _).

leq-tensor : leq R (tensor P Q)
              -> eq R (tensor P' Q')
              -> leq P' P
              -> leq Q' Q -> type.
%mode leq-tensor +X1 -X2 -X3 -X4.

%worlds () (leq-tensor _ _ _ _).
%trustme %total D (leq-tensor D _ _ _).

narrow : true A P -> leq Q P -> true A Q' -> eq Q Q' -> type.
%mode narrow +X1 +X2 -X3 -X4.

- : narrow top-i _ top-i eq/refl.

- : narrow
        (and-i
            ([d:hyp A] DtrueB d : true B Q)
            (DtrueA : true A P))
        (Dleq : leq R (tensor P Q))
        (and-i DtrueB'' DtrueA'')
        DeqRtens'
     <- leq-tensor Dleq
        (DeqRtens : eq R (tensor P' Q'))
        (DleqP' : leq P' P)
        (DleqQ' : leq Q' Q)
     <- narrow DtrueA DleqP'
        (DtrueA'' : true A P'')
        (DeqP'' : eq P' P'')
     <- ({d:hyp A}
            narrow
            (DtrueB d) DleqQ'
            (DtrueB'' d : true B Q'')
            (DeqQ'' : eq Q' Q''))
     <- tensor-compat-eq DeqP'' DeqQ''
        (Deqtens' : eq (tensor P' Q') (tensor P'' Q''))
     <- eq-trans DeqRtens Deqtens'
        (DeqRtens' : eq R (tensor P'' Q'')).

- : narrow
        (and-e1 (DtrueAandB : true (and A B) P))
        (Dleq : leq Q P)
        (and-e1 DtrueAandB')
```

```
        DeqQ
     <- narrow DtrueAandB Dleq
        (DtrueAandB' : true (and A B) Q')
        (DeqQ : eq Q Q').

- : narrow
        (and-e2 (DtrueAandB : true (and A B) P))
        (Dleq : leq Q P)
        (and-e2 DtrueAandB')
        DeqQ
     <- narrow DtrueAandB Dleq
        (DtrueAandB' : true (and A B) Q')
        (DeqQ : eq Q Q').

- : narrow (var Dhyp) _ (var Dhyp) eq/refl.

%worlds (assm) (narrow _ _ _ _).
%total D (narrow D _ _ _).
```