Entailment Among Probabilistic Implications*

Albert Atserias and José L. Balcázar
Department of Computer Science
Universitat Politècnica de Catalunya

May 19, 2015

Abstract

We study a natural variant of the implicational fragment of propositional logic. Its formulas are pairs of conjunctions of positive literals, related together by an implicational-like connective; the semantics of this sort of implication is defined in terms of a threshold on a conditional probability of the consequent, given the antecedent: we are dealing with what the data analysis community calls confidence of partial implications or association rules. Existing studies of redundancy among these partial implications have characterized so far only entailment from one premise and entailment from two premises. By exploiting a previously noted alternative view of this entailment in terms of linear programming duality, we characterize exactly the cases of entailment from arbitrary numbers of premises. As a result, we obtain decision algorithms of better complexity; additionally, for each potential case of entailment, we identify a critical confidence threshold and show that it is, actually, intrinsic to each set of premises and antecedent of the conclusion.

1 Introduction

The quite deep issue of how to represent human knowledge in a way that is most useful for applications has been present in research for decades now. Often, knowledge representation is necessary in a context of incomplete information, whereby inductive processes are required in addition. As a result, two facets that are common to a great number of works in knowledge representation, and particularly more so in contexts of inductive inference, machine learning, or data analysis, are logic and probability.

Adding probability-based mechanisms to already expressive logics enhances their expressiveness and usefulness, but pays heavy prices in terms of computational difficulty. Even

*First author partially supported by MINECO project TASSAT2 (TIN2013-48031-C4-1-P). Second author partially supported by SGR2014-890 (MACDA) of the Generalitat de Catalunya, MICINN project BASMATI (TIN2011-27479-C04-03) and MINECO project APCOM (TIN2014-57226-P).

without probability, certain degrees of expressivity and computational feasibility are known to be incompatible, and this is reflected in the undecidability results for many logics. In other cases, the balance between expressivity and feasibility hinges on often open complexity-theoretic statements. To work only within logics known to be polynomially tractable may imply serious expressiveness limitations.

Literally hundreds of studies have explored this difficult balance. Even limiting ourselves somewhat to the machine learning perspective, we could mention a large number of references such as those cited in the book [1], for one.

Both in machine learning and in data mining, one particularly well-studied knowledge representation mechanism is given by relaxed implication connectives: a relatively natural abstract concept which can be made concrete in various ways. The common idea is to relax the semantics of the implication connective so as to allow for exceptions, a feature actually mandatory in all applications in data analysis or machine learning. However, this can be done in any of a number of ways; and each form of endowing relaxed implications with a precise meaning yields a different notion with, often, very different properties. See the survey [2].

This paper focuses on one of the simplest forms of relaxed implication, endowed with its most natural semantics: the one given by conditional probability. Syntactically, these partial implications are pairs of conjunctions of positive propositional literals. For sets X and Y of propositional variables, we write the corresponding implication as $X \to Y$. Now, instead of the classical semantics, whereby a model satisfies the implication if it either fails the antecedent or fulfills the consequent, we want to quantify exceptions; hence, instead of individual propositional models, our semantic structures are, then, so-called "transactional datasets", that is, multisets of propositional models. By mere counting, we find, on each dataset, a frequentist probability for X and Y seen as conjunctions (or, equivalently, as events): then, the meaning of the implication is simply that the conditional probability of the consequent, given the antecedent, exceeds some fixed threshold, here denoted $\gamma \in (0, 1)$. In application-aware works, very often that quantity, the frequentist conditional probability, is called *confidence* of the partial implication. We also use this name here.

This probabilistic version of implications has been proposed in different research communities. For instance, [3] introduced them as "partial implications"; much later, [4] defined "association rules" (see also [5] and the survey [6]): these are partial implications that impose the additional condition that the consequent is a single propositional variable, and where additional related parameters are used to assess their interest.

Actually, confidence does not seem to be the best choice in practice for the meaning of a partial implication, as discussed e.g. in [2]. However, it is clearly the most natural choice and the obvious step to start the logical study of partial implications, many other preferable options being themselves, actually, variations or sophistications of confidence.

