A Logic for Reasoning about Upper Probabilities

Joseph Y. Halpern Riccardo Pucella

Department of Computer Science
Cornell University

Abstract

We present a propositional logic to reason about the uncertainty of events, where the uncertainty is modeled by a set of probability measures assigning an interval of probability to each event. We give a sound and complete axiomatization for the logic, and show that the satisfiability problem is NP-complete, no harder than satisfiability for propositional logic.

1 Introduction

Various measures exist that attempt to quantify uncertainty. For anyone familiar with probability theory, probability measures are an obvious choice. However, probability measures have difficulties dealing with certain situations of interest. Most notably, whereas probabilities can deal with the direct uncertainty of an event happening, there is no way for probabilities to model the uncertainty related to the probabilities themselves. Consider a simple example: suppose we have a bag of 100 marbles; we know 30 are red and we know the remaining 70 are either blue or yellow, although we do not know the exact proportion of blue and yellow. If we are modeling the situation where we pick a ball from the bag at random, we need to assign a probability to three different events: picking up a red ball (*red-event*), picking up a blue ball (*blue-event*), and picking up a yellow ball (*yellow-event*). We can clearly assign a probability of .3 to *red-event*, but there is no clear probability to assign to *blue-event* or *yellow-event*.

One way to approach this problem is to represent the uncertainty using a set of probability measures, with a probability measure for each possible proportion of blue and yellow balls. For instance, we could use the set of probabilities $\mathcal{P} = \{\mu_\alpha : \alpha \in [0, .7]\}$, where μ_α gives *red-event* probability $.3$, *blue-event* probability α, and *yellow-event* probability $.7 - \alpha$. To any set of probabilities $\mathcal{P}$ we can assign a pair of functions, the upper and lower probability measure, that for an event X give the supremum (respectively, the infimum) of the probability of X according to the probability measures in $\mathcal{P}$. These measures can be used to deal with uncertainty in the manner described above, where the lower and upper probability of an event defines a range of probability for that event[1]. Note that this is not the only way to model the situation. An alternative approach, using inner measures, is studied in [6].

[1] An alternate view of upper probabilities originates from assigning subjective probabilities to events by testing and finding at what odds a person is prepared to bet against them. This gives rise, given suitable assumptions, to an equivalent formulation of lower and upper probability measures [16, 17].

Given a measure of uncertainty, one can define a logic for reasoning about it. Fagin, Halpern and Megiddo [7] introduce a logic for reasoning about probabilities, with a possible-worlds semantics that assigns a probability to each possible world. They provide an axiomatization for the logic, which they prove sound and complete with respect to the semantics. They also show that the satisfiability problem for the logic, somewhat surprisingly, is NP-complete, and hence no harder than the satisfiability problem for propositional logic. They moreover show how their logic can be extended to other notions of uncertainty, such as inner measures [6] and Dempster-Shafer belief functions [14].

In this paper, we describe a logic for reasoning about upper probability measures, along the lines of the logic introduced in [7]. The main challenge is to derive a provably complete axiomatization of the logic; to do this, we need a characterization of upper probability measures in terms of properties that can be expressed in the logic. Many semantic characterizations of upper probability measures have been proposed in the literature. The characterization of Anger and Lembcke [1] turns out to be best suited for our purposes. Even though we are reasoning about potentially infinite sets of probability measures, the satisfiability problem for our logic remains NP-complete. Intuitively, we need guess only a small number of probability measures to satisfy any given formula, polynomially many in the size of the formula. Moreover, these probability measures can be taken to be defined on a finite state space, again polynomial in the size of the formula. Thus, we need to basically determine polynomially many values—a value for each probability measure at each state—to decide the satisfiability of a formula.

The rest of this paper is structured as follows. In Section 2, we review the required material from probability theory and the theory of upper probabilities. In Section 3, we present the logic and an axiomatization. In Section 4, we prove that the axiomatization is sound and complete with respect to the natural semantic models expressed in terms of upper probability spaces. Finally, in Section 5, we prove that the decision problem for the logic is NP-complete. The proofs of the more technical results are given in the appendices.

2 Characterizing upper probability measures

We start with a brief review of the relevant definitions. Recall that a probability measure is a function $\mu : \Sigma \rightarrow [0,1]$ for Σ an algebra of subsets of Ω (that is Σ is closed under complements and unions), satisfying $\mu(\emptyset) = 0$, $\mu(\Omega) = 1$ and $\mu(A \cup B) = \mu(A) + \mu(B)$ for all disjoint sets A, B in Σ. [2] A probability space is a tuple (Ω, Σ, μ), where Ω is a set, Σ is an algebra of subsets of Ω (the measurable sets), and μ is a probability measure defined on Σ. Given a set $\mathcal{P}$ of probability measures , let $\mathcal{P}^*$ be the upper probability measure[3] defined by $\mathcal{P}^*(X) = \sup\{\mu(X) \ : \ \mu \in \mathcal{P}\}$ for $X \in \Sigma$. Similarly, $\mathcal{P}_*(X) = \inf\{\mu(X) \ : \ \mu \in \mathcal{P}\}$ is the lower probability of $X \in \Sigma$. A straightforward derivation shows that the relationship $\mathcal{P}_*(X) = 1 - \mathcal{P}^*(\overline{X})$ holds between upper

[2] If Ω is infinite, we could also require that Σ be a σ-algebra (i.e., closed under countable unions) and that μ be countably additive. Requiring countable additivity would not affect our results, since we show that we can take Ω to be finite. For ease of exposition, we have not required it.

[3] In the literature, the term upper probability is sometimes used in a more restricted sense than here. For example, Dempster [4] uses the term to denote a class of measures which were later characterized as Dempster-Shafer belief functions [14]; belief functions are in fact upper probability measures in our sense, but the converse is not true [10]. In the measure theory literature, what we call upper probability measures correspond to *upper envelopes* of measures, which are defined as the sup of sets of general measures, not just probability measures.

and lower probabilities, where $\overline{X}$ is the complement of X in Ω. Because of this duality, we restrict the discussion to upper probability measures in this paper, with the understanding that results for lower probabilities can be similarly derived. Finally, an *upper probability space* is a tuple $(\Omega, \Sigma, \mathcal{P})$ where $\mathcal{P}$ is a set of probability measures on Σ.

We would like a set of properties that completely characterizes upper probability measures. In other words, we would like a set of properties that allow us to determine if a function $f : \Sigma \to [0, 1]$ (for an algebra Σ of subsets of Ω) is an upper probability measure, that is, whether there exists a set $\mathcal{P}$ of probability measures such that for all $X \in \Sigma$, $\mathcal{P}^*(X) = f(X)$.

One approach to the characterization of upper probability measures is to adapt the characterization of Dempster-Shafer belief functions; these functions are known to be the lower envelope of the probability measures that dominate them, and thus form a subclass of the class of lower probability measures. By the duality noted earlier, a characterization of lower probability measures would yield a characterization of upper probability measures. The characterization of belief functions is derived from a generalization of the following inclusion-exclusion principle for probabilities (by replacing the equality with an inequality):

$$\mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} (-1)^{i-1} (\sum_{\substack{J \subseteq \{1,\dots,n\} \\ |J|=i}} \bigcap_{j \in J} A_j).$$

It seems reasonable that a characterization of lower (or upper) probability measures could be derived along similar lines. As we now show, most properties derivable from the inclusion-exclusion principle (which include most of the properties reported in the literature) are insufficient to characterize upper probability measures.

Consider the following "inclusion-exclusion"–style properties (mainly taken from [17]). To simplify the statement of these properties, let $\mathcal{P}^{-1} = \mathcal{P}^*$ and $\mathcal{P}^{+1} = \mathcal{P}_*$.

(1) $\mathcal{P}^*(A_1 \cup \cdots A_n) \leq \sum_{i=1}^{n} \sum_{|I|=i} (-1)^{i+1} \mathcal{P}^{(-1)^i}(\bigcap_{j \in I} A_j)$,

(2) $\mathcal{P}_*(A_1 \cup \cdots A_n) \geq \sum_{i=1}^{n} \sum_{|I|=i} (-1)^{i+1} \mathcal{P}^{(-1)^{i+1}}(\bigcap_{j \in I} A_j)$,

(3) $\mathcal{P}_*(A \cup B) + \mathcal{P}_*(A \cap B) \leq \mathcal{P}_*(A) + \mathcal{P}^*(B) \leq \mathcal{P}^*(A \cup B) + \mathcal{P}^*(A \cap B)$,

(4) $\mathcal{P}_*(A) + \mathcal{P}_*(B) \leq \mathcal{P}_*(A \cup B) + \mathcal{P}^*(A \cap B) \leq \mathcal{P}^*(A) + \mathcal{P}^*(B)$,

(5) $\mathcal{P}_*(A) + \mathcal{P}_*(B) \leq \mathcal{P}_*(A \cap B) + \mathcal{P}^*(A \cup B) \leq \mathcal{P}^*(A) + \mathcal{P}^*(B)$.

It is easily verified that the above properties hold for upper probability measures. The issue is whether they completely characterize the class of upper probability measures. We show the inherent incompleteness of these properties by proving that they are all derivable from the following simple property, which is by itself insufficient to characterize upper probability measures.

(6) If $A \cap B = \emptyset$, then $\mathcal{P}^*(A) + \mathcal{P}_*(B) \leq \mathcal{P}^*(A \cup B) \leq \mathcal{P}^*(A) + \mathcal{P}^*(B)$.

Proposition 2.1: *Property (6) implies properties (1)-(5).*

Proof: See Appendix A. ∎

The following example shows the inusfficiency of Propoerty (6). Let $\mathcal{P}$ be the set of probability measures $\{\mu_1, \mu_2, \mu_3, \mu_4\}$ over $\Omega = \{a, b, c, d\}$ (with Σ containing all subsets of Ω) defined on singletons by:

$$\begin{array}{llll} \mu_1(a) = \frac{1}{4} & \mu_1(b) = \frac{1}{4} & \mu_1(c) = \frac{1}{4} & \mu_1(d) = \frac{1}{4} \\ \mu_2(a) = 0 & \mu_2(b) = \frac{1}{8} & \mu_2(c) = \frac{3}{8} & \mu_2(d) = \frac{1}{2} \\ \mu_3(a) = \frac{1}{8} & \mu_3(b) = \frac{3}{8} & \mu_3(c) = 0 & \mu_3(d) = \frac{1}{2} \\ \mu_4(a) = \frac{3}{8} & \mu_4(b) = 0 & \mu_4(c) = \frac{1}{8} & \mu_4(d) = \frac{1}{2} \end{array}$$

and extended by additivity to all of Σ. This defines an upper probability measure $\mathcal{P}^*$ over Σ. Consider the function $\upsilon_\epsilon : \Sigma \rightarrow [0, 1]$ defined by:

$$\upsilon_\epsilon(X) = \begin{cases} \mathcal{P}^*(X) + \epsilon & \text{if } X = \{a, b, c\} \\ \mathcal{P}^*(X) & \text{otherwise} \end{cases}$$

We claim that the function υ_ϵ, for small enough $\epsilon > 0$, satisfies property (6), but cannot be an upper probability measure.

