Annals of Pure and Applied Logic 177 (2026) 103714

Contents lists available at ScienceDirect

Annals of Pure and Applied Logic

journal homepage: www.elsevier.com/locate/apal

Full Length Article

A decidable class of inferences in first-order objective Bayesian inductive logic

Jürgen Landes *, Soroush Rafiee Rad, Jon Williamson

ARTICLE INFO

Article history:
Received 12 November 2024
Received in revised form 30 December 2025
Accepted 31 December 2025
Available online 7 January 2026

MSC:
03B48
03B25
03C10
03C75
94A17
68T37

Keywords:
Decidability
Inductive logic
Objective Bayesianism
Maximum entropy
Truth table
Bayesian network

ABSTRACT

We show that while standard first-order inductive logic is not decidable, a large class of inferences in objective Bayesian inductive logic is decidable. Decidability is achieved by reducing the general inference problem to a quantifier-free problem. We show that for any inference, if the quantifier-free reduction of the premisses is satisfiable, then the original inference is decidable. We go on to show that Bayesian networks offer the potential to provide a computationally tractable inference procedure for objective Bayesian inductive logic. We also consider inferences with infinitely many premisses and explore some properties of the logic.

1. Introduction

Is there a computable procedure for deciding whether any given inference from finitely many premisses is valid? Hilbert described this decision problem to be 'the main problem of mathematical logic' [13, p. 113]. Here, we consider the decision problem in the context of inductive logic.

In propositional deductive logic, the truth-table method provides an effectively computable procedure for deciding whether any inference from premisses $\varphi_1, \dots, \varphi_k$ to conclusion ψ is valid—i.e., for deciding whether the entailment relationship $\varphi_1, \dots, \varphi_k \models \psi$ holds. However, there is no such procedure for first-order deductive logic: in first-order deductive logic, the class of inferences from finitely many premisses is undecidable [34]; [24, Theorem 16.52]. There are decidable fragments of first-order deductive logic, such

* Corresponding author.
E-mail address: J.Landes@lrz.uni-muenchen.de (J. Landes).

https://doi.org/10.1016/j.apal.2025.103714
0168-0072/

as the special case in which all the predicate symbols are unary [22, §3]. Standard fragments can now be classified as to their decidability, and at the turn of the present century it was observed that 'the work on the classical decision problem is by and large completed' for first-order deductive logic [3, p. 8].

Inductive logic generalises deductive logic to the situation in which premiss and conclusion sentences may be less than certain: i.e., to entailment relationships of the form $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx \psi^Y$, where $X_1, \ldots, X_k, Y$ are representations of the uncertainty that attaches to the corresponding sentences $\varphi_1, \ldots, \varphi_k, \psi$ of the logic, and where $\mid\!\approx$ is an inductive entailment relation [9]. Given that first-order inductive logic is a generalisation of first-order deductive logic, the prospects for decidability are dim. Indeed, thanks to the undecidability of first-order deductive logic, first-order inductive logic is undecidable when endowed with the 'standard semantics,' i.e., when $X_1, \ldots, X_k, Y$ are sets of probabilities and one deems an entailment relationship to hold just when all probability functions that satisfy the premisses also satisfy the conclusion (§3).

In this paper, we consider objective Bayesian inductive logic (OBIL) [17,31,39], which provides an alternative to the standard semantics. For OBIL, premisses inductively entail a conclusion, written $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx^{\circ} \psi^Y$, just when the probability functions with maximal entropy, from all those probability functions that satisfy the premisses, satisfy the conclusion. Such functions can be regarded as probability functions that satisfy the premisses but which are maximally non-committal with respect to other propositions.

We show that a truth-table method can be used to determine the validity of a surprisingly large class of inferences of OBIL. Indeed, this class of inferences is decidable.

In §2 we outline the formal framework. In §3 we show that the standard semantics for probabilistic logic is not decidable but show how a truth-table method can be used to test for the validity of those inferences in OBIL that involve only quantifier-free sentences. In §4 we consider the more general case of quantified sentences and introduce a quantifier-free 'support' problem that is associated with the more general problem. In §5 we show that the general problem can often be reduced to the support problem. §6 shows that a large class of inferences in OBIL is therefore decidable. §7 provides a more computationally tractable method for solving the associated support problem, which appeals to Bayesian networks. §8 considers the extent to which these results can be generalised to inferences that involve infinitely many premisses. Finally, §10 develops a more detailed understanding of the class of decidable inferences identified in this paper.

2. Inductive logic

In this section, we provide the background on inductive logic to which we shall appeal throughout the paper.

2.1. Logic

We shall work in pure first-order logic, i.e., first-order logic without function symbols or equality. We take language $\mathcal{L}$ to have finitely many relation symbols, countably many constant symbols and countably many variable symbols. By default, we shall use $U_1, \ldots, U_l$ for the relation symbols, $t_1, t_2, \ldots$ for the constant symbols, and $x_1, x_2, \ldots$ for the variable symbols, but we shall occasionally use other symbols where convenient. The sentences $S\mathcal{L}$ of $\mathcal{L}$ are formed in the usual way from the atomic sentences $Ut_{i_1} \ldots t_{i_k}$ using the standard connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$, quantifiers $\exists, \forall$ and variables.

Suppose $a_1, a_2, \ldots$ is an ordering of the atomic sentences such that atomic sentences involving only $t_1, \ldots, t_n$ appear before those that involve t_{n+1}, for each $n \geq 1$. For any n, let $\mathcal{L}_n$ be the finite sublanguage of $\mathcal{L}$ that has $t_1, \ldots, t_n$ as its only constant symbols. The atomic sentences of $\mathcal{L}_n$ are $a_1, \ldots a_{r_n}$ for some $r_n \geq n$.

Example 1. Suppose $\mathcal{L}$ has just a binary relation symbol U and a unary relation symbol V. $\mathcal{L}_1$ has the atomic propositions $a_1 = Vt_1$ and $a_2 = Ut_1t_1$, so $r_1 = 2$. $\mathcal{L}_2$ also involves $a_3 = Vt_2$, $a_4 = Ut_1t_2$, $a_5 =$

$Ut_2t_1,\ a_6 = Ut_2t_2$, so $r_2 = 6$. $\mathcal{L}_3$ also involves $a_7 = Vt_3,\ a_8 = Ut_1t_3,\ a_9 = Ut_2t_3,\ a_{10} = Ut_3t_1,\ a_{11} = Ut_3t_2,\ a_{12} = Ut_3t_3$, so $r_3 = 12$, and so on. For this language, $r_n = r_{n-1} + 2n = n(n+1)$.

Definition 1 *(N_φ).* For any sentence $\varphi \in S\mathcal{L}$, let N_φ be the greatest index of all the constants that appear in φ. If φ has no constants, we adopt the convention that $N_\varphi = 1$.

A crucial role in the following analysis is played by the sentences that are called the *n-states* or *state-descriptions* of $\mathcal{L}$:

Definition 2 *(n-states).* For any $n \geq 1$, the set Ω_n of n-states is the set of sentences of the form $\pm a_1 \wedge \cdots \wedge \pm a_{r_n}$, where $+a_i$ is just a_i and $-a_i$ is $\neg a_i$, for $i = 1, \ldots, r_n$.

2.2. Probability

Probability functions on the language $\mathcal{L}$ (or more accurately on $S\mathcal{L}$) are defined as follows:

Definition 3 *(Probability).* A *probability function* P on $\mathcal{L}$ is a function $P : S\mathcal{L} \longrightarrow \mathbb{R}_{\geq 0}$ such that:

P1. If $\models \tau$, then $P(\tau) = 1$.
P2. If $\models \neg(\theta \wedge \varphi)$, then $P(\theta \vee \varphi) = P(\theta) + P(\varphi)$.
P3. $P(\exists x \theta(x)) = \sup_m P\left(\bigvee_{i=1}^m \theta(t_i)\right)$.

Remark 1. Axiom P3, which is due to [7], requires the presupposition that every member of the domain is named by at least one constant symbol [28, p. 162]. This therefore restricts us to interpretations with countable domains.

Remark 2. A probability function is determined by the values it gives to the n-states—see, e.g., [41, §2.6.3]. On the other hand, an assignment P of values to the n-states generates a probability function if the following conditions hold: $\sum_{\omega \in \Omega_n} P(\omega) = 1$ and $P(\omega) = \sum_{\zeta \in \Omega_{n+1}, \zeta \models \omega} P(\zeta)$ for all $\omega \in \Omega_n$ and $n \geq 1$.

We denote the set of probability functions by $\mathbb{P}$. Of particular importance will be the *equivocator* function, $P_= \in \mathbb{P}$, which gives the same probability to each n-state, for each n:

Definition 4 *(Equivocator function).* The *equivocator function* is the probability function $P_=$ defined by:

$$P_=(\omega_n) = \frac{1}{2^{r_n}} = \frac{1}{|\Omega_n|}$$

for each n-state $\omega_n \in \Omega_n$ and each $n \geq 1$.

Definition 5 *(Measure).* The *measure* of a sentence θ is the probability given to it by the equivocator function. In particular, θ has *positive measure* if and only if $P_=(\theta) > 0$.

Probabilities on first-order languages are similar to probabilities on finite domains since the axioms P1 – P3 have a number of simple and intuitive but very important consequences—see [28, Proposition 2.1], [29, Lemma 3.8] and [41, §2.3.2] for example:

Proposition 1. *For sentences $\theta, \varphi, \psi \in S\mathcal{L}$*

1. $P(\neg\theta) = 1 - P(\theta)$.

2. *If* $\models \theta$, *then* $P(\neg\theta) = 0$.
3. *If* $\theta \models \varphi$, *then* $P(\theta) \leq P(\varphi)$.
4. *If* $\theta \equiv \varphi$, *then* $P(\theta) = P(\varphi)$.
5. $P(\varphi) \in [0, 1]$.

2.3. Probabilistic logics

An inductive logic posits entailment relationships between premisses and conclusion sentences that may be uncertain. In a probabilistic inductive logic, this uncertainty is expressed using probabilities. We shall consider probabilistic logics that posit entailment relationships of the following form [9]:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!\!\models \psi^Y \quad .$$

Here, $\varphi_1, \ldots, \varphi_k, \psi \in S\mathcal{L}$ and $X_1, \ldots, X_k, Y \subseteq [0, 1]$. This entailment relationship should be interpreted as saying: $\varphi_1, \ldots, \varphi_k$ having probabilities in $X_1, \ldots, X_k$ respectively inductively entails that ψ has probability in Y. An absence of premisses, $k = 0$, provides the set of tautologies of the inductive logic.

Definition 6 *(Feasible region).* Let $\mathcal{A}$ be the set of expressions of the form θ^W where $\theta \in S\mathcal{L}$ and $W \subseteq [0, 1]$. For any $A \subseteq \mathcal{A}$, let $\mathbb{P}[A]$ be the set of probability functions satisfying all the expressions in A:

$$\mathbb{P}[A] \stackrel{\text{df}}{=} \{P \in \mathbb{P} : P(\theta) \in W \text{ for all } \theta^W \in A\}.$$

Given premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$, we define the feasible region to be

$$\mathbb{E} \stackrel{\text{df}}{=} \mathbb{P}[\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}] \stackrel{\text{df}}{=} \mathbb{P}[\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}].$$

In a probabilistic logic, models of a set of probabilistic expressions are probability functions specified by some function $[\![\cdot]\!] : \mathscr{P}\mathcal{A} \longrightarrow \mathscr{P}\mathbb{P}$. This function assigns to every set of probabilistic expressions of the form θ^W a set of probability functions, and satisfies the following condition:

$$[\![A]\!] \subseteq \mathbb{P}[A] \text{ for any consistent } A \subseteq \mathcal{A}.$$

This function can be used to provide semantics for the entailment relation:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!\!\models \psi^Y \text{ if and only if } [\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] \subseteq \mathbb{P}[\psi^Y].$$

What is sometimes called the *standard semantics* for probabilistic logic [9,26,10] considers the entire set of probability functions that satisfy the premisses, i.e.,

$$[\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] = \mathbb{E} = \mathbb{P}[\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}].$$

In the standard semantics, then,

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!\!\models \psi^Y \text{ if and only if } \mathbb{E} \subseteq \mathbb{P}[\psi^Y].$$

Note that if $\mathbb{E} = \emptyset$ then any conclusion ψ^Y follows.

2.4. OBIL

Objective Bayesian inductive logic (OBIL) provides an alternative to the standard semantics. The objective Bayesian approach interprets probabilities as rational degrees of belief [39, Chapter 7]. It takes the premisses on the left-hand side of the entailment relationship to capture all the constraints on rational degrees of belief that are imposed by evidence, and it asks: what probabilities are given to the conclusion sentence ψ by the maximally non-committal probability functions that satisfy the premisses? The idea is to consider probability functions that best represent the premisses in the sense that they satisfy the premisses but go as little beyond the premisses as possible. Entropy is standardly used to measure the extent to which a probability function is non-committal, i.e., the extent to which it equivocates between the basic expressible possibilities. Hence, OBIL considers those probability functions that satisfy the premisses which have maximal entropy, in the following sense.

Definition 7 *(n-entropy).* The *n-entropy* of a probability function P is defined as

$$H_n(P) \stackrel{\text{df}}{=} -\sum_{\omega\in\Omega_n} P(\omega)\log P(\omega) \ .$$

We adopt the usual convention that $0 \cdot \log 0 = 0$. We shall sometimes use P^n to refer to an *n-entropy maximiser*, i.e., a probability function in $\mathbb{E}$ that maximises n-entropy.

Remark 3. We take the logarithm in the previous definition to have base 2, which is the natural base from an information-theoretic perspective. Using any other base $b > 1$ would instead give:

$$H_{n,b}(P) := -\sum_{\omega\in\Omega_n} P(\omega)\log_b P(\omega) = -\sum_{\omega\in\Omega_n} P(\omega)\frac{\log_e P(\omega)}{\log_e b} = \frac{\log_e 2}{\log_e b} H_n(P) \ .$$

Since $\log_e(b) > 0$, $H_n(P) \geq H_n(Q)$ iff $H_{n,b}(P) \geq H_{n,b}(Q)$. Since in this paper we are only interested in comparing n-entropies to one another, the choice of the base $b > 1$ is inconsequential for our purposes, and we suppress the base in the notation.

The n-entropies, which only take into account the probabilities on finitely many n-states, are then used to define a notion of comparative entropy on the infinite language $\mathcal{L}$ as a whole:

Definition 8 *(Comparative entropy).* Probability function $P \in \mathbb{P}$ *has greater entropy than* $Q \in \mathbb{P}$ if and only if the n-entropy of P dominates that of Q for sufficiently large n, i.e., if and only if there is an $N \in \mathbb{N}$ such that for all $n \geq N$, $H_n(P) > H_n(Q)$.

The greater entropy relation defines a partial order $\prec_H$ on the probability functions on $\mathcal{L}$. We shall focus on functions in $\mathbb{E} = \mathbb{P}[\varphi_1^{X_1}, \dots, \varphi_k^{X_k}]$ that are maximal with respect to this partial ordering:

Definition 9 *(Maximal entropy functions).* The set of maximal entropy functions on $\mathbb{E}$, $\operatorname{maxent}\mathbb{E}$, is defined as

$$\operatorname{maxent}\mathbb{E} := \{P \in \mathbb{E} \ : \ \forall Q \in \mathbb{E}, P \not\prec_H Q\} \ .$$

Where $\operatorname{maxent}\mathbb{E}$ is non-empty, we shall often use $P^\dagger$ or $P^\dagger_{\mathbb{E}}$, to refer to some member of $\operatorname{maxent}\mathbb{E}$. In this case, we set $[\![\varphi_1^{X_1}, \dots, \varphi_k^{X_k}]\!] = \operatorname{maxent}\mathbb{E}$ in order to provide semantics for OBIL [41, §5.3]:

Definition 10 *(Objective Bayesian inductive entailment).* Suppose $\mathrm{maxent}\,\mathbb{E} \neq \emptyset$. Premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ inductively entail ψ^Y, denoted by $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx} \psi^Y$, if and only if $P(\psi) \in Y$ for all $P \in \mathrm{maxent}\,\mathbb{E}$. I.e.,

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx} \psi^Y \text{ if and only if } \mathrm{maxent}\,\mathbb{E} \subseteq \mathbb{P}[\psi^Y].$$

Something needs to be said about the case in which $\mathrm{maxent}\,\mathbb{E}$ is empty.[1] Given the objective Bayesian semantics for OBIL, it is natural to avoid explosion—i.e., the claim that it is reasonable to believe any conclusion statement to any degree [39]. For the purposes of this paper, we shall say that if $\mathrm{maxent}\,\mathbb{E} = \emptyset$ but $\mathbb{E} \neq \emptyset$, the entailment relationship holds when $P(\psi) \in Y$ for every $P \in \mathbb{E}$. If $\mathbb{E} = \emptyset$, we shall take the entailment relationship to hold when $P(\psi) \in Y$ for $P \in \mathrm{maxent}\,\mathbb{P} = \{P_=\}$.[2] In sum, in OBIL,

$$[\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] = \begin{cases} \mathrm{maxent}\,\mathbb{E} & : \quad \mathrm{maxent}\,\mathbb{E} \neq \emptyset \\ \mathbb{E} & : \quad \mathrm{maxent}\,\mathbb{E} = \emptyset \neq \mathbb{E} \\ \mathrm{maxent}\,\mathbb{P} & : \quad \mathbb{E} = \emptyset. \end{cases} \tag{1}$$

We shall focus on the first of these three cases in this paper.

In the context of objective Bayesianism, constraints on rational degrees of belief are convex.[3] Hence, in the context of OBIL, we shall take the $X_1, \ldots, X_k, Y$ to be intervals of probabilities.[4] Moreover, to simplify our exposition, we shall suppose that these intervals are closed intervals. This simplifies the exposition because it ensures that, for each n, a satisfiable set of quantifier-free premisses has an n-entropy maximiser P^n that is uniquely determined on the sentences of $\mathcal{L}_n$ (because $\mathbb{E}$ is closed and convex and H_n is strictly concave).[5]

We write φ^{c_i} to abbreviate $\varphi^{[c_i,c_i]}$, which attaches a single probability $c_i \in [0,1]$ to sentence φ_i, and we identify the interval $[c_i, c_i]$ with c_i. We abbreviate a statement of the form θ^1 by θ, for $\theta \in S\mathcal{L}$, and call such a statement 'categorical'.

In the absence of any premisses $\mathbb{E} = \mathbb{P}$, so $\mathrel{|\!\!\approx} \psi^Y$ holds if and only if $P_=(\psi) \in Y$, since $\mathrm{maxent}\,\mathbb{P} = \{P_=\}$.

Definition 11. A sentence ψ is an *inductive tautology* if $\mathrel{|\!\!\approx} \psi$, i.e., if it has measure 1. ψ is an *inductive contradiction* if $\mathrel{|\!\!\approx} \neg\psi$, i.e., if it has measure 0. ψ is *inductively consistent* if $\not\mathrel{|\!\!\approx} \neg\psi$, i.e., if it has positive measure. Sentences ψ and θ are *inductively equivalent* if $\mathrel{|\!\!\approx} \psi \leftrightarrow \theta$.

3. Decidability and truth tables in OBIL

In this section, we see that the prospects for the decidability of a first-order probabilistic logic are dim. However, we go on to informally describe a truth-table method, originally introduced by [41], and we show that this method can be used to decide whether inferences that invoke quantifier-free sentences of $\mathcal{L}$ are valid in OBIL. In later sections, we show that there is a surprisingly large class of inferences that involve quantified sentences and that are decidable by means of this truth-table method.

[1] Note that since $\prec_H$ is a partial order on an infinite set, it may contain an infinite chain with no maximal element. For example, the premiss $\exists x \forall y Uxy^1$ is satisfiable ($\mathbb{E} \neq \emptyset$) but $\mathrm{maxent}\,\mathbb{E} = \emptyset$ [20, Proposition 53].

[2] There are more sophisticated approaches that one can take here. In the former case, one can restrict attention to probability functions in $\mathbb{E}$ that are *sufficiently* equivocal. In the latter case, one can consider probability functions that satisfy some maximal consistent subset of the premisses. See [39] and [16, §9] for further discussion.

[3] For the rationale behind convexity, see [39, Chapter 3], [35] and [40].

[4] Although we allow the possibility that one or more of these intervals is empty, such a possibility is of little interest because any proposition of the form $\theta^\emptyset$ will be unsatisfiable.

[5] If the premisses are not satisfiable then, as stipulated above, $[\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] = \mathrm{maxent}\,\mathbb{P}$. $\mathrm{maxent}\,\mathbb{P} = \{P_=\}$, so inferences are drawn using the equivocator function $P_=$.

3.1. Decidability and precision

Before proceeding, we should note that undecidability can arise in probabilistic logic in two ways. The kind of undecidability we are concerned with in this paper is undecidability that arises from the logical structure of the sentences that occur in the premisses of an inductive inference. But undecidability also arises trivially in a second way in probabilistic logic: when trying to determine the equality of two non-terminating decimals. Suppose for example that we have an entailment relationship

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^c$$

and the probabilistic logic in question gives a single model of the premisses, $[\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] = \{P\}$. To decide whether the entailment relationship holds, we need to determine whether $P(\psi) = c$. Now suppose that we have a procedure for determining successive digits of $P(\psi)$ and that the decimal expansion of c is non-terminating. If indeed $P(\psi) = c$, the comparison between $P(\psi)$ and c will not terminate in a finite time. Hence, there is a trivial—and rather uninteresting—sense in which there is no effective procedure for deciding whether an inductive entailment holds, if the probabilities in question include real numbers with non-terminating decimal expansions.

In order to focus on the first, logical kind of undecidability we eliminate this second, numerical kind of undecidability by imposing two restrictions. Firstly, we take all probability intervals in OBIL to be finitely represented:

Definition 12 *(Finitely represented).* A closed interval is *finitely represented* if it is represented as $[l, u]$ where l and u are terminating decimal fractions, i.e., are of the form 1.0 or $0.d_1 d_2 \ldots d_s$, where $s \in \mathbb{N}$ and $d_i \in \{0, 1, \ldots 9\}$ for $i = 1, \ldots, s$. An expression of the form θ^Z, where $\theta \in S\mathcal{L}$ and Z is an interval, is *finitely represented* if the interval Z is finitely represented. An inference is *finitely represented* if its premisses and its conclusion are finitely represented.

This restriction is not enough on its own to eliminate numerical undecidability: if c is say 0.479, our procedure for generating successive digits of $P(\psi)$ might yield $0.47900000\ldots$, in which case it will still not be possible to determine that $P(\psi) = c$ in a finite amount of time. Hence, we also presuppose a given level of precision with which to perform numerical comparisons. Thus if it is sufficient to perform comparisons to 20 decimal places, we need only determine that $P(\psi) = 0.47900000000000000000$ to 20 decimal places in order to decide that the entailment relationship holds.

Without further explicit mention, then, we consider only finitely represented inferences—entailment relationships of the form $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx^{\circ} \psi^Y$ in which the sets of probabilities $X_1, \ldots, X_k, Y$ are finitely represented closed intervals—and we suppose some fixed finite level of precision with which to perform numerical comparisons. This will allow us to focus on logical decidability, i.e., decidability modulo comparison of real numbers.

3.2. Decidability and deductive logic

Any probabilistic logic generalises deductive logic in the following sense:

Proposition 2. *In any probabilistic logic, if $\varphi_1, \ldots, \varphi_k$ are jointly consistent then*

$$\varphi_1, \ldots, \varphi_k \approx \psi \text{ if } \varphi_1, \ldots, \varphi_k \models \psi.$$

Proof. If $[\![\varphi_1, \ldots, \varphi_k]\!] = \emptyset$, the inductive entailment relationship holds trivially.

Otherwise, suppose $P \in [\![\varphi_1, \ldots, \varphi_k]\!]$.