Motivated by practical issues, several works have analyzed notions of redundancy among partial implications: two proposals in [7] and [8] turned out to be equivalent among them and were, in turn, as described in [9], equivalent to the natural notion of logical entailment of one partial implication by another (modulo minor details such as allowing or disallowing empty antecedents or consequents). This entailment means that any dataset in which the premise reaches confidence at least γ must assign confidence at least γ as well to the conclusion.

The contributions of [9] that are relevant to the present paper are chiefly syntactic characterizations of one partial implication entailing another, and of two partial implications entailing another. Further details are provided below; for the time being, we simply indicate that, whereas the case of one premise is quite natural, the case of two premises is quite complex. For perspective, let's briefly consider here the case of transitivity. In contrast with full implications, which obey it, here transitivity fails: it is not difficult to see that, if $X \to Y$ has confidence over γ, and $Y \to Z$ as well, still most occurrences of Y could be without X, leaving low or even zero confidence for $X \to Z$. Even if we consider $X \to Y$ and $XY \to Z$, the probabilities multiply together and leave just $\gamma^2 < \gamma$ as provable threshold. (Cf. [9].)

A tempting intuition is to generalize the observation and jump to the statement that no nontrivial consequence follows from two partial implications; however, this statement is wrong, and an explicit example of proper entailment from two premises is given in the same reference and restated below in Section 3.1.

Besides offering this observation, [9] goes beyond, and generalizes the example into a precise characterization of when a partial implication is entailed by two partial implications. The proof is not deep, using just basic set-theoretic constructions; but it is long, cumbersome, and of limited intuitive value. Attempts at generalizing it directly to more than two premises rapidly reach unmanageable difficulties, among which the most important one is the lack of hints at a crucial property that we will explain below in Section 5.1.

Here, we identify an alternative, quite different approach, that turns out to be successful in finding the right generalization. The new ingredient is a connection with linear programming that is almost identical to a technical lemma in [10]. Stated in our language, the lemma asserts that k partial implications entail another if and only if the dual of a natural linear program associated to the entailment is feasible. We develop this tool and use it to get our main results:

1) for low enough values of the confidence threshold γ, we use this connection to show that k partial implications never entail nontrivially another one;

2) for high enough values of γ, we use it also to provide a characterization of the cases in which k partial implications entail another one, but this one purely in terms of elementary Boolean algebraic conditions among the sets of attributes that make the partial implications;

3) for the intermediate values of γ, we explain how to compute the exact threshold, if any, at which a specific set of k partial implications entails another one.

The characterizations provide algorithms to decide whether a given entailment holds. More concretely, under very general conditions including the case that γ is large, the connection to linear programming gives an algorithm that is polynomial in the number of premises k, but exponential in the number of attributes n. Our subsequent characterization reverses the situation: it gives an algorithm that is polynomial in n but exponential in k. This may sound surprising since the proof of this characterization is based on the previous LP-based take; but it merely reflects the fact that, in our proof of our main characterization, the theory of linear programming was just used as a technical tool.

At any rate, our main characterization also shows that the decision problem for entailments at large γ is in NP, and this does not seem to follow from the linear programming formulation by itself (since the program is exponentially big in n), let alone the definition of entailment (since the number of datasets on n attributes is infinite). We discuss this in Section 7.

2 Preliminaries and notation

Our expressions involve propositional variables, which receive Boolean values from propositional models; we define their semantics through data-sets: simply, multisets of propositional models. However, we mostly follow a terminology closer to the standard one in the data analysis community, where our propositional variables are called attributes or, sometimes, items; likewise, a set of attributes (that is, a propositional model), seen as an element of a dataset, is often called a transaction.

Thus, attributes take Boolean values, true or false, and a transaction is simply a subset of attributes, those that would be set to true if we thought of it as a propositional model. Typically, our set of attributes is simply $[n] := \{1, \ldots, n\}$, for a natural number n, so transactions are subsets of $[n]$. Fix now such a set of attributes.

If Z is a transaction and X is a set of attributes, we say that Z covers X if $X \subseteq Z$. A data-set, as a multi-set of transactions, is formally a mapping from the set of all transactions to the natural numbers: their multiplicities as members of the data-set (alternative formalizations exist in the literature). If $\mathcal{D}$ is a data-set and X is a set of attributes, we write $\mathrm{C}_{\mathcal{D}}[\,X\,]$ for the number of transactions in $\mathcal{D}$ that cover X, counted with multiplicity.