Proposition 2.2: *For $0 < \epsilon < \frac{1}{8}$, the function υ_ϵ satisfies property (6), but is not an upper probability measure. That is, we cannot find a set $\mathcal{P}'$ of probability measures such that $\upsilon_\epsilon = (\mathcal{P}')^*$.*

Proof: See Appendix B. ∎

This example clearly illustrates the need to go beyond the inclusion-exclusion principle to find properties that characterize upper probability measures. As it turns out, various complete characterizations have been described in the literature [8, 9, 18, 19, 1]. While all are equivalent in spirit, we focus on the characterization given by Anger and Lembcke [1], because it is particularly well-suited to the logic presented in the next section. The characterization is based on the notion of *set cover*: a set A is said to be covered n times by a multiset $\{\{A_1, \ldots, A_m\}\}$ of sets if every element of A appears in at least n sets from $A_1, \ldots, A_m$: for all $x \in A$, there exists $i_1, \ldots, i_n$ in $\{1, \ldots, m\}$ such that for all $j \leq n$, $x \in A_{i_j}$. It is important to note here that $\{\{A_1, \ldots, A_m\}\}$ is a multiset, not a set; the A_i's are not necessarily distinct. (We use the $\{\{\ \}\}$ notation to denote multisets.) An (n, k)-*cover of* (A, Ω) is a multiset $\{\{A_1, \ldots, A_m\}\}$ that covers Ω k times and covers A $n+k$ times.

The notion of (n, k)-cover is the key concept in Anger and Lembcke's characterization of upper probability measures.

Theorem 2.3:[1] *Let Ω be a set, Σ an algebra of subsets of Ω, and υ a function $\upsilon : \Sigma \rightarrow [0, 1]$. There exists a set $\mathcal{P}$ of probability measures with $\upsilon = \mathcal{P}^*$ if and only if υ satisfies the following three properties:*

UP1. $\upsilon(\emptyset) = 0$,

UP2. $\upsilon(\Omega) = 1$,

UP3. *for all integers m, n, k and all subsets $A_1, \ldots, A_m$ in Σ, if $\{\{A_1, \ldots, A_m\}\}$ is an (n, k)-cover of (A, Ω), then $k + n\upsilon(A) \leq \sum_{i=1}^m \upsilon(A_i)$.*

Proof: We reproduce the proof of this result in Appendix C. ∎

We need to strengthen Theorem 2.3 in order to prove the main result of this paper, namely, the completeness of the axiomatization of the logic we introduce in the next section. We show that if the cardinality of the state space Ω is finite, then we need only finitely many instances of property **UP3**. Notice that we cannot derive this from Theorem 2.3 alone: even if $|\Omega|$ is finite, **UP3** does not provide any bound on m, the number of sets to consider in an (n, k) cover of a set A. Indeed, there does not seem to be any *a priori* reason why the value of m, n, and k can be bounded. Bounding this value of m (and hence of n and k, since they are no larger than m) is the one of the key technical results of this paper, and the necessary foundation of our work.

Theorem 2.4: *There exists constants $B_0, B_1, \ldots$ such that if Σ is an algebra of subsets of Ω and υ is a function $\upsilon : \Sigma \rightarrow [0, 1]$, then there exists a set $\mathcal{P}$ of probability measures such that $\upsilon = \mathcal{P}^*$ if and only if υ satisfies the following properties:*

UPF1. $\upsilon(\emptyset) = 0$,

UPF2. $\upsilon(\Omega) = 1$,

UPF3. *for all integers $m, n, k \leq B_{|\Omega|}$ and all sets $A_1, \ldots, A_m$, if $\{\{A_1, \ldots, A_m\}\}$ is an (n, k)-cover of (A, Ω), then $k + n\upsilon(A) \leq \sum_{i=1}^{m} \upsilon(A_i)$.*

Proof: See Appendix D. ∎

Property **UPF3** is significantly weaker than **UP3**. In principle, checking that **UP3** holds for a given function requires checking that it holds for arbitrarily large collections of sets, even if the underlying set Ω is finite. On the other hand, **UPF3** guarantees that it is in fact sufficient to look at collections of size at most $B_{|\Omega|}$. This observation is key to the completeness result.

It is not important for our purposes (namely to get completeness of the axiomatization of the logic introduced in the next section) what the actual values of $B_0, B_1, \ldots$ are; it is sufficient for them to exist and be finite. The proof of Theorem 2.4 found in Appendix D relies on a Ramsey-theoretic argument that does not provide a bound on the B_is.

3 The logic

The syntax for the logic is straightforward, and is taken from [7]. We fix a set $\Phi_0 = \{p_1, p_2, \ldots\}$ of *primitive propositions.* The set Φ of *propositional formulas* is the closure of Φ_0 under $\wedge$ and $\neg$. We assume a special propositional formula *true*, and abbreviate $\neg$*true* as *false*. We use p to represent primitive propositions, and ϕ and ψ to represent propositional formulas. A *term* is an expression of the form $\theta_1 l(\phi_1) + \cdots + \theta_k l(\phi_k)$, where $\theta_1, \ldots, \theta_k$ are reals and $k \geq 1$. A *basic likelihood formula* is a statement of the form $t \geq \alpha$, where t is a term and α is a real. A *likelihood formula* is a boolean combination of basic likelihood formulas. We use f and g to represent likelihood formulas. We use obvious abbreviations where needed, such as $l(\phi) - l(\psi) \geq a$ for $l(\phi) + (-1)l(\psi) \geq a$, $l(\phi) \geq l(\psi)$ for $l(\phi) - l(\psi) \geq 0$, $l(\phi) \leq a$ for $-l(\phi) \geq -a$, $l(\phi) < a$ for $\neg(l(\phi) \geq a)$ and $l(\phi) = a$ for $(l(\phi) \geq a) \wedge (l(\phi) \leq a)$. Define the length $|f|$ of the likelihood formula f to be the number of symbols required to write f, where each coefficient is counted as one symbol.

We assign a semantics to likelihood formulas through an upper probability space, as defined in Section 2. Formally, an *upper probability structure* is a tuple $M = (\Omega, \Sigma, \mathcal{P}, \pi)$ where $(\Omega, \Sigma, \mathcal{P})$

is an upper probability space and π associates with each state (or world) in Ω a truth assignment on the primitive propositions in Φ_0. Thus, $\pi(s)(p) \in \{\textbf{true}, \textbf{false}\}$ for $s \in \Omega$ and $p \in \Phi_0$. Let $[\![p]\!]_M = \{s \in \Omega : \pi(s)(p) = \textbf{true}\}$. We call M *measurable* if for each $p \in \Phi_0$, $[\![p]\!]_M$ is measurable. If M is measurable then $[\![\phi]\!]_M$ is measurable for all propositional formulas ϕ. In this paper, we restrict our attention to measurable upper probability structures. Extend $\pi(s)$ to a truth assignment on all propositional formulas in a standard way, and associate with each propositional formula the set $[\![\phi]\!]_M = \{s \in \Omega : \pi(s)(\phi) = \textbf{true}\}$. An easy structural induction shows that $[\![\phi]\!]_M$ is a measurable set. If $M = (\Omega, \Sigma, \mathcal{P}, \pi)$, let

$$
\begin{array}{l}
M \models \theta_1 l(\phi_1) + \cdots + \theta_k l(\phi_k) \geq \alpha \text{ iff } \theta_1 \mathcal{P}^*([\![\phi_1]\!]_M) + \cdots + \theta_k \mathcal{P}^*([\![\phi_k]\!]_M) \geq \alpha \\
M \models \neg f \text{ iff } M \not\models f \\
M \models f \wedge g \text{ iff } M \models f \text{ and } M \models g.
\end{array}
$$

Note that the logic can express lower probabilities: it follows from the duality between upper and lower probabilities that $M \models -l(\neg\phi) \geq \beta - 1$ iff $\mathcal{P}_*([\![\neg\phi]\!]_M) \geq \beta$.[4]

Consider the following axiomatization $\mathbf{AX}^{up}$ for likelihood formulas, which we prove sound and complete in the next section. As in [7], $\mathbf{AX}^{up}$ is divided into three parts, dealing respectively with propositional reasoning, reasoning about linear inequalities, and reasoning about upper probabilities.

Propositional reasoning

Taut. All instances of propositional tautologies,

MP. From f and $f \implies g$ infer g.

Reasoning about linear inequalities

Ineq. All instances of valid formulas about linear inequalities (see below).

Reasoning about upper probabilities

L1. $l(\textit{false}) = 0$,

L2. $l(\textit{true}) = 1$,

L3. $l(\phi) \geq 0$,

L4. $l(\phi_1) + \cdots + l(\phi_m) - nl(\phi) \geq k$ if $\phi \Rightarrow \bigvee_{J \subseteq \{1,\ldots,m\}, |J|=k+n} \bigwedge_{j \in J} \phi_j$ and $\bigvee_{J \subseteq \{1,\ldots,m\}, |J|=k} \bigwedge_{j \in J} \phi_j$ are propositional tautologies.

L5. $l(\phi) = l(\psi)$ if $\phi \Leftrightarrow \psi$ is a propositional tautology.

The only difference between $\mathbf{AX}^{up}$ and the axiomatization for reasoning about probability given in [7] is that the axiom $l(\phi \wedge \psi) + l(\phi \wedge \neg\psi) = l(\phi)$ in [7], which expresses the additivity of probability, is replaced by **L4**. Although it may not be immediately obvious, **L4** is the logical analogue of **UP3**. To see this, first note that $\{\{A_1, \ldots, A_m\}\}$ covers A m times if and only if $A \subseteq \bigcup_{J \subseteq \{1,\ldots,m\}, |J|=n} \bigcap_{j \in J} A_j$. Thus, the formula $\phi \Rightarrow \bigvee_{J \subseteq \{1,\ldots,m\}, |J|=k+n} \bigwedge_{j \in J} \phi_j$ says that

[4]Another approach, more in keeping with [7], would be to interpret l as a lower probability measure. On the other hand, interpreting l as an upper probability measure is more in keeping with the literature on upper probabilities.

ϕ (more precisely, the set of worlds where ϕ is true) is covered $k+n$ times by $\{\{\phi_1, \dots, \phi_n\}\}$, while $\bigvee_{J \subseteq \{1,\dots,m\}, |J|=k} \bigwedge_{j \in J} \phi_j$ says that the whole space is covered k times by $\{\{\phi_1, \dots, \phi_n\}\}$; roughly speaking, is an (n, k)-cover of $([\![\phi]\!], [\![true]\!])$. The conclusion of **L4** thus corresponds to the conclusion of **UP3**.

Instances of **Taut** include all formulas of the form $f \vee \neg f$, where f is a likelihood formula. We could replace **Taut** by a simple collection of axioms that characterize propositional reasoning (see, for example, [11]), but we have chosen to focus on aspects of reasoning about upper probability.

As in [7], the axiom **Ineq** includes "all valid formulas about linear inequalities." Roughly speaking, an inequality formula is a formula of the form $a_1x_1 + \cdots + a_nx_n \geq c$, over variables $x_1, \dots, x_n$. The formula is said to be true if we satisfy the resulting inequality when we assign a real number to each variable of the formula. As usual, a formula is valid if it is true under every possible assignment of real numbers to variables. To get an instance of **Ineq**, we replace each variable x_i that occurs in a valid formula about linear inequalities by a primitive likelihood term of the form $l(\phi_i)$ (naturally each occurence of the variable x_i must be replaced by the same primitive likelihood term $l(\phi_i)$). As with **Taut**, we can replace **Ineq** by a sound and complete axiomatization for boolean combinations of linear inequalities. One such axiomatization is given in [7].