If $\varphi_1, \ldots, \varphi_k \models \psi$ then $\models \neg(\varphi_1 \wedge \ldots \wedge \varphi_k) \vee \psi$, so by axiom P1,

$$P(\neg(\varphi_1 \wedge \ldots \wedge \varphi_k) \vee \psi) = 1.$$

Now $\varphi_1 \wedge \ldots \wedge \varphi_k$ and $\neg\psi$ are mutually exclusive, so by axiom P2,

$$P(\neg(\varphi_1 \wedge \ldots \wedge \varphi_k)) + P(\psi) = P(\neg(\varphi_1 \wedge \ldots \wedge \varphi_k) \vee \psi) = 1.$$

But $P \in [\![\varphi_1, \ldots, \varphi_k]\!] \subseteq \mathbb{P}[\varphi_1, \ldots, \varphi_k]$, since $\varphi_1, \ldots, \varphi_k$ are jointly consistent. So $P(\varphi_1 \wedge \ldots \wedge \varphi_k) = 1$ and $P(\neg(\varphi_1 \wedge \ldots \wedge \varphi_k)) = 0$. Thus $P(\psi) = 1$, i.e., $P \in \mathbb{P}[\psi]$. Hence, $\varphi_1, \ldots, \varphi_k \approx\!\!\!\!\!\mid\;\, \psi$. □

In the case of the standard semantics for probabilistic logic, we can say more:

Proposition 3. *With the standard semantics,*

$$\varphi_1, \ldots, \varphi_k \mid\!\approx \psi \text{ if and only if } \varphi_1, \ldots, \varphi_k \models \psi.$$

Proof. Consider first the claim that $\varphi_1, \ldots, \varphi_k \models \psi$ implies $\varphi_1, \ldots, \varphi_k \mid\!\approx \psi$. Given Proposition 2, we need only consider the case in which $\varphi_1, \ldots, \varphi_k$ are jointly inconsistent. In that case, $[\![\varphi_1, \ldots, \varphi_k]\!] = \emptyset$ and any conclusion follows. In particular, $\varphi_1, \ldots, \varphi_k \mid\!\approx \psi$.

It remains to show that $\varphi_1, \ldots, \varphi_k \mid\!\approx \psi$ implies $\varphi_1, \ldots, \varphi_k \models \psi$.

If the premisses are inconsistent then both the inductive and the deductive entailment relationships hold vacuously, so it is trivially the case that $\varphi_1, \ldots, \varphi_k \mid\!\approx \psi$ implies $\varphi_1, \ldots, \varphi_k \models \psi$.

Consider next the case in which the premisses are consistent. If $\varphi_1, \ldots, \varphi_k \not\models \psi$ then there is some interpretation of $\mathcal{L}$ under which the premisses are true and the conclusion false—i.e., a model of the premisses together with the negation of the conclusion. By the Löwenheim-Skolem Theorem, there is such a model with a countable domain. Without loss of generality, we can suppose that each member of this countable domain is named by at least one constant symbol: otherwise, add new constant symbols to the language to refer to previously unnamed members of the domain and revise the interpretation to specify the referents of the new names, leading to an expansion of the original model. This interpretation thus satisfies the requirements outlined in Remark 1.

This interpretation yields a truth assignment ν to the sentences of $\mathcal{L}$ such that $\nu(\varphi_i) = 1$ (i.e., $\nu \models \varphi_i$) for $i = 1, \ldots, k$, and $\nu(\psi) = 0$ (i.e., $\nu \models \neg\psi$).

Note that ν is also a probability function:

P1. If $\models \tau$, then $\nu(\tau) = 1$.
P2. If $\models \neg(\theta \wedge \chi)$, then there are two possible cases: either $\nu(\theta) = \nu(\chi) = \nu(\theta \vee \chi) = 0$ or ν models precisely one of θ and χ and gives $\nu(\theta \vee \chi) = 1$. Either way, $\nu(\theta \vee \chi) = \nu(\theta) + \nu(\chi)$.
P3. If $\nu(\theta(t_i)) = 0$ for all i then by induction on P2, $\nu\left(\bigvee_{i=1}^m \theta(t_i)\right) = 0$ for all m, and since each member of the domain is named by some constant symbol, $\nu\left(\exists x \theta(x)\right) = \sup_m \nu\left(\bigvee_{i=1}^m \theta(t_i)\right) = 0$. Otherwise $\nu(\theta(t_j)) = 1$ for some j, $\nu\left(\exists x \theta(x)\right) = 1$, and by induction on P2, $\nu\left(\bigvee_{i=1}^m \theta(t_i)\right) = 1$ for all $m \geq j$, so $\sup_m \nu\left(\bigvee_{i=1}^m \theta(t_i)\right) = 1$.

Since $\nu(\varphi_i) = 1$ for each $i = 1, \ldots, k$ and $\nu(\psi) = 0$, $\nu \in \mathbb{E}$ but $\nu \notin \mathbb{P}[\psi]$. Hence $\varphi_1, \ldots, \varphi_k \not\mid\!\approx \psi$, as required. □

This feature enables the use of the standard semantics to provide semantics for deductive logic as well as inductive logic [6]. It also has important consequences for decidability:

Definition 13 *(Decidable inferences).* A class of inferences of a given logic is *decidable* if there is an effective procedure for deciding whether any given inference lies within the class and, if so, whether the inference is valid in the logic.[6] Otherwise it is *undecidable*.

Corollary 1. *Suppose $\mathcal{L}$ contains at least one relation symbol of arity at least 2. Then for the standard semantics for probabilistic logic, the class of entailment relationships from finitely many premisses is undecidable.*

Proof. Suppose for contradiction that the class of inductive entailment relationships of the form $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!\!| \;\psi^Y$ were decidable, with respect to the standard semantics. Then there would be an effective procedure for deciding, in particular, whether $\approx\!\!\!\!| \;\psi$ for any $\psi \in S\mathcal{L}$. By Proposition 3, this procedure would decide whether $\models \psi$ for any $\psi \in S\mathcal{L}$. But the class of logically valid sentences of first-order deductive logic is undecidable when there are relation symbols that are at least binary (see [34]; [24, Theorem 16.52]; [3, p. 10]). This gives the required contradiction. □

On the other hand, it is possible to define a decidable probabilistic logic. Consider the *trivial probabilistic logic*, defined by setting $[\![A]\!] = \emptyset$ for all $A \in \mathcal{A}$. In the trivial probabilistic logic, every entailment relationship holds and the class of all inferences in this logic is clearly decidable. Notwithstanding this fact, Corollary 1 might lead to pessimism about the decidability of any *reasonable* probabilistic logic. The fact is that the class of inferences from finitely many premisses in first-order deductive logic is undecidable, and a first-order inductive logic generalises first-order deductive logic to cover cases in which the premisses are uncertain. It is hard to see how any reasonable generalisation could be decidable.[7]

3.3. Truth tables

The aim of this paper is to show that there is a wide class of inferences in OBIL that is decidable using a truth-table method. Truth tables are usually introduced in the context of propositional deductive logic, which is decidable. Indeed, the truth-table method provides perhaps the best known decision procedure for the class of deductive inferences of a finite propositional logic. The lines (rows) of the truth table run through all the truth assignments to the propositional variables that occur in the inference. The truth value of each premiss and the conclusion of the inference are calculated on each line, and the entailment relationship holds just when the conclusion is true at all lines of the truth table at which the premisses are true.

Consider for example the truth table for a simple deductive entailment claim:

$$a \rightarrow b, b \models a$$

a	b	$a \rightarrow b$	b	a
T	T	T	T	T
T	F	F	F	T
F	T	$\boxed{\text{T}}$	$\boxed{\text{T}}$	$\boxed{\text{F}}$
F	F	T	F	F

On the third row, the premisses are true and the conclusion false, so the inference is invalid.

[6] Here we appeal to the standard notion of 'effective procedure': informally, a mechanical procedure that terminates to give the correct answer up to an arbitrarily close approximation after a finite number of steps. This notion is usually formally explicated by appeal to recursive functions or Turing machines [24, Part I].

[7] [33] shows, for example, that a probabilistic logic based on the theory of PAC learning is undecidable. See [1] for more general pessimism about the decidability of probabilistic logics.

Exactly the same method can be used to decide whether an inference in first-order deductive logic (from finitely many premisses) is valid in the special case in which the premisses and conclusion are all quantifier-free sentences of $\mathcal{L}$. One can simply build a truth table around the atomic propositions $a_{i_1}, \ldots, a_{i_m}$ that occur in the inference and the premisses and conclusion. Thus by setting $a = U_1t_4t_6$ and $b = U_2t_9$, the same truth table can be used to test the inference:

$$U_1t_4t_6 \to U_2t_9, U_2t_9 \models U_1t_4t_6.$$

Again, the third line of the truth table tells us that the inference is invalid.

Moreover, as we shall see now, the same truth table can be used to determine whether the following inference holds in OBIL:

$$U_1t_4t_6 \to U_2t_9, U_2t_9 \mid\!\approx^{\circ} U_1t_4t_6{}^{1/2} \quad . \tag{2}$$

In OBIL, when an inference involves categorical (i.e., certain) and consistent quantifier-free premisses, the probability that attaches to a quantifier-free conclusion sentence is the proportion of all those lines of the truth table at which the premisses are true where the conclusion is also true [41, Chapter 1 and §6.1]. (Note that the question of the consistency of the premisses is decidable here, because the truth table can also be used to check that there is a truth assignment to the atomic propositions, i.e., a line of the truth table, at which all the premisses are true.) In the above truth table, there are two lines at which the premisses are true, one of which makes the conclusion true, so the probability that attaches to the conclusion sentence is $\frac{1}{2}$. Thus, the entailment relationship (2) does indeed hold.

Recall that a probability function on $S\mathcal{L}$ is determined by its values on the n-states. This fact allows us to extend the truth table method for OBIL to handle non-categorical quantifier-free premisses. The idea is to attach a probability to each line of the truth table: this is the probability that is induced by the maximal entropy function. Consider

$$a \to b, b^{2/5} \mid\!\approx^{\circ} a^{1/5},$$

where, as before, a is $U_1t_4t_6$ and b is U_2t_9. We can build the following augmented truth table: The premiss

$P^{\dagger}$	a	b	$a \to b$	b	a
$\frac{1}{5}$	T	T	T	T	T
0	T	F	F	F	T
$\frac{1}{5}$	F	T	T	T	F
$\frac{3}{5}$	F	F	T	F	F

$a \to b$ forces the second line to have probability 0. The premiss $b^{2/5}$ ensures that probability 2/5 is distributed between lines 1 and 3 of the truth table; the maximal entropy function will distribute this probability equally in the absence of further information concerning b. The remaining probability, 3/5, must attach to line 4. The probability that attaches to the conclusion a is the sum of the probabilities attached to lines 1 and 2, i.e., 1/5. Thus the entailment relationship does indeed hold.

This approach generalises as follows. Suppose that in the context of a particular inference we have premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ in which the premiss sentences $\varphi_1, \ldots, \varphi_k$ are all quantifier-free. (Recall that $X_1, \ldots, X_k$ are assumed to be closed and convex.) Let $a_{i_1}, \ldots, a_{i_m}$ be the atomic propositions that occur in $\varphi_1, \ldots, \varphi_k$ and Ξ be the set of states of $a_{i_1}, \ldots, a_{i_m}$. For any $n \geq 1$ let $\bar{\Xi}$ be the set of states of the atomic propositions, other than $a_{i_1}, \ldots, a_{i_m}$, that are in $\mathcal{L}_n$; if there are no such atomic propositions, take $\bar{\Xi}$ to contain just an arbitrary tautology. Let P_{Ξ} be a probability function on $\mathcal{L}$ that satisfies the premisses, and maximises the entropy on Ξ,

$$P_\Xi \in \arg\max_{P \in \mathbb{E}} -\sum_{\xi \in \Xi} P(\xi) \log P(\xi),$$

if $\mathbb{E} \neq \emptyset$. Note that all such entropy maximisers agree on Ξ: on a finite domain, a closed, convex set of probability functions has a unique entropy maximiser because the entropy function is strictly concave. If $\mathbb{E} = \emptyset$, let $P_\Xi \stackrel{\text{df}}{=} P_=$. Thus P_Ξ is uniquely determined on Ξ, whether or not the constraints are satisfiable.

We are now in a position to identify a unique maximal entropy function on $\mathcal{L}$ itself:

Proposition 4 *(Quantifier-free entropy maximisation). Suppose premiss sentences $\varphi_1, \dots, \varphi_k$ are quantifier-free. Then $[\![\varphi_1^{X_1}, \dots, \varphi_k^{X_k}]\!] = \{P^\dagger\}$, where the probability function $P^\dagger$ is characterised by*

$$P^\dagger(\omega) \stackrel{df}{=} P_\Xi(\xi) P_=(\zeta), \tag{3}$$

for all $\omega \in \Omega_n$ and $n \geq 1$, and where $\xi \in \Xi$ and $\zeta \in \bar{\bar{\Xi}}$ are states induced by ω, i.e., $\omega \equiv \xi \wedge \zeta$.

Proof. If the premisses are unsatisfiable, i.e., $\mathbb{E} = \emptyset$, then by definition $[\![\varphi_1^{X_1}, \dots, \varphi_k^{X_k}]\!] = \text{maxent}\,\mathbb{P} = \{P_=\}$. In this case, $P_\Xi \stackrel{\text{df}}{=} P_=$, so $P^\dagger = P_=$, as required.

If the premisses are satisfiable, $[\![\varphi_1^{X_1}, \dots, \varphi_k^{X_k}]\!] = \text{maxent}\,\mathbb{E}$. Now $P_\Xi \in \mathbb{E}$, by construction. Consequently, $P^\dagger$, as defined above, is in $\mathbb{E}$:

$$P^\dagger(\varphi_i) = \sum_{\substack{\xi \in \Xi \\ \xi \models \varphi_i}} P^\dagger(\xi) = \sum_{\substack{\xi \in \Xi \\ \xi \models \varphi_i}} P_\Xi(\xi) = P_\Xi(\varphi_i) \in X_i,$$

for $i = 1, \dots, k$.

Consider n large enough that the premiss sentences $\varphi_1, \dots, \varphi_k$ can all be expressed in $\mathcal{L}_n$. $P^\dagger$ is an n-entropy maximiser, as can be seen as follows [41, Theorem 5.13]. By the chain rule for entropy [5, Theorem 2.2.1], for any probability function $Q \in \mathbb{E}$,

$$\begin{aligned} H_n(Q) &= -\sum_{\xi \in \Xi} Q(\xi) \log Q(\xi) - \sum_{\substack{\omega \in \Omega_n \\ \xi \wedge \zeta \equiv \omega}} Q(\omega) \log Q(\zeta|\xi) \\ &\leq -\sum_{\xi \in \Xi} P^\dagger(\xi) \log P^\dagger(\xi) - \sum_{\substack{\omega \in \Omega_n \\ \xi \wedge \zeta \equiv \omega}} P^\dagger(\omega) \log P^\dagger(\zeta|\xi) \\ &= H_n(P^\dagger), \end{aligned}$$

with equality if and only if Q coincides with $P^\dagger$ on all sentences of $\mathcal{L}_n$. The above inequality holds because P_Ξ is the entropy maximiser on Ξ and $P^\dagger$ is defined as $P^\dagger(\omega) = P_\Xi(\xi) P_=(\zeta)$, so

$$-\sum_{\xi \in \Xi} Q(\xi) \log Q(\xi) \leq -\sum_{\xi \in \Xi} P_\Xi(\xi) \log P_\Xi(\xi) = -\sum_{\xi \in \Xi} P^\dagger(\xi) \log P^\dagger(\xi)$$

with equality if and only if Q coincides with P_Ξ on Ξ, and

$$-\sum_{\substack{\omega \in \Omega_n \\ \xi \wedge \zeta \equiv \omega}} Q(\omega) \log Q(\zeta|\xi) \leq -\sum_{\substack{\omega \in \Omega_n \\ \xi \wedge \zeta \equiv \omega}} P^\dagger(\omega) \log P^\dagger(\zeta|\xi),$$

with equality if and only if $Q(\zeta|\xi) = P_=(\zeta|\xi)$ for all $\xi \in \Xi$ and $\zeta \in \bar{\bar{\Xi}}$. To see why this last inequality obtains, note first that $P^\dagger(\zeta|\xi) = P_=(\zeta) = \frac{|\Xi|}{|\Omega_n|}$ for each $\xi \in \Xi$ and $\zeta \in \bar{\bar{\Xi}}$, so,

$$
\begin{aligned}
-\sum_{\substack{\omega\in\Omega_n\\ \xi\wedge\zeta\equiv\omega}} Q(\omega)\log Q(\zeta|\xi) &= \sum_{\substack{\omega\in\Omega_n\\ \xi\wedge\zeta\equiv\omega}} Q(\omega)\log\frac{Q(\xi)}{Q(\omega)}\\
&\le \log\sum_{\substack{\omega\in\Omega_n\\ \xi\wedge\zeta\equiv\omega}} Q(\omega)\frac{Q(\xi)}{Q(\omega)}\\
&= \log\sum_{\substack{\omega\in\Omega_n\\ \xi\wedge\zeta\equiv\omega}} Q(\xi)\\
&= \log\left(\frac{|\Omega_n|}{|\Xi|}\sum_{\xi\in\Xi} Q(\xi)\right)\\
&= \log\frac{|\Omega_n|}{|\Xi|}\\
&= \log\frac{|\Omega_n|}{|\Xi|}\left(\sum_{\omega\in\Omega_n} P^\dagger(\omega)\right)\\
&= -\sum_{\omega\in\Omega_n} P^\dagger(\omega)\log\frac{|\Xi|}{|\Omega_n|}\\
&= -\sum_{\substack{\omega\in\Omega_n\\ \xi\wedge\zeta\equiv\omega}} P^\dagger(\omega)\log P^\dagger(\zeta|\xi),
\end{aligned}
$$

where the second line of the above equation is an instance of Jensen's inequality [5, Theorem 2.6.2].

We have seen that $P^\dagger$ maximises n-entropy for sufficiently large n and that any function Q that maximises n-entropy for sufficiently large n agrees with $P^\dagger$ on $\mathcal{L}_n$ for each sufficiently large n, and so coincides with $P^\dagger$ on $\mathcal{L}$. Hence, maxent $\mathbb{E} = \{P^\dagger\}$. $\square$

This result enables the use of a truth table to represent the maximal entropy probability function $P^\dagger$, given quantifier-free premisses $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$. Each line of the truth table needs to be augmented by the probability $P^\dagger(\xi)$ of the state ξ that is satisfied by the truth valuation on that line, which is found by first maximising entropy on Ξ to get P_Ξ and then equivocating beyond Ξ, i.e., by the construction $P^\dagger(\omega) = P_\Xi(\xi)P_=(\zeta)$ of Equation (3). (Note that if the premisses are not jointly satisfiable, $[\![\varphi_1^{X_1}, \dots, \varphi_k^{X_k}]\!]$ = maxent $\mathbb{P} = \{P_=\}$, and each line of the truth table is given the same probability.)

We thus have:

Proposition 5. *If the premiss sentences $\varphi_1, \dots, \varphi_k$ are all quantifier-free then the truth-table method can be used to determine $P^\dagger$ on $S\mathcal{L}$.*

Proof. The (augmented) truth table determines $P^\dagger$ via Equation (3), because a probability function is determined by its values on the n-states. $\square$

Proposition 6 *(Quantifier-free satisfiability). Whether quantifier-free premisses $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$ are jointly satisfiable is decidable.*

Proof. The existence of a probability function that satisfies the premisses is equivalent to the existence of a solution for a system of linear inequalities with unknowns $P(\varphi_i) = \sum_{\xi\models\varphi_i} P(\xi)$. That this problem is decidable follows from the Tarski–Seidenberg theorem and the decidability of first order theory of closed real fields.

To see this, take an inequality constraint of the form $x_{i_1} + \ldots + x_{i_r} \leq c_i$, partly expressing the premiss $\varphi_i^{[b_i,c_i]}$. Remember that c_i is finitely represented which implies $c_i \in \mathbb{Q}$, and hence c_i is an algebraic number. That is, c_i can be expressed as the root of a polynomial in one variable with integer coefficients and thus it is definable in the first order theory of closed real fields. Let $\kappa_{c_i}(y)$ be the polynomial defining c_i. Then the above constraints can be expressed by the formula

$$\Psi(\vec{x}) := \exists y(\kappa_{c_i}(y) = 0 \wedge x_{i_1} + \ldots + x_{i_r} \leq y),$$

and the existence of a solution for the system of linear inequalities $\Psi_1(\vec{x}), \ldots, \Psi_k(\vec{x})$ can be expressed as the first order sentence

$$\exists \vec{x} \bigwedge_{i=1}^{k} \Psi_i(\vec{x}),$$

in the first order language of closed real fields. $\square$

Theorem 1 *(Quantifier-free decidability). The class of quantifier-free inferences is decidable in OBIL.*

Proof. Since the proof is somewhat long, we split it into parts.

Set-up of the problem. Take any inference in the class of quantifier-free inferences. Since the premiss sentences are quantifier free, the truth-table method can be used to fully determine $P^\dagger$. Since the conclusion sentence ψ is quantifier free, $\psi \equiv \bigvee_{\substack{\omega \in \Omega_n \\ \omega \models \psi}} \omega$ for sufficiently large n, so $P^\dagger(\psi) = \sum_{\substack{\omega \in \Omega_n \\ \omega \models \psi}} P^\dagger(\omega)$. Since $P^\dagger(\psi)$ can be effectively determined from the truth table and Equation (3), the key task then is to fill in the probability values in the truth table: i.e., to find $P^\dagger(\xi)$ for each $\xi \in \Xi$.

By Proposition 6, whether the premisses are jointly satisfiable is decidable. If they are not satisfiable then $P^\dagger = [\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] = \operatorname{maxent} \mathbb{P} = \{P_=\}$, and each line of the truth table is given the same probability $1/|\Xi|$.

Otherwise, the task is to determine $P^\dagger(\xi)$ for each $\xi \in \Xi$ where $P^\dagger$ is the function in $\mathbb{E} \neq \emptyset$ which maximises entropy. Given Proposition 4, we can focus on probability functions defined over Ξ, rather than on the whole language $\mathcal{L}$. We shall use $\mathbb{X}$ to denote the set of probability distributions defined over Ξ that satisfy the constraints imposed by the premisses. The task is to determine the unique probability function $x^\dagger$ on Ξ such that $\operatorname{maxent} \mathbb{X} = \{x^\dagger\}$. Recall that we are working to some degree of precision, so the task is to determine, for any given $\varepsilon > 0$, some $x^* \in \mathbb{X}$ such that $|x^* - x^\dagger| \stackrel{\mathrm{df}}{=} \sup_{\xi \in \Xi} | x^*(\xi) - x^\dagger(\xi) | < \varepsilon$.

Proof sketch. Let us first sketch how to find such an approximation x^*. We shall consider a closed region $\mathbb{X}' \subseteq \mathbb{X}$ within which $x^\dagger$ is known to lie and an effectively specifiable tessellation $\mathcal{T}$ of $\mathbb{X}'$ involving finitely many closed convex polytopes (henceforth called 'tiles').[8] Given any $\delta > 0$, one can find $\mathbb{X}'$, a tessellation $\mathcal{T}$ of $\mathbb{X}'$, and rational functions H^+ and H^- on $\mathcal{T}$ such that

1. $H^+(\tau)$ (respectively, $H^-(\tau)$) is an upper (respectively, lower) bound on the entropy of the entropy maximiser, $x_\tau^\dagger \stackrel{\mathrm{df}}{=} \arg\max_{x \in \tau} H(x)$ within the tile τ, and
2. for all tiles $\tau \in \mathcal{T}$, $H^+(\tau) - H^-(\tau) < \delta$.

Let τ^* be some tile that maximises H^+. Note that τ^* can be found effectively because there are only finitely many tiles in $\mathcal{T}$. Then we have that:

[8] A tessellation is a cover of $\mathbb{X}_i$ such that the intersection of two different tiles contains none of their interior points.

$$H^-(\tau^*) \leq H(x^\dagger) \leq H^+(\tau^*)$$

and

$$H^+(\tau^*) - H^-(\tau^*) < \delta$$

so

$$H^+(\tau^*) - H(x^\dagger) < \delta.$$

Thus we can approximate $H(x^\dagger)$ as close as we like by means of $H^+(\tau^*)$. Now consider some effectively specifiable probability function $x^* \in \tau^*$. x^* may not yet be close to $x^\dagger$ in the sense that $|x^* - x^\dagger \,| < \varepsilon$. However, δ can be reduced until x^* provides a close enough approximation to $x^\dagger$.

In more detail, a suitable approximation x^* to $x^\dagger$ can be found as follows.

Determining $\mathbb{X}$. If $x_= \stackrel{\mathrm{df}}{=} P_{=\restriction\Xi} \in \mathbb{X}$, then $x^\dagger = x_=$. Hence we first check whether $x_=$ satisfies all constraints (this is easily computable). If so, each line of the truth table is given the same probability $1/|\Xi|$ and we can simply set $x^* = x^\dagger$.