A partial or probabilistic implication is made of a pair of finite subsets X and Y of attributes. We write them as $X \to Y$. If X and Y are sets of attributes, we write XY to denote their union $X \cup Y$. This is fully customary and very convenient notation in this context. Let $X \to Y$ be a partial implication with all its attributes in $[n]$. If $\mathcal{D}$ is a data-set on the set of attributes $[n]$, and γ is a real parameter in the interval $[0, 1]$, we write $\mathcal{D} \models_\gamma X \to Y$ if either $\mathrm{C}_{\mathcal{D}}[\,X\,] = 0$, or else $\mathrm{C}_{\mathcal{D}}[\,XY\,]/\mathrm{C}_{\mathcal{D}}[\,X\,] \geq \gamma$. Thus, if we think of $\mathcal{D}$ as specifying the probability distribution on the set of transactions that assigns probabilities proportionally to their multiplicity in $\mathcal{D}$, then $\mathcal{D} \models_\gamma X \to Y$ if and only if the conditional probability of Y given X is at least γ.

If $X_0 \to Y_0, \ldots, X_k \to Y_k$ are partial implications, we write

$$X_1 \to Y_1, \ldots, X_k \to Y_k \models_\gamma X_0 \to Y_0 \tag{1}$$

if for every data-set $\mathcal{D}$ for which $\mathcal{D} \models_\gamma X_i \to Y_i$ holds for every $i \in [k]$, it also holds that $\mathcal{D} \models_\gamma X_0 \to Y_0$. Note that the symbol $\models_\gamma$ is overloaded much in the same way that the symbol $\models$ is overloaded in propositional logic. In case Expression (1) holds, we say that the entailment holds, or that the set $X_1 \to Y_1, \ldots, X_k \to Y_k$ entails $X_0 \to Y_0$ at confidence threshold γ. If Σ is a set of partial implications for which $\Sigma \models_\gamma X_0 \to Y_0$ holds but $\Gamma \models_\gamma X_0 \to Y_0$ does not hold for any proper subset $\Gamma \subset \Sigma$, then we say that the entailment holds properly. Note that entailments without premises vacuously hold properly when they hold. The real number γ is often referred to as the confidence parameter.

A linear program (LP) is the following optimization problem: $\min\{c^{\mathrm{T}}x : Ax \geq b,\ x \geq 0\}$, where x is a vector of n real variables, b and c are vectors in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and A is a matrix in $\mathbb{R}^{m\times n}$. The program is feasible if there exists an $x \in \mathbb{R}^n$ such that $Ax \geq b$ and $x \geq 0$. The program is unbounded if there exist feasible solutions with arbitrarily small values of the objective function $c^{\mathrm{T}}x$. If the goal were max instead of min, unboundedness would refer to arbitrarily large values of the objective function. The dual LP is $\max\{b^{\mathrm{T}}y : A^{\mathrm{T}}y \leq c,\ y \geq 0\}$, where y is a vector of m real variables. Both LPs together are called a primal-dual pair. The duality theorem of linear programming states that exactly one of the following holds: either both primal and dual are infeasible, or one is unbounded and the other is infeasible, or both are feasible and have optimal points with the same optimal value. See [11] [Corollary 25 and Theorem 23].

3 Related Work and Technical Basis

We review here connected existing work. We describe first the results from [9] on entailments among partial implications with one or two premises. The study there starts with a detailed comparison of entailment as defined in Section 2 with the notions of redundancy among partial implications previously considered in the literature. Here we go directly to the point and consider entailment as defined in Section 2 from the start. Then, we develop a variant of a result in [10], adapted to our context and notation, on which our main results are based, plus additional properties related to that variant.

3.1 Entailment with up to two premises

We discuss here Expression (1) for $k \leq 2$. For this subsection and most of the paper we assume that the confidence parameter γ is strictly positive; otherwise everything holds everywhere, and strictly below 1; otherwise we fall back to classical implication.