4 Soundness and completeness

A likelihood formula f is *provable from* F for F a set of formulas if it can be proven using the axioms and rules of inferences, along with the formulas in F. In the special case where F is empty, we say that f is simply *provable*. An axiom system is *sound* if every provable formula is valid. An axiom system is *complete* if every valid formula is provable.

Our goal is to prove that $\mathbf{AX}^{up}$ is a sound and complete axiomatization for reasoning about upper probability. The soundness of $\mathbf{AX}^{up}$ is immediate from our earlier disscussion. Completeness is, as usual, harder. Unfortunately, the standard technique for proving completeness in modal logic, which involves considering maximal consistent sets and canonical structures (see, for example, [12]) does not work. We briefly review the approach, just to point out the difficulties.

The standard approach uses the following definitions. A formula σ is *consistent* with an axiom system **AX** if $\neg\sigma$ is not provable from **AX**. A finite set of formulas $\{\sigma_1, \dots, \sigma_n\}$ is consistent with **AX** if the formula $\sigma_1 \wedge \cdots \wedge \sigma_n$ is consistent with **AX**; an infinite set of formulas is consistent with **AX** if all its finite subsets are consistent with **AX**. A *maximal* **AX**-consistent set of formulas F is a set of formulas consistent with **AX** with the property that for any formula $\sigma \notin F$, $F \cup \{\sigma\}$ is not consistent with **AX**. Using just axioms of proposiitonal logic, it is not hard to show that a **AX**-consistent set of formulas can be extended to a maximal **AX**-consistent set of formulas. To show that **AX** is a complete axiomatization with respect to some class of structures $\mathcal{M}$, we must show that every formula that is valid in every structure in $\mathcal{M}$ is provable in **AX**. To do this, it is sufficient to show that every **AX**-consistent formula σ is satisfiable in $\mathcal{M}$. Typically, this is done by constructing what is called a *canonical structure* M^c in $\mathcal{M}$ whose states are the maximal **AX**-consistent sets, and then showing that a formula σ is satisfied in a world w in M^c iff σ is one of the formulas in the canonical set associated with world w.

Unfortunately, this approach cannot be used to prove completeness here. To see this, consider the set of formulas:

$$F' = \{l(\phi) \leq \frac{1}{n},\ n = 1, 2, \dots\} \cup \{l(\phi) > 0\}.$$

This set is clearly $\mathbf{AX}^{up}$–consistent according to our definition, since every finite subset is satisfiable and $\mathbf{AX}^{up}$ is sound. It thus can be extended to a maximal $\mathbf{AX}^{up}$–consistent set F. However, the set F' of formulas is not satisfiable: it is not possible to assign $l(\phi)$ a value that will satisfy all the formulas at the same time. Hence, F is not satisfiable. Thus, the canonical model approach, at least applied naively, simply will not work.

We take a different approach here, similar to the one taken in [7]. Specifically, we show that if a formula f is $\mathbf{AX}^{up}$–consistent, then it is satisfiable in an upper probability structure. By a simple argument, we can easily reduce the problem to the case where f is a conjunction of basic likelihood formulas and negations of basic likelihood formulas. Let $p_1, \dots, p_N$ be the primitive propositions that appear in f. Observe that there are 2^{2^N} inequivalent propositional formulas over $p_1, \dots, p_N$. The argument goes as follow. Let an *atom over* $p_1, \dots, p_N$ be a formula of the form $q_1 \wedge \dots \wedge q_N$, where q_i is either p_i or $\neg p_i$. There are clearly 2^N atoms over $p_1, \dots, p_N$. Moreover, it is easy to see that any formula over $p_1, \dots, p_N$ can be written in a unique way as a disjunction of atoms. There are 2^{2^N} such disjunctions, so the claim follows.

Let $\rho_1, \dots, \rho_{2^{2^N}}$ be some canonical listing of the inequivalent formulas over $p_1, \dots, p_N$. Without loss of generality, we assume that ρ_1 is equivalent to *true*, and $\rho_{2^{2^N}}$ is equivalent to *false*. Since every propositional formula over $p_1, \dots, p_N$ is provably equivalent to some ρ, it follows that f is provably equivalent to a formula f' where each conjunct of f' is of the form $\theta_1 l(\rho_1) + \cdots + \theta_{2^{2^N}} l(\rho_{2^{2^N}}) \geq \beta$. Note that the negation of such a formula has the form $\theta_1 l(\rho_1) + \cdots + \theta_{2^{2^N}} l(\rho_{2^{2^N}}) < \beta$ or, equivalently, $(-\theta_1) l(\rho_1) + \cdots + (-\theta_{2^{2^N}}) l(\rho_{2^{2^N}}) > -\beta$. Thus, the formula f gives rise in a natural way to a system of inequalities of the form:

$$\begin{array}{rcl}
\theta_{1,1} l(\rho_1) + \cdots + \theta_{1,2^{2^N}} l(\rho_{2^{2^N}}) & \geq & \alpha_1 \\
\vdots & \vdots & \vdots \\
\theta_{r,1} l(\rho_1) + \cdots + \theta_{r,2^{2^N}} l(\rho_{2^{2^N}}) & \geq & \alpha_r \\
\theta'_{1,1} l(\rho_1) + \cdots + \theta'_{1,2^{2^N}} l(\rho_{2^{2^N}}) & > & \beta_1 \\
\vdots & \vdots & \vdots \\
\theta'_{s,1} l(\rho_1) + \cdots + \theta'_{s,2^{2^N}} l(\rho_{2^{2^N}}) & > & \beta_s.
\end{array} \quad (1)$$

We can express (1) as a conjunction of inequality formulas, by replacing each occurrence of $l(\rho_i)$ in (1) by x_i. Call this inequality formula $\overline{f}$.

If f is satisfiable in some upper probability structure M, then we can take x_i to be the upper probability of ρ_i in M; this gives a solution of $\overline{f}$. However, $\overline{f}$ may have a solution without f being satisfiable. For example, if f is the formula $l(p) = 1/2 \wedge l(\neg p) = 0$, then $\overline{f}$ has an obvious solution; f, however, is not satisfiable in an upper probability structure, because the upper probability of the set corresponding to p and the upper probability of the set corresponding to $\neg p$ must sum to at least 1 in all upper probability structures. Thus, we must add further constraints to the solution to force it to act like an upper probability.

UP1–**UP3** or, equivalently, the axioms **L1**–**L4**, describe exactly what additional constraints are needed. The constraint corresponding to **L1** (or **UP1**) is just $x_1 = 0$, since we have assumed ρ_1 is the formula *false*. Similarly, the constraint corresponding to **L2** is $x_{2^{2^N}} = 1$. The constraint corresponding to **L3** is $x_i \geq 0$, for $i = 1, \dots, 2^{2^N}$. What about **L4**? This seems to require an infinite collection of constraints, just as **UP3** does.[5]

[5]Although we are dealing with only finitely many formulas here, $\rho_1, \dots, \rho_{2^{2^N}}$, recall that the formulas $\phi_1, \dots, \phi_m$

This is where **UPF3** comes into play. It turns out that, if f is satisfiable at all, it is satisfiable in a structure with at most 2^N worlds, one for each atom over $p_1, \ldots, p_N$. Thus, we need to add only instances of **L4** where $k, m, n < B_{2^N}$ and $\phi_1, \ldots, \phi_m, \phi$ are all among $\rho_1, \ldots, \rho_{2^{2^N}}$. Although this is a large number of formulas (in fact, we do not know exactly how large, since it depends on B_{2^N}, which we have not computed), it suffices for our purposes that it is a finite number. For each of these instances of **L4**, there is an inequality of the form $a_1 x_1 + \cdots + a_{2^{2^N}} x_{2^{2^N}} \geq k$. Let $\hat{f}$, the *inequality formula corresponding to* f, be the conjunction consisting of $\overline{f}$, together with all the inequalities corresponding to the relevant instances of **L4**, and the equations and inequalitites $x_1 = 0$, $x_{2^{2^N}} = 1$, and $x_i \geq 0$ for $i = 1, \ldots, 2^{2^N}$, corresponding to axioms **L1**–**L3**.

Proposition 4.1: *The formula f is satisfiable in an upper probability structure iff the inequality formula $\hat{f}$ has a solution. Moreover, if $\hat{f}$ has a solution, then f is satisfiable in an upper probability structure with at most $2^{|f|}$ worlds.*

Proof: See Appendix E. ∎

Theorem 4.2: *The axiom system **AX**^{up} is sound and complete for upper probability structures.*

Proof: For soundness, it is easy to see that every axiom is valid for upper probability structures, including **L4**, which represents **UP3**.

For completeness, we proceed as in the discussion above. Assume that formula f is not satisfiable in an upper probability structure; we must show that f is **AX**^{up}–inconsistent. We first reduce f to a canonical form. Let $g_1 \vee \cdots \vee g_r$ be a disjunctive normal form expression for f (where each g_i is a conjunction of basic likelihood formulas and their negations). Using propositional reasoning, we can show that f is provably equivalent to this disjunction. Since f is unsatisfiable, each g_i must also be unsatisfiable. Thus, it is sufficient to show that any unsatisfiable conjunction of basic likelihood formulas and their negations is inconsistent. Assume that f is such a conjunction. Using propositional reasoning and axiom **L5**, f is equivalent to a likelihood formula f' that refers to formulas $\rho_1, \ldots, \rho_{2^{2^N}}$. Since f is unsatisfiable, so is f'. By Proposition 4.1, the the inequality formula $\hat{f}'$ corresponding to f' has no solution. Thus, by **Ineq**, the formula $\neg f''$ that results by replacing each instance of x_i in $\hat{f}'$ by $l(\rho_i)$ is **AX**^{up}–provable. All the conjuncts of f'' that are instances of axioms **L1**–**L4** are **AX**^{up}–provable. It follows that $\neg f'$ is **AX**^{up}–provable, and hence so is $\neg f$. ∎

5 Decision procedure

Having settled the issue of the soundness and completeness of the axiom system **AX**^{up}, we turn to the problem of the complexity of deciding satisfiability. Recall the problem of satisfiability: given a likelihood formula f, we want to determine if there exists an upper probability structure M such that $M \models f$. As we now show, the satisfiability problem is NP-complete, and thus no harder than satisfiability for propositional logic.

For the decision problem to make sense, we need to restrict our language slightly. If we allow real numbers as coefficients in likelihood formulas, we have to carefully discuss the issue of representation of such numbers. To avoid these complications, we restrict our language to allow

in **L4** need not be distinct, so there are potentially infinitely many instances of **L4** to deal with.

only integer coefficients. Note that we can still express rational coefficients by the standard trick of "clearing the denominator". For example, we can express $\frac{2}{3}l(\phi) \geq 1$ by $2l(\phi) \geq 3$ and $l(\phi) \geq \frac{2}{3}$ by $3l(\phi) \geq 2$. Define $||f||$ to be the length of the longest coefficient appearing in f, when written in binary. The size of a rational number $\frac{a}{b}$, denoted $||\frac{a}{b}||$, where a and b are relatively prime, is defined to be $||a|| + ||b||$.

A preliminary result required for the analysis of the decision procedure shows that if a formula is satisfied in some upper probability structure, it is satisfied in a structure $(\Omega, \Sigma, \mathcal{P}, \pi)$, which is "small" in terms of the number of states in Ω, the cardinality of the set $\mathcal{P}$ of probability measures, and the size of the coefficients in f.