If $x_= \notin \mathbb{X}$, the next step is to effectively determine $\mathbb{X}$ by computing the vertices of $\mathbb{X}$. We note that

$$\begin{aligned}\mathbb{X} = \{x \ : \ & \sum_{\xi\in\Xi} x(\xi) \leq 1, \sum_{\xi\in\Xi} x(\xi) \geq 1, \\ & x(\xi) \geq 0, \text{ for all } \xi \in \Xi, \\ & \sum_{\substack{\xi\in\Xi \\ \xi\models\varphi_1}} x(\xi) \leq X_1^+, \sum_{\substack{\xi\in\Xi \\ \xi\models\varphi_1}} x(\xi) \geq X_1^-, \\ & \ldots \\ & \sum_{\substack{\xi\in\Xi \\ \xi\models\varphi_k}} x(\xi) \leq X_k^+, \sum_{\substack{\xi\in\Xi \\ \xi\models\varphi_k}} x(\xi) \geq X_k^-\}\end{aligned}$$

with $X_i = [X_i^-, X_i^+]$. We can use the Fourier–Motzkin elimination algorithm to compute this set [14]. The algorithm is effectively computable on a Turing machine since it only requires addition and multiplication of rational numbers.

We next repeatedly eliminate superfluous constraints by checking whether an application of the Fourier–Motzkin elimination algorithm to all but one of the constraints gives the same result. If so, then the omitted constraint is superfluous and can be dropped. If not, then the constraint is relevant and cannot be dropped. Eventually, we arrive at a minimal set of constraints $\mathcal{C}$ that cannot be further simplified.

Next, turn $|\Xi|$-many constraints in $\mathcal{C}$ into equality constraints by replacing $\leq, \geq$ by $=$ to yield new sets $\mathcal{C}'$ of constraints. In this way each equality constraint serves as the border that divides the space of probability functions into two disjoint regions; one in which the inequality is satisfied, and one in which it is violated. The set $\mathbb{X}$ will then be the region enclosed by these borders. Vertices of $\mathbb{X}$ will be where these borders intersect at a point.

In order to find these vertices, we check whether each such set $\mathcal{C}'$ of constraints has a unique solution via the Fourier–Motzkin elimination algorithm. Consider those $\mathcal{C}'$ that do have a unique solution. The unique solutions of these subsets of constraints are the vertices of $\mathbb{X}$, since $\mathcal{C}$ is minimal. $\mathbb{X}$ is then effectively characterised as the set of convex combinations of these vertices.

If the feasible region $\mathbb{X}$ consists of a single element, then this element is $x^\dagger$ and we can simply take $x^* = x^\dagger$. In the following, then, we assume that the feasible region has at least two elements. By the convexity of the feasible region, this entails that the feasible region contains uncountably many points.

Iterative approach. Our approach is to iteratively produce an ever smaller region $\mathbb{X}_i$ of $\mathbb{X}$ which contains $x^\dagger$. Initially, $i = 1$ and we set $\mathbb{X}_1 = \mathbb{X}$.

For all $i \in \mathbb{N}$ we split $\mathbb{X}_i$ into a tessellation $\mathcal{T}_i$ of finitely many closed, convex polytopes such that: (i) all vertices of each tile have rational coordinates, and (ii) for every tile $\tau \in \mathcal{T}_i$, $|\tau| \stackrel{\text{df}}{=} \sup_{x,y\in\tau} |x - y| < 1/2^i|\Xi|$.

We define an upper bound H_i^+ associated with $\mathbb{X}_i$, that satisfies the requirements introduced above, as follows. Consider the L_1 bound on entropy [5, Theorem 17.3.3]: if

$$\|x - y\|_1 \stackrel{\text{df}}{=} \sum_{x\in\Xi} | x(\xi) - y(\xi)| \leq \frac{1}{2},$$

then

$$|H(x) - H(y)| \leq - \| x - y \|_1 \log \frac{\|x - y\|_1}{|\Xi|}.$$

Applying this to $x, y \in \tau$, since $\sup_{x,y\in\tau} |x - y| < \frac{1}{2^i|\Xi|}$, we have that

$$\sum_{\xi\in\Xi} | x(\xi) - y(\xi)| \leq \frac{|\Xi|}{2^i|\Xi|} = \frac{1}{2^i} \leq \frac{1}{2},$$

so

$$|H(x) - H(y)| \leq -\frac{1}{2^i} \log \frac{1}{2^i|\Xi|} = \frac{i+m}{2^i} \log 2,$$

since $|\Xi| = 2^m$. Let x_τ^c be the centre of mass of tile τ, assuming uniform density. x_τ^c can be effectively determined as a convex combination of the vertices of τ. Thus, the centre of mass of a convex and non-empty set with a dimension of at least 1 lies in the interior (with respect to the norm topology of the dimension of the convex set) of this set.

For any i, then,

$$\begin{aligned} H(x_\tau^\dagger) &\leq H(x_\tau^c) + |H(x_\tau^\dagger) - H(x_\tau^c)| \\ &\leq H(x_\tau^c) + \frac{i+m}{2^i} \log 2. \end{aligned}$$

Thus we can let

$$H_i^+(\tau) \stackrel{\text{df}}{=} h_i^+(x_\tau^c) + \frac{i+m}{2^i} \log 2,$$

where an upper estimate $h_i^+(x_\tau^c)$ of $H(x_\tau^c)$ is found by calculating $H(x_\tau^c)$ to $d + i$ decimal places (e.g., by using a Taylor approximation) and incrementing the final digit, and where d is the number of decimal places needed to represent numbers at the required accuracy ε. Note that the upper bound improves as i increases, but this procedure does not tell us exactly how good the upper bound is.

Next, define the lower bound H_i^-. For each tile $\tau \in \mathcal{T}_i$, $H_i^-(\tau)$ is defined by computing a lower estimate $h_i^-(x_\tau^c)$ of $H(x_\tau^c)$, e.g., by calculating $H(x_\tau^c)$ to $t + i$ decimal places and decrementing the final digit:

$$H_i^-(\tau) \stackrel{\text{df}}{=} h_i^-(x_\tau^c).$$

Again, the lower bound improves as i increases, but this procedure does not tell us how good the lower bound is.

Note that $H_i^+(\tau)$ and $H_i^-(\tau)$ become arbitrarily close, since for all large enough i and all $\tau \in \mathcal{T}_i$

$$\begin{aligned} H_i^+(\tau) - H_i^-(\tau) &= h_i^+(x_\tau^c) + \frac{i+m}{2^i}\log 2 - h_i^-(x_\tau^c) \\ &= h_i^+(x_\tau^c) - h_i^-(x_\tau^c) + \frac{i+m}{2^i}\log 2 \\ &< \delta. \end{aligned}$$

For tiles $\sigma, \tau \in \mathcal{T}_i$, define a partial order $\succ_i$ by:

$$\sigma \succ_i \tau \iff H_i^-(\sigma) > H_i^+(\tau).$$

If $\sigma \succ_i \tau$ then clearly $H(x_\sigma^\dagger) > H(x_\tau^\dagger)$ and the overall entropy maximiser $x^\dagger$ cannot lie in τ.

Let $\mathbb{X}_{i+1} \stackrel{\text{df}}{=} \bigcup\{\tau \in \mathcal{T}_i :$ there is no $\sigma \in \mathcal{T}_i$ such that $\sigma \succ_i \tau\}$, i.e., the union of all τ that are maximal with respect to $\succ_i$. $\mathbb{X}_{i+1}$ is a subset of $\mathbb{X}_i$ within which $x^\dagger$ is guaranteed to lie. We can then define a new, finer tessellation $\mathcal{T}_{i+1}$ of $\mathbb{X}_{i+1}$ such that $\mathcal{T}_{i+1}$ is a refinement of $\mathcal{T}_i$ restricted to $\mathbb{X}_{i+1}$, and we can define approximations H_{i+1}^-, H_{i+1}^+ on $\mathcal{T}_{i+1}$ using the definitions provided above. Iterating, we refine the tessellation and compute new bounds. By construction, $x^\dagger \in \mathbb{X}_i$, for each i.

We claim that after finitely many iterations we find a set $\mathbb{X}_n$ such that $\sup_{x\in\mathbb{X}_n} |x - x^\dagger| < \varepsilon$, the required precision. Hence, any $x^* \in \mathbb{X}_n$ approximates $x^\dagger$ sufficiently closely. To be concrete, we can take $x^* = x_\tau^c$ for some $\tau \in \mathcal{T}_n$, as this element of $\mathbb{X}_n$ is effectively specifiable.

Termination of the algorithm. We can see that $|x - x^\dagger| < \varepsilon$ for all $x \in \mathbb{X}_i$ and sufficiently large i, as follows.

For every $i \in \mathbb{N}$, let $\mathcal{S}_i(x) \subseteq \mathcal{T}_i$ be the set of tiles in $\mathcal{T}_i$ that contain x. Furthermore, let $\mathcal{R}_i = \mathcal{S}_i(x^\dagger)$. Notice that if $x^\dagger$ is an interior point of a tile in $\tau \in \mathcal{T}_i$ then $\mathcal{R}_i$ is the singleton $\{\tau\}$, while if $x^\dagger$ lies on the boundary of a tile in $\mathcal{T}_i$ then $\mathcal{R}_i$ will have as elements all the tiles that share that part of the boundary.

Since the diameters of the tiles go to zero as i increases and since the tiles in $\mathcal{R}_i$ are adjacent, there is some $N \in \mathbb{N}$ such that for all $i \geq N$ and all x, y that feature in tiles in $\mathcal{R}_i$,

$$|x - y| < \epsilon.$$

That is, for all $i \geq N$ the region consisting of the set of tiles of $\mathcal{T}_i$ that contain the entropy maximiser has diameter less than our given precision ϵ, and thus for any $i \geq N$ any point in a tile in $\mathcal{R}_i$ (and in particular $x^\dagger$) can be suitably approximated by any other point in (some tile in) $\mathcal{R}_i$, given our threshold of precision.

Consider $\mathcal{T}_N$ and some $\sigma \in \mathcal{T}_N$ such that $\sigma \notin \mathcal{R}_N$. We next show that there is some $M \geq N$ such that $\sigma \cap \mathbb{X}_{M+1} = \emptyset$. That is, after $M + 1 - N$ more iterations, all the points in σ have been eradicated from the feasible region $\mathbb{X}_{M+1}$.

To see this, let $\delta \stackrel{\text{df}}{=} H(x^\dagger) - H(x_\sigma^\dagger)$. By the construction of the upper bound H_i^+, there is some $M_1 \in \mathbb{N}$ such that for all $i \geq M_1$, all $x \in \mathbb{X}_i$, and all tiles $\tau_x \in \mathcal{S}_i(x)$,

$$|H_i^+(\tau_x) - H(x_{\tau_x}^c)| < \frac{\delta}{2}.$$

By the construction of the lower bound H_i^-, there is some $M_2 \in \mathbb{N}$ such that for all $i \geq M_2$ and all tiles $\tau^\dagger \in \mathcal{R}_i = \mathcal{S}_i(x^\dagger)$,

$$|H_i^-(\tau_{x^\dagger}) - H(x_{\tau^\dagger}^c)| < \frac{\delta}{2}.$$

Let $M = \max\{M_1, M_2, N\}$ and consider the tessellation $\mathcal{T}_M$. Suppose tiles $\tau_1, \ldots, \tau_m$ are the refinements of $\sigma \in \mathcal{T}_N$ in $\mathcal{T}_M$, that is $\sigma \cap \mathbb{X}_M = \bigcup_{j=1}^m \tau_j$.

Now consider an arbitrary such tile τ_j, where $j \in \{1, \ldots, m\}$ and any tile $\tau^\dagger \in \mathcal{R}_M$. We have that:

$$H_M^+(\tau_j) < H(x^c_{\tau_j}) + \frac{\delta}{2} \leq H(x^\dagger_\sigma) + \frac{\delta}{2} = H(x^\dagger) - \frac{\delta}{2} < H_M^-(\tau^\dagger).$$

So $\tau^\dagger \succ_M \tau_j$ and $\tau_j \cap \mathbb{X}_{M+1} = \emptyset$. Since this holds for all $j = 1, \ldots, m$, it is indeed the case that $\sigma \cap \mathbb{X}_{M+1} = \emptyset$, as claimed above.

Denote this particular choice of M by M_σ and note that the tessellation $\mathcal{T}_N$ is finite. Consider $L = \max\{M_\sigma + 1 \,|\, \sigma \in \mathcal{T}_N, \sigma \notin \mathcal{R}_N\}$. Then for all such tiles σ, $\sigma \cap \mathbb{X}_L = \emptyset$. Hence, $\mathcal{T}_L = \mathcal{R}_L$. By construction, for all x, y in (tiles in) $\mathcal{R}_L$ we have that $|x - y| < \epsilon$, so any point in $\mathbb{X}_L$ can be taken as a suitable approximation x^* to $x^\dagger$, as required. $\square$

This result is perhaps surprising in the light of recent research that suggests that, for many important optimisation problems, determining the optimiser is in fact undecidable [21]. Note that while the algorithm that we provide in the above proof offers an effective procedure to obtain the entropy maximiser, and thus can be used to demonstrate decidability, we do not suggest that it is efficient enough to be used in practice to fill in a truth table. In practice, standard convex optimisation methods, such as gradient ascent methods [4] or Lagrange multiplier methods (see the Appendix), work perfectly well to find the entropy-maximising values that are required for the truth-table method. Moreover, the truth-table method is itself not the most efficient method for determining the probability that attaches to the conclusion sentence in an OBIL inference, because the number of rows of a truth table increases exponentially in the number of atomic propositions in the inference. In §7 we introduce an inference procedure that employs probabilistic graphical models and that is potentially much more efficient.

As an aside, we note that while the above result requires that the premiss and conclusion sentences are quantifier-free, the truth-table method can also be used to determine the probability $P^\dagger(\psi)$ of a conclusion sentence $\psi \in S\mathcal{L}$ that contains quantifiers. That this is the case will follow from a later result, Theorem 4; here it suffices to provide a couple of illustrative examples:

Example 2.

$$U_1t_4t_6 \to U_2t_9, {U_2t_9}^{2/5} \mathrel{|\!\!\approx^{\circ}} \exists x U_1 t_4 x$$

since, by axiom P3,

$$\begin{aligned}
P^\dagger(\exists x U_1t_4x) &= \sup_{m\to\infty} P^\dagger(\bigvee_{i=1}^m U_1t_4t_i) \\
&= \lim_{m\to\infty} P^\dagger(\bigvee_{i=1}^m U_1t_4t_i) \\
&= \lim_{m\to\infty} P^\dagger(\neg \bigwedge_{i=1}^m \neg U_1t_4t_i) \\
&= \lim_{m\to\infty} 1 - P^\dagger(\bigwedge_{i=1}^m \neg U_1t_4t_i) \\
&= 1 - \lim_{m\to\infty} P^\dagger(\bigwedge_{i=1}^m \neg U_1t_4t_i) \\
&= 1 - \lim_{m\to\infty} \frac{4}{5}\left(\frac{1}{2}\right)^{m-1} \\
&= 1
\end{aligned}$$

where the penultimate equation is obtained by equivocating beyond the truth table.

Furthermore,

Example 3.

$$U_1t_4t_6 \to U_2t_9, U_2t_9{}^{2/5} \mathrel{\approx\!\!\!\!\mid} \forall y\exists x U_1yx$$

because

$$P^\dagger(\forall y\exists x U_1yx) = \lim_{m\to\infty} P^\dagger(\bigwedge_{i=1}^{m} \exists x U_1t_ix) = \lim_{m\to\infty} 1^m = 1.$$

The key goal of the rest of the paper is extend the above decidability result to a much richer class of inferences. We next introduce a generalisation of a concept from [41, §5.5]:

Definition 14 *(Finitely generated consequences).* A set $A \subseteq \mathcal{A}$ of statements has *finitely generated consequences* if there are quantifier-free sentences $\varphi_1, \dots, \varphi_k$ and closed intervals $X_1, \dots, X_k \subseteq [0,1]$ such that $[\![A]\!] = [\![\varphi_1^{X_1}, \dots, \varphi_k^{X_k}]\!]$. $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$ are *generating statements* for A.

Definition 15 *(Finitely reducible).* A set $A \subseteq \mathcal{A}$ of statements is *finitely reducible* if it has finitely generated consequences and there is an effectively computable procedure for determining the generating statements for A. A class of inferences is *finitely reducible* iff

1. it is effectively determinable whether any given inference lies within the class of inferences,
2. each inference in the class has premisses with finitely generated consequences, and
3. there is an effectively computable procedure for determining the generating statements for the premisses of each inference in the class.

The task of the following sections is to show that there is a large class of inferences of OBIL that has finitely reducible consequences. By determining the generating statements for the premisses and using the truth-table method, this large class of inferences is then decidable.

4. The support inference

In this section, we consider premisses $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$ with $\varphi_1, \dots, \varphi_k$ being arbitrary sentences of $\mathcal{L}$ (not assumed to be quantifier-free), and $X_1, \dots, X_k$ closed subintervals of the unit interval as usual. We shall associate quantifier-free sentences $\check{\varphi}_1, \dots, \check{\varphi}_k$ with the premiss sentences $\varphi_1, \dots, \varphi_k$. In the next section, we shall specify conditions under which these yield generating statements for the premisses.

Definition 16 *(Support).* Suppose $a_{i_1}, \dots, a_{i_m}$ include all the atomic propositions that appear in sentence φ of $\mathcal{L}$, and let $\Xi_\varphi \stackrel{\text{df}}{=} \{\pm a_{i_1} \wedge \dots \wedge \pm a_{i_m}\}$ be the set of states of these atomic propositions. If φ contains no atomic propositions, we take $\Xi_\varphi \stackrel{\text{df}}{=} \{a_1, \neg a_1\}$.

The *support* $\check{\varphi}$ of φ is the disjunction of states in Ξ_φ that are inductively consistent with φ, i.e., the disjunction of $\xi \in \Xi_\varphi$ such that $\not\mathrel{\approx\!\!\!\!\mid} \neg(\xi \wedge \varphi)$. Equivalently,

$$\check{\varphi} \stackrel{\text{df}}{=} \bigvee\{\xi \in \Xi_\varphi : P_=(\xi \wedge \varphi) > 0\}.$$

If φ is a tautology, then so is $\check{\varphi}$. If φ is a contradiction, then so is $\check{\varphi}$.

Example 4. If $\varphi = \exists x(U_1t_1 \wedge U_2t_1x)$ then $\check{\varphi} = U_1t_1$. Note that φ does not mention an atomic proposition containing U_2.

Example 5. If $\varphi = U_1t_1 \vee (\forall x U_2x \wedge \neg U_1t_1)$ then $\check{\varphi} = U_1t_1$ because

$$P_=(\forall x U_2 x) = \lim_{m\to\infty} P_=(\bigwedge_{i=1}^{m} U_2t_i) = \lim_{m\to\infty} 2^{-m} = 0 \ .$$

Definition 17 *(Support inference).* Given an inference

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx^{\circ}} \psi^Y,$$

we shall consider an associated *support inference*,

$$\check{\varphi}_1^{X_1}, \ldots, \check{\varphi}_k^{X_k} \mathrel{|\!\!\approx^{\circ}} \check{\psi}^Y.$$

One can think of the support inference as a quantifier-free simulation of the original inference. We show in this section that if the support premisses $\check{\varphi}_1^{X_1}, \ldots, \check{\varphi}_k^{X_k}$ are satisfiable then they are generating statements for the original premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$, i.e., the original inference has finitely generated consequences. In the next subsection, we shall demonstrate that this is the case when $X_1, ..., X_k$ are point values in $[0, 1]$ (Theorem 2). We subsequently generalise the key result to the situation in which $X_1, ..., X_k$ are non-empty subintervals of $[0, 1]$ (Theorem 3).

It turns out that the construction of the support inference from the original inference is effectively computable (Proposition 10). The premisses of the support inference are quantifier-free, so, as we shall see in Theorem 5, the support inference is decidable by means of the truth-table method outlined in the previous section.

In what follows, in order to clearly distinguish the support inference from the original inference we shall adopt some notational conventions:

	Original inference	Support inference
Premisses	$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$	$\check{\varphi}_1^{X_1}, \ldots, \check{\varphi}_k^{X_k}$
Feasible region	$\mathbb{E} \overset{\text{df}}{=} \mathbb{P}[\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]$	$\check{\mathbb{E}} \overset{\text{df}}{=} \mathbb{P}[\check{\varphi}_1^{X_1}, \ldots, \check{\varphi}_k^{X_k}]$
n-entropy maximiser	P^n	$\check{P}^n$
Maximal entropy function	$P^\dagger \in \operatorname{maxent} \mathbb{E}$	$\check{P}^\dagger \in \operatorname{maxent} \check{\mathbb{E}}$

We now introduce a concept that is key to the results of this paper.

Definition 18 *(Support-satisfiability).* Premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ are said to have *satisfiable support* or to be *support-satisfiable* if and only if $\check{\mathbb{E}} \neq \emptyset$, i.e., if and only if there exists a probability function $P \in \mathbb{P}$ such that for all $i = 1, \ldots, k$, $P(\check{\varphi}_i) \in X_i$. An inference has *satisfiable support* or is *support-satisfiable* if its premisses have satisfiable support.

As we shall see in Proposition 13, premisses have satisfiable support as long as they do not force an inductive tautology to have probability less than one, or equivalently, an inductive contradiction to have probability greater than zero. Recall from Definition 11 that the inductive tautologies include not just the deductive tautologies but all the sentences with measure 1, e.g., $\exists x Ux$. Similarly, the inductive contradictions are the measure-zero sentences, e.g., $\forall x Ux$.

If an inference has satisfiable support then the support problem, being quantifier-free, admits a unique maximal entropy function $\check{P}^\dagger \in \operatorname{maxent} \check{\mathbb{E}}$ (Proposition 4) and the class of such inferences with quantifier-free conclusions is decidable (Theorem 1). The main task of the paper is to show that this phenomenon carries over to the original inference itself: if the original inference has satisfiable support then it is reducible to the support inference, $P^\dagger = \check{P}^\dagger$ (Theorem 3), and moreover, the class of all inferences with satisfiable support is decidable (Theorem 5).

This is perhaps surprising, because the feasible regions of the two inferences can be very different, even where the original inference is support-satisfiable:

Example 6 *($\mathbb{E}$ and $\check{\mathbb{E}}$).* Consider the single premiss $\varphi^{0.5} = (Ut_1 \vee \forall x Vx)^{0.5}$. Then $\check{\varphi} = Ut_1$, so $P \in \check{\mathbb{E}}$ if and only if $P(Ut_1) = 0.5$. But $P(Ut_1) = 0.5$ does not entail that $P(\varphi) = 0.5$. So, $\check{\mathbb{E}} \not\subseteq \mathbb{E}$. Furthermore, for $Q \in \mathbb{P}$ with $Q(Ut_1) = 0$ and $Q(\forall x Vx) = 0.5$, we have that $Q \in \mathbb{E}$. However, $Q \notin \check{\mathbb{E}}$. So, $\mathbb{E} \not\subseteq \check{\mathbb{E}}$.

In the remainder of this section, we explore some properties of the support propositions $\check{\varphi}$. In particular, in Proposition 7 we see that $\check{\varphi}$ is logically equivalent to $\varphi^{N_\varphi} \stackrel{\mathrm{df}}{=} \bigvee\{\omega_{N_\varphi} \in \Omega_{N_\varphi} : P_=(\omega_{N_\varphi} \wedge \varphi) > 0\}$, so any probability function gives these two propositions the same probability. Before proceeding to Proposition 7, we require a definition and two lemmas.

Definition 19 *(Constant exchangeability).* Let $\theta(x_1, x_2, \ldots, x_l)$ be a formula of $\mathcal{L}$ that does not contain constants. A probability function P on $\mathcal{SL}$ satisfies *constant exchangeability* if and only if for all such θ and all sets of pairwise distinct constants $t_1, t_2, \ldots, t_l$, and $t'_1, t'_2, \ldots, t'_l$,

$$P(\theta(t_1, t_2, \ldots, t_l)) = P(\theta(t'_1, t'_2, \ldots, t'_l)) \ .$$

Equivalently, constant exchangeability holds if and only if for all $n \in \mathbb{N}$ and all n-states $\omega, \nu \in \Omega_n$, if ω can be obtained from ν by a permutation of the first n constants then $P(\omega) = P(\nu)$.

Lemma 1. *Suppose probability function P satisfies constant exchangeability. If the following identity holds for all quantifier-free sentences then it holds for all sentences $\varphi, \psi \in S\mathcal{L}$:*

$$P(\varphi \wedge \psi|\lambda) = P(\varphi|\lambda) \cdot P(\psi|\lambda), \tag{4}$$

where λ is any contingent conjunction of closed literals that contains all the atomic propositions that occur in both φ and ψ.

Proof. The result follows by a straightforward adaptation of the proof of [29, Corollary 6.2] and proceeds by induction on the quantifier complexity of $\varphi \wedge \psi$ when written in prenex normal form.