The case of zero premises, i.e. tautological partial implications, trivializes to the classical case: $\models_\gamma X_0 \to Y_0$ if and only if $Y_0 \subseteq X_0$, at any positive confidence threshold γ. The first interesting case is thus the entailment from one partial implication $X_1 \to Y_1$ to another $X_0 \to Y_0$. If $X_0 \to Y_0$ is tautological by itself, there is nothing else to say. Otherwise, entailment is still characterized by a simple Boolean algebraic condition on the sets X_0, Y_0, X_1, and Y_1 as stated in the following theorem:

Theorem 1 ([9]). *Let γ be a confidence parameter in $(0, 1)$ and let $X_0 \to Y_0$ and $X_1 \to Y_1$ be two partial implications. Then the following are equivalent:*

1. $X_1 \to Y_1 \models_\gamma X_0 \to Y_0$,
2. *either* $Y_0 \subseteq X_0$, *or* $X_1 \subseteq X_0$ *and* $X_0Y_0 \subseteq X_1Y_1$.

Note that the second statement is independent of γ. This shows that entailment at confidence γ below 1 differs from classical entailment. An example shows this equally well: although it is obvious that $A \to B$ classically entails $AC \to BC$, the entailment fails badly when both the premise and the conclusion are considered as partial implications at some confidence γ in $(0, 1)$: any data-set with many ocurrences of AB, only one occurrence of AC, and none at all of BC, ruins everything. Of course, what fails is that X_0Y_0 is not included in X_1Y_1.

The case of two partial implications entailing a third was also solved in [9]. The starting point for that study was a specific example of a non-trivial entailment:

$$A \to BC,\, A \to BD \models_{1/2} ACD \to B. \tag{2}$$

Indeed, this entailment holds true at any γ in the interval $[1/2, 1)$. This is often found counterintuitive. The intuition of many is that combining two partial implications that only guarantee the threshold $\gamma < 1$ would lead to arithmetic operations leading to values unavoidably below γ. Classical transitivity as discussed in the introduction is a good example. However, this intuition is incorrect, as (2) shows. The good news is that a similar statement, when appropriately generalized, covers all the cases of entailment from two partial implication premises. We omit the proof of (2) as it follows from the next theorem, which will be generalized in our main result.

Theorem 2 ([9]). *Let γ be a confidence parameter in $(0, 1)$ and let $X_0 \to Y_0$, $X_1 \to Y_1$ and $X_2 \to Y_2$ be three partial implications. If $\gamma \geq 1/2$, then the following are equivalent:*

1. $X_1 \to Y_1,\, X_2 \to Y_2 \models_\gamma X_0 \to Y_0$,
2. *either $Y_0 \subseteq X_0$, or $X_i \subseteq X_0$ and $X_0Y_0 \subseteq X_iY_i$ for some $i \in \{1, 2\}$, or all seven inclusions below hold simultaneously:*
   - *(a) $X_1 \subseteq X_2Y_2$ and $X_2 \subseteq X_1Y_1$,*
   - *(b) $X_1 \subseteq X_0$ and $X_2 \subseteq X_0$,*
   - *(c) $X_0 \subseteq X_1X_2Y_1Y_2$.*
   - *(d) $Y_0 \subseteq X_0Y_1$ and $Y_0 \subseteq X_0Y_2$,*

Indeed, the characterization is even tighter than what this statement suggests: whenever $\gamma < 1/2$, it can be shown that entailment from two premises holds only if it holds from one or zero premises. This was also proved in [9], thus fully covering all cases of entailment with two premises and all confidence parameters γ. Note, finally, that all conditions stated in the theorem are easy to check by an algorithm running in time $O(n)$, where n is the number of attributes, if the sets are given as bit vectors, say.

The proof of Theorem 2 in [9] is rather long and somewhat involved, although it uses only elementary Boolean algebraic manipulation. For instance, several different counterexamples to the entailment are built ad hoc depending on which of the seven set-inclusion conditions fail. Its intuition-building value is pretty limited, and a generalization to the case of more than two premises remained elusive. A somewhat subtle point about Theorem 2 is that the seven inclusion conditions alone do not characterize proper entailment (even if $\gamma \geq 1/2$, that is): they are only necessary conditions for that. But when these necessary conditions for proper entailment are disjuncted with the necessary and sufficient conditions for improper entailment, what results is an *if and only if* characterization of entailment. That is why the theorem is stated as it is, with the two escape clauses at the beginning of part *2*. Our main result will have a similar flavour, but with fewer cases to consider.