Theorem 5.1: *Suppose f is a likelihood formula that is satisfied in some upper probability structure. Then f is satisfied in a structure $(\Omega, \Sigma, \mathcal{P}, \pi)$, where $|\Omega| \leq |f|^2$, $\Sigma = 2^\Omega$ (every subset set of Ω is measurable), $|\mathcal{P}| \leq |f|$, $\mu(w)$ is a rational number such that $||\mu(w)||$ is $O(|f|^2||f|| + |f|^2 \log(|f|))$ for every world $w \in \Omega$ and $\mu \in \mathcal{P}$, and $\pi(w)(p) =$ **false** for every world $w \in \Omega$ and every primitive proposition p not appearing in f.*

Proof: See Appendix F. ∎

Theorem 5.2: *The problem of deciding whether a likelihood formula is satisfiable in an upper probability structure is NP-complete.*

Proof: For the lower bound, it is clear that a given propositional formula ϕ is satisfiable iff the likelihood formula $l(\phi) > 0$ is satisfiable, therefore the satisfiability problem is NP-hard. For the upper bound, given a likelihood formula f, we guess a "small" satisfying structure $M = (\Omega, \Sigma, \mathcal{P}, \pi)$ for f of the form guaranteed to exist by Theorem 5.1. We verify that $M \models f$ as follows. Let $l(\psi)$ be an arbitrary likelihood term in f. We compute $[\![\psi]\!]_M$ by checking the truth assignment of each $s \in \Omega$ and seeing whether this truth assignment makes ψ true. We then replace each occurence of $l(\psi)$ in f by $\max_{\mu \in \mathcal{P}}\{\sum_{s \in S_\psi} \mu(s)\}$ and verify that the resulting expression is true. ∎

6 Conclusion

We have considered a logic with the same syntax as the logic for reasoning about probability, inner measures, and belief presented in [7], with uncertainty interpreted as the upper probability of a set of probability measures. Under this interpretation, we have provided a sound and complete axiomatization for the logic. We further showed that the satisfiability problem is NP-complete (as it is for reasoning about probability, inner measures, and beliefs [7]), despite having to deal with probability structures with possibility infinitely many states and infinite sets of probability measures. The key step in the axiomatization involves finding a characterization of upper probability measures that can be captured in the logic. The key step in the complexity result involves showing that if a formula was satisfiable at all, it is satisfiable in a "small" structure, where the size of the state space as well as the size of the set of probability measures and the size of all probabilities involved, is polynomial in the length of the formula.

Given the similarity in spirit of the results for the various interpretations of the uncertainty operator (as a probability, inner measure, belief function, and upper probability), we conjecture that there is some underlying result from which all these results should follow. It would be interesting to make that precise.

Acknowledgments. Thanks to Dexter Kozen, Jon Kleinberg, and Hubie Chen for discussions concerning set covers. Vicky Weissman read a draft of this paper and provided numerous helpful comments.

A Proof of Proposition 2.1

Proposition 2.1: *Property (6) implies properties (1)-(5).*

Proof: We introduce the following auxiliary properties to help derive the implications:

(7) $\mathcal{P}_*(A) + \mathcal{P}_*(B) \leq \mathcal{P}_*(A \cup B) + \mathcal{P}^*(A \cap B)$

(8) $\mathcal{P}_*(A) + \mathcal{P}_*(B) \leq \mathcal{P}_*(A \cap B) + \mathcal{P}^*(A \cup B)$

(9) $\mathcal{P}_*(A \cup B) + \mathcal{P}_*(A \cap B) \leq \mathcal{P}_*(A) + \mathcal{P}^*(B)$

(10) If $A \cap B = \emptyset$, then

$$\mathcal{P}_*(A) + \mathcal{P}_*(B) \leq \mathcal{P}_*(A \cup B) \leq \mathcal{P}_*(A) + \mathcal{P}^*(B) \leq \mathcal{P}^*(A \cup B) \leq \mathcal{P}^*(A) + \mathcal{P}^*(B).$$

Using these properties, we show the following chain of implications:

$$(6) \implies (10) \qquad \begin{array}{l} (10) \implies (9) \implies (3) \\ (10) \implies (7) \implies (4) \\ (10) \implies (8) \implies (5) \end{array} \qquad (4), (5) \implies (1), (2)$$

The implication (4), (5) $\implies$ (1), (2) follows easily by mutual induction on n. The base case is the following instances of properties (4) and (5): $\mathcal{P}_*(A \cup B) \geq \mathcal{P}_*(A) + \mathcal{P}_*(B) - \mathcal{P}^*(A \cap B)$ and $\mathcal{P}^*(A \cup B) \leq \mathcal{P}^*(A) + \mathcal{P}^*(B) - \mathcal{P}_*(A \cap B)$. The details are left to the reader.

We now prove the remaining implications.

(9) $\implies$ (3): Since (9) is already one of the inequalities in (3), it remains to show that it implies the other inequality in (3), that is, $\mathcal{P}_*(A) + \mathcal{P}^*(B) \leq \mathcal{P}^*(A \cup B) + \mathcal{P}^*(A \cap B)$.

$$\begin{array}{rcl} \mathcal{P}^*(A \cup B) + \mathcal{P}^*(A \cap B) & = & 1 - \mathcal{P}_*(\overline{A \cup B}) + 1 - \mathcal{P}_*(\overline{A \cap B}) \\ & = & 1 - \mathcal{P}_*(\overline{A} \cap \overline{B}) + 1 - \mathcal{P}_*(\overline{A} \cup \overline{B}) \\ & = & 2 - (\mathcal{P}_*(\overline{A} \cap \overline{B}) + \mathcal{P}_*(\overline{A} \cup \overline{B})) \\ & = & 2 - (\mathcal{P}_*(\overline{B} \cap \overline{A}) + \mathcal{P}_*(\overline{B} \cup \overline{A})) \\ & \geq & 2 - (\mathcal{P}_*(\overline{B}) + \mathcal{P}^*(\overline{A})) \\ & = & 1 - \mathcal{P}_*(\overline{B}) + 1 - \mathcal{P}^*(\overline{A}) \\ & = & \mathcal{P}^*(B) + \mathcal{P}_*(A). \end{array}$$

(7) $\implies$ (4): Since (7) is already one of the inequalities in (4), it remains to show that it implies the other inequality in (4), that is, $\mathcal{P}_*(A \cup B) + \mathcal{P}^*(A \cap B) \le \mathcal{P}^*(A) + \mathcal{P}^*(B)$.

$$\begin{aligned}
\mathcal{P}^*(A) + \mathcal{P}^*(B) &= 1 - \mathcal{P}_*(\overline{A}) + 1 - \mathcal{P}_*(\overline{B}) \\
&= 2 - (\mathcal{P}_*(\overline{A}) + \mathcal{P}_*(\overline{B})) \\
&\ge 2 - (\mathcal{P}_*(\overline{A} \cup \overline{B}) + \mathcal{P}^*(\overline{A} \cap \overline{B})) \\
&= 1 - \mathcal{P}_*(\overline{A} \cup \overline{B}) + 1 - \mathcal{P}^*(\overline{A} \cap \overline{B}) \\
&= 1 - \mathcal{P}_*(\overline{A \cap B}) + 1 - \mathcal{P}^*(\overline{A \cup B}) \\
&= \mathcal{P}^*(A \cap B) + \mathcal{P}_*(A \cup B).
\end{aligned}$$

(8) $\implies$ (5): Since (8) is already one of the inequalities in (5), it remains to show that it implies the other inequality in (5), that is, $\mathcal{P}_*(A \cap B) + \mathcal{P}^*(A \cup B) \le \mathcal{P}^*(A) + \mathcal{P}^*(B)$.

$$\begin{aligned}
\mathcal{P}^*(A) + \mathcal{P}^*(B) &= 1 - \mathcal{P}_*(\overline{A}) + 1 - \mathcal{P}_*(\overline{B}) \\
&= 2 - (\mathcal{P}_*(\overline{A}) + \mathcal{P}_*(\overline{B})) \\
&\ge 2 - (\mathcal{P}_*(\overline{A} \cap \overline{B}) + \mathcal{P}^*(\overline{A} \cup \overline{B})) \\
&= 1 - \mathcal{P}_*(\overline{A} \cap \overline{B}) + 1 - \mathcal{P}^*(\overline{A} \cup \overline{B}) \\
&= 1 - \mathcal{P}_*(\overline{A \cup B}) + 1 - \mathcal{P}^*(\overline{A \cap B}) \\
&= \mathcal{P}^*(A \cup B) + \mathcal{P}_*(A \cap B).
\end{aligned}$$

For the next implications, given A, B, let $Z = A \cap B$:

(10) $\implies$ (9):

$$\begin{aligned}
\mathcal{P}_*(A \cup B) &= \mathcal{P}_*((A - Z) \cup B) \\
&\le \mathcal{P}_*(A - Z) + \mathcal{P}^*(B) \quad \text{[since } (A - Z) \cap B = \emptyset\text{]} \\
&\le \mathcal{P}_*((A - Z) \cup Z) - \mathcal{P}_*(Z) + \mathcal{P}^*(B) \\
&= \mathcal{P}_*(A) + \mathcal{P}^*(B) - \mathcal{P}_*(A \cap B).
\end{aligned}$$

(10) $\implies$ (7):

$$\begin{aligned}
\mathcal{P}_*(A \cup B) &= \mathcal{P}_*((A - Z) \cup B) \\
&\ge \mathcal{P}_*(A - Z) + \mathcal{P}_*(B) \\
&\ge \mathcal{P}_*((A - Z) \cup Z) - \mathcal{P}^*(Z) + \mathcal{P}_*(B) \\
&= \mathcal{P}_*(A) + \mathcal{P}_*(B) - \mathcal{P}^*(A \cap B).
\end{aligned}$$

(10) $\implies$ (8):

$$\begin{aligned}
\mathcal{P}^*(A \cup B) &= \mathcal{P}^*((A - Z) \cup B) \\
&\ge \mathcal{P}^*(A - Z) + \mathcal{P}_*(B) \\
&\ge \mathcal{P}_*((A - Z) \cup Z) - \mathcal{P}_*(Z) + \mathcal{P}_*(B) \\
&= \mathcal{P}_*(A) + \mathcal{P}_*(B) - \mathcal{P}_*(A \cap B).
\end{aligned}$$

(6) $\implies$ (10): Again, since (6) already comprises two of the inequalities in (10), it remains to show that it implies the other two, that is, if $A \cap B = \emptyset$, then

$$\mathcal{P}_*(A) + \mathcal{P}_*(B) \leq \mathcal{P}_*(A \cup B) \leq \mathcal{P}^*(A) + \mathcal{P}_*(B).$$

First, we show that $\mathcal{P}_*(A) + \mathcal{P}_*(B) \leq \mathcal{P}_*(A \cup B)$. Using (6), we know that

$$\mathcal{P}^*(\overline{A} \cap \overline{B}) + \mathcal{P}_*(A) \leq \mathcal{P}^*((\overline{A} \cap \overline{B}) \cup A) = \mathcal{P}^*(\overline{B}).$$