The result holds by assumption when $\varphi \wedge \psi$ is quantifier free. For the induction step it is sufficient to consider

$$\exists x_1, \ldots, x_r \theta(x_1, \ldots, x_r, \vec{t}) \wedge \exists x_1, \ldots, x_s \eta(x_1, \ldots, x_s, \vec{t'}) \tag{5}$$

where all constants appearing in both $\vec{t}$ and $\vec{t'}$ are included in $\{t_1, \ldots, t_l\}$. To see that this is sufficient notice that if (4) holds for sentences of this form then,

$$\begin{aligned}
P(\exists \vec{x}\theta \wedge \forall \vec{y}\eta \,|\, \lambda) &= P(\exists \vec{x}\theta \,|\, \lambda) - P(\exists \vec{x}\theta \wedge \neg\forall \vec{y}\eta \,|\, \lambda) \\
&= P(\exists \vec{x}\theta \,|\, \lambda) - P(\exists \vec{x}\theta \wedge \exists \vec{y}\neg\eta \,|\, \lambda) \\
&= P(\exists \vec{x}\theta \,|\, \lambda) - (P(\exists \vec{x}\theta \,|\, \lambda) \cdot P(\exists \vec{y}\neg\eta \,|\, \lambda)) \\
&= P(\exists \vec{x}\theta \,|\, \lambda) - (P(\exists \vec{x}\theta \,|\, \lambda) \cdot (1 - P(\forall \vec{y}\eta \,|\, \lambda))) \\
&= P(\exists \vec{x}\theta \,|\, \lambda) - P(\exists \vec{x}\theta \,|\, \lambda) + P(\exists \vec{x}\theta \,|\, \lambda) \cdot P(\forall \vec{y}\eta \,|\, \lambda) \\
&= P(\exists \vec{x}\theta \,|\, \lambda) \cdot P(\forall \vec{y}\eta \,|\, \lambda)
\end{aligned}$$

and,

$$\begin{aligned}
P(\forall\vec{x}\theta \wedge \forall\vec{y}\eta \mid \lambda) &= 1 - P(\exists\vec{x}\neg\theta \vee \exists\vec{y}\neg\eta \mid \lambda)\\
&= 1 - P(\exists\vec{x}\neg\theta \mid \lambda) - P(\exists\vec{y}\neg\eta \mid \lambda) + P(\exists\vec{x}\neg\theta \wedge \exists\vec{y}\neg\, \eta \mid \lambda)\\
&= 1 - P(\exists\vec{x}\neg\theta \mid \lambda) - P(\exists\vec{y}\neg\eta \mid \lambda) + P(\exists\vec{x}\neg\theta \mid \lambda)\cdot P(\exists\vec{y}\neg\eta \mid \lambda)\\
&= P(\forall\vec{x}\theta \mid \lambda) + P(\forall\vec{y}\eta \mid \lambda) - 1 + (1 - P(\forall\vec{x}\theta \mid \lambda))\cdot(1 - P(\forall\vec{y}\eta \mid \lambda))\\
&= P(\forall\vec{x}\theta \mid \lambda)\cdot P(\forall\vec{y}\eta \mid \lambda)\ .
\end{aligned}$$

To show (4) for sentences of the form in (5), let $u_1, u_2, u_3, \ldots$ be distinct constants containing those in $\vec{t}$ and $u_1', u_2', u_3', \ldots$ distinct constants containing those in $\vec{t'}$ such that $\{u_1, u_2, u_3, \ldots\}$ and $\{u_1', u_2', u_3', \ldots\}$ are disjoint except for the constants shared between $\vec{t}$ and $\vec{t'}$.

By [29, Lemma 6.1],

$$\lim_{n\to\infty} P\left(\left(\bigvee_{i_1,\ldots,i_r\le n}\theta(u_{i_1}, u_{i_2}, \ldots, u_{i_r}, \vec{t})\right) \leftrightarrow \exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\,|\lambda\right) = 1$$

and

$$\lim_{n\to\infty} P\left(\left(\bigvee_{i_1,\ldots,i_s\le n}\eta(u'_{i_1}, u'_{i_2}, \ldots, u'_{i_s}, \vec{t'})\right) \leftrightarrow \exists x_1,\ldots,x_s\eta(x_1,\ldots,x_s,\vec{t'})\,|\lambda\right) = 1.$$

Then for every $\epsilon > 0$ there is N large enough such that for all $n \ge N$

$$P\left(\left(\bigvee_{i_1,\ldots,i_r\le n}\theta(u_{i_1}, u_{i_2}, \ldots, u_{i_r}, \vec{t})\right) \leftrightarrow \exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\,|\lambda\right) > 1 - \frac{\epsilon}{4}$$

and

$$P\left(\left(\bigvee_{i_1,\ldots,i_s\le n}\eta(u'_{i_1}, u'_{i_2}, \ldots, u'_{i_s}, \vec{t'})\right) \leftrightarrow \exists x_1,\ldots,x_s\eta(x_1,\ldots,x_t,\vec{t'})\,|\lambda\right) > 1 - \frac{\epsilon}{4}$$

by [29, Lemma 3.7],

$$\begin{gathered}
P\left(\exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t}) \wedge \exists x_1,\ldots,x_s\eta(x_1,\ldots,x_s,\vec{t'})\,|\lambda\right) -\\
P\left(\bigvee_{i_1,\ldots,i_r\le n}\theta(u_{i_1}, u_{i_2}, \ldots, u_{i_r}, \vec{t}) \wedge \bigvee_{i_1,\ldots,i_s\le n}\eta(u'_{i_1}, u'_{i_2}, \ldots, u'_{i_s}, \vec{t'})\,|\lambda\right) < \frac{\epsilon}{2}.
\end{gathered}$$

But

$$P\left(\bigvee_{i_1,\ldots,i_r\le n}\theta(u_{i_1}, u_{i_2}, \ldots, u_{i_r}, \vec{t}) \wedge \bigvee_{i_1,\ldots,i_s\le n}\eta(u'_{i_1}, u'_{i_2}, \ldots, u'_{i_s}, \vec{t'})\,|\lambda\right)$$

equals

$$P\left(\bigvee_{i_1,\ldots,i_r\leq n}\theta(u_{i_1},u_{i_2},\ldots,u_{i_r},\vec{t})\,|\lambda\right)\cdot P\left(\bigvee_{i_1,\ldots,i_s\leq n}\eta(u'_{i_1},u'_{i_2},\ldots,u'_{i_s},\vec{t'})\,|\lambda\right)$$

by the induction hypothesis, and taking n large enough we have:

$$P\left(\bigvee_{i_1,\ldots,i_r\leq n}\theta(u_{i_1},u_{i_2},\ldots,u_{i_r},\vec{t})\,|\lambda\right)\cdot P\left(\bigvee_{i_1,\ldots,i_s\leq n}\eta(u'_{i_1},u'_{i_2},\ldots,u'_{i_s},\vec{t'})\,|\lambda\right)-$$
$$P\left(\exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\,|\lambda\right)\cdot P\left(\exists x_1,\ldots,x_s\eta(x_1,\ldots,x_s,\vec{t'})\,|\lambda\right)<\frac{\epsilon}{2}$$

and thus

$$\begin{aligned}&P\left(\exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\wedge\exists x_1,\ldots,x_s\eta(x_1,\ldots,x_s,\vec{t'})\,|\lambda\right)\\&\quad-P\left(\exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\,|\lambda\right)\cdot P\left(\exists x_1,\ldots,x_s\eta(x_1,\ldots,x_s,\vec{t'})\,|\lambda\right)\\&<\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon,\end{aligned}$$

which gives the required result. □

In particular, Lemma 1 applies to the equivocator function:

Corollary 2. *For all $\varphi,\psi\in S\mathcal{L}$,*

$$P_{=}(\varphi\wedge\psi|\lambda)=P_{=}(\varphi|\lambda)\cdot P_{=}(\psi|\lambda),$$

where λ is any contingent conjunction of closed literals that contains all the atomic propositions that occur in both φ and ψ.

Consequently, for all $\varphi\in S\mathcal{L}$ and all $\omega\in\Omega_{N_\varphi}$,

$$P_{=}(\varphi|\omega)=P_{=}(\varphi\wedge\varphi|\omega)=P_{=}(\varphi|\omega)^2\in\{0,1\}\ .$$

Proof. The first part follows immediately if φ or ψ has measure zero; the second part is also trivial for measure-zero φ. So suppose otherwise.

Consider $\varphi,\psi\in S\mathcal{L}$ and λ as in Lemma 1 and let $M:=\max\{N_\varphi,N_\psi,N_\lambda\}$. Since λ is contingent we do not divide by zero in the following equation,

$$P_{=}(\varphi\wedge\psi|\lambda)=\frac{|\{\omega\in\Omega_M\ :\ \omega\vDash\varphi\wedge\psi\wedge\lambda\}|/|\Omega_M|}{|\{\omega\in\Omega_M\ :\ \omega\vDash\lambda\}|/|\Omega_M|}\ .$$

Let us now split M-states ω into four conjunctions, one mentioning the atomic propositions of λ, one mentioning those unique to φ, one mentioning those unique to ψ and one mentioning the remainder, $\omega_\varphi,\omega_\psi,\omega^+$. For all $\omega\in\Omega_M$ we shall consider here we hence have $\omega\equiv\lambda\wedge\omega_\varphi\wedge\omega_\psi\wedge\omega^+$.[9]

Also put

$$\Omega_{\varphi\wedge\lambda}:=\{\nu\in\Xi_{\omega_\varphi}\ :\ \nu\wedge\lambda\vDash\varphi\wedge\lambda\}$$

[9] Note that $\varphi\equiv\lambda$ is possible. If φ had measure zero, so would be the measure of λ and we could not conditionalise on λ. If $\lambda\vDash\varphi$, then ω_λ is an empty conjunction, a tautology.

$$\Omega_{\psi\wedge\lambda} := \{\nu \in \Xi_{\omega_\psi} \ : \ \nu \wedge \lambda \vDash \psi \wedge \lambda\}$$
$$\Omega_{\varphi\wedge\psi\wedge\lambda} := \{\nu \in \Xi_{\omega_\varphi \wedge \omega_\psi} \ : \ \nu \wedge \lambda \vDash \varphi \wedge \psi \wedge \lambda\} \ .$$

Let us now observe that $|\Omega_{\varphi\wedge\psi\wedge\lambda}| = |\Omega_{\varphi\wedge\lambda}||\Omega_{\psi\wedge\lambda}|$. Let $l_\varphi, l_\chi, l_{\omega^+}$ denote the number of conjuncts in these conjunctions and $|\omega_\varphi| := 2^{l_\varphi}, |\omega_\psi| := 2^{l_\psi}, |\omega^+| := 2^{l_{\omega^+}}$. Then,

$$\begin{aligned}
P_=(\varphi \wedge \psi|\lambda) =& \frac{|\{\omega \in \Omega_M \ : \ \omega \vDash \varphi \wedge \psi \wedge \lambda\}|/|\Omega_M|}{|\{\omega \in \Omega_M \ : \ \omega \vDash \lambda\}|/|\Omega_M|} \\
=& \frac{|\omega^+| \cdot |\Omega_{\varphi\wedge\psi\wedge\lambda}|/|\Omega_M|}{|\omega^+| \cdot |\omega_\varphi| \cdot |\omega_\psi|/|\Omega_M|} \\
=& \frac{|\omega^+|^2 \cdot |\Omega_{\varphi\wedge\lambda}| \cdot |\omega_\psi| \cdot |\Omega_{\psi\wedge\lambda}| \cdot |\omega_\varphi|/|\Omega_M|}{|\omega^+|^2 \cdot |\omega_\varphi|^2 \cdot |\omega_\psi|^2/|\Omega_M|} \\
=& \frac{|\omega^+| \cdot |\Omega_{\varphi\wedge\lambda}| \cdot |\omega_\psi|/|\Omega_M|}{|\omega^+| \cdot |\omega_\varphi| \cdot |\omega_\psi|/|\Omega_M|} \cdot \frac{|\omega^+| \cdot |\Omega_{\psi\wedge\lambda}| \cdot |\omega_\varphi|/|\Omega_M|}{|\omega^+| \cdot |\omega_\varphi| \cdot |\omega_\psi|/|\Omega_M|} \\
=& \frac{|\{\omega \in \Omega_M \ : \ \omega \vDash \varphi \wedge \lambda\}|/|\Omega_M|}{|\{\omega \in \Omega_M \ : \ \omega \vDash \lambda\}|/|\Omega_M|} \cdot \frac{|\{\omega \in \Omega_M \ : \ \omega \vDash \psi \wedge \lambda\}|/|\Omega_M|}{|\{\omega \in \Omega_M \ : \ \omega \vDash \lambda\}|/|\Omega_M|} \\
=& P_=(\varphi|\lambda) P_=(\psi|\lambda) \ .
\end{aligned}$$

Since φ, ψ were arbitrary, this holds for all quantifier-free sentences φ, ψ and all such λ. Consequently, the assumptions of Lemma 1 hold for $P_=$.

Letting $\psi = \varphi$ and recalling that λ may be a state completes the proof. □

Example 7. Note that $\varphi = \exists x Ux \wedge Vt_1$ and $\psi = \exists x Ux \wedge \neg Vt_1$ share the atomic proposition Vt_1. In particular, they do not share an atomic proposition mentioning U, since the literal Ux mentions a variable. So, $\Xi_\psi = \{Vt_1\}$. Observe furthermore that, $P_=(\varphi \wedge \psi) = 0 < 0.25 = P_=(\varphi) \cdot P_=(\psi)$. But for both $\lambda = Vt_1$ and $\lambda = \neg Vt_1$ we have $P_=(\varphi \wedge \psi|\lambda) = 0 = P_=(\varphi|\lambda) \cdot P(\psi|\lambda)$. Note that λ may contain further literals such as Vt_2, $\neg Ut_1$ and Ut_2.

Some properties of probability 1 sentences will be useful:

Lemma 2. *For any $\psi, \chi \in S\mathcal{L}$:*

1. *If $P(\psi) = 1$, then $P(\alpha \wedge \psi) = P(\alpha)$ for all $\alpha \in S\mathcal{L}$.*
2. *If $P(\psi \leftrightarrow \chi) = 1$, then $P(\psi) = P(\chi)$.*
3. *If $P(\psi \leftrightarrow \chi) = 1$ and $P(\psi) > 0$, then $P(\cdot|\psi) = P(\cdot|\chi)$.*
4. *If $P(\psi \leftrightarrow \chi) = 1$, then $P(\alpha \wedge \psi) = P(\alpha \wedge \chi)$ for all $\alpha \in S\mathcal{L}$.*

Proof. (1) We have that:

$$P(\alpha) = P(\alpha \wedge \psi) + P(\alpha \wedge \neg\psi) \leq P(\alpha \wedge \psi) + P(\neg\psi) = P(\alpha \wedge \psi) \ .$$

Since $P(\alpha) \geq P(\alpha \wedge \psi)$ must also hold, we find that $P(\alpha \wedge \psi) = P(\alpha)$.

(2) Since $P(\psi \wedge \neg\chi) = 0 = P(\neg\psi \wedge \chi)$,

$$P(\psi) = P(\psi \wedge \chi) + P(\psi \wedge \neg\chi) + P(\neg\psi \wedge \chi) = P(\chi) + P(\psi \wedge \neg\chi) = P(\chi) \ .$$

(3) We first observe that

$$\begin{aligned}P(\alpha \wedge \psi) &= P(\alpha \wedge \psi \wedge \chi) + P(\alpha \wedge \psi \wedge \neg\chi) + P(\alpha \wedge \neg\psi \wedge \chi) \\ &= P(\alpha \wedge \chi) + P(\alpha \wedge \psi \wedge \neg\chi) = P(\alpha \wedge \chi) \ .\end{aligned}$$

Since $P(\psi) = P(\chi)$ by (2), $P(\alpha|\psi) = P(\alpha|\chi)$.

(4) Let us first assume that $P(\psi) = 0$. Then

$$1 = P(\psi \leftrightarrow \chi) = P(\psi \wedge \chi) + P(\neg\psi \wedge \neg\chi) = P(\neg\psi \wedge \neg\chi) \stackrel{(1)}{=} P(\neg\chi) \ .$$

So, $P(\chi) = 0$. This establishes that $P(\psi \wedge \alpha) = 0 = P(\chi \wedge \alpha)$ for all $\alpha \in S\mathcal{L}$.

Finally, assume that $P(\psi) > 0$. $P(\alpha|\psi) = P(\alpha|\chi)$ for all $\alpha \in S\mathcal{L}$ by (3). Hence, for all $\alpha \in S\mathcal{L}$

$$P(\psi \wedge \alpha) = P(\alpha|\psi)P(\psi) = P(\alpha|\chi)P(\psi) = P(\alpha|\chi)P(\chi) = P(\chi \wedge \alpha) \ ,$$

where the penultimate equality follows from $P(\psi \leftrightarrow \chi) = 1$ entailing $P(\psi) = P(\chi)$ (2). □

Proposition 7. *For all $\varphi \in S\mathcal{L}$ and $n \geq N_\varphi$, the following two sentences are logically equivalent:*

$$\check{\varphi} := \bigvee\{\xi \in \Xi_\varphi : P_=(\xi \wedge \varphi) > 0\} \qquad \varphi^n := \bigvee\{\omega \in \Omega_n : P_=(\omega \wedge \varphi) > 0\} \ . \tag{6}$$

Since $\check{\varphi}$ and φ^n are logically equivalent, each probability function must give them the same probability. We can thus switch freely between them in the sense of the above Lemma. $\check{\varphi}$ is the most economical representative of this class of equivalent propositions insofar as it involves fewest atomic propositions. This provides computational advantages that we shall exploit in §7.

Proof. If φ has zero measure, both disjunctions are empty and the result follows trivially.

Let us now assume that φ has positive measure and that $n \geq N_\varphi$.

If $P_=(\varphi \wedge \xi) = 0$, then $P_=(\omega \wedge \xi) \leq P_=(\varphi \wedge \xi) = 0$ for all n-states ω with $\omega \vDash \xi$. So, if $\xi \in \Xi_\varphi$ then no n-state $\omega \in \Omega_n$ entailing ξ is such that $P_=(\varphi \wedge \omega) > 0$.

Now let $P_=(\varphi \wedge \xi) > 0$ and ω^+ be the conjunction of the conjuncts $\omega \vDash \xi$ that are not entailed by ξ, so $\omega \equiv \omega^+ \wedge \xi$. Then,

$$\begin{aligned}0 < P_=(\varphi \wedge \xi \wedge \omega^+) &\stackrel{P_=(\xi)>0}{=} P_=(\varphi \wedge \omega^+|\xi) \cdot P_=(\xi) \\ &\stackrel{\text{Cor. 2}}{=} P_=(\varphi|\xi) \cdot P_=(\omega^+|\xi) \cdot P_=(\xi) \\ &= P_=(\varphi \wedge \xi) \cdot \frac{2^m}{|\Omega_n|} \ ,\end{aligned}$$

where m is the number of atomic propositions that feature in ξ.

Note that $P_=(\varphi \wedge \xi) \cdot \frac{2^m}{|\Omega_n|}$ does not depend on ω^+: it is a constant. And since

$$0 < P_=(\varphi \wedge \xi) = \sum_{\substack{\omega^+ \\ \xi \wedge \omega^+ \in \Omega_n}} P_=(\varphi \wedge \xi \wedge \omega^+) = P_=(\varphi \wedge \xi) \cdot \frac{2^m}{|\Omega_n|} \ ,$$

this constant cannot be zero. This shows that for all ω^+, $P_=(\varphi \wedge \xi \wedge \omega^+) > 0$. This in turn implies that the n-state $\xi \wedge \omega^+$ entails $\bigvee\{\omega \in \Omega_n : P_=(\omega \wedge \varphi) > 0\}$, as claimed. □

Lemma 3. *For all $\varphi, \psi \in S\mathcal{L}$,*

$$\begin{aligned}
(i)\ & \varphi \check{\wedge} \psi = \check{\varphi} \wedge \check{\psi} \\
(ii)\ & \varphi \check{\vee} \psi = \check{\varphi} \vee \check{\psi} \\
(iii)\ & \check{\neg\varphi} = \neg\check{\varphi} \\
(iv)\ & P_=(\check{\varphi}) = P_=(\varphi)\ .
\end{aligned}$$

Proof. (i) Let $M = \max\{N_\varphi, N_\psi\}$. By [16, Lemma 27], for all $n \geq M$ and $\omega \in \Omega_n$, we have $P_=(\varphi \wedge \psi|\omega) = P_=(\varphi|\omega) \cdot P_=(\psi|\omega)$. Hence $P_=(\omega \wedge \varphi \wedge \psi) > 0$ if and only if $P_=(\varphi \wedge \omega) > 0$ and $P_=(\psi \wedge \omega) > 0$. Then $\{\omega \in \Omega_M : P_=(\omega \wedge \varphi \wedge \psi) > 0\} = \{\omega \in \Omega_M : P_=(\omega \wedge \varphi) > 0\} \cap \{\omega \in \Omega_M : P_=(\omega \wedge \psi) > 0\}$. Since for distinct $\omega_i, \omega_j \in \Omega_M$, $\models \neg(\omega_i \wedge \omega_j)$, we have:

$$\begin{aligned}
\varphi \check{\wedge} \psi &= \bigvee\{\omega \in \Omega_M : P_=(\omega \wedge \varphi \wedge \psi) > 0\} \\
&= \bigvee\{\omega \in \Omega_M : P_=(\omega \wedge \varphi) > 0\} \wedge \bigvee\{\omega \in \Omega_M : P_=(\omega \wedge \psi) > 0\} \\
&= \check{\varphi} \wedge \check{\psi}\ .
\end{aligned}$$

(ii) Notice that

$$\begin{aligned}
\{\omega \in \Omega_M : P_=(\omega \wedge (\varphi \vee \psi)) > 0\} &= \{\omega \in \Omega_M : P_=((\omega \wedge \varphi) \vee (\omega \wedge \psi)) > 0\} \\
&= \{\omega \in \Omega_M : P_=(\omega \wedge \varphi) > 0\} \cup \{\omega \in \Omega_M : P_=(\omega \wedge \psi) > 0\}
\end{aligned}$$

and so,

$$\begin{aligned}
\varphi \check{\vee} \psi &= \bigvee\{\omega \in \Omega_M : P_=(\omega \wedge (\varphi \vee \psi)) > 0\} \\
&= \bigvee\{\omega \in \Omega_M : P_=(\omega \wedge \varphi) > 0\} \vee \bigvee\{\omega \in \Omega_M : P_=(\omega \wedge \psi) > 0\} \\
&= \check{\varphi} \vee \check{\psi}\ .
\end{aligned}$$

(iii) See [16, Proposition 40].
(iv) Observe that

$$\begin{aligned}
P_=(\varphi) &= \sum_{\substack{\xi \in \Xi_\varphi \\ P_=(\xi \wedge \varphi) > 0}} P_=(\xi \wedge \varphi) \\
&\leq \sum_{\substack{\xi \in \Xi_\varphi \\ P_=(\xi \wedge \varphi) > 0}} P_=(\xi) \\
&= P_=(\check{\varphi})
\end{aligned}$$

and then similarly find that

$$P_=(\neg\varphi) \leq P_=(\check{\neg\varphi}) \overset{(iii)}{=} P_=(\neg\check{\varphi})$$

Since $\langle \varphi, \neg\varphi \rangle$ and $\langle \check{\varphi}, \neg\check{\varphi} \rangle$ are partitions it must be the case that $P_=(\varphi) = P_=(\check{\varphi})$. $\square$

5. The reduction theorem

The main purpose of this section is to show that, where the premisses have satisfiable support, $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx} \psi^Y$ if and only if $\check{\varphi}_1^{X_1}, \ldots, \check{\varphi}_k^{X_k} \mathrel{|\!\!\approx} \check{\psi}^Y$ (Theorem 4). The latter problem is quantifier-free and decidable.

In the remainder of the paper, we shall move freely between the representation of $\check{P}^\dagger$ of Equation (3) and the following representation of $\check{P}^\dagger$, which follows directly from Proposition 4:

Observation 1. If $\check{\mathbb{E}} \neq \emptyset$ for quantifier-free premisses, then maxent $\check{\mathbb{E}} = \{\check{P}^\dagger\}$ with

$$\check{P}^\dagger(\cdot) \stackrel{\text{df}}{=} \sum_{\omega \in \Omega_M} \check{P}^M(\omega) \cdot P_{=}(\cdot|\omega) \ ,$$

where $M = \max\{N_{\varphi_i} : i = 1, \ldots, k\}$, and $\check{P}^M$ is any M-entropy maximiser in $\check{\mathbb{E}}$.