Before we move on to larger numbers of premises, one more comment is in order. Among the seven set-inclusion conditions in the statement of Theorem 2, those in the first item $X_1 \subseteq X_2Y_2$ and $X_2 \subseteq X_1Y_1$ are by far the least intuitive. Discovering the right generalization of this turned out to be the key to getting our results. This is discussed in Sections 5.1 and 5.3. Before that, however, we need to discuss a characterization of entailment in terms of linear programming duality. Interestingly, LP will end up disappearing altogether from the statement that generalizes Theorem 2; its use will merely be a (useful) technical detour.

3.2 Entailment in terms of LP duality

The goal in this section is to characterize the valid entailments as in Expression (1),

$$X_1 \to Y_1, \ldots, X_k \to Y_k \models_\gamma X_0 \to Y_0,$$

where each $X_i \to Y_i$ is a partial implication on the set of attributes $[n]$. The characterization can be seen as a variant, stated in the standard form of linear programming and tailored to our setting, of Proposition 4 in [10], where it applies to deduction rules of probabilistic consequence relations in general propositional logics. Our linear programming formulation makes it easy to check a number of simple properties of the solutions of the dual linear program at play, which are necessary for our application (Lemma 5). Before we state the characterization, we want to give some intuition for what to expect. At the same time we introduce some notation and terminology.

Following standard usage in full implications (see e.g. [12]), we say that a transaction $Z \subseteq [n]$ covers $X \to Y$ if $X \subseteq Z$, and that it violates it if $X \subseteq Z$ but $Y \not\subseteq Z$. If Z covers $X \to Y$ without violating it, that is, $XY \subseteq Z$, we say that Z witnesses $X \to Y$. For each partial implication $X \to Y$ and each transaction Z we define a weight $w_Z(X \to Y)$ that, intuitively, measures the extent to which Z witnesses $X \to Y$. Moreover, since we are aiming to capture confidence threshold γ we assign the weight proportionally:

$$\begin{array}{ll} w_Z(X \to Y) = 1 - \gamma & \text{if } Z \text{ witnesses } X \to Y, \\ w_Z(X \to Y) = -\gamma & \text{if } Z \text{ violates } X \to Y, \\ w_Z(X \to Y) = 0 & \text{if } Z \text{ does not cover } X \to Y. \end{array}$$

With these weights in hand we give a quantitative interpretation to the entailment in Expression (1).

First note that the weights are defined in such a way that, as long as $\gamma > 0$, a transaction Z satisfies the implication $X \to Y$ interpreted classically if and only if $w_Z(X \to Y) \geq 0$. With this in mind the entailment in Expression (1) interpreted classically would read as follows: for all Z, whenever all weights on the left are non-negative, the weight on the right is also non-negative. Of course, a sufficient condition for this to hold would be that the weights on the right are bounded below by some non-negative linear combination of the weights on the left, uniformly over Z. What the characterization below says is that this sufficient condition for classical entailment is indeed necessary and sufficient for entailment at confidence threshold γ, if the weights are chosen proportionally to γ as above. Formally:

Theorem 3. *Let γ be a confidence parameter in $[0,1]$ and let $X_0 \to Y_0, \dots, X_k \to Y_k$ be a set of partial implications. The following are equivalent:*

1. $X_1 \to Y_1, \dots, X_k \to Y_k \models_\gamma X_0 \to Y_0$
2. *There is a vector $\lambda = (\lambda_1, \dots, \lambda_k)$ of real non-negative components such that for all $Z \subseteq [n]$*

$$\sum_{i=1}^{k} \lambda_i \cdot w_Z(X_i \to Y_i) \le w_Z(X_0 \to Y_0) \tag{3}$$

Towards the proof of Theorem 3, let us state a useful lemma. This gives an alternative understanding of the weights $w_Z(X \to Y)$ than the one given above:

Lemma 4. *Let γ be a confidence parameter in $[0,1]$, let $X \to Y$ be a partial implication, let $\mathcal{D}$ be a transaction multiset, and for each $Z \subseteq [n]$ let x_Z be the multiplicity of Z in $\mathcal{D}$, that is, the number of times that Z appears (as a complete transaction) in $\mathcal{D}$. Then, $\mathcal{D} \models_\gamma X \to Y$ if and only if $\sum_{Z \subseteq [n]} w_Z(X \to Y) \cdot x_Z \ge 0$.*