In other words, $\mathcal{P}^*(\overline{A} \cap \overline{B}) \leq \mathcal{P}^*(\overline{B}) + \mathcal{P}_*(A)$. From this, we derive:

$$\begin{array}{rcl}
\mathcal{P}_*(A \cup B) & = & 1 - \mathcal{P}^*(\overline{A \cup B}) \\
& = & 1 - \mathcal{P}^*(\overline{A} \cap \overline{B}) \\
& \geq & 1 - (\mathcal{P}^*(\overline{B}) - \mathcal{P}_*(A)) \\
& = & 1 - \mathcal{P}^*(\overline{B}) + \mathcal{P}_*(A) \\
& = & \mathcal{P}_*(B) + \mathcal{P}_*(A).
\end{array}$$

Second, we show that $\mathcal{P}_*(A \cup B) \leq \mathcal{P}^*(A) + \mathcal{P}_*(B)$. Using (6), we know that

$$\mathcal{P}^*(\overline{A} \cap \overline{B}) + \mathcal{P}^*(A) \geq \mathcal{P}^*((\overline{A} \cap \overline{B}) \cup A) = \mathcal{P}^*(\overline{B}).$$

(The last equality follows from the fact that $(\overline{A} \cap \overline{B}) \cup A = \overline{B}$ when $A \cap B = \emptyset$.) In other words, $\mathcal{P}^*(\overline{A} \cap \overline{B}) \geq \mathcal{P}^*(\overline{B}) - \mathcal{P}^*(A)$. From this, we derive:

$$\begin{array}{rcl}
\mathcal{P}_*(A \cup B) & = & 1 - \mathcal{P}^*(\overline{A \cup B}) \\
& = & 1 - \mathcal{P}^*(\overline{A} \cap \overline{B}) \\
& \leq & 1 - (\mathcal{P}^*(\overline{B}) - \mathcal{P}^*(A)) \\
& = & 1 - \mathcal{P}^*(\overline{B}) + \mathcal{P}^*(A) \\
& = & \mathcal{P}_*(B) + \mathcal{P}^*(A). \quad \blacksquare
\end{array}$$

B Proof of Proposition 2.2

Proposition 2.2: *For $0 < \epsilon < \frac{1}{8}$, the function υ_ϵ satisfies property (6), but is not an upper probability measure. That is, we cannot find a set $\mathcal{P}'$ of probability measures such that $\upsilon_\epsilon = (\mathcal{P}')^*$.*

Proof: We are given $0 < \epsilon < \frac{1}{8}$. It is easy to check mechanically that υ_ϵ satisfies (6).

We now show that there is no set $\mathcal{P}'$ such that $\upsilon_\epsilon = (\mathcal{P}')^*$. By way of contradiction, assume there is such a $\mathcal{P}'$. By the properties of sup, this means that there is a $\mu \in \mathcal{P}'$ such that $\mu(\{a,b,c\}) > \frac{3}{4}$, since $\upsilon_\epsilon(\{a,b,c\}) = \frac{3}{4}+\epsilon > \frac{3}{4}$. Consider this μ in detail. Since $\mu \in \mathcal{P}$, we must have for all $X \in \Sigma$, $X \neq \{a,b,c\}$, that $\mu(X) \leq (\mathcal{P}')^*(X) = \mathcal{P}^*(X)$. In particular, $\mu(\{a,b\}), \mu(\{b,c\}), \mu(\{a,c\}) \leq \frac{1}{2}$. Therefore,

$$\mu(\{a,b\}) + \mu(\{b,c\}) + \mu(\{a,c\}) \leq \frac{3}{2}. \tag{2}$$

However, from standard properties of probability, it follows that

$$\mu(\{a,b\}) + \mu(\{b,c\}) + \mu(\{a,c\}) = 2\mu(\{a,b,c\}) > 2 \times \frac{3}{4} = \frac{3}{2},$$

which contradicts (2). Therefore, μ, and therefore $\mathcal{P}'$ cannot exist, and υ_ϵ is not an upper probability measure. $\blacksquare$

C Proof of Theorem 2.3

To make this paper self-contained, in this appendix we give a proof of Theorem 2.3, the characterization of upper probability measures due to Anger and Lembcke from [1].

Theorem 2.3: *Let Ω be a set, Σ an algebra of subsets of Ω, and υ a function $\upsilon : \Sigma \to [0,1]$. There exists a set $\mathcal{P}$ of probability measures with $\upsilon = \mathcal{P}^*$ if and only if υ satisfies the following three properties:*

UP1. $\upsilon(\emptyset) = 0$,

UP2. $\upsilon(\Omega) = 1$,

UP3. *for all integers m, n, k and all subsets $A_1, \ldots, A_m$ in Σ, if $\{\{A_1, \ldots, A_m\}\}$ is an (n,k)-cover of (A, Ω), then $k + n\upsilon(A) \leq \sum_{i=1}^{m} \upsilon(A_i)$.*

Proof: The "if" direction of the characterization is straightforward. Given $\mathcal{P} = \{\mu_i\}_{i \in I}$ a set of probability measures, we show $\mathcal{P}^*$ satisfies **UP1-UP3**.

UP1: $\mathcal{P}^*(\emptyset) = \sup\{\mu_i(\emptyset)\} = \sup\{0\} = 0$

UP2: $\mathcal{P}^*(\Omega) = \sup\{\mu_i(\Omega)\} = \sup\{1\} = 1$

UP3: Given $A_1, \ldots, A_m$ and A such that $A \subseteq \bigcup_{J \subseteq \{1,\ldots,m\}, |J|=k+n} \bigcap_{j \in J} A_{i_j}$ and $\Omega \subseteq \bigcup_{J \subseteq \{1,\ldots,m\}, |J|=k} \bigcap_{j \in J} A_{i_j}$, then for any i we have $k\mu_i(\Omega) + n\mu_i(A) \leq \sum_{j=1}^{m} \mu_i(A_j)$, that is $k + n\mu_i(A) \leq \sum_{j=1}^{m} \mu_i(A_j) \leq \sup_i\{\sum_{j=1}^{m} \mu_i(A_j)\} \leq \sum_{j=1}^{m} \sup_i\{\mu_i(A_j)\} = \sum_{j=1}^{m} \mathcal{P}^*(A_j)$. But $\sup_i\{k + n\mu_i(A)\} = k + n\sup_i\{\mu_i(A)\} = k + n\mathcal{P}^*(A)$, so $k + n\mathcal{P}^*(A) \leq \sum_{j=1}^{m} \mathcal{P}^*(A_j)$, as required.

As for the "only if" direction, we first prove a general lemma relating the problem to the Hahn-Banach Theorem. Some general definitions are needed. Suppose that we are given a space W and an algebra $\mathcal{F}$ of subsets of W. Let $\mathcal{K}$ be the vector space generated by the indicator functions 1_X defined by

$$1_X(x) = \begin{cases} 0 & \text{if } x \notin X \\ 1 & \text{if } x \in X, \end{cases}$$

for $X \in \mathcal{F}$. A *sublinear functional on* $\mathcal{K}$ is a mapping $c : \mathcal{K} \to \mathbb{R}$ such that $c(\alpha h) = \alpha c(h)$ for $\alpha \geq 0$ and $c(h_1 + h_2) \leq c(h_1) + c(h_2)$ for all h_1, h_2. A sublinear functional is *increasing* if $h \geq 0$ implies $c(h + h') \geq c(h')$ for all $h' \in \mathcal{K}$. The following result is a formulation of the well-known Hahn-Banach Theorem (see, for example, [3]).

Theorem (Hahn-Banach): *Let $\mathcal{K}$ be a vector space over $\mathbb{R}$, and let g be a sublinear functional on $\mathcal{K}$. If $\mathcal{M}$ is a linear subspace in $\mathcal{K}$ and $\lambda : \mathcal{M} \to \mathbb{R}$ is a linear functional such that $\lambda(x) \leq g(x)$ for all x in $\mathcal{M}$, then there is a linear functional $\lambda' : \mathcal{K} \to \mathbb{R}$ such that $\lambda'|_{\mathcal{M}} = \lambda$ and $\lambda'(x) \leq g(x)$ for all x in $\mathcal{K}$.*

Lemma C.1: *Let $g : \mathcal{F} \to [0,1]$ be such that $g(W) = 1$ and suppose that there is an increasing sublinear functional $\tilde{g}$ on $\mathcal{K}$ such that*

1. *$\tilde{g}(1_K) = g(K)$ for $K \in \mathcal{F}$;*

2. $\tilde{g}(h) \leq 0$ *if* $h \leq 0$;

3. $\tilde{g}(-1) \leq -1$ *(where* $\tilde{g}(\alpha)$ *is identified with* $\tilde{g}(\alpha 1_W)$*).*

Then g *is an upper probability measure.*

Proof: We show that g is an upper probability by exhibiting a set $\{\mu_X : X \in \Sigma\}$ of probability measures, with the property that $\mu_X(X) = g(X)$ and $\mu_X(Y) \leq g(X)$ for $Y \neq X$. Each probability measure μ_X is constructed through an application of the Hahn-Banach Theorem.

Given $X \in \mathcal{F}$, define the linear functional λ on the subspace generated by 1_X by $\lambda(\alpha 1_X) = \alpha\tilde{g}(1_X)$. We claim that $\lambda(h) \leq \tilde{g}(h)$ for all h in the subspace. Since the elements of the subspace have the form $\alpha 1_X$, there are two cases to consider: $\alpha \geq 0$ and $\alpha < 0$. If $\alpha \geq 0$, then $\lambda(\alpha 1_X) = \alpha\tilde{g}(1_X) = \tilde{g}(\alpha 1_X)$, since $\tilde{g}$ is sublinear. Moreover, $0 = \tilde{g}(0) = \tilde{g}(-1_X + 1_X) \leq \tilde{g}(-1_X) + \tilde{g}(1_X)$, so $\tilde{g}(-1_X) \geq -\tilde{g}(1_X)$. Thus, if $\alpha > 0$, then

$$\lambda(-\alpha 1_X) = -\alpha\tilde{g}(1_X) \leq \alpha\tilde{g}(-1_X) = \tilde{g}(-\alpha 1_X).$$

Now, by the Hahn-Banach Theorem, we can extend λ to a linear functional λ' on all of $\mathcal{K}$ such that $\lambda'(h) \leq \tilde{g}(h)$ for all h. We claim that (a) $\lambda'(1_Y) \geq 0$ for all $Y \in \mathcal{K}$ and (b) $\lambda'(1) = 1$. For (a), note that $\lambda'(-1_Y) \leq \tilde{g}(-1_Y) \leq 0$ by assumption, so $\lambda'(1_Y) \geq 0$. For (b), note that $\lambda'(1) \leq \tilde{g}(1) = g(W) = 1$ and that $\lambda'(1) = -\lambda'(-1) \geq -\tilde{g}(-1) \geq 1$ (since $\tilde{g}(-1) \leq -1$, by assumption).

Define $\mu_X(Y) = \lambda'(1_Y)$. Since $\lambda'(1_W) = 1$, $\mu_X(W) = 1$. If Y and Y' are disjoint, it is immediate from the linearity of λ that $\mu_X(Y \cup Y') = \mu_X(Y) + \mu_X(Y')$. By construction, $\mu_X(Y) \leq \tilde{g}(1_Y) = g(Y)$ for any $Y \neq X$, and $\mu_X(X) = \lambda(1_X) = \tilde{g}(1_X) = g(X)$. Bottom line: there is a probability measure μ_X dominated by g such that $\mu_X(X) = g(X)$.