The use of this latter representation will allow us to assess the entropy of $\check{P}^\dagger$ more directly and will allow us to apply results of [16]. Note that $\check{P}^\dagger$ equivocates beyond M, in the sense of the following definition:

Definition 20 *(Equivocation beyond N).* Given some $N \in \mathbb{N}$, we say that a probability function $P \in \mathbb{P}$ *equivocates beyond N* if and only if for all $n \geq N$ and all $\omega_n \in \Omega_n$, $P(\omega_n) = P(\omega_N) \cdot \frac{|\Omega_N|}{|\Omega_n|}$, where ω_N is the restriction of ω_n to $\mathcal{L}_N$, that is the unique N-state such that $\omega_n \vDash \omega_N$.

We now show that equivocation beyond the support problem fixes conditional probabilities of quantified sentences.

Lemma 4. *If P equivocates beyond N, then for all $\omega_N \in \Omega_N$ such that $P(\omega_N) > 0$ and for all sentences $\varphi \in S\mathcal{L}$, $P(\varphi|\omega_N) = P_{=}(\varphi|\omega_N)$.*

Proof. First note that $P(\nu|\omega_N) = P_{=}(\nu|\omega_N)$ for all N-states $\nu \in \Omega_N$ with $P(\nu) > 0$. Since these probability functions also both equivocate beyond N, they agree on all quantifier-free sentences. By Remark 2, they are thus equal. □

5.1. Point-valued premisses

In this subsection we consider inferences of the form

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx} \psi^Y,$$

where $X_i = c_i \in [0, 1]$ for $i = 1 \ldots k$. For the rest of this subsection let $\mathbb{E} = \{\varphi_1^{c_1}, \ldots, \varphi_k^{c_k}\}$ and $M = \max\{N_{\varphi_i} : i = 1, \ldots, k\}$.

Example 8 *(Support-satisfiability of a single premiss).* If $P_{=}(\varphi) > 0$, then the premiss φ^1 is support-satisfiable, since $P_{=}(\varphi) = P_{=}(\check{\varphi})$ (Lemma 3 (iv)) and $P_{=}(\check{\varphi}|\check{\varphi}) = 1$.

If $0 < P_{=}(\varphi) < 1$, then the premiss φ^c for $0 < c < 1$ is support-satisfiable, since $Q(\check{\varphi}) := c \cdot P_{=}(\check{\varphi}|\check{\varphi}) + (1 - c) \cdot P_{=}(\check{\varphi}|\neg\check{\varphi}) = c + 0 = c$.

Consequently, given categorical premisses $\varphi_1^1, \ldots, \varphi_k^1$, if $P_{=}(\varphi) > 0$ for $\varphi := \bigwedge_{i=1}^k \varphi_i$, then the inference is support-satisfiable.

Proposition 8. *If $\check{\mathbb{E}} \neq \emptyset$, then $\check{P}^\dagger \in \mathbb{E}$.*

In particular, if the premisses are support satisfiable then they are also satisfiable.

Proof. For all premiss sentences φ_i,

$$\begin{aligned}
\check{P}^{\dagger}(\varphi_i) &= \sum_{\omega\in\Omega_M} \check{P}^M(\omega)\cdot P_=(\varphi_i|\omega) \\
&= \sum_{\substack{\omega\in\Omega_M \\ P_=(\omega\wedge\varphi_i)>0}} \check{P}^M(\omega)\cdot P_=(\varphi_i|\omega) \\
&= \sum_{\substack{\omega\in\Omega_M \\ P_=(\omega\wedge\varphi_i)>0}} \check{P}^M(\omega)\cdot 1 \\
&= \sum_{\substack{\omega\in\Omega_M \\ \omega\models\check{\varphi}_i}} \check{P}^{\dagger}(\omega) \\
&= \check{P}^M(\check{\varphi}_i) \\
&= c_i \ ,
\end{aligned}$$

where the third equality follows from [16, Proposition 28] $(P_=(\varphi_i|\omega) = P_=(\omega|\omega))$. □

Proposition 9 *(Entropy of $\check{P}^{\dagger}$). For all $n \geq M$,*

$$H_n(\check{P}^{\dagger}) = H_M(\check{P}^{\dagger}) + \log(|\Omega_n|) - \log(|\Omega_M|) \ .$$

Proof.

$$\begin{aligned}
H_n(\check{P}^{\dagger}) &= -\sum_{\omega\in\Omega_M}\sum_{\substack{\zeta\in\Omega_n \\ \zeta\models\omega}} \check{P}^M(\omega)\cdot P_=(\zeta|\omega)\cdot\log(\check{P}^M(\omega)\cdot P_=(\zeta|\omega)) \\
&= -\sum_{\omega\in\Omega_M} \check{P}^M(\omega)\cdot \sum_{\substack{\zeta\in\Omega_n \\ \zeta\models\omega}} \frac{|\Omega_M|}{|\Omega_n|}\cdot\left[\log(\check{P}^M(\omega)) + \log\left(\frac{|\Omega_M|}{|\Omega_n|}\right)\right] \\
&= -\sum_{\omega\in\Omega_M} \check{P}^M(\omega)\cdot\left[\log(\check{P}^M(\omega)) + \log\left(\frac{|\Omega_M|}{|\Omega_n|}\right)\right] \\
&= \log(|\Omega_n|) - \log(|\Omega_M|) - \sum_{\omega\in\Omega_M} \check{P}^M(\omega)\cdot\log(\check{P}^M(\omega)) \\
&= H_M(\check{P}^{\dagger}) + \log(|\Omega_n|) - \log(|\Omega_M|) \ . \quad \square
\end{aligned}$$

Lemma 5. *Let $P \in \mathbb{E}\setminus\check{\mathbb{E}}$. Then there are constants $g_1, g_2 > 0$, a strictly positive sequence v_n diverging to infinity, and some N such that $H_n(P) < g_1 + \left(1 - \frac{g_2\cdot v_n}{\log(|\Omega_n|)}\right)\log(|\Omega_n|)$ for all $n \geq N$.*

Proof. Since $P \notin \check{\mathbb{E}}$ there has to be at least one premiss φ_i such that $P(\check{\varphi}_i) \notin X_i$ and $P(\varphi_i) \in X_i$. Considering $\neg\varphi_i^{[1-X_i^+,1-X_i^-]}$ if necessary, there then exists some i such that $P(\check{\varphi}_i)$ is less than the minimum of the interval X_i, X_i^-. So,

$$X_i^- \leq P(\varphi_i) = \sum_{\omega\in\Omega_M} P(\omega\wedge\varphi_i) = \sum_{\substack{\omega\in\Omega_M \\ P_=(\varphi_i\wedge\omega)>0}} P(\omega\wedge\varphi_i) + \sum_{\substack{\omega\in\Omega_M \\ P_=(\varphi_i\wedge\omega)=0}} P(\omega\wedge\varphi_i)$$

Table 1
Example of the $S_{n,l}$ and the N_l. The shaded cells indicate the element in the sequence $(S_{n,l})_{n\in\mathbb{N}}$ after which $\frac{|S_{n,l}|}{|\Omega_n|} < \frac{1}{l}$. Since $N_2 = 2$ we have $S_1 = \Omega_1$ and $S_2 = \Omega_2$. The **bold font** indicates that this cell is part of S_n, for these cells $S_n = S_{n,l}$. The sequence $(S_{n,3})_{n\in\mathbb{N}}$ is skipped and does not appear in S_n.

l	$S_{n,l}$	N_l	$\forall n \geq N_l$	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7
2	$S_{n,2}$	$N_2 = 2$	$\frac{\|S_{n,2}\|}{\|\Omega_n\|} < \frac{1}{2}$	$S_{1,2}$	$S_{2,2}$	$\mathbf{S_{3,2}}$	$S_{4,2}$	$S_{5,2}$	$S_{6,2}$	$S_{7,2}$
3	$S_{n,3}$	$N_3 = 3$	$\frac{\|S_{n,3}\|}{\|\Omega_n\|} < \frac{1}{3}$	$S_{1,3}$	$S_{2,3}$	$S_{3,3}$	$S_{4,3}$	$S_{5,3}$	$S_{6,3}$	$S_{7,3}$
4	$S_{n,4}$	$N_4 = 3$	$\frac{\|S_{n,4}\|}{\|\Omega_n\|} < \frac{1}{4}$	$S_{1,4}$	$S_{2,4}$	$S_{3,4}$	$\mathbf{S_{4,4}}$	$\mathbf{S_{5,4}}$	$\mathbf{S_{6,4}}$	$S_{7,4}$
5	$S_{n,5}$	$N_5 = 6$	$\frac{\|S_{n,5}\|}{\|\Omega_n\|} < \frac{1}{5}$	$S_{1,5}$	$S_{2,5}$	$S_{3,5}$	$S_{4,5}$	$S_{5,5}$	$S_{6,5}$	$\mathbf{S_{7,5}}$
6	$S_{n,6}$	$N_6 = 100$	$\frac{\|S_{n,6}\|}{\|\Omega_n\|} < \frac{1}{6}$	$S_{1,6}$	$S_{2,6}$	$S_{3,6}$	$S_{4,6}$	$S_{5,6}$	$S_{6,6}$	$S_{7,6}$
7	$S_{n,7}$	$N_7 = 101$	$\frac{\|S_{n,7}\|}{\|\Omega_n\|} < \frac{1}{7}$	$S_{1,7}$	$S_{2,7}$	$S_{3,7}$	$S_{4,7}$	$S_{5,7}$	$S_{6,7}$	$S_{7,7}$

$$= P(\check{\varphi}_i) + \sum_{\substack{\omega\in\Omega_M \\ P_=(\varphi_i\wedge\omega)=0}} P(\omega \wedge \varphi_i) \ .$$

Since $P(\check{\varphi}_i) < X_i^-$, there is thus some $\omega \in \Omega_M$ such that $P(\omega \wedge \varphi_i) > 0$ and $P_=(\omega \wedge \varphi_i) = 0$. Let $\chi = \omega \wedge \varphi_i$. In particular, $0 < P(\chi)$.

From [16, Lemma 32] it follows that for any $\epsilon \in (0, 1)$ there is some $N_\epsilon \in \mathbb{N}$ such that for all $n \geq N_\epsilon$ there exists some set S'_n of n-states with $\frac{|S'_n|}{|\Omega_n|} < \epsilon$ such that for all $n \geq N_\epsilon$, P concentrates at least probability $(1 - \epsilon) \cdot P(\chi)$ on S'_n.

So, for all natural numbers $l \geq 2$ there is $N_l \in \mathbb{N}$ such that for all $n \geq N_l$, there exists some set $S_{n,l}$ of n-states with $\frac{|S_{n,l}|}{|\Omega_n|} < \frac{1}{l}$ and P concentrates at least probability $(1 - \frac{1}{l}) \cdot P(\chi)$ on $S_{n,l}$. (We shall refer to this as Condition $*l$.) We assume without loss of generality that $N_l \leq N_{l+1}$ for all l. Clearly, no such set $S_{n,l}$ can be empty. Let us now define a sequence S_n of n-states

$$S_n := \begin{cases} \Omega_n, & \text{if } 1 \leq n \leq N_2 \\ S_{n,l}, & \text{if } N_l < n \leq N_{l+1} \text{ for all } l \geq 2 \end{cases} .$$

Our assumption of $N_l \leq N_{l+1}$ uniquely determines the value of l given fixed n.

Example 9 *(Illustrating the definition of S_n).*

$$\begin{aligned}
S_3 &= S_{3,2} & N_2 = 2 < 3 = n \leq 3 = N_3 \\
S_4 &= S_{4,4} & N_4 = 3 < 4 = n \leq 6 = N_5 \\
S_5 &= S_{5,4} & N_4 = 3 < 5 = n \leq 6 = N_5 \\
S_6 &= S_{6,4} & N_4 = 3 < 6 = n \leq 6 = N_5 \\
S_7 &= S_{7,5} & N_5 = 6 < n = 7 \leq 100 = N_6 \\
S_g &= S_{g,5} & N_5 = 6 < g \leq 100 = N_6 \text{ for all } 8 \leq g \leq 100 \\
S_{101} &= S_{101,6} & N_6 = 100 < 101 = n \leq 101 = N_7 \ .
\end{aligned}$$

Intuitively, S_n starts with Ω_n and then moves to $S_{n,2}$ as soon as $\frac{|S_{n,2}|}{|\Omega_n|} < \frac{1}{2}$ and P concentrates at least probability $0.5 \cdot P(\chi)$ on $S_{n,2}$ (see Table 1). (Call this Condition $*2$.) We move to $S_{n,l+1}$ as soon as $\frac{|S_{n,l+1}|}{|\Omega_n|} < \frac{1}{l+1}$ and P concentrates at least probability $(1 - \frac{1}{l+1}) \cdot P(\chi)$ on $S_{n,l+1}$. (Condition $*l + 1$.) Note that we might skip some l. In the above example, cells with a bold font can never appear on the left of an orange cell. So, the ratio of $|S_n|$ and $|\Omega_n|$ is never zero but falls below every strictly positive upper bound; $\left(\frac{|S_n|}{|\Omega_n|}\right)_{n\in\mathbb{N}}$ converges to zero. We define furthermore the level l of the sequence S_n, for $n \geq N_2 + 1$, by

$$\Lambda_n := \max_{l \in \mathbb{N}} \{ l \ : \ N_l < n \} \ .$$

So, $S_n = S_{n,\Lambda_n}$. Intuitively, Λ_n tells us how far in the sequences of the $S_{n,l}$ we need to go in order to define S_n. The ratio of S_n and Ω_n is thus less than $\frac{1}{\Lambda_n}$ and P concentrates at least probability $(1 - \frac{1}{\Lambda_n}) \cdot P(\chi)$ on S_n (Condition $*l$ for ever larger l.)

Notice that $\frac{|S_n|}{|\Omega_n|} \to 0$ as $n \to \infty$. Thus, dividing as much of the probability mass among n-states in $\Omega_n \setminus S_n$ and as little as possible among n-states in S_n increases n-entropy. By assumption, S_n should get at least $(1 - \frac{1}{\Lambda_n}) \cdot P(\chi)$ of probability mass. Thus the greatest entropy would be achieved by dividing the minimum permissible probability mass of $(1 - \frac{1}{\Lambda_n}) \cdot P(\chi)$ equally among n-states in S_n and the remaining probability mass equally among $\Omega_n \setminus S_n$:

$$\begin{aligned}
H_n(P) \leq & - \sum_{\omega_n \in \Omega_n \setminus S_n} \frac{P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)}{|\Omega_n| - |S_n|} \log \left(\frac{P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)}{|\Omega_n| - |S_n|} \right) \\
& - \sum_{\omega_n \in S_n} \frac{(1 - \frac{1}{\Lambda_n}) \cdot P(\chi)}{|S_n|} \log \left(\frac{(1 - \frac{1}{\Lambda_n}) \cdot P(\chi)}{|S_n|} \right) \\
\leq & - \sum_{\omega_n \in \Omega_n} \frac{P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)}{|\Omega_n|} \log \left(\frac{P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)}{|\Omega_n|} \right) \\
& - \sum_{\omega_n \in S_n} \frac{(1 - \frac{1}{\Lambda_n}) \cdot P(\chi)}{|S_n|} \log \left(\frac{(1 - \frac{1}{\Lambda_n}) \cdot P(\chi)}{|S_n|} \right) \\
= & - (P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)) \cdot \log \left(\frac{P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)}{|\Omega_n|} \right) \\
& - (1 - \frac{1}{\Lambda_n}) \cdot P(\chi) \cdot \log \left(\frac{(1 - \frac{1}{\Lambda_n}) \cdot P(\chi)}{|S_n|} \right) \ .
\end{aligned}$$

Let $g_0 := -(P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)) \cdot \log(P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)) - (1 - \frac{1}{\Lambda_n}) \cdot P(\chi) \cdot \log((1 - \frac{1}{\Lambda_n}) \cdot P(\chi))$. Note that g_0 is positive but less than the constant $g_1 := -\log(P(\neg\chi)) + 1 > 0$ (since $-x\log(x) < 1$ for all $x \in [0,1]$),[10] which only depends on $P(\chi)$ (and thus also on $P(\neg\chi)$) but not on $\frac{1}{\Lambda_n}$ nor on n. For the others terms, which depend on n, we find

$$\begin{aligned}
& (P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)) \cdot \log(|\Omega_n|) + (1 - \frac{1}{\Lambda_n}) \cdot P(\chi) \cdot \log(|S_n|) \\
= & \log(|\Omega_n|) \cdot [P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi) + (1 - \frac{1}{\Lambda_n}) \cdot P(\chi) \cdot \frac{\log(|S_n|)}{\log(|\Omega_n|)}] \\
= & \log(|\Omega_n|) \cdot [P(\neg\chi) + P(\chi)(\frac{1}{\Lambda_n} + (1 - \frac{1}{\Lambda_n}) \frac{\log(|S_n|)}{\log(|\Omega_n|)})] \ .
\end{aligned}$$

Letting $u_n := \frac{|S_n|}{|\Omega_n|}$ we find that for all large enough $n \in \mathbb{N}$,

$$\begin{aligned}
& (P(\neg\chi) + \frac{1}{\Lambda_n} P(\chi)) \cdot \log(|\Omega_n|) + (1 - \frac{1}{\Lambda_n}) \cdot P(\chi) \cdot \log(|S_n|) \\
= & \log(|\Omega_n|) \cdot [P(\neg\chi) + P(\chi)(\frac{1}{\Lambda_n} + (1 - \frac{1}{\Lambda_n}) \frac{\log(|S_n|)}{\log(|\Omega_n|)})]
\end{aligned}$$

[10] If a different base of the logarithm is chosen, the upper bound of $-x\log(x)$ for all $x \in [0,1]$ may change. In the following only the existence of a fixed upper bound matters. The choice of the base of the logarithm thus remains inconsequential.

$$
\begin{aligned}
&= \log(|\Omega_n|) \cdot [P(\neg\chi) + P(\chi)(\frac{1}{\Lambda_n} + (1 - \frac{1}{\Lambda_n})\frac{\log(u_n|\Omega_n|)}{\log(|\Omega_n|)})] \\
&= \log(|\Omega_n|) \cdot [P(\neg\chi) + P(\chi)(\frac{1}{\Lambda_n} + (1 - \frac{1}{\Lambda_n})\left(1 + \frac{\log(u_n)}{\log(|\Omega_n|)}\right))] \\
&= \log(|\Omega_n|) \cdot [P(\neg\chi) + P(\chi)(1 + (1 - \frac{1}{\Lambda_n})\frac{\log(u_n)}{\log(|\Omega_n|)})] \\
&= \log(|\Omega_n|) \cdot [1 + P(\chi)(1 - \frac{1}{\Lambda_n})\frac{\log(u_n)}{\log(|\Omega_n|)}] \\
&= \log(|\Omega_n|) \cdot [1 - P(\chi)(1 - \frac{1}{\Lambda_n})\frac{\log(u_n^{-1})}{\log(|\Omega_n|)}] \ .
\end{aligned}
$$

Overall, with $v_n := \log(u_n^{-1})$ and $g_2 := P(\chi)/2$,

$$H_n(P) < g_1 + \left(1 - g_2 \frac{v_n}{\log(|\Omega_n|)}\right) \log(|\Omega_n|) \ .$$

Since u_n is a null sequence, v_n diverges to infinity; from some point onwards all v_n exceed any fixed lower bound. □

Theorem 2 *(Support-satisfiability theorem: point-valued case). If the premisses $\varphi_1^{c_1}, \dots, \varphi_k^{c_k}$ have satisfiable support then $P \prec_H \check{P}^\dagger$ for all $P \in \mathbb{E} \setminus \{\check{P}^\dagger\}$. Hence,*

$$\operatorname{maxent} \mathbb{E} = \{\check{P}^\dagger\}.$$

Proof. By Proposition 8 we have $\check{P}^\dagger \in \mathbb{E}$. Next we show that for all other probability functions in $P \in \mathbb{E} \setminus \{\check{P}^\dagger\}$, $P \prec_H \check{P}^\dagger$.

Consider a probability function $P \in \mathbb{E} \setminus \{\check{P}^\dagger\}$, and let $M = \max\{N_{\varphi_1}, \dots, N_{\varphi_k}\}$ as before. Then P must satisfy one of the three mutually exclusive and exhaustive cases below:

1. $P_{\restriction M} = \check{P}^\dagger{}_{\restriction M}$,
2. $P_{\restriction M} \neq \check{P}^\dagger{}_{\restriction M}$ and for all $1 \leq i \leq k$, $P(\check{\varphi}_i) = \check{P}^\dagger(\check{\varphi}_i)$ or
3. there exists a premiss sentence φ_i such that $P(\check{\varphi}_i) \neq \check{P}^\dagger(\check{\varphi}_i)$.

The first case is that $P_{\restriction M}$ and $\check{P}^\dagger$ agree on Ω_M. The second case is that the restrictions of both $P_{\restriction M}$ and $\check{P}^\dagger$ to Ω_M are in $\check{\mathbb{E}}$ but they differ on Ω_M. In the third case, $P_{\restriction M}$ is not a solution to the support problem, i.e., the restriction of P to Ω_M is not in $\check{\mathbb{E}}$. We go on to show that in all cases $\check{P}^\dagger$ has greater entropy than P, $P \prec_H \check{P}^\dagger$.

We first provide some intuition for thinking that $P \prec_H \check{P}^\dagger$ in these three cases.
In the <u>first case</u>, $P_{\restriction M}$ and $\check{P}^\dagger$ have the same M-entropy (they agree on Ω_M). However, $\check{P}^\dagger$ is maximally equivocal beyond M and hence $\check{P}^\dagger$ has strictly greater j-entropy than P for all j greater than some threshold $L \geq M$.
In the <u>second case</u>, $\check{P}^\dagger$ has greater M-entropy than $P_{\restriction M}$, since $\check{P}^\dagger$ is M-entropy maximiser for the finite constraints. Since $\check{P}^\dagger$ is also maximally equivocal beyond M, the $M + j$-entropy of $\check{P}^\dagger$ must be strictly greater than the $M + j$-entropy of P for all $j \geq 1$.
In the <u>third case</u>, P might have greater M-entropy than $\check{P}^\dagger$. However, P must concentrate some probability on sets of $M + j$-states that are very small compared to $|\Omega_{M+j}|$ ([16, Lemma 32]). This entails that $H_n(P)$ grows at most like $\log(|\Omega_n|) - v_n$ with v_n diverging to infinity. Since $H_n(\check{P}^\dagger)$ grows like $1 \cdot \log(|\Omega_n|) - c$ for some *constant* c, the n-entropy of $\check{P}^\dagger$ is eventually greater than the n-entropy of P.

Case 1: $P_{\downharpoonright M} = \check{P}^\dagger{}_{\downharpoonright M}$. Since $P \neq \check{P}^\dagger$, there has to exist some $L \in \mathbb{N}$ with $L \geq M$ such that $P_{\downharpoonright L+j} \neq \check{P}^\dagger{}_{\downharpoonright L+j}$ for all $j \geq 0$. Since $\check{P}^\dagger$ equivocates beyond M, it straightforwardly follows that $H_{L+j}(P) < H_{L+j}(\check{P}^\dagger)$ for all $j \geq 0$. Hence, $P \prec_H \check{P}^\dagger$.

Case 2: $P_{\downharpoonright M} \neq \check{P}^\dagger{}_{\downharpoonright M}$ and for all $1 \leq i \leq k,\ P(\check{\varphi}_i) = \check{P}^\dagger(\check{\varphi}_i)$. Since $\check{P}^\dagger$ has greatest M-entropy among all those functions that agree with P on all φ_i, $H_M(\check{P}^\dagger) > H_M(P)$. It now suffices to observe that for all $j \geq 1$, (Proposition 9)

$$\begin{aligned} H_{M+j}(P) &\leq H_M(P) + \log(|\Omega_{M+j}|) - \log(|\Omega_M|) \\ &< H_M(\check{P}^\dagger) + \log(|\Omega_{M+j}|) - \log(|\Omega_M|) \\ &= H_{M+j}(\check{P}^\dagger) \ . \end{aligned}$$

Hence, $P \prec_H \check{P}^\dagger$.

Case 3: There exists some $1 \leq i \leq k$ such that $P(\check{\varphi}_i) \neq \check{P}^\dagger(\check{\varphi}_i)$.

This follows directly from Lemma 5 since n-entropy of P grows less quickly than the n-entropy of $P^\dagger$: for all $P \in \mathbb{E} \setminus \check{\mathbb{E}}$ we find for all large enough $n \in \mathbb{N}$

$$\begin{aligned} &H_n(P^\dagger) - H_n(P) \\ &\quad \geq H_M(P^\dagger) + \log(|\Omega_n|) - \log(|\Omega_M|) - g_1 - \left(1 - \frac{g_2 v_n}{\log(|\Omega_n|)}\right)\log(|\Omega_n|) \\ &\quad = H_M(P^\dagger) - \log(|\Omega_M|) - g_1 + g_2 v_n \ . \end{aligned}$$

Since v_n diverges to infinity, $H_n(P^\dagger) > H_n(P)$ for all large enough $n \in \mathbb{N}$. □

5.2. Interval-valued premisses

We now drop the assumption that only single probabilities can attach to premisses sentences and consider the general case where intervals $\emptyset \neq X_i \subseteq [0,1]$ are attached to the premisses. We treat point-valued premisses, i.e., premisses in which $X_i = c_i$, by setting $X_i = [c_i, c_i]$. Without loss of generality we may assume that no premiss is of the form φ^0, since we can equivalently replace it by $(\neg\varphi)^1$.