Proof. Let $\mathcal{U}$ denote the set of transactions in $\mathcal{D}$ that cover $X \to Y$, let $\mathcal{V}$ denote those that violate $X \to Y$, and $\mathcal{W}$ those that witness $X \to Y$. Observe that $\mathcal{U} = \mathcal{V} \cup \mathcal{W}$ and that this union is a partition. By definition, $\mathcal{D} \models_\gamma X \to Y$ means that either $\sum_{Z \in \mathcal{U}} x_Z = 0$, or else $\left(\sum_{Z \in \mathcal{W}} x_Z\right) / \left(\sum_{Z \in \mathcal{U}} x_Z\right) \ge \gamma$. Recalling that $\mathcal{V} \cup \mathcal{W} = \mathcal{U}$ is a partition, this is equivalent to $\sum_{Z \in \mathcal{W}} x_Z \ge \gamma \cdot \left(\sum_{Z \in \mathcal{W}} x_Z + \sum_{Z \in \mathcal{V}} x_Z\right)$. Rearranging we get $\sum_{Z \in \mathcal{W}} (1-\gamma) \cdot x_Z - \sum_{Z \in \mathcal{V}} \gamma \cdot x_Z \ge 0$, from which the result follows by recalling that $w_Z(X \to Y) = 1 - \gamma$ for each $Z \in \mathcal{W}$ and $w_Z(X \to Y) = -\gamma$ for each $Z \in \mathcal{V}$, and that $w_Z(X \to Y) = 0$ for every other Z. □

This lemma is parallel to the first part of the proof of Proposition 4 in [10]. With this lemma in hand we can prove Theorem 3. We resort to duality here, while the version in [10] uses instead the closely related Farkas' Lemma.

Proof of Theorem 3. The statement of Lemma 4 leads to a natural linear program: for every Z let x_Z be a non-negative real variable, impose on these variables the inequalities from Lemma 4 for $X_1 \to Y_1$ through $X_k \to Y_k$, and check if the corresponding inequality for $X_0 \to Y_0$ can be falsified by minimizing its left-hand side:

$$\begin{aligned} P\colon \ \min\ & \textstyle\sum_{Z \subseteq [n]} w_Z(X_0 \to Y_0) \cdot x_Z & \\ \text{s.t.}\ & \textstyle\sum_{Z \subseteq [n]} w_Z(X_i \to Y_i) \cdot x_Z \ge 0 & \text{all } i \in [k], \\ & x_Z \ge 0 & \text{all } Z. \end{aligned}$$

The dual D of P has one non-negative variable y_i for every $i \in [k]$ and one inequality constraint for each non-negative variable x_Z. Since the objective function of D would just be the trivial constant function 0 we write it directly as a linear programming feasibility problem:

$$D\colon \textstyle\sum_{i\in[k]} w_Z(X_i \to Y_i) \cdot y_i \leq w_Z(X_0 \to Y_0) \quad \text{all } Z,$$
$$y_1, \ldots, y_k \geq 0$$

Note that this is really the characterization statement in the theorem that we are trying to prove, with y_i in place of λ_i. Thus, the theorem will be proved if we show that the following are equivalent:

(1) $X_1 \to Y_1, \ldots, X_k \to Y_k \models_\gamma X_0 \to Y_0$,

(2) the primal P is feasible and bounded below,

(3) the dual D is feasible.

(1) $\Rightarrow$ (2). We prove the contrapositive. Assume that P is unbounded below; it is certainly feasible since the all-zero vector satisfies all constraints. Let x be a feasible solution with $\sum_{Z\subseteq[n]} w_Z(X_0 \to Y_0) \cdot x_Z < 0$. Since the rationals are dense in the reals and linear maps are surely continuous, we may assume that x has rational components with a positive common denominator N, while preserving feasibility and a negative value for the objective function. Then $N \cdot x$ is still a feasible solution and its components are natural numbers. Let $\mathcal{D}$ be the transaction multiset that has $N \cdot x_Z$ copies of Z for every $Z \subseteq [n]$. By feasibility we have $\sum_{Z\subseteq[n]} w_Z(X_i \to Y_i) \cdot N \cdot x_Z \geq 0$ and therefore $\mathcal{D} \models_\gamma X_i \to Y_i$ for every $i \in [k]$ by Lemma 4. On the other hand $\sum_{Z\subseteq[n]} w_Z(X_0 \to Y_0) \cdot N \cdot x_Z < 0$ from which it follows that $\mathcal{D} \not\models_\gamma X_0 \to Y_0$, again by Lemma 4.