Take $\mathcal{P} = \{\mu_X : X \in \Sigma\}$. Since for any X we have that $\mu_X(X) = g(X)$ and $\mu_X(Y) \leq g(X)$ (if $Y \neq X$), we have $\mathcal{P}^*(X) = \mu_X(X) = g(X)$. Therefore, $g = \mathcal{P}^*$. ∎

The main result follows by showing how to construct, from a function υ satisfying the properties of Theorem 2.3, a sublinear functional c on $\mathcal{K}$ with the required properties.

Suppose that $g : \Sigma \to [0,1]$ is a function satisfying **UP1-UP3**. Since g satisfies **UP3**, if $\{\{K_1, \ldots, K_m\}\}$ is an (n,k)-cover of (K, Ω), we have $k + ng(K) \leq \sum_{i=1}^m K_i$. This is equivalent to saying that $k + n1_K \leq \sum_{i=1}^m 1_{K_i}$. Hence, for all $K_1, \ldots, K_m$ such that $k + n1_K \leq \sum_{i=1}^m 1_{K_i}$, we have $k + ng(K) \leq \sum_{i=1}^m g(K_i)$, or equivalently

$$-\frac{k}{n} + \frac{1}{n}\sum_{i=1}^m g(K_i) \geq g(K). \tag{3}$$

This observation motivates the following definition of the functional $\tilde{g} : \mathcal{K} \to \mathbb{R} \cup \{-\infty, \infty\}$:

$$\tilde{g}(h) = \inf\left\{-\frac{k}{n} + \frac{1}{n}\sum_{i=1}^m g(K_i) : m,n,k \in \mathbb{N},\ m,n > 0,\ K_1, \ldots, K_m \in \mathcal{F},\ -\frac{k}{n} + \frac{1}{n}\sum_{i=1}^m 1_{K_i} \geq h\right\}.$$

Our goal now is to show that $\tilde{g}$ satisfies the conditions of Lemma C.1.

- It is almost immediate from the definitions that $\tilde{g}$ is increasing: if $h \geq 0$ and $-\frac{k}{n} + \frac{1}{n}\sum_{i=1}^m 1_{K_i} \geq h + h'$, then $-\frac{k}{n} + \frac{1}{n}\sum_{i=1}^m 1_{K_i} \geq h'$.

- To see that $\tilde{g}$ is sublinear, note that it is easy to see using the properties of inf that $\tilde{g}(h_1+h_2) \leq \tilde{g}(h_1) + \tilde{g}(h_2)$. To show that $\tilde{g}(\alpha h) = \alpha\tilde{g}(h)$ for $\alpha \geq 0$, first observe that the definition of $\tilde{g}$ is equivalent to

$$\inf\left\{-\beta + \sum_{i=1}^{m} \beta_i g(K_i) : m \in \mathbb{N}, \beta, \beta_i \in \mathbb{R}_+, K_1, \ldots, K_m \in \mathcal{F} \; -\beta + \sum_{i=1}^{m} \beta_i 1_{K_i} \geq h\right\}.$$

  Consider first the case $\alpha > 0$. Then

$$\begin{aligned}
\tilde{g}(\alpha h) &= \inf\left\{-\beta + \sum_{i=1}^{m} \beta_i g(K_i) \;:\; -\beta + \sum_{i=1}^{m} \beta_i 1_{K_i} \geq \alpha h\right\} \\
&= \inf\left\{-\beta + \sum_{i=1}^{m} \beta_i g(K_i) \;:\; -\frac{\beta}{\alpha} + \frac{1}{\alpha}\sum_{i=1}^{m} \beta_i 1_{K_i} \geq h\right\} \\
&= \alpha \inf\left\{-\frac{\beta}{\alpha} + \frac{1}{\alpha}\sum_{i=1}^{m} \beta_i g(K_i) \;:\; -\frac{\beta}{\alpha} + \frac{1}{\alpha}\sum_{i=1}^{m} \beta_i 1_{K_i} \geq h\right\} \\
&= \alpha\tilde{g}(h).
\end{aligned}$$

  For $\alpha = 0$, it is clear from the definition of $\tilde{g}$ that $\tilde{g}(1_\emptyset) \leq g(\emptyset)$. From (3) it follows that $\tilde{g}(1_\emptyset) \geq g(\emptyset)$, and hence $\tilde{g}(0) = \tilde{g}(1_\emptyset) = g(\emptyset) = 0$.

- It is immediate from the definition of $\tilde{g}$ that $\tilde{g}(1_K) \leq g(K)$ for $K \in \mathcal{F}$; the fact that $\tilde{g}(1_K) = g(K)$ now follows from (3).

- It is immediate from the definition that $\tilde{g}(-1) \leq -1$.

- If $h \leq 0$, then $-h \geq 0$; since $\tilde{g}$ is increasing, $\tilde{g}(h) \leq \tilde{g}(-h+h) = \tilde{g}(0)$, and since $\tilde{g}$ is sublinear, $\tilde{g}(0) = 0$.

Since the conditions of Lemma C.1 are satisfied, g is an upper probability measure. ∎

D Proof of Theorem 2.4

Theorem 2.4: *There exists constants $B_0, B_1, \ldots$ such that if Σ is an algebra of subsets of Ω and υ is a function $\upsilon : \Sigma \to [0,1]$, then there exists a set $\mathcal{P}$ of probability measures such that $\upsilon = \mathcal{P}^*$ if and only if υ satisfies the following properties:*

UPF1. $\upsilon(\emptyset) = 0$,

UPF2. $\upsilon(\Omega) = 1$,

UPF3. *for all integers $m, n, k \leq B_{|\Omega|}$ and all sets $A_1, \ldots, A_m$, if $\{\{A_1, \ldots, A_m\}\}$ is an (n,k)-cover of (A, Ω), then $k + n\upsilon(A) \leq \sum_{i=1}^{m} \upsilon(A_i)$.*

Proof: In view of Theorem 2.3, we need only show that there exist constant $B_0, B_1, \ldots$ such that a function υ satisfies **UP3** iff it satisfies **UPF3**. Clearly, **UP3** always implies **UPF3**, so it is sufficient to show that there exists $B_0, B_1, \ldots$ such that **UPF3** implies **UP3**.

We need some terminology before proceeding. An *exact* (n, k)*-cover* of (A, Ω) is a cover C of A with the property that every element of A appears in exactly $n+k$ sets in C, and every element of $\Omega - A$ appears in exactly k sets in C. Thus, while an (n, k)-cover of (A, Ω) can have many extra sets, as long as the sets cover A at least $n + k$ times and Ω k times, an exact cover has only the necessary sets, with the right total number of elements. An exact (n, k)-cover C of (A, Ω) is *decomposable* if there exists an exact (n_1, k_1)-cover C_1 and an exact (n_2, k_2)-cover C_2 of (A, Ω) such that C_1 and C_2 form a nontrivial partition of C, with $n = n_1 + n_2$ and $k = k_1 + k_2$. Intuitively, an exact cover C is decomposable if it can be split into two exact covers. It follows easily by induction that for any exact (n, k)-cover, there exists a (not necessarily unique) finite set of nondecomposable exact covers $C_1, \ldots, C_m$, with C_i an exact (n_i, k_i)-cover, such that the C_i's a nontrivial partition of C with $n = \sum_{i=1}^{m} n_i$ and $k = \sum_{i=1}^{m}$. (If C is itself nondecomposable, we can take $m = 1$ and $C_1 = C$.) One can easily verify that if C is an exact (n, k)-cover of (A, Ω) and $C' \subseteq C$ is an exact (n', k')-cover of (A, Ω) with $n' + k' < n + k$, then C is decomposable.

The following lemma highlights the most important property of exact covers from our perspective. It says that for any set $A \in \Sigma$, there cannot be a "large" nondecomposable exact cover of (A, Ω).

Lemma D.1: *There exists a sequence $B'_1, B'_2, B'_3, \ldots$ such that for all $A \subseteq \Omega$, every exact (n, k)-cover of (A, Ω) with $n > B'_{|\Omega|}$ or $k > B'_{|\Omega|}$ is decomposable.*

Proof: Given $A \subseteq \Omega$, we first show that there exists N_A such that if $n > N_A$ or $k > N_A$, every exact (n, k)-cover of (A, Ω) is decomposable. Suppose for the sake of contradiction that this is not the case. This means that we can find an infinite sequence $C_1, C_2, \ldots$ such that C_i is a nondecomposable exact (n_i, k_i)-cover of (A, Ω), with either $n_1 < n_2 < \ldots$ or $k_1 < k_2 < \ldots$.

To derive a contradiction, we use the following lemma (known as Dickson's Lemma [5]).

> **Lemma D.2:** *Every infinite sequence of d-dimensional vectors over the natural numbers contains a monotonically nondecreasing subsequence in the pointwise ordering (where $x \leq y$ in the pointwise ordering iff $x_i \leq y_i$ for all i).*
>
> **Proof:** It is straightforward to prove by induction on k that if $k \leq d$, then every infinite sequence of vectors $x^1, x^2, \ldots$ contains a subsequence $x^{i_1}, x^{i_2}, \ldots$ such that $x_j^{i_1}, x_j^{i_2}, \ldots$ is a nondecreasing sequence of natural numbers for all $j \leq k$. The base case is immediate from the observation that every infinite sequence of natural numbers contains a nondecreasing subsequence. For the inductive step, observe that if $x^{i_1}, x^{i_2}, \ldots$ is a subsequence such that $x_j^{i_1}, x_j^{i_2}, \ldots$ is a nondecreasing sequence of natural numbers for all $j \leq k$, then the sequence $x_{k+1}^{i_1}, x_{k+1}^{i_2}, \ldots$ of natural numbers must have a nondecreasing subsequence. This determines a subsequence of the original sequence with the appropriate property for all $j \leq k + 1$. ∎

Let $S_1, \ldots, S_{2^{|\Omega|}}$ be an arbitrary ordering of the $2^{|\Omega|}$ subsets of Ω. We can associate to any cover C a $2^{|\Omega|}$-dimensional vector $x^C = (x_1^C, \ldots, x_{2^{|\Omega|}}^C)$, where x_i^C is the number of times the subset S_i of Ω appears in the multiset C. The key property of this association is that if C' and C are multisets, then $C' \subseteq C$ iff $x^{C'} \leq x^C$ in the pointwise ordering.

Consider the sequence of vectors $x^{C_1}, x^{C_2}, \ldots$ associated with the sequence $C_1, C_2, \ldots$ of nondecomposable exact covers of (A, Ω). By Lemma D.2, there is a nondecreasing subsequence of vectors, $x^{C_{i_1}} \leq x^{C_{i_2}} \leq \cdots$. But this means that $C_{i_1} \subseteq C_{i_2} \subseteq \cdots$. Since $n_1 < n_2 < \ldots$ or $k_1 < k_2 < \ldots$, every cover in the chain must be distinct. But any pair of exact covers in the chain is such that $C_i \subseteq C_{i+1}$, meaning C_{i+1} is decomposable, contradicting our assumption. Therefore, there must exist an N_A such that any exact (n,k)-cover of A with $n > N_A$ or $k > N_A$ is decomposable.

Now define $B'_N = \max\{N_A : A \subseteq \{1, \ldots, N\}\}$. It is easy to see that this choice works. ∎

Let $B_N = 2NB'_N$, where B'_N is as in Lemma D.1. We now show that **UPF3** implies **UP3** with this choice of B_N. Assume that **UPF3** holds. Fix Ω. Suppose that $C = \{\{A_1, \ldots, A_m\}\}$ is an (n,k)-cover of (A,Ω) with $|C| = m$. We want to show that $k + n\upsilon(A) \leq \sum_{i=1}^{m} \upsilon(A_i)$. We proceed as follows.