Definition 21 *($P^\dagger_{\vec{c}}$).* Let the premisses $\varphi_1^{c_1}, \ldots, \varphi_k^{c_k}$ be support-satisfiable. Let $P^\dagger_{\vec{c}} = \check{P}^\dagger_{\vec{c}}$ denote the unique maximal entropy function (uniqueness follows from Theorem 2).

Definition 22 *($P^\dagger_{\vec{X}}$).* Given support-satisfiable premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$, let $P^\dagger_{\vec{X}}$ be a probability function in

$$\{P^\dagger_{\vec{c}} \ : \ \vec{c} \in \vec{X} \text{ such that } \varphi_1^{c_1}, \ldots, \varphi_k^{c_k} \text{ are support-satisfiable}\}$$

with maximal M-entropy, where $M = \max\{N_{\varphi_i} : i = 1, \ldots, k\}$.

Note that $P^\dagger_{\vec{c}}$ equivocates beyond M.

Theorem 3 *(Support-satisfiability theorem). If the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ are support-satisfiable then $P^\dagger_{\vec{X}}$ is the unique $P^\dagger_{\vec{c}}$ with maximal M-entropy and $P \prec_H P^\dagger_{\vec{X}}$ for all other probability functions in $P \in \mathbb{E} \setminus \{P^\dagger_{\vec{X}}\}$. Hence,*

$$\operatorname{maxent} \mathbb{E} = \{P^\dagger_{\vec{X}}\}.$$

Proof. Since $\check{\mathbb{E}} \neq \emptyset$ there is some function in $\mathbb{E}$ such that $P(\check{\varphi}_i) \in X_i$ which equivocates beyond M. The n-entropy for $n \geq M$ of such a function is (Proposition 9):

$$H_n(P) = H_M(P) + \log(|\Omega_n|) - \log(|\Omega_M|) \ .$$

From Theorem 2 we obtain that for all other $Q \in \mathbb{E}$ with $Q(\check{\varphi}_i) = P(\check{\varphi}_i)$ for all i, $Q \prec_H P$.

From Lemma 5 we obtain that for $Q \in \mathbb{E} \setminus \check{\mathbb{E}}$ the n-entropy is eventually strictly less than the n-entropy of such a P.

Denote the functions in $\check{\mathbb{E}}$ which equivocate beyond M by $\mathbb{E}^+$,

$$\mathbb{E}^+ := \{P : P(\check{\varphi}_i) \in X_i \text{ for all } 1 \leq i \leq k \text{ and } P \text{ equivocates beyond } M\} \neq \emptyset \ .$$

Then every function in $Q \in \mathbb{E} \setminus \mathbb{E}^+$ has less entropy than some function $P \in \mathbb{E}^+$.

Next let us consider $R, S \in \mathbb{E}^+$ such that $H_M(R) = H_M(S)$. Then define $Q := \frac{R+S}{2}$ and note that $Q(\check{\varphi}_i) = \frac{R(\check{\varphi}_i) + S(\check{\varphi}_i)}{2} \in X_i$ for all i since the X_i are convex. So, $Q \in \mathbb{E}^+$, $R \prec_H Q$ and $S \prec_H Q$. Since n-entropies of the probability functions P in $\mathbb{E}^+$ are, for large n, determined by their M-entropies, the maximal entropy function in $\mathbb{E}^+$ (and hence in $\mathbb{E}$) is unique, if it exists.

Note that every function in $\mathbb{E}^+$ is represented by the $|\Omega_M|$ probabilities it assigns to the M-states. The set of these $|\Omega_M|$-tuples representing the probability functions in $\mathbb{E}^+$ is compact since (i) it is bounded (probabilities lie in the unit interval) and (ii) it is closed (the intervals X_i are closed, hence the condition defining $\mathbb{E}^+$ is closed). Thus every convergent sequence of probability functions $(P_n)_{n\in\mathbb{N}}$ with $P_n \in \mathbb{E}^+$ has a limit in $\mathbb{E}^+$. So, the supremum of $H_M(P)$ with $P \in \mathbb{E}^+$ exists, is unique and obtains for a unique probability function $P^\dagger \in \mathbb{E}^+ \subset \mathbb{E}$. This function has maximal entropy. $\square$

5.3. Reduction to the finite problem

We now arrive at the main result of this section.

Theorem 4 *(Reduction theorem). If the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ have satisfiable support then*

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\!\approx^{\circ}} \psi^Y \text{ if and only if } \check{\varphi}_1^{X_1}, \ldots, \check{\varphi}_k^{X_k} \mathrel{|\!\!\!\approx^{\circ}} \check{\psi}^Y \ .$$

Proof. Theorem 2 and Theorem 3 have already established that the original problem and the support problem have the same solution, $P^\dagger$. It only remains to show that $P^\dagger(\psi) = P^\dagger(\check{\psi})$, dropping the cumbersome index $\vec{X}$.

We first make a small observation. For all $\psi \in S\mathcal{L}$ and all $\omega_M \in \Omega_M$,

$$\begin{aligned} P^\dagger(\psi \wedge \omega_M) > 0 &\iff P^\dagger(\psi|\omega_M) > 0 \ \& \ P^\dagger(\omega_M) > 0 \\ &\iff P_=(\psi|\omega_M) > 0 \ \& \ P^\dagger(\omega_M) > 0 \\ &\implies P_=(\psi \wedge \omega_M) > 0 \\ &\iff P_=(\check{\psi} \wedge \omega_M) > 0 \ . \end{aligned}$$

Using the fact that $P^\dagger$ equivocates beyond M (Theorem 3) we thus note,

$$P^\dagger(\psi) = \sum_{\omega_M \in \Omega_M} P^\dagger(\psi \wedge \omega_M) = \sum_{\substack{\omega_M \in \Omega_M \\ P^\dagger(\psi \wedge \omega_M) > 0}} P^\dagger(\psi \wedge \omega_M)$$

$$
\begin{aligned}
&= \sum_{\substack{\omega_M \in \Omega_M \\ P^\dagger(\psi \wedge \omega_M)>0}} P_=(\psi|\omega_M) \cdot P^\dagger(\omega_M) \\
&\leq \sum_{\substack{\omega_M \in \Omega_M \\ P^\dagger(\check{\psi} \wedge \omega_M)>0}} P_=(\check{\psi}|\omega_M) \cdot P^\dagger(\omega_M) \\
&\overset{\text{Lemma 4}}{=} \sum_{\substack{\omega_M \in \Omega_M \\ P^\dagger(\check{\psi} \wedge \omega_M)>0}} P^\dagger(\check{\psi}|\omega_M) \cdot P^\dagger(\omega_M) \\
&= \sum_{\substack{\omega_M \in \Omega_M \\ P^\dagger(\check{\psi} \wedge \omega_M)>0}} P^\dagger(\check{\psi} \wedge \omega_M) \\
&= P^\dagger(\check{\psi}) \ .
\end{aligned}
$$

Replacing ψ by $\neg\psi$ and exploiting that $\check{\neg\psi} = \neg\check{\psi}$ ((iii) of Lemma 3) we obtain $P^\dagger(\neg\psi) \leq P^\dagger(\neg\check{\psi})$. This entails that $P^\dagger(\psi) = P^\dagger(\check{\psi})$. $\square$

6. Decidability

In this section, we establish that the class of support-satisfiable inferences is decidable.

First, we present an important lemma.

Lemma 6 *(Computability of the measure of φ). For any sentence $\varphi \in S\mathcal{L}$, $P_=(\varphi)$ is computable and $P_=(\varphi) \in \{0, \frac{1}{|\Omega_N|}, \frac{2}{|\Omega_N|}, \dots, 1\}$, where $N = N_\varphi$.*

Proof. First, note that for any sentence $\varphi \in S\mathcal{L}$,

$$P_=(\varphi) = \sum_{\omega \in \Omega_N} P_=(\varphi \wedge \omega) = \sum_{\omega \in \Omega_N} P_=(\varphi|\omega) P_=(\omega) \ .$$

By Corollary 2, we observe that for all $\omega \in \Omega_N$,

$$P_=(\varphi|\omega) = P_=(\varphi \wedge \varphi|\omega) = P_=(\varphi|\omega)^2 \in \{0, 1\} \ .$$

So, $P_=(\omega \wedge \varphi) \in \{0, \frac{1}{|\Omega_N|}\}$ and consequently $P_=(\varphi) \in \{0, \frac{1}{|\Omega_N|}, \frac{2}{|\Omega_N|}, \dots, 1\}$. To conclude the proof, it hence suffices show that we can compute whether $P_=(\omega \wedge \varphi)$ is zero or $\frac{1}{|\Omega_N|}$ for all $\varphi \in S\mathcal{L}$ and every $\omega \in \Omega_N$.

We shall show this by induction on the number of quantifiers in the prenex normal form of φ.

Base Case. The claim follows immediately for quantifier-free sentences φ, which are logically equivalent to a disjunction of N-states:

$$P_=(\omega \wedge \varphi) = P_=(\omega \wedge \bigvee_{\substack{\omega' \in \Omega_N \\ \omega' \models \varphi}} \omega') = \begin{cases} P_=(\omega \wedge \omega) = P_=(\omega) = \frac{1}{|\Omega_N|} & : \quad \omega \models \varphi \\ 0 & : \quad \omega \not\models \varphi \end{cases}$$

Induction step. Now suppose that φ has $q \geq 1$ quantifiers and assume that the induction hypothesis, i.e., the statement of the lemma for φ, holds for any sentence in prenex normal form with fewer than q quantifiers.

Suppose first that φ is in prenex normal form and $\varphi = \forall x \theta(x)$. Consider for $\omega \in \Omega_N$

$$\chi := \omega \wedge \forall x \theta(x) \ .$$

By P3,

$$P_=(\chi) = P_=(\omega \wedge \forall x\theta(x)) = \lim_{n\to\infty} P_=(\omega \wedge \bigwedge_{i=1}^{n} \theta(t_i))$$
$$\leq P_=(\omega \wedge \bigwedge_{i=1}^{j+1} \theta(t_i)) \leq P_=(\omega \wedge \bigwedge_{i=1}^{j} \theta(t_i)) \ ,$$

for all $j \geq 0$.

$\omega \wedge \theta(t_i)$ is not in prenex normal form, but by moving the quantifiers of $\theta(t_i)$ to the front of the sentence $\omega \wedge \theta(t_i)$, we obtain a (deductively) logically equivalent sentence $\psi_\omega(t_i)$ that is in prenex normal form and has the same number of quantifiers. Since the two sentences are deductively logically equivalent, $P_=(\psi_\omega(t_i)) = P_=(\omega \wedge \theta(t_i))$.

By the induction hypothesis, for all $\omega \in \Omega_N$, and all $i = 1, \dots, N+1$,

$$P_=(\psi_\omega(t_i)) \in \{0, \frac{1}{|\Omega_{N+1}|}, \frac{2}{|\Omega_{N+1}|}, \dots, 1\}$$

is computable. In fact, $P_=(\omega) = 1/|\Omega_N|$ and the probability of a conjunction cannot exceed the probability of either conjunct, so

$$P_=(\psi_\omega(t_i)) = P_=(\omega \wedge \theta(t_i)) \in \{0, \frac{1}{|\Omega_{N+1}|}, \frac{2}{|\Omega_{N+1}|}, \dots, \frac{1}{|\Omega_N|}\}$$

and this value is computable.

If $P_=(\omega \wedge \theta(t_i)) < 1/|\Omega_N|$, for any $i = 1, \dots, N+1$ then $P_=(\omega \wedge \bigwedge_{i=1}^{N+1} \theta(t_i)) < 1/|\Omega_N|$, and since $P_=(\chi) \leq P_=(\omega \wedge \bigwedge_{i=1}^{N+1} \theta(t_i))$ and $P_=(\chi) \in \{0, \frac{1}{|\Omega_N|}\}$, we must have $P_=(\chi) = 0$.

On the other hand, if $P_=(\omega \wedge \theta(t_i)) = 1/|\Omega_N|$ for all $i = 1, \dots, N+1$, then $P_=(\omega \wedge \bigwedge_{i=1}^{N+1} \theta(t_i)) = 1/|\Omega_N|$. We show that this implies $P_=(\chi) = 1$. Since $P_=$ satisfies the principle of Constant Exchangeability (CX) (Definition 19), we have that for all $j \geq 1$,

$$P_=(\omega \wedge \theta(t_{N+j}) \wedge \bigwedge_{i=1}^{N} \theta(t_i)) = P_=(\omega \wedge \theta(t_{N+1}) \wedge \bigwedge_{i=1}^{N} \theta(t_i)) = \frac{1}{|\Omega_N|} \ .$$

By the definition of conditional probability,

$$P_=(\theta(t_{N+j}) \wedge \bigwedge_{i=1}^{N} \theta(t_i)|\omega) = \frac{P_=(\omega \wedge \theta(t_{N+j}) \wedge \bigwedge_{i=1}^{N} \theta(t_i))}{P_=(\omega)} = \frac{|\Omega_N|}{|\Omega_N|} = 1.$$

Since $P_=(\cdot|\omega)$ is a probability function, by Lemma 2(1) we can add $\theta(t_{N+j}) \wedge \bigwedge_{i=1}^{N} \theta(t_i)$, for any j, as a conjunct to any sentence and the conditional measure remains unchanged. So, for all $s \geq 2$,

$$1 = P_=(\theta(t_{N+1}) \wedge \bigwedge_{i=1}^{N} \theta(t_i)|\omega)$$
$$= P_=(\theta(t_{N+1}) \wedge \theta(t_{N+2}) \wedge \bigwedge_{i=1}^{N} \theta(t_i)|\omega)$$
$$= P_=(\bigwedge_{j=1}^{s} \theta(t_{N+j}) \wedge \bigwedge_{i=1}^{N} \theta(t_i)|\omega)$$

$$
\begin{aligned}
&= \lim_{n\to\infty} P_=(\bigwedge_{i=1}^{n} \theta(t_i)|\omega) \\
&= \lim_{n\to\infty} \frac{P_=(\omega \wedge \bigwedge_{i=1}^{n} \theta(t_i))}{P_=(\omega)} \\
&= |\Omega_N| \cdot \lim_{n\to\infty} P_=(\omega \wedge \bigwedge_{i=1}^{n} \theta(t_i)) \\
&\overset{\text{P3}}{=} |\Omega_N| \cdot P_=(\chi) \ .
\end{aligned}
$$

Thus, $P_=(\chi) = \frac{1}{|\Omega_N|}$.

The case of an existential quantifier, $\varphi = \exists x\theta(x)$, is proved by noting that

$$
P_=(\omega \wedge \exists x\theta(x)) = P_=(\omega) - P_=(\omega \wedge \forall x\neg\theta(x)) \ ,
$$

which is (computably) verifiably equal to zero or $\frac{1}{|\Omega_N|}$. $\square$

In particular, we can compute $\check{\varphi}$ because we can compute the measure of $\varphi \wedge \omega$ for all $\omega \in \Omega_{N_\varphi}$.

This result gives the following immediate corollary.

Proposition 10. *Determining the support problem from the original problem is effectively computable.*

Our key result is that the support-satisfiable inferences are finitely reducible (Definition 15), and hence decidable:

Theorem 5 *(Support-satisfiable Decidability). The class of support-satisfiable inferences is decidable in OBIL.*

Proof. We shall show that the class of support-satisfiable inferences is finitely reducible (Definition 15). Since testing for support-satisfiability is decidable (Proposition 6), it follows that the class of support-satisfiable inferences is decidable.

Given an inference $\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \approx\!\!\!\mid \psi^Y$, one can effectively construct the support inference $\check{\varphi}_1^{X_1}, \dots, \check{\varphi}_k^{X_k} \approx\!\!\!\mid \check{\psi}^Y$ by Proposition 10. By Proposition 6, we can effectively test for satisfiability of the premisses of the support inference, thereby determining whether the original inference is within the class of support-satisfiable inferences.

If it is support-satisfiable, then by the Reduction Theorem (Theorem 4),

$$
\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \approx\!\!\!\mid \psi^Y \text{ if and only if } \check{\varphi}_1^{X_1}, \dots, \check{\varphi}_k^{X_k} \approx\!\!\!\mid \check{\psi}^Y \ .
$$

The support inference is decidable by Theorem 1. $\square$

We have the following corollaries:

Corollary 3 *(Decidability of premiss-free inferences). The class of inferences from no premisses is decidable. In particular, it is decidable whether any given sentence is an inductive tautology.*

This stands in marked contrast to the situation with deductive logic: as we observed in §1, there is no effective procedure for deciding whether any given sentence is a deductive tautology.

Corollary 4 *(Decidability of inferences from an inductively consistent premiss). If φ is inductively consistent, i.e., $P_=(\varphi) > 0$, then the premiss φ^1 is support-satisfiable (Example 8). Hence, it is decidable whether $\varphi \approx\!\!\!\mid \psi^Y$ for any sentence $\psi \in S\mathcal{L}$ and any interval Y.*

If $0 < P_=(\varphi) < 1$, then the premiss φ^c for $0 < c < 1$ is support-satisfiable (Example 8). Hence, it is decidable whether $\varphi^c \mid\!\approx \psi^Y$ for all sentences $\psi \in S\mathcal{L}$ and all intervals Y.

Consequently, given categorical premisses $\varphi_1^1, \ldots, \varphi_k^1$, if $P_=(\varphi) > 0$ for $\varphi := \bigwedge_{i=1}^k \varphi_i$, any inference is decidable. I.e., the class of inferences from inductively consistent categorical premisses is decidable. In fact in this case, $P^\dagger(\psi) = P_=(\psi|\varphi_1 \wedge \ldots \wedge \varphi_k)$ [16, Theorem 34].

7. Objective Bayesian networks

The truth-table method introduced in §3 serves to highlight the decidability of the class of finitely reducible inferences. The truth-table method is not particularly computationally tractable, however: the number of rows in a truth table increases exponentially with the number of atomic propositions that feature in an inference. Furthermore, the fact that computing the maximum entropy function on a finite domain has a high worst-case complexity [28, Chapter 10] has raised worries about the practical feasibility of entropy maximisation [30, p. 463].[11] While the focus of this paper is on decidability rather than computational complexity, it is worth observing that there is a method for inference that is tractable in many cases. This is the graphical modelling approach of *objective Bayesian networks*. Hitherto, this approach has been applied to the case of finite propositional inductive logic [18,37,38]. Finite reducibility allows its use also for predicate inductive logic. In this section, we briefly sketch the approach and provide an example.

Suppose, as above, that the task is to verify an entailment relationship of the form

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx \psi^Y, \tag{7}$$

and that the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ are support-satisfiable. Let $a_{i_1}, \ldots, a_{i_m}$ be the atomic propositions that occur in $\check{\varphi}_1, \ldots, \check{\varphi}_k$ and let Ξ be the set of states of $a_{i_1}, \ldots, a_{i_m}$. As before, $P^\dagger$ is the maximal entropy function in $\mathbb{P}[\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]$, which can be found by maximising entropy subject to $\check{\varphi}_1^{X_1}, \ldots, \check{\varphi}_k^{X_k}$ and equivocating beyond Ξ. The atomic propositions $a_{i_1}, \ldots, a_{i_m}$ will be the nodes in our Bayesian network:

Definition 23 *(Objective Bayesian network).* An *objective Bayesian network* or OBN for $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ consists of (i) a directed acyclic graph $\mathcal{H}$ whose nodes are the atomic propositions $a_{i_1}, \ldots, a_{i_m}$, and (ii) the probability distribution, induced by $P^\dagger$, of each node conditional on its parents in $\mathcal{H}$, such that for each $\xi \in \Xi$,

$$P^\dagger(\xi) = \prod_{j=1}^m P^\dagger(a_{i_j}^\xi \mid par_{i_j}^\xi),$$

where $a_{i_j}^\xi$ is the state of a_{i_j} (i.e., a_{i_j} or $\neg a_{i_j}$) that is consistent with ξ, and $par_{i_j}^\xi$ is the state of its parents that is consistent with ξ.

Thus an objective Bayesian network for $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ is a means of representing the maximal entropy function in $\mathbb{P}[\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]$. The OBN directly represents $P^\dagger$ on Ξ and this is extended to the whole of $\mathcal{L}$ by equivocating elsewhere, as per Proposition 4.

An OBN can be constructed by means of the following procedure:

1. Construct an undirected graph $\mathcal{G}$ that represents independencies of $P^\dagger$: for $i = 1, \ldots, k$, connect atomic propositions that occur in the same support sentence $\check{\varphi}_i$ by undirected edges. $\mathcal{G}$ can be thought of

[11] Tractable entropy optimisation is an active sub-field of optimisation theory [2,8,11,12,23,27].

as a Markov network structure for $P^\dagger$: for any sets X, Y, Z of atomic propositions, if Z separates X from Y in $\mathcal{G}$ then $P^\dagger$ is guaranteed to render X and Y probabilistically independent conditional on Z, $X \perp\!\!\!\perp_{P^\dagger} Y \mid Z$ [36, §5.6]; [19, Appendix A].

2. Construct a minimal triangulation $\mathcal{G}^T$ of $\mathcal{G}$ and transform this into a directed acyclic graph $\mathcal{H}$ that represents the independencies captured by $\mathcal{G}^T$. This transformation can be performed as follows—see [36, §5.7] and [25] for further discussion. (i) Order the vertices of $\mathcal{G}^T$ with vertex set V according to maximum cardinality search: at each step select a vertex which is adjacent to the largest number of previously numbered vertices. (ii) Let $D_1, \dots, D_l$ be the cliques of $\mathcal{G}^T$, ordered according to the highest labelled vertex. (iii) Let $E_j := D_j \cap (\bigcup_{i=1}^{j-1} D_i)$ and $F_j := D_j \setminus E_j$. (iv) Add an arrow from each vertex in E_j to each vertex in F_j. (v) Add further arrows to ensure there is an arrow between each pair of vertices in D_j such that the resulting directed graph $\mathcal{H}$ is acyclic.
3. Solve an optimisation problem to determine the associated conditional probability parameters $P(a_{i_j}^\xi \mid par_{i_j}^\xi)$ that maximise entropy:

$$H(P) = -\sum_{j=1}^{m} \sum_{\xi \in \Xi} \left(\prod_{l=1}^{m} P(a_{i_l}^\xi \mid par_{i_l}^\xi) \right) \cdot \log(P(a_{i_j}^\xi \mid par_{i_j}^\xi)),$$

subject to the constraints imposed by the premisses [36, §5.7].

In the worst case, an OBN of $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$ requires as many parameters as there are lines in the truth table for $\varphi_1^{X_1}, \dots, \varphi_k^{X_k}$. This worst case occurs when there is some support premiss that mentions every atomic sentence that occurs in the inference.

More typically, however, each support sentence will mention only a small subset of the atomic propositions $a_{i_1}, \dots, a_{i_m}$. In such a scenario, the OBN will require far fewer parameters than there are lines of the corresponding truth table. It is in this sense that OBNs can be more computationally tractable than the truth table method.

Let us consider an example, based around the following premisses:

$$\exists x U t_1 x, (Vt_2 \vee \forall x R x)^{0.9}, Vt_1 \to Ut_1t_3, (Vt_1 \vee (\exists x Uxt_3 \to Vt_2))^{[0.95,1]}$$

The language of this inference is the language $\mathcal{L}$ of Example 1. We can enumerate the atomic propositions as follows:

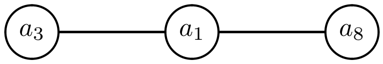
$$a_1 : Vt_1, \quad a_2 : Ut_1t_1, \quad a_3 : Vt_2, \quad a_4 : Ut_1t_2,$$
$$a_5 : Ut_2t_1, \quad a_6 : Ut_2t_2, \quad a_7 : Vt_3, \quad a_8 : Ut_1t_3, \dots$$

Then the support of each premiss sentence is as follows:

i	φ_i	$\check{\varphi}_i$
1	$\exists x U t_1 x$	$a_1 \vee \neg a_1$
2	$Vt_2 \vee \forall x R x$	a_3
3	$Vt_1 \to Ut_1t_3$	$a_1 \to a_8$
4	$Vt_1 \vee (\exists x Uxt_3 \to Vt_2)$	$a_1 \vee a_3$

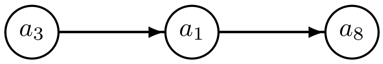
Note in particular for φ_1, i.e., $\exists x U t_1 x$, we have that $\check{\varphi}_1$ is $a_1 \vee \neg a_1$ because φ_1 mentions no atomic propositions and $P_=(a_1 \wedge \exists x U t_1 x) = P_=(\neg a_1 \wedge \exists x U t_1 x) = 1/2 > 0$. Strictly speaking, $\check{\varphi}_3$ is defined as $(a_1 \wedge a_8) \vee (\neg a_1 \wedge a_8) \vee (\neg a_1 \wedge \neg a_8)$ but we abbreviate this sentence by the logically equivalent $a_1 \to a_8$. Similarly for $\check{\varphi}_4$.