(2) $\Rightarrow$ (3). This is a direct consequence of the duality theorem for linear programming: if P is feasible and bounded below, D is feasible; see the preliminaries and the references there.

(3) $\Rightarrow$ (1). Assume D is feasible and let y be a feasible solution. Let $\mathcal{D}$ be a transaction multiset such that $\mathcal{D} \models_\gamma X_i \to Y_i$ for every $i \in [k]$. For every $Z \subseteq [n]$, let x_Z be the number of times that Z appears (alone, as a complete transaction) in $\mathcal{D}$. By dual feasibility of y and positivity of x_Z we get

$$\sum_{Z\subseteq[n]} w_Z(X_0 \to Y_0) \cdot x_Z \geq \sum_{Z\subseteq[n]} \Big(\sum_{i\in[k]} w_Z(X_i \to Y_i) \cdot y_i\Big) \cdot x_Z.$$

Distributing, exchanging the order of summation, and refactoring, the right-hand side reads

$$\sum_{i\in[k]} y_i \cdot \Big(\sum_{Z\subseteq[n]} w_Z(X_i \to Y_i) \cdot x_Z\Big).$$

Note that this is non-negative since the y_i are non-negative and $\sum_{Z\subseteq[n]} w_Z(X_i \to Y_i) \cdot x_Z \geq 0$ for every $i \in [k]$ by the assumption on $\mathcal{D}$ and Lemma 4. This proves that $\sum_{Z\subseteq[n]} w_Z(X_0 \to Y_0) \cdot x_Z \geq 0$, from which $\mathcal{D} \models_\gamma X_0 \to Y_0$ follows by one more call to Lemma 4. $\square$

3.3 Properties of the LP characterization

Whenever an entailment as in Expression (1) holds properly, the characterization in Theorem 3 gives a good deal of information about the inclusion relationships that the sets satisfy, and about the values that the λ_i can take. In this section we discuss this. Note that, from now on, the confidence parameter γ is in $(0, 1)$ instead of $[0, 1]$.

Lemma 5. *Let γ be a confidence parameter in $(0, 1)$ and let $X_0 \to Y_0, \ldots, X_k \to Y_k$ be a set of partial implications with $k \geq 1$. Assume that the entailment $X_1 \to Y_1, \ldots, X_k \to Y_k \models_\gamma X_0 \to Y_0$ holds properly. In particular, $Y_0 \not\subseteq X_0$. Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ denote any vector as promised to exist by Theorem 3 for this entailment. The following hold:*

1. *$\lambda_i > 0$ for every $i \in [k]$.*
2. *$X_0Y_0 \subseteq X_1Y_1 \cdots X_kY_k$.*
3. *$\sum_{i\in[k]} \lambda_i \leq 1$.*
4. *$X_i \subseteq X_0$ for every $i \in [k]$.*
5. *$X_iY_i \not\subseteq X_0$ for every $i \in [k]$.*
6. *$\sum_{i\in[k]} \lambda_i = 1$.*
7. *$Y_0 \subseteq X_0Y_i$ for every $i \in [k]$.*

Proof. The order in which we state them is the one that we deem best to follow smoothly the flow of proofs, as some of them are proved jointly and/or depend on previous ones. In what follows, for every Z, define:

$$U_Z = \{i \in [k] : Z \text{ covers } X_i \to Y_i\},$$

$$V_Z = \{i \in [k] : Z \text{ violates } X_i \to Y_i\},$$

$$W_Z = \{i \in [k] : Z \text{ witnesses } X_i \to Y_i\}.$$

Note that $U_Z = V_Z \cup W_Z$ and that this union is a partition.

1. For every $i \in [k]$, if $\lambda_i = 0$, then the inequalities in Expression (3) reduce to the same inequalities for the entailment without the i-th premise, and the remaining λ_j would still be a solution. Then, by Theorem 3 itself the