The first step is to show that, without loss of generality, C is an exact (n,k)-cover of (A,Ω). Let B_i consist of those states $s \in A_i$ such that either $s \in A$ and s appears in more than $n+k$ sets in $A_1, \ldots, A_{i-1}$ or $s \in \Omega - A$ and s appears in more than k sets in $A_1, \ldots, A_{i-1}$. Let $A'_i = A_i - B_i$. Let $C' = \{\{A'_1, \ldots, A'_m\}\}$. It is easy to check that C' is an exact (n,k)-cover of (A,Ω). For if $s \in A$, then s appears in exactly $n+k$ sets in C' (it appears in A'_j iff A_j is among the first $n+k$ sets in C in which s appeared) and, similarly, if $s \in \Omega - A$, then s appears in exactly k sets in C'. Clearly if **UP3** holds for C', then it holds for C, since $\upsilon(A'_i) \leq \upsilon(A_i)$ for $i = 1, \ldots, m$. Thus, we can assume without loss of generality that C is an exact (n,k)-cover of A.

We can also assume without loss of generality that no set in C is empty (otherwise, we can simply remove the empty sets in C; the resulting set is still an (n,k)-cover of (A,Ω)). There are now two cases to consider. If $\max(m,n,k) \leq B_{|\Omega|}$, the desired result follows from **UPF3**. If not, consider a decomposition of C into multisets $C_1, \ldots, C_p$, where C_h is an exact (n_h,k_h)-cover of (A,Ω) and is not further decomposable. We claim that $\max(|C_h|, n_h, k_h) \leq B_{|\Omega|}$ for $h = 1, \ldots, p$. If $n_h > B_{|\Omega|}$ or $k_h > B_{|\Omega|}$, then it is immediate from Lemma D.1 that C_h can be further decomposed, contradicting the fact that C_h is not decomposable. And if $|C_h| > B_{|\Omega|}$, then observe that $\sum_{X \in C_h} |X| \geq |C_h|$. Since $|C_h| > B_{|\Omega|} = 2|\Omega|B'_{|\Omega|}$, there must be some $s \in \Omega$ which appears in at least $2B'_{|\Omega|}$ sets in C_h. Since C_h is an exact (n_h,k_h)-cover, it follows that either $n_h > B'_{|\Omega|}$ or $k_h > B'_{|\Omega|}$. But then, by Lemma D.1, C_h is decomposable, again a contradiction.

Now we can apply **UPF3** to each of $C_1, \ldots, C_k$ to get

$$\sum_{X \in C_h} \upsilon(X) - n_h\upsilon(A) \geq k_h.$$

Since the C_h's form a decomposition of C, we have

$$\begin{aligned} & \sum_{h=1}^{p} \left(\sum_{X \in C_h} \upsilon(X) - n_h\upsilon(X) \right) \geq \sum_{h=1}^{p} k_h \\ \Rightarrow\ & \sum_{h=1}^{p} \left(\sum_{X \in C_h} \upsilon(X) \right) - \sum_{h=1}^{p} n_h\upsilon(A) \geq \sum_{h=1}^{p} k_h \\ \Rightarrow\ & \sum_{i=1}^{m} \upsilon(A_i) - (\sum_{h=1}^{p} n_h)\upsilon(A) \geq \sum_{h=1}^{p} k_h \end{aligned}$$

By decomposition, $n = \sum_{h=1}^{p} n_h$ and $k = \sum_{h=1}^{p} k_h$, and therefore $\sum_{i=1}^{m} \upsilon(A_i) - n\upsilon(A) \geq k$, showing that **UP3** holds, as desired. ∎

E Proof of Proposition 4.1

Proposition 4.1: *The formula f is satisfiable in an upper probability structure iff the inequality formula $\hat{f}$ has a solution. Moreover, if $\hat{f}$ has a solution, then f is satisfiable in an upper probability structure with at most $2^{|f|}$ worlds.*

Proof: Assume first that f is satisfiable. Thus there is some upper probability structure $M = (\Omega, \Sigma, \mathcal{P}, \pi)$ such that $M \models f$. Define the vector x^* by letting $x^*_i = \mathcal{P}^*([\![\rho_i]\!]_M)$, for $1 \leq i \leq 2^{2^N}$. Since $M \models f$, it is immediate that x^* is a solution to the inequality formula $\overline{f}$. Moreover, since $\rho_1 = \mathit{false}$ and $\rho^{2^{2^N}} = \mathit{true}$, it follows that $x^*_1 = 0$ (since $\mathcal{P}^*([\![\mathit{false}]\!]_M) = \mathcal{P}^*(\emptyset) = 0$) and $x^*_{2^{2^N}} = 1$ (since $\mathcal{P}^*([\![\mathit{true}]\!]_M) = \mathcal{P}^*(\Omega) = 1$). Finally, consider a conjunct of $\hat{f}$ corresponding to an instance of **L4**; suppose it has the form $x_{i_1} + \cdots x_{i_m} - n x_{i_{m+1}} \geq k$. Since this conjunct appears in $\hat{f}$, it must be the case that $(\rho_{i_{m+1}} \Rightarrow \bigvee_{J \subseteq \{1,\ldots,m\},\, |J|=k+n} \bigwedge_{J \subseteq \{1,\ldots,m\},\, j \in J} \rho_{i_j}) \wedge (\bigvee_{|J|=k} \bigwedge_{j \in J} \rho_{i_j})$ is a propositional tautology. Thus, it follows that $[\![\rho_{i_1}]\!]_M, \ldots, [\![\rho_{i_m}]\!]_M$ is an (n,k)-cover for $([\![\rho_{i_{m+1}}]\!]_M, [\![\mathit{true}]\!]_M)$. It follows from UP3 that

$$\mathcal{P}^*([\![\rho_{i_1}]\!]_M) + \cdots + \mathcal{P}^*([\![\rho_{i_m}]\!]_M) - n\mathcal{P}^*([\![\rho]\!]_M) \geq k.$$

Thus, x^* is a solution to the inequality formulas corresponding to **L4**. Hence, x^* is a solution to $\hat{f}$.

For the converse, assume that x^* is a solution to $\hat{f}$. We construct an upper probability structure $M = (S, E, \mathcal{P}, \pi)$ such that $M \models f$ as follows. Let $p_1, \ldots, p_N$ be the primitive propositions appearing in f. Let $S = \{\delta_1, \ldots, \delta_{2^N}\}$ be the atoms over $p_1, \ldots, p_N$. Let E be the set of all subsets of S. As observed earlier, every propositional formula over $p_1, \ldots, p_n$ is equivalent to a unique disjunction of atoms. Thus, we can get a canonical collection $\rho_1, \ldots, \rho_{2^{2^N}}$ of inequivalent formulas over $p_1, \ldots, p_n$ by identifying each formula ρ_i with a different element of E, where ρ_1 corresponds to the empty set and $\rho_{2^{2^N}}$ corresponds to all of S. Define a set function υ by taking $\upsilon(\{\delta_{i_1}, \ldots, \delta_{i_j}\}) = x^*_i$ if ρ_i is the disjunction of the atoms $\delta_{i_1}, \ldots, \delta_{i_j}$. Let $\pi(\delta)(\rho) = \mathbf{true}$ iff $\delta \Rightarrow \rho$.

It is now sufficient to show that υ is an upper probability (of a set $\mathcal{P}$ of probability measures), since then it is clear that $(S, E, \mathcal{P}, \pi) \models f$ (since x^* is a solution to $\hat{f}$, the system of inequalities derived from formula f). To do this, by Theorem 2.4, it suffices to verify **UPF1**, **UPF2**, and **UPF3**, using B_{2^N} in **UPF3**, since $|S| = 2^N$.

UPF1: $\upsilon(\emptyset) = x^*_1 = 0$.

UPF2: $\upsilon(S) = x^*_{2^{2^N}} = 1$.

UPF3: Suppose that A and $A_1, \ldots, A_m$ are in E and satisfy the premises of property **UPF3**, with $k, m, n \leq B_{2^N}$. Let $\rho_{i_1}, \ldots, \rho_{i_m}, \rho_{i_{m+1}}$ be the canonical formulas corresponding to $A_1, \ldots, A_m, A$, respectively. Clearly, $A \subseteq \bigcup_{J \subseteq \{1,\ldots,m\},\, |J|=k+n} \bigcap_{j \in J} A_{i_j}$ iff $\rho_{i_{m+1}} \Rightarrow \bigvee_{J \subseteq \{1,\ldots,m\},\, |J|=k+n} \bigwedge_{j \in J} \rho_{i_j}$ is a propositional tautology and $\Omega \subseteq \bigcup_{J \subseteq \{1,\ldots,m\},\, |J|=k} \bigcap_{j \in J} A_{i_j}$ iff $\bigvee_{J \subseteq \{1,\ldots,m\},\, |J|=k} \bigwedge_{j \in J} \rho_{i_j}$ is a propositional tautology. Thus, $\sum_{j=1}^{m} x_{i_j} - x_{i_{m+1}} \geq k$ is one of the inequality formulas in $\hat{f}$. Thus, it follows that $\sum_{j=1}^{m} x^*_{i_j} - x^*_{i_{m+1}} \geq k$, as desired. By our definition of υ, we therefore have $k + n\upsilon(A) \leq \sum_{i=1}^{m} \upsilon(A_i)$, and so **UPF3** holds. ∎

F Proof of Theorem 5.1

Theorem 5.1: *Suppose f is a likelihood formula that is satisfied in some upper probability structure. Then f is satisfied in a structure $(\Omega, \Sigma, \mathcal{P}, \pi)$ where $|\Omega| \leq |f|^2$, $\Sigma = 2^\Omega$ (so that every subset set of Ω is measurable), $|\mathcal{P}| \leq |f|$, for every $w \in \Omega$ and $\mu \in \mathcal{P}$, $\mu(w)$ is a rational number such that $||\mu(w)||$ is $O(|f|^2||f|| + |f|\log(|f|))$.*

Proof: The first step in the proof involves showing that if $\mathcal{P}$ is a set of probability measures defined on an algebra Σ of a finite space Ω, we can assume without loss of generality that for each set $X \in \Sigma$, there is a probability measure $\mu_X \in \mathcal{P}$ such that $\mu_X(X) = \mathcal{P}^*(X)$ (rather than $\mathcal{P}^*(X)$ just being the sup of $\mu(X)$ for $\mu \in \mathcal{P}$).

Lemma F.1: *Let $\mathcal{P}$ be a set of probability measures defined on an algebra Σ over a finite set Ω. Then there exists a set $\mathcal{P}'$ of probability measures such that, for each $X \in \Sigma$, $\mathcal{P}^*(X) = (\mathcal{P}')^*(X)$; moreover, there is a probability measure $\mu_X \in \mathcal{P}'$ such that $\mu_X(X) = \mathcal{P}^*(X)$. In addition, for any interpretation π, if $M = (\Omega, \Sigma, \mathcal{P}, \pi)$ and $M = (\Omega, \Sigma, \mathcal{P}', \pi)$, then for all likelihood formulas f, $M \models f$ iff $M' \models f$.*

Proof: Since Σ is finite, to show that $\mathcal{P}'$ exists, it clearly suffices to show that, for each $X \in \Sigma$, there is a probability measure μ_X such that $\mu_X(X) = \mathcal{P}^*(X)$ and, if $\mathcal{P}' = \mathcal{P} \cup \{\mu_X\}$, then $\mathcal{P}^*(Y) = (\mathcal{P}')^*(Y)$ for all $Y \in \Sigma$.