To construct the corresponding OBN we first construct the undirected graph $\mathcal{G}$. We take a_1, a_3 and a_8 as nodes because they are the atomic propositions that feature in the supports. We include an edge between a_1 and a_8 because they both feature in the third support premiss, and an edge between a_1 and a_3 because they both feature in the fourth support premiss:



Separation in $\mathcal{G}$ provably corresponds to a conditional probabilistic independence of the maximal entropy function $P^\dagger$, so we can conclude that $P^\dagger$ renders a_3 and a_8 probabilistically independent conditional on a_1.

We next transform $\mathcal{G}$ into a directed acyclic graph $\mathcal{H}$ that preserves as many of the conditional independencies of $\mathcal{G}$ as possible. For example, we can set $\mathcal{H}$ to be:



D-separation in $\mathcal{H}$ also implies that $P^\dagger$ renders a_3 and a_8 probabilistically independent conditional on a_1.[12]

We parameterise the OBN by finding the values of the following parameters that maximise entropy:

$$P(a_3), P(a_1|a_3), P(a_1|\neg a_3), P(a_8|a_1), P(a_8|\neg a_1).$$

A simple numerical optimisation subject to the constraints $P(a_3) = 0.9, P(a_1 \to a_8) = 1$ and $P(a_1 \vee a_3) \in [0.95, 1]$ yields:

$$P(a_3) = 0.9, P(a_1|a_3) = 1/3, P(a_1|\neg a_3) = 1/2, P(a_8|a_1) = 1, P(a_8|\neg a_1) = 1/2.$$

The OBN can then be used to perform inference. For example,

$$\begin{aligned}
&\exists x U t_1 x, (Vt_2 \vee \forall x Rx)^{0.9}, Vt_1 \to Ut_1t_3, (Vt_1 \vee (\exists x Uxt_3 \to Vt_2))^{[0.95,1]} \\
&\mathrel{|}\!\!\!\approx^{\circ} (\neg(Vt_1 \vee Vt_3) \wedge \exists x Uxx \wedge Ut_1t_3)^{0.1625}.
\end{aligned}$$

To see this, note that the support of the conclusion sentence is $\neg a_1 \wedge \neg a_7 \wedge a_8$ and that a_7 is not mentioned by any of the premisses so $P^\dagger$ renders a_7 probabilistically independent of a_1 and a_8 and $P^\dagger(a_7) = 1/2$. Hence,

$$\begin{aligned}
P^\dagger(\neg a_1 \wedge \neg a_7 \wedge a_8) &= 1/2 \cdot P(a_8|\neg a_1)\,(P(\neg a_1|a_3) \cdot P(a_3) + P(\neg a_1|\neg a_3) \cdot P(\neg a_3)) \\
&= 1/2 \cdot 1/2\,(2/3 \cdot 9/10 + 1/2 \cdot 1/10) = 0.1625\ .
\end{aligned}$$

The main advantage of this OBN over the truth table method is a reduction in the number of parameters required to specify the maximal entropy function. The truth table for the premisses can be written down as follows:

[12] Subset Z D-separates subsets X from Y of nodes if each path between a node in X and a node in Y contains either (i) some node a_i in Z at which the arrows on the path meet head-to-tail ($\longrightarrow a_i \longrightarrow$) or tail-to-tail ($\longleftarrow a_i \longrightarrow$), or (ii) some node a_j at which the arrows on the path meet head-to-head ($\longrightarrow a_j \longleftarrow$) and neither a_j nor any of its descendants are in Z. The key result is that if Z D-separates X from Y in $\mathcal{H}$ then the maximal entropy function renders X and Y probabilistically independent conditional on Z [36, Theorem 5.3].

$P^\dagger$	a_3	a_1	a_8	$a_1 \vee \neg a_1$	a_3	$a_1 \to a_8$	$a_1 \vee a_3$
0.025	F	F	F	T	F	T	F
0.025	F	F	T	T	F	T	F
0	F	T	F	T	F	F	T
0.05	F	T	T	T	F	T	T
0.3	T	F	F	T	T	T	T
0.3	T	F	T	T	T	T	T
0	T	T	F	T	T	F	T
0.3	T	T	T	T	T	T	T

We see then that although the support problem only involves three atomic propositions, the truth table requires 8 parameters while the OBN requires only 5. Typically, this reduction in the number of parameters becomes more marked as the number of atomic propositions in the premisses increases.

8. Infinitely many premisses

Thus far, we have considered inductive inferences involving finitely many premisses. In this section, we consider inferences involving infinitely many premisses and show that it is possible to obtain a reduction theorem. Handling infinite objects is often a difficult endeavour in practice. Our main point here is that there are some cases which we can treat as if they were finite, namely those that are finitely support-satisfiable (Theorem 6). However, some complications can also arise, as we point out in Examples 11 and 12 as well as Observations 2 and 3. As we shall see, some results in Section 9 also hold for infinitely many premisses.

Consider the inductive inference:

$$(\varphi_i^{X_i})_{i\in I} \approx\!\!\!\!\mid\;\; \psi^Y$$

where I is an index set of arbitrary size. We again define the set of probability functions consistent with the premisses to be:

$$\mathbb{E} = \mathbb{P}[(\varphi_i^{X_i})_{i\in I}].$$

As before, the maximal entropy functions are:

$$\operatorname{maxent}\mathbb{E} := \{P \in \mathbb{E} \;:\; \text{there is no } Q \in \mathbb{E} \text{ with } P \prec_H Q\},$$

and we define objective Bayesian inductive entailment in the usual way. We consider the support of φ as defined in Definition 16,

$$\check{\varphi} \stackrel{\text{df}}{=} \bigvee\{\xi \in \Xi_\varphi : P_=(\xi \wedge \varphi) > 0\}.$$

Notational conventions remain in line with those adopted earlier:

	Original inference	Support inference
Premisses	$(\varphi_i^{X_i})_{i\in I}$	$(\check{\varphi}_i^{X_i})_{i\in I}$
Feasible region	$\mathbb{E} \stackrel{\text{df}}{=} \mathbb{P}[(\varphi_i^{X_i})_{i\in I}]$	$\check{\mathbb{E}} \stackrel{\text{df}}{=} \mathbb{P}[(\check{\varphi}_i^{X_i})_{i\in I}]$
n-entropy maximisers	P^n	$\check{P}^n$
Models	$P^\dagger \in \operatorname{maxent}\mathbb{E}$	$\check{P}^\dagger \in \operatorname{maxent}\check{\mathbb{E}}$

Definition 24. We let M be the supremum of the indices of the constants that appear in the premisses: $M := \sup\{N_{\varphi_i} : i \in I\}$. Note that M may be infinite.

Definition 25 *(Finite support-satisfiability).* The premisses are *finitely support-satisfiable* if and only if

1. they are support-satisfiable, $\check{\mathbb{E}} \neq \emptyset$, and
2. there exists a $K \in \mathbb{N}$ with the following property: for all $P \in \check{\mathbb{E}}$, there exists a $Q \in \check{\mathbb{E}}$ such that $Q(\check{\varphi}_i) = P(\check{\varphi}_i)$ for all $i \in I$ and Q equivocates beyond K.

Let K^* denote the minimal such K. Note that for finitely many premisses, the second condition follows from the first, so the above definition extends Definition 18.

As the following example shows, there are some cases in which the premisses are finitely satisfiable, but where every constant appears in at least one $\check{\varphi}_i$.

Example 10 *(Finite support-satisfiability of infinitely many premisses).* If $\varphi_i := U_1t_1 \vee (\forall x U_2x \wedge U_3t_i)$ for all $i \in \mathbb{N}$, then the premisses $(\varphi_i)_{i\in\mathbb{N}}$ are finitely support-satisfiable. Here $\check{\varphi}_i = [(U_1t_1 \wedge U_3t_i) \vee (U_1t_1 \wedge \neg U_3t_i)]$ and $\check{\mathbb{E}} = \{P \in \mathbb{P} \ : \ P(U_1t_1) = 1\}$, $K^* = 1$. However, every constant is mentioned by an atomic proposition in at least one premiss sentence, $M = \sup\{N_{\varphi_1}, \dots\} = \infty$.

Example 11 *(Infinitely many categorical premisses).* Let the premisses be $(Ut_{2i})_{i\in\mathbb{N}}$. Unlike in the finite case, it is not possible to collect all these categorical premisses as a single premiss.

Example 12. Let the premisses be $(Ut_i)_{i\in\mathbb{N}}$. Then $\check{\mathbb{E}} = \mathbb{E} \neq \emptyset$, because it contains a probability function with $P(\forall x Ux) = 1$. Note that the premisses are not finitely support-satisfiable, since no $Q \in \mathbb{E}$ equivocates beyond some fixed $K \in \mathbb{N}$.

Theorem 6 *(The maximal entropy function). If the premisses $(\varphi_i^{X_i})_{i\in I}$ are finitely support-satisfiable, then $\check{P}^\dagger$ is the unique function that has maximal K^*-entropy and that equivocates beyond K^*. $P \prec_H \check{P}^\dagger$ for all other probability functions $P \in \mathbb{E} \setminus \{\check{P}^\dagger\}$. Thus,* $\operatorname{maxent}\mathbb{E} = \{\check{P}^\dagger\}$.

Proof. This is analogous to the proofs of Theorem 2 and 3. None of the above proofs use the fact that the number of premisses is finite; only support-satisfiability and equivocation beyond some $N \in \mathbb{N}$ is assumed in these proofs. The second condition of Definition 25 guarantees that there is a fixed number K^* for which it is sufficient to maximise K^*-entropy. $\square$

Definition 26 *(Finitely presented support).* An inference has *finitely presented support* iff the support inference is represented by means of finitely many finitely represented, quantifier-free premisses, i.e., if and only if premisses $\theta_1^{W_1}, \dots, \theta_k^{W_k}$ are provided where $\theta_1, \dots, \theta_k$ are quantifier-free, $W_1, \dots, W_k$ are finitely represented and $\check{\mathbb{E}} = \{P \in \mathbb{P} : P(\theta_i) \in W_i, i = 1, \dots, k\}$.

Note that an inference with finitely presented, satisfiable support is finitely support-satisfiable.

Effectively computable decision procedures for infinitely many premisses are hard to come by, since we need to ensure that the conclusions are compatible with *all*, i.e., infinitely many, premisses. In order to isolate a class of decidable inferences with infinitely many premisses we hence assume a given finite presentation of the support premisses—this does not need to be effectively computed since it is given. Putting Definition 26 together with Theorem 5 we obtain:

Corollary 5. *The class of inferences with finitely presented, satisfiable support is decidable in OBIL.*

Inferences with infinitely many premisses may not be well-behaved in other respects:

Observation 2 *(Non-compactness 1).* Given categorical premisses $\varphi_i := Ut_i$ for $i \in \mathbb{N}$, we have that $P^\dagger(\forall x Ux) = 1$. However, all finite subsets of premisses $(J \subset I)$ are finitely support-satisfiable and have a maximal entropy function $P_J^\dagger$ with $P_J^\dagger(\cdot) = P_=(\cdot | \bigwedge_{j\in J} \varphi_j)$. Thus, $P_J^\dagger(\forall x Ux) = 0$.

Observation 3 *(Non-compactness 2).* The premisses $\exists x \neg Ux, Ut_1, Ut_2, \dots$ are not satisfiable. However, every finite subset has a well-defined maximal entropy function $P^\dagger$ with $P^\dagger(\exists x \neg Ux) = 1$.

9. Preservation of inductive tautologies

Learning new information changes what we are in a position to infer. On finite domains, maximum entropy inference from consistent premisses preserves measure one and measure zero propositions. In other words, any inductive tautology and any inductive contradiction (the empty event) are assigned probability one and zero, respectively, after learning consistent information. This inferential property was called Preservation of Inductive Tautologies (PIT) by [16, Section 7]. PIT states that inductive tautologies (i.e., probability 1 inferences in the absence of any premisses) are preserved on learning new information. In OBIL, PIT can be stated as

$$\{\psi \in S\mathcal{L} \ : \approx\!\!\!\!|\ \psi^1\} \subseteq \{\psi \in S\mathcal{L} \ : \ \varphi_1^{X_1}, \varphi_2^{X_2}, \dots \approx\!\!\!\!|\ \psi^1\}.$$

Proposition 11. *If the premisses are finitely support-satisfiable then OBIL satisfies PIT:*

$$\{\psi \in S\mathcal{L} \ : \approx\!\!\!\!|\ \psi^1\} \subseteq \{\psi \in S\mathcal{L} \ : \ \varphi_1^{X_1}, \varphi_2^{X_2}, \dots \approx\!\!\!\!|\ \psi^1\}.$$

Proof. Let φ be an inductive tautology and $P^\dagger$ be the maximal entropy function for given support-satisfiable premisses which equivocates beyond some number. Then the assumptions of Lemma 4 hold. Recall that we showed in Lemma 6 that $P_=(\varphi \wedge \omega) \in \{0, \frac{1}{|\Omega_N|}\}$. So, for all large enough n we find

$$\begin{aligned}
P^\dagger(\varphi) &= \sum_{\omega_n \in \Omega_n} P^\dagger(\varphi \wedge \omega_n) \\
&= \sum_{\substack{\omega_n \in \Omega_n \\ P^\dagger(\omega_n) > 0}} P^\dagger(\varphi \wedge \omega_n) \\
&= \sum_{\substack{\omega_n \in \Omega_n \\ P^\dagger(\omega_n) > 0}} P^\dagger(\varphi | \omega_n) \cdot P^\dagger(\omega_n) \\
&\overset{\text{Lemma 4}}{=} \sum_{\substack{\omega_n \in \Omega_n \\ P^\dagger(\omega_n) > 0}} P_=(\varphi | \omega_n) \cdot P^\dagger(\omega_n) \\
&= \sum_{\substack{\omega_n \in \Omega_n \\ P^\dagger(\omega_n) > 0}} \frac{P_=(\varphi \wedge \omega_n)}{P_=(\omega_n)} \cdot P^\dagger(\omega_n) \\
&\overset{\text{Lemma 6}}{=} \sum_{\substack{\omega_n \in \Omega_n \\ P^\dagger(\omega_n) > 0}} \frac{P_=(\omega_n)}{P_=(\omega_n)} \cdot P^\dagger(\omega_n) \\
&= \sum_{\substack{\omega_n \in \Omega_n \\ P^\dagger(\omega_n) > 0}} P^\dagger(\omega_n) \\
&= 1 \ . \qquad \square
\end{aligned}$$

Note that PIT holds for a class of premisses if and only if inductive contradictions are preserved by premisses in that class:

$$\{\psi \in S\mathcal{L} \, : \, \approx\!\!\!\mid\!\!\!\sim \psi^0\} \subseteq \{\psi \in S\mathcal{L} \, : \, \varphi_1^{X_1}, \varphi_2^{X_2}, \ldots \approx\!\!\!\mid\!\!\!\sim \psi^0\}.$$

We can also say something about when these inclusions are strict:

Theorem 7. *If the premisses are finitely support-satisfiable, then*

$$\{\psi \in S\mathcal{L} \, : \, \approx\!\!\!\mid\!\!\!\sim \psi^1\} \subseteq \{\psi \in S\mathcal{L} \, : \, \varphi_1^{X_1}, \varphi_2^{X_2}, \ldots \approx\!\!\!\mid\!\!\!\sim \psi^1\}$$
$$\{\psi \in S\mathcal{L} \, : \, \approx\!\!\!\mid\!\!\!\sim \psi^0\} \subseteq \{\psi \in S\mathcal{L} \, : \, \varphi_1^{X_1}, \varphi_2^{X_2}, \ldots \approx\!\!\!\mid\!\!\!\sim \psi^0\} \ .$$

The inclusions are strict if and only if there exists some N and some N-state $\omega_N \in \Omega_N$ such that $P^\dagger(\omega_N) = 0$.

Proof. The non-strict inclusion relationships follow directly from PIT (Proposition 11).

If there exists some $\omega_N \in \Omega_N$ such that $P^\dagger(\omega_N) = 0$, then $\neg\omega_N$ follows inductively from the premisses and ω_N is ruled out by the premisses.

For the converse, consider an arbitrary $\psi \in S\mathcal{L}$ such that $0 < P_=(\psi) < 1$. Then there exists some $\omega'_N \in \Omega_N$ such that $P_=(\psi \wedge \omega'_N) > 0$. We may assume that $N > M$ (or K^* in the case of infinitely many premisses).

We assume that $P^\dagger(\omega_N) > 0$ for all $\omega_N \in \Omega_N$. We now show that $P^\dagger(\psi) > 0$ using the fact that $P^\dagger$ is unique and equivocates beyond M (K^* in the case of infinitely many premisses) (Theorem 6):

$$\begin{aligned} P^\dagger(\psi) &\geq P^\dagger(\psi \wedge \omega'_N) = P^\dagger(\psi|\omega'_N) \cdot P^\dagger(\omega'_N) \overset{\text{Lemma 4}}{=} P_=(\psi|\omega'_N) \cdot P^\dagger(\omega'_N) \\ &> 0 \ . \end{aligned}$$

Replacing ψ by $\neg\psi$ we note that $0 < P_=(\psi) < 1$ entails that $0 < P^\dagger(\psi) < 1$.

If $P^\dagger(\omega_N) > 0$ for all $\omega_N \in \Omega_N$, there are hence no new inductive tautologies and no new inductive impossibilities. □

Note that inductive non-contradictions that are consistent with the premisses may be given zero probability in OBIL:

Corollary 6 *(Failure of open-mindedness). There exist finitely many support-satisfiable premisses, an M-state ω_M and $P \in \mathbb{E}$ such that $P(\omega_M) > 0$ but $P^\dagger(\omega_M) = 0$.*

Proof. Let

$$\begin{aligned} \varphi_1^{c_1} &:= ([Rt_1 \wedge Rt_2] \vee \forall x Ux)^{\frac{1}{3}} \\ \varphi_2^{c_2} &:= ([Rt_1 \wedge \neg Rt_2] \vee \forall x Ux)^{\frac{1}{3}} \\ \varphi_3^{c_3} &:= ([\neg Rt_1 \wedge Rt_2] \vee \forall x Ux)^{\frac{1}{3}} \ . \end{aligned}$$

Then $P^\dagger(\neg Rt_1 \wedge \neg Rt_2) = 0$, since $P^\dagger(\forall x Ux) = 0$. However, every probability function with $P(\forall x Ux) = \frac{1}{3} = P(\forall x Ux \wedge \neg Rt_1 \wedge \neg Rt_2)$ satisfies all the premisses (P assigns the conjunctions in the square brackets of the φ_i probability zero). So, as claimed $P(Ut_1 \wedge Ut_2 \wedge \neg Rt_1 \wedge \neg Rt_2) \geq \frac{1}{3} > 0$. □

Note that this differs from the case in which the premisses are quantifier-free, where Open-Mindedness does hold [28, Chapter 7]: if there exists a probability function satisfying quantifier-free premisses which gives some state positive probability, then so does the maximal entropy function.

10. Support-satisfiability

Having shown that OBIL is decidable for finitely many support-satisfiable premisses or finitely presented, satisfiable support, we now investigate the notion of support-satisfiability. Firstly, we contrast the notion of support-satisfiability with satisfiability in more detail. We then show that this contrast has a pronounced effect on maximal entropy functions.

10.1. Support-satisfiability and satisfiability

In the following, we suppose without loss of generality that there is a single categorical premiss φ_1. In the absence of any categorical premisses, one may simply let $\varphi_1 := Ut_1 \vee \neg Ut_1$. If there are multiple categorical premisses, φ_1 can be taken to be their conjunction.

Proposition 12 *(Connection between support-satisfiability and satisfiability). Suppose the premisses take the form $\varphi_1^{c_1}, \ldots, \varphi_k^{c_k}$ with $c_1 = 1$, $k \geq 1$, $c_2, \ldots, c_k \in (0,1)$. The premisses are support-satisfiable if and only if*

1. *the premisses are satisfiable and*
2. *there exists a probability function $P \in \mathbb{E}$ such that whenever $P_{=}(\varphi_1 \wedge \pm\varphi_2 \wedge \ldots \wedge \pm\varphi_k) = 0$, then $P(\varphi_1 \wedge \pm\varphi_2 \wedge \ldots \wedge \pm\varphi_k) = 0$.*

In words, condition 2 says that there exists a probability function $P \in \mathbb{E}$ assigning every measure zero conjunction of premiss sentences or their negations probability zero. Note that every probability function Q solving the support problem must also assign such a conjunction probability zero since there is no M-state which has positive measure taken together with such a conjunction, where $M := \max\{N_{\varphi_1}, \ldots, N_{\varphi_k}\}$.

Proof. Let us first recall that support-satisfiability entails satisfiability (Proposition 8).

In case of $k = 1$, there are no premisses given with non-extreme probability. φ_1^1 being support-satisfiable ($P(\check{\varphi}_1) = 1$ is satisfiable), entails that $\check{\varphi}_1$ is not a contradiction. Hence, there is no such conjunction in 2 and 2 follows trivially.

Now consider the case $k \geq 2$.

Assume that the premisses are support satisfiable with maximal entropy function $P^\dagger$. $P^\dagger$ is a convex combination of probability functions of the form $P_{=}(\cdot|\omega_M)$, where $M := \max\{N_{\varphi_1}, \ldots, N_{\varphi_k}\}$. These functions all assign zero measure sentences zero probability, since they assign all measure one sentences probability 1 [16, Theorem 45]. Hence, so does $P^\dagger$, which is in $\mathbb{E}$.

Consider the converse implication. Suppose condition 1 holds and let $P \in \mathbb{E}$ satisfy condition 2. For the remainder of this proof and all $i \in \{2, \ldots, k\}$ we use φ_i^1 to denote φ_i and φ_i^0 to denote $\neg\varphi_i$. Note first that the $\varphi_1^1 \wedge \varphi_2^{\epsilon_2} \wedge \ldots \wedge \varphi_k^{\epsilon_k}$ are pairwise inconsistent. Note also that ${\check{\varphi}_1}^1 \wedge {\check{\varphi}_2}^{\epsilon_2} \wedge \ldots \wedge {\check{\varphi}_k}^{\epsilon_k}$ is a contradiction if and only if ${\check{\varphi}_1}^1 \wedge {\check{\varphi}_2}^{\epsilon_2} \wedge \ldots \wedge {\check{\varphi}_k}^{\epsilon_k}$ has measure zero (this sentence is quantifier-free) if and only if ${\varphi_1}^1 \wedge {\varphi_2}^{\epsilon_2} \wedge \ldots \wedge \varphi_k^{\epsilon_k}$ has measure zero (Lemma 3 (iv) shows that $P_{=}(\varphi) = P_{=}(\check{\varphi})$).

Then define a function Q on the M-states as follows:

$$Q({\check{\varphi}_1}^{\epsilon_1} \wedge {\check{\varphi}_2}^{\epsilon_2} \wedge \ldots \wedge {\check{\varphi}_k}^{\epsilon_k}) := P(\varphi_1^{\epsilon_1} \wedge \varphi_2^{\epsilon_2} \wedge \ldots \wedge \varphi_k^{\epsilon_k}) \ ,$$

and let Q equivocate between those M-states which entail the same ${\check{\varphi}_1}^{\epsilon_1} \wedge {\check{\varphi}_2}^{\epsilon_2} \wedge \ldots \wedge {\check{\varphi}_k}^{\epsilon_k}$. In particular, $Q(\neg\check{\varphi}_1 \wedge {\check{\varphi}_2}^{\epsilon_2} \wedge \ldots \wedge {\check{\varphi}_k}^{\epsilon_k}) = 0$ for all $\epsilon_2, \ldots, \epsilon_k$ since $P(\neg\varphi_1) = 0$. In particular, $Q(\check{\varphi}_1) = 1$ and $Q(\neg\check{\varphi}_1) = 0$.

We next observe

$$\sum_{\omega \in \Omega_M} Q(\omega) = \sum_{\vec{\epsilon} \in \{0,1\}^{k-1}} Q(\check{\varphi}_1 \wedge {\check{\varphi}_2}^{\epsilon_2} \ldots \wedge {\check{\varphi}_k}^{\epsilon_k})$$

$$= \sum_{\vec{\epsilon}\in\{0,1\}^{k-1}} P(\varphi_1 \wedge \varphi_2^{\epsilon_2} \wedge \ldots \wedge \varphi_k^{\epsilon_k})$$
$$= P(\varphi_1) = 1 \ .$$

This means that Q is a probability function on Ω_M. In other words, there exists some $P \in \mathbb{P}$ which agrees with Q on Ω_M.

Let us now extend Q to the subsets of Ω_M as usual: for all $S \subseteq \Omega_M$ let $Q(S) := \sum_{\omega\in S} Q(\omega)$. We note that for all $2 \le i \le k$,

$$\begin{aligned} Q(\check{\varphi}_i) &= \sum_{\vec{\epsilon}\in\{0,1\}^{k-2}} Q(\check{\varphi_1}^1 \wedge \check{\varphi_2}^{\epsilon_2} \wedge \ldots \wedge \check{\varphi_{i-1}}^{\epsilon_{i-1}} \wedge \check{\varphi}_i \wedge \check{\varphi_{i+1}}^{\epsilon_{i+1}} \wedge \check{\varphi_k}^{\epsilon_k}) \\ &= \sum_{\vec{\epsilon}\in\{0,1\}^{k-2}} P(\varphi_1{}^1 \wedge \varphi_2{}^{\epsilon_2} \wedge \ldots \wedge \varphi_{i-1}^{\epsilon_{i-1}} \wedge \varphi_i \wedge \varphi_{i+1}^{\epsilon_{i+1}} \wedge \varphi_k^{\epsilon_k}) \\ &= P(\varphi_i) \\ &= c_i \ . \end{aligned}$$

Hence, Q is a probability function on Ω_M satisfying $Q(\varphi_i) = c_i$ for all $1 \le i \le k$. Hence, the premisses are support-satisfiable. □

There are sets of premisses that are jointly satisfiable, and where every premiss sentence has positive measure, but that are not support-satisfiable:

Observation 4. Satisfiability of the premisses does not entail their support-satisfiability.

This emphasises that the second condition of the equivalence in Proposition 12 cannot be omitted.

Proof. Consider premisses $\varphi_1^{c_1} := (Ut_1 \vee \forall x V x)^{.9}, \varphi_2^{c_2} := (\neg Ut_1 \vee \forall x V x)^{.9}$ [16, Example 43]. Clearly, these two premisses are jointly satisfiable. Support-satisfiability holds if and only if $P(Ut_1) = 0.9$ and $P(\neg Ut_1) = 0.9$ are jointly satisfiable. This is obviously not the case. □

Support-satisfiability with non-extreme c_is entails that all premiss sentences have non-extreme measures:

Observation 5. If $X_i = c_i \in (0, 1)$ for some $i \in I$ and the premisses are support-satisfiable, then $P_=(\varphi_i) \in (0, 1)$ for this $i \in I$.

Proof. Let us suppose for contradiction that $P_=(\varphi_i) = 1$ (replace φ_i by $\neg\varphi_i$ if necessary). This means that $\check{\varphi}_i$ is the disjunction of all N-states (i.e., a tautology) and $\neg\check{\varphi}_i$ is the empty disjunction (a contradiction). Support-satisfiability requires that $P(\check{\varphi}_i) = c_i \in (0, 1)$ and $P(\neg\check{\varphi}_i) = 1 - c_i \in (0, 1)$. However, $P(\check{\varphi}_i) = 1$ and $P(\neg\check{\varphi}_i) = 0$. Contradiction. □

This shows that while sets of support-satisfiable premisses are such that every premiss sentence and its negation must have positive measure, a conjunction of premiss sentences may have measure zero.

10.2. Support-satisfiability and inductive tautologies

The question arises as to why finitely many premisses that are support-satisfiable are so well-behaved in OBIL. After all, we know that there is (i) a satisfiable premiss ($\varphi = \exists x\forall y Uxy$) that does not yield a maximal entropy function [16, Proposition 53] and (ii) a premiss without a maximal entropy function which nevertheless yields a well-defined unique maximal entropy function after adding a further categorical measure-one premiss [15, Proposition 5].

We now give another characterisation of the key notion of support-satisfiability for finitely many premisses. This observation exploits our above result on the preservation of inductive tautologies. This characterisation will be used to elucidate similarities between OBIL and entropy maximisation on finite domains.

Proposition 13 *(Characterisation of support-satisfiability). Let the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ be satisfiable. The following are equivalent:*

1. $\check{\mathbb{E}} = \emptyset$.
2. *For all $P \in \mathbb{E}$ there exists a sentence $\varphi \in S\mathcal{L}$ such that $P(\varphi) \in (0,1)$ and $P_=(\varphi) \in \{0,1\}$.*

Proof. Let us first assume the negation of the first condition, i.e., $\check{\mathbb{E}} \neq \emptyset$. Applying Proposition 11 we conclude that $P^\dagger$ exists and satisfies PIT. So, $P_=(\varphi) \in \{0,1\}$ entails that $P^\dagger(\varphi) \notin (0,1)$. So, the second condition fails ($P^\dagger \in \mathbb{E}$). Hence, the second condition implies the first.

Now assume the first condition. Since the premisses are satisfiable ($\mathbb{E} \neq \emptyset$) but not support-satisfiable ($\check{\mathbb{E}} = \emptyset$), for all $P \in \mathbb{E}$ there must exist a premiss φ_i such that $P(\varphi_i) \neq P(\check{\varphi}_i)$. Swapping φ_i with $\neg\varphi_i$ if necessary, we may assume that $P(\varphi_i) > P(\check{\varphi}_i)$. We now note that

$$\begin{aligned} P(\varphi_i) &= \sum_{\substack{\omega_N \in \Omega_N \\ P_=(\omega_N \wedge \varphi_i) > 0}} P(\omega_N \wedge \varphi) + \sum_{\substack{\omega_N \in \Omega_N \\ P_=(\omega_N \wedge \varphi_i) = 0}} P(\omega_N \wedge \varphi) \\ &= P(\check{\varphi}_i) + \sum_{\substack{\omega_N \in \Omega_N \\ P_=(\omega_N \wedge \varphi_i) = 0}} P(\omega_N \wedge \varphi) \ . \end{aligned}$$

Since $P(\varphi_i) > P(\check{\varphi}_i)$, the last term must have non-zero probability. Since all sentences $\omega_N \wedge \varphi$ have measure zero, at least one measure zero sentence is assigned non-zero probability by P. □

A failure of condition 2 in Proposition 13 is equivalent to PIT. If condition 2 fails then by Proposition 13, condition 1 fails, i.e., support satisfiability holds, and by Proposition 11 this implies PIT. On the other hand, if condition 2 holds, then it holds for the maximal entropy function in particular (if it exists), so an inductive tautology is not preserved, i.e., PIT fails. If there is no maximal entropy function of $\mathbb{E}$, then OBIL uses every function of $\mathbb{E}$ for inference and PIT fails, too.

We see, then, that satisfiable premisses naturally sub-divide into two mutually exclusive cases. The first case comprises the non-support-satisfiable premisses where all probability functions in the feasible region give some inductive tautologies probability less than 1 and give some inductive contradictions positive probability ($P_=(\varphi) = 0 < P(\varphi)$ and $P_=(\neg\varphi) = 1 > P(\neg\varphi)$). The second case comprises support-satisfiable premisses which possess a unique maximal entropy function $P^\dagger$ that satisfies PIT (inductive tautologies and inductive contradictions remain as such).

Let us compare this with maximum entropy reasoning on quantifier-free languages. There, the only inductive tautologies are deductive tautologies and the only inductive contradictions are deductive contradictions. Hence, every satisfiable set of premisses is as in the second case. The first case can only arise when quantifiers are introduced.

11. Conclusion

The undecidability of first-order deductive logic carries over to first-order inductive logic under the standard semantics (Proposition 3). It is therefore interesting and surprising that a large class of inferences in first-order objective Bayesian inductive logic is decidable: namely the class of support-satisfiable inferences (Theorem 5).[13] In particular, in OBIL the class of inferences from no premisses is decidable (Corollary 3); this does not hold of first-order deductive logic, and hence it does not hold of first-order inductive logic with the standard semantics (by Proposition 3).

The main question for further research concerns the extent to which this decidable class of inferences can be expanded. We saw that the class of quantifier-free inferences is decidable (Theorem 1). Moreover, any class of inferences from unsatisfiable premisses is trivially decidable, as long as one can effectively determine that the premisses are unsatisfiable. This is because the equivocator function is used for inference whenever the premisses are unsatisfiable. (Examples of such classes include the class of inferences from unsatisfiable premisses that involve only unary predicates [22], and the class of inferences from unsatisfiable Σ_2 premisses [32].) Hence any class of inferences involving premisses that are quantifier-free, support-satisfiable or decidably unsatisfiable is decidable. In addition, there are decidable inferences from premisses that are satisfiable but not support-satisfiable. For example, the class of inferences with a single premiss of the form $\forall x U x^c$ is decidable [16, Example 17]. A key task for further research is to find other decidable subclasses of the class of inferences that are satisfiable but do not have satisfiable support.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Stephen Boyd and Natarajan Sukumar for very useful advice on convex optimisation. We are also very grateful to the comments of the anonymous referees, which led to substantive improvements to the paper. Jon Williamson's research was supported by grants from the Leverhulme Trust (RPG-2022-336) and UKRI (APP25130). Jürgen Landes gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 528031869. Soroush Rafiee Rad's research was partially funded by the Italian Ministry of University and Research (MUR) under the PRIN 2022 Project "CORTEX" (cod. 2022ZLLR3T), and partly by the Leverhulme Trust Research Project Grant RPG-2023-236, What If...? Knowing by Imagining [WIKI]: the Logic and Rationality of Imagination.

Appendix A. Lagrange multipliers for determining entropy maximisers

In this appendix, we sketch how the maximal entropy function $\check{P}^\dagger$ can be obtained using Lagrange-multiplier optimisation methods.

Recall that $\check{P}^\dagger$ is the (unique) maximal entropy function with respect to the support problem, $\text{maxent}\{\check{\varphi}_1^{X_1}, \ldots, \check{\varphi}_k^{X_k}\} = \{\check{P}^\dagger\}$ (Theorems 2 and 3). For ease of exposition, we take $X_1 = c_1, \ldots, X_k = c_k$ where $c_1, \ldots, c_k \in [0, 1]$, but the approach can be straightforwardly generalised to the case in which $X_1, \ldots, X_k$ are intervals by using inequality constraints instead of equality constraints. Throughout this appendix we work with the natural logarithm for convenience.

Fix $n \geq \max\{N_{\check{\varphi}_1}, \ldots, N_{\check{\varphi}_k}\}$. Let $x_\omega \stackrel{\text{df}}{=} \check{P}^\dagger(\omega)$ for each $\omega \in \Omega_n$.

The task is to use Lagrange multipliers to solve an optimisation problem to find the x_ω subject to $k+1$ constraints. We have an additivity constraint with multiplier $\mu \in \mathbb{R}$:

[13] Recall that we focus throughout on decidability modulo comparison of real numbers, by assuming that inferences are finitely represented and that numerical comparisons are made to some given fixed precision.

$$\sum_{\omega\in\Omega_n} x_\omega = 1 \quad . \tag{8}$$

We also have k premiss constraints with multipliers $\lambda_i \in \mathbb{R}$ for $1 \le i \le k$:

$$\sum_{\omega\models\check{\varphi}_i} x_\omega = c_i \quad . \tag{9}$$

Call these latter constraints $f_1, \ldots, f_k$. The Lagrange function L is

$$L = -\sum_{\omega\in\Omega_n} x_\omega \log x_\omega + \mu(-1 + \sum_{\omega\in\Omega_n} x_\omega) + \sum_{i=1}^{k} \lambda_i(-c_i + \sum_{\omega\models\check{\varphi}_i} x_\omega) \quad . \tag{10}$$

The Lagrange equations are obtained by taking partial derivatives of L with respect to the unknown x_ω, for each $\omega \in \Omega_n$:

$$\frac{\partial H_n}{\partial x_\omega} + \mu + \sum_{i=1}^{k} \lambda_i \frac{\partial f_i}{\partial x_\omega} = 0 \quad .$$

Here,

$$\frac{\partial H_n}{\partial x_\omega} = -1 - \log x_\omega$$

and

$$\frac{\partial f_i}{\partial x_\omega} = \begin{cases} 1 & : \quad \omega \models \check{\varphi}_i \\ 0 & : \quad \text{otherwise.} \end{cases}$$

So we have:

$$\log x_\omega = -1 + \mu + \sum_{i:\omega\models\check{\varphi}_i} \lambda_i \tag{11}$$

and thus

$$x_\omega = e^{-1+\mu} \cdot e^{\sum_{i:\omega\models\check{\varphi}_i} \lambda_i} \quad . \tag{12}$$

Since probabilities sum to one, we find

$$1 = \sum_{\omega\in\Omega_n} x_\omega = \sum_{\omega\in\Omega_n} e^{-1+\mu} \cdot e^{\sum_{i:\omega\models\check{\varphi}_i} \lambda_i} = e^{-1+\mu} \cdot \sum_{\omega\in\Omega_n} e^{\sum_{i:\omega\models\check{\varphi}_i} \lambda_i} \quad . \tag{13}$$

Let us now consider the k constraints arising from the premisses (9),

$$c_i = \sum_{\omega\models\check{\varphi}_i} x_\omega = \sum_{\omega\models\check{\varphi}_i} e^{-1+\mu} \cdot e^{\sum_{i:\omega\models\check{\varphi}_i} \lambda_i} = \frac{\sum_{\omega\models\check{\varphi}_i} e^{\sum_{i:\omega\models\check{\varphi}_i} \lambda_i}}{\sum_{\omega\in\Omega_n} e^{\sum_{i:\omega\models\check{\varphi}_i} \lambda_i}} \quad . \tag{14}$$

This is a set of k equations in the unknowns $\lambda_1, \ldots, \lambda_k$. Once these λ are determined we can use (13) to determine μ and thus the x_ω from (12).

Substituting (11) into (10) gives using (13) in the last step

$$L = -\sum_{\omega\in\Omega_n} x_\omega[-1+\mu+\sum_{i:\omega\models\check{\varphi}_i}\lambda_i] + \mu(-1+\sum_{\omega\in\Omega_n} x_\omega) + \sum_{i=1}^k \lambda_i(-c_i+\sum_{\omega\models\check{\varphi}_i} x_\omega)$$

$$= 1-\mu-\sum_{\omega\in\Omega_n} x_\omega \sum_{i:\omega\models\check{\varphi}_i}\lambda_i + \sum_{i=1}^k \lambda_i(-c_i+\sum_{\omega\models\check{\varphi}_i} x_\omega)$$

$$= 1-\mu-\sum_{i=1}^k \lambda_i c_i$$

$$= \frac{1}{\log\left(\sum_{\omega\in\Omega_n} e^{\sum_{i:\omega\models\check{\varphi}_i}\lambda_i}\right)} - \sum_{i=1}^k \lambda_i c_i \ .$$

Since the original problem of maximising n-entropy is a convex minimisation problem (minimise $-H_n$) and H_n is continuous, maximising n-entropy is equivalent to maximising the above equation, called the dual problem. This dual problem is a convex optimisation problem [4, p. 215].

We have hence three ways of maximising n-entropy:

1. Numerically solve the problem of maximising $H_n(P)$.
2. Solve (14) for the unknown λ_i.
3. Solve the dual optimisation problem.

In practice, the choice of method will depend on circumstances.

1. Maximising $H_n(P)$ is a convex optimisation problem, that hence has a unique solution. One may use one of the many gradient descent (hill climbing) algorithms to find arbitrarily good approximations of $\check{P}^\dagger$ and $H_n(\check{P}^\dagger)$. The number of unknowns to be determined is $|\Omega_n|$. Such algorithms are hence likely to perform well if $|\Omega_n|$ is small.
2. (14) is a system of k multilinear equations. Since a solution of this system provides a solution to the original problem, this system must have at least one solution. The number of unknowns to be determined is k, the number of premisses. A solution is such more likely to be found quickly if k is small.
3. Solving the dual problem is again a convex optimisation, which can again be tackled by a gradient descent algorithm (hill climbing). The number of unknowns to be determined is k, such algorithms are hence likely to perform well if k is small.

What is the n-entropy of $\check{P}^\dagger$? Write $x_\omega = z_0\prod_{i:\omega\models\check{\varphi}_i} z_i$, where $z_0 \stackrel{\text{df}}{=} e^{\mu-1}$ and $z_i \stackrel{\text{df}}{=} e^{\lambda_i}$. Then,

$$H_n(\check{P}^\dagger) = -\sum_{\omega\in\Omega_n} x_\omega \log x_\omega$$

$$= -\sum_{\omega\in\Omega_n} x_\omega \log\left(z_0 \prod_{i:\omega\models\check{\varphi}_i} z_i\right)$$

$$= -\sum_{\omega\in\Omega_n} x_\omega \left(\log z_0 + \sum_{i:\omega\models\check{\varphi}_i}\log z_i\right)$$

$$
\begin{aligned}
&= -\sum_{\omega\in\Omega_n} x_\omega \left(\mu - 1 + \sum_{i:\omega\models\check{\varphi}_i} \lambda_i\right) \\
&= -(\mu-1)\sum_{\omega\in\Omega_n} x_\omega - \sum_{i=1}^{k} \lambda_i \sum_{\omega\models\check{\varphi}_i} x_\omega \\
&= 1-\mu-\sum_{i=1}^{k}\lambda_i \check{P}^\dagger(\check{\varphi}_i) \\
&= 1-\mu-\sum_{i=1}^{k}\lambda_i c_i \ .
\end{aligned}
$$

Thus, the n-entropy of $\check{P}^\dagger$ can be straightforwardly determined from the values of the Lagrange multipliers.

Data availability

No data was used for the research described in the article.

References

[1] M. Abadi, J.Y. Halpern, Decidability and expressiveness for first-order logics of probability, Inf. Comput. 112 (1) (1994) 1–36.
[2] A. Balestrino, A. Caiti, E. Crisostomi, Efficient numerical approximation of maximum entropy estimates, Int. J. Control 79 (9) (2006) 1145–1155.
[3] Egon Börger, Erich Grädel, Yuri Gurevich, The Classical Decision Problem, Perspectives in Mathematical Logic, Springer, Berlin, 1997.
[4] Stephen Boyd, Lieven Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, 2004.
[5] Thomas M. Cover, Joy A. Thomas, Elements of Information Theory, John Wiley and Sons, New York, 1991, second (2006) edition.
[6] Hartry H. Field, Logic, meaning, and conceptual role, J. Philos. 74 (7) (1977) 379.
[7] Haim Gaifman, Concerning measures in first order calculi, Isr. J. Math. 2 (1) (1964) 1–18.
[8] Sally A. Goldman, Ronald L. Rivest, A non-iterative maximum entropy algorithm, in: L.N. Kanal, J.F. Lemmer (Eds.), Uncertainty in Artificial Intelligence, Elsevier, North-Holland, 1988, pp. 133–148.
[9] Rolf Haenni, Jan-Willem Romeijn, Gregory Wheeler, Jon Williamson, Probabilistic Logics and Probabilistic Networks, Synthese Library. Springer, Dordrecht, 2011.
[10] Theodore Hailperin, Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications, Lehigh University Press, Bethlehem, Penn, 1996.
[11] Diego Havenstein, Peter Lysakovski, Norman May, Guido Moerkotte, Gabriele Steidl, Fast entropy maximization for selectivity estimation of conjunctive predicates on cpus and gpus, in: Angela Bonifati, Yongluan Zhou, Marcos Antonio Vaz Salles, Alexander Böhm, Dan Olteanu, George Fletcher, Arijit Khan, Bin Yang (Eds.), Proceedings of EDBT, Konstanz, 2020. OpenProceedings.
[12] Elad Hazan, Sham Kakade, Karan Singh, Abby Van Soest, Provably efficient maximum entropy exploration, in: Kamalika Chaudhuri, Ruslan Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 97, PMLR, 2019, pp. 2681–2691.
[13] David Hilbert, Wilhelm Ackermann, Principles of Mathematical Logic, Chelsea Publishing Company, New York, 1928, 1950 edition.
[14] Leonid Khachiyan, Fourier-Motzkin elimination method, in: Christodoulos A. Floudas, Panos M. Pardalos (Eds.), Encyclopedia of Optimization, Kluwer, Dordrecht, 2001, pp. 695–699.
[15] Jürgen Landes, Rules of proof for maximal entropy inference, Int. J. Approx. Reason. 153 (2023) 144–171.
[16] Jürgen Landes, Soroush Rafiee Rad, Jon Williamson, Determining maximal entropy functions for objective Bayesian inductive logic, J. Philos. Log. 52 (2023) 555–608.
[17] Jürgen Landes, Jon Williamson, Justifying objective Bayesianism on predicate languages, Entropy 17 (4) (2015) 2459–2543.
[18] Jürgen Landes, Jon Williamson, Objective Bayesian nets from consistent datasets, in: Adom Giffin, Kevin H. Knuth (Eds.), Proceedings of MaxEnt, vol. 1757, AIP, 2016, pp. 020007–1–020007–8.
[19] Jürgen Landes, Jon Williamson, Objective Bayesian nets for integrating consistent datasets, J. Artif. Intell. Res. 74 (2022) 393–458.
[20] Jürgen Landes, Soroush Rafiee Rad, Jon Williamson, Towards the entropy-limit conjecture, Ann. Pure Appl. Log. 172 (2) (2021) 102870.
[21] Yunseok Lee, Holger Boche, Gitta Kutyniok, Computability of optimizers, IEEE Trans. Inf. Theory 70 (2024) 2967–2983.
[22] Leopold Löwenheim, On possibilities in the calculus of relatives, Math. Ann. 76 (1915) 447–470.

[23] V. Markl, P.J. Haas, M. Kutsch, N. Megiddo, U. Srivastava, T.M. Tran, Consistent selectivity estimation via maximum entropy, VLDB J. 16 (1) (2006) 55–76.
[24] J. Donald Monk, Mathematical Logic, Graduate Texts in Mathematics, Springer, New York, 1976.
[25] Richard E. Neapolitan, Probabilistic Reasoning in Expert Systems: Theory and Algorithms, Wiley, New York, 1990.
[26] Nils J. Nilsson, Probabilistic logic, Artif. Intell. 28 (1) (1986) 71–87.
[27] Dirk Ormoneit, Halbert White, An efficient algorithm to compute maximum entropy densities, Econom. Rev. 18 (2) (1999) 127–140.
[28] Jeff B. Paris, The Uncertain Reasoner's Companion, Cambridge University Press, Cambridge, 1994.
[29] Jeff B. Paris, Alena Vencovská, Pure Inductive Logic, Cambridge University Press, Cambridge, 2015.
[30] Judea Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[31] Soroush Rafiee Rad, Probabilistic characterisation of models of first-order theories, Ann. Pure Appl. Log. 172 (1) (2021) 102875.
[32] Frank P. Ramsey, On a problem of formal logic, Proc. Lond. Math. Soc. s2–30(1) (1930) 264–286.
[33] Sebastiaan A. Terwijn, Decidability and undecidability in probability logic, in: Sergei Artemov, Anil Nerode (Eds.), Logical Foundations of Computer Science, Springer, Berlin, Heidelberg, 2009, pp. 441–450.
[34] Boris A. Trakhtenbrot, On recursive separability, Dokl. Akad. Nauk SSSR 88 (6) (1953) 953–956.
[35] Gregory Wheeler, Objective Bayesian calibration and the problem of non-convex evidence, Br. J. Philos. Sci. 63 (4) (2012) 841–850.
[36] Jon Williamson, Bayesian Nets and Causality: Philosophical and Computational Foundations, Oxford University Press, Oxford, 2005.
[37] Jon Williamson, Objective Bayesian nets, in: Sergei Artemov, Howard Barringer, Artur S. d'Avila Garcez, Luis C. Lamb, John Woods (Eds.), We Will Show Them! Essays in Honour of Dov Gabbay, vol. 2, College Publications, London, 2005, pp. 713–730.
[38] Jon Williamson, Objective Bayesian probabilistic logic, J. Algorithms 63 (4) (2008) 167–183.
[39] Jon Williamson, in: Defence of Objective Bayesianism, Oxford University Press, Oxford, 2010.
[40] Jon Williamson, Calibration and convexity: response to Gregory Wheeler, Br. J. Philos. Sci. 63 (4) (2012) 851–857.
[41] Jon Williamson, Lectures on Inductive Logic, Oxford University Press, Oxford, 2017.