Given X, if there exists $\mu \in \mathcal{P}$ such that $\mu(X) = \mathcal{P}^*(X)$, then we are done. Otherwise, we construct a sequence $\mu_1, \mu_2, \ldots$ of probability measures in $\mathcal{P}$ such that $\lim_i \mu_i(X) = \mathcal{P}^*(X)$ and, for all $Y \in \Sigma$, the sequence $\mu_i(Y)$ converges to some limit. Let $X_1, \ldots, X_n$ be an enumeration of the sets in Σ, with $X_1 = X$. We inductively construct a sequence of measures $\mu_{m1}, \mu_{m2}, \ldots$ in $\mathcal{P}$ for $m \leq n$ such that $\mu_{mi}(X_j)$ converges to a limit for $i \leq k$ and $\lim_{i\to\infty} \mu_{mi}(X) = \mathcal{P}^*(X)$. For $m = 1$, we know there must be a sequence $\mu_{11}, \mu_{12}, \ldots$ of measures in $\mathcal{P}$ such that $\mu_{1i}(X)$ converges to $\mathcal{P}^*(X)$. For the inductive step, if $m < n$, suppose we have constructed an appropriate sequence $\mu_{m1}, \mu_{m2}, \ldots$. Consider the sequence of real numbers $\mu_{mi}(X_{m+1})$. Using the Bolzano-Weierstrass theorem [13] (which says that every sequence of real numbers has a convergent subsequence), this sequence has a convergent subsequence. Let $\mu_{(m+1)1}, \mu_{(m+1)2}, \ldots$ be the subsequence of $\mu_{m1}, \mu_{m2}, \ldots$ which generates this convergent subsequence. This sequence of probability measures clearly has all the required properties. This completes the inductive step.

Define $\mu_X(Y) = \lim_{i\to\infty} \mu_{ni}(Y)$. It is easy to check that that μ_X is indeed a probability measure, that $\mu_X(X) = \mathcal{P}^*(X)$, and if $\mathcal{P}' = \mathcal{P} \cup \{\mu_X\}$, that $\mathcal{P}^*(Y) = (\mathcal{P}')^*(Y)$ for all $Y \in \Sigma$. This shows that an appropriate set $\mathcal{P}'$ exists.

Now, given π, let $M = (\Omega, \Sigma, \mathcal{P}, \pi)$ and $M' = (\Omega, \Sigma, \mathcal{P}', \pi)$. A straightforward induction on the structure of f shows that $M \models f$ iff $M' \models f$. For the base case:

$$\begin{array}{ll} & (\Omega, \Sigma, \mathcal{P}, \pi) \models a_1 l(\phi_1) + \cdots + a_n l(\phi_n) \geq a \\ \Leftrightarrow & a_1 \mathcal{P}^*([\![\phi_1]\!]_M) + \cdots + a_n \mathcal{P}^*([\![\phi_n]\!]_M) \geq a \\ \Leftrightarrow & a_1 (\mathcal{P}')^*([\![\phi_1]\!]_M) + \cdots + a_n (\mathcal{P}')^*([\![\phi_n]\!]_M) \geq a \\ \Leftrightarrow & (\Omega, \Sigma, \mathcal{P}', \pi) \models a_1 l(\phi_1) + \cdots + a_n l(\phi_n) \geq a. \end{array}$$

The others cases are trivial. ∎

Just as in [7], to prove Theorem 5.1, we make use of the following lemma which can be derived from Cramer's rule [15] and simple estimates on the size of the determinant (see also [2] for a simpler variant):

Lemma F.2: *If a system of r linear equalities and/or inequalities with integer coefficients each of length at most l has a nonnegative solution, then it has a nonnegative solution with at most r entries positive, and where the size of each member of the solution is $O(rl + r\log(r))$.*

Continuing with the proof of Theorem 5.1, suppose that f is satisfiable in an upper probability structure. By Proposition 4.1, the system $\hat{f}$ of equality formulas has a solution, so f is satisfied in a upper probability structure with a finite state space. Thus, by Lemma F.1, f is satisfied in a structure $(\Omega, \Sigma, \mathcal{P}, \pi)$ such that for all $X \in \Sigma$, there exists $\mu_X \in \mathcal{P}$ such that $\mu_X(X) = \mathcal{P}^*(X)$.

As in the completeness proof, we can write f in disjunctive normal form. Each disjunct g is a conjunction of at most $|f| - 1$ basic likelihood formulas and their negations. Since $M \models f$, there must be some disjunct g such that $M \models g$. Suppose that g is the conjunction of r basic likelihood formulas and s negations of basic likelihood formulas. Let $p_1, \ldots, p_N$ be the primitive formulas appearing in f. Let $\delta_1, \ldots, \delta_{2^N}$ be the atoms over $p_1, \ldots, p_N$. As in the proof of completeness, we derive a system of equalities and inequalities from g. It is a slightly more complicated system, however. Recall that each propositional formula over $p_1, \ldots, p_N$ is a disjunction of atoms. Let $\phi_1, \ldots, \phi_k$ be the propositional formulas that appear in g. Notice that $k < |f|$ (since there are some symbols in f, such as the coefficients, that are not in the propositional formulas). The system of equations and inequalities we construct involve variables x_{ij}, where $i = 1, \ldots, k$ and $j = 1, \ldots, 2^N$. Intuitively, x_{ij} represents $\mu_{[\![\phi_i]\!]_M}([\![\delta^j]\!]_M)$, where $\mu_{[\![\phi_i]\!]_M} \in \mathcal{P}$ is such that $\mu_{[\![\phi_i]\!]_M}([\![\phi_i]\!]_M) = \mathcal{P}^*([\![\phi_i]\!]_M)$. Thus, the system includes $k < |f|$ equations of the following form,

$$x_{i1} + \cdots + x_{i2^N} = 1,$$

for $i = 1, \ldots, k$. Since $\mu_{[\![\phi_i]\!]_M}([\![\phi_i]\!]_M) \geq \mu([\![\phi_i]\!]_M)$ for all $\mu \in \mathcal{P}$, if E_i is the subset of $\{1, \ldots, 2^N\}$ such that $\phi_i = \bigvee_{j \in E_i} \delta_j$, the system includes $k^2 - k$ inequalities of the form

$$\sum_{j \in E_i} x_{ij} \geq \sum_{j \in E_i} x_{i'j},$$

for each pair i, i' such that $i \neq i'$. For each conjunct in g of the form $\theta_1 l(\phi_1) + \cdots + \theta_n l(\phi_k) \geq \alpha$, there is a corresponding inequality where, roughly speaking, we replace $l(\phi_i)$ by $\mu_{[\![\phi_i]\!]_M}([\![\phi]\!]_M)$.[6] Since $\mu_{[\![\phi_i]\!]_M}$ corresponds to $\sum_{j \in E_i} x_{ij}$, the appropriate inequality is

$$\sum_{i=1}^{k} \theta_i \sum_{j \in E_i} x_{ij} \geq \alpha.$$

Negations of such formulas correspond to a negated inequality formula; as before, this is equivalent to a formula of the form

$$-(\sum_{i=1}^{k} \theta_i \sum_{j \in E_i} x_{ij}) > -\alpha.$$

[6]For simplicity here, we are implicitly assuming that each of the formulas ϕ_i appears in each conjunct of g. This is without loss of generality, since if ϕ_i does not appear, we can put it in, taking $\theta_i = 0$.

Notice that there are at most $|f|$ inequalities corresponding to the conjuncts of g. Thus, altogether, there are at most $k(k-1)+2|f| < |f|^2$ equations and inequalities in the system (since $k < |f|$). We know that the system has a nonnegative solution (taking x_j^i to be $\mu_{[\![\phi_i]\!]_M}([\![\delta^j]\!]_M)$). Thus, by Lemma F.2, the system has a solution $x^* = (x_{11}^*, \ldots, x_{12^N}^*, \ldots, x_{k1}^*, \ldots, x_{k2^N}^*)$ with $t \leq |f|^2$ entries positive, and with each entry of size $O(|f|^2||f|| + |f|^2 \log(|f|))$.

We use this solution to construct a small structure satisfying f. Let $I = \{i : x_{ij}^* \text{ is positive, for some } j\}$; suppose that $I = \{i_1, \ldots, i_{t'}\}$, for some $t' \leq t$. Let $M = (S, E, \mathcal{P}, \pi)$ where S has t' states, say $s_1, \ldots, s_{t'}$, and E consists of all subsets of S. Let $\pi(s_h)$ be the truth assignment corresponding to the formula δ_{i_h} (and where $\pi(s_h)(p) =$ **false** if p does not appear in f). Define $\mathcal{P} = \{\mu_j : 1 \leq i \leq k\}$, where $\mu_j(s_h) = x_{i_h j}^*$. It is clear from the construction that $M \models f$. Since $|\mathcal{P}| = k < |f|$, $|S| = t' \leq t \leq |f|^2$ and $\mu_j(s_h) = x_{i_h j}^*$, where, by construction, the size of $x_{i_h j}^*$ is $O(|f|^2||f|| + |f|^2 \log(|f|))$, the theorem follows. ∎

References

[1] B. Anger and J. Lembcke. Infinitely subadditive capacities as upper envelopes of measures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 68:403–414, 1985.

[2] V. Chvátal. *Linear Programming*. W. Freeman and Co., San Francisco, Calif., 1983.

[3] J. B. Conway. *A Course in Functional Analysis*. Number 96 in Graduate Texts in Mathematics. Springer-Verlag, second edition, 1990.

[4] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967.

[5] L. E. Dickson. Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. *American Journal of Mathematics*, 35(4):413–422, 1913.

[6] R. Fagin and J. Y. Halpern. Uncertainty, belief and probability. *Computational Intelligence*, 7(3):160–173, 1991.

[7] R. Fagin, J. Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87(1,2):78–128, 1990.

[8] P. J. Huber. Kapazitäten statt Wahrscheinlichkeiten? Gedanken zur Grundlegung der Statistik. *Jber. Deutsch. Math.-Verein*, 78:81–92, 1976.

[9] P. J. Huber. *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley Interscience, 1981.

[10] H. E. Kyburg, Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[11] E. Mendelson. *Introduction to Mathematical Logic*. Van Nostrand, New York, 1964.

[12] S. Popkorn. *First Steps in Modal Logic*. Cambridge University Press, Cambridge; New York, 1994.

[13] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, third edition, 1976.

[14] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ, 1976.

[15] T. Shores. *Applied Linear Algebra and Matrix Analysis.* McGraw-Hill, second edition, 1999.

[16] C. A. B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series B*, 23:1–25, 1961.

[17] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, 1991.

[18] P. M. Williams. Indeterminate probabilities. In M. Przelecki, K. Szaniawski, and E. Wójciki, editors, *Formal Methods in the Methodology of Empirical Sciences*, pages 229–246, 1976.

[19] G. Wolf. *Obere und Untere Wahrscheinlichkeiten.* Doctoral dissertation, Eidgenössischen Technischen Hochschule, Zurich, 1977. (Diss. ETH 5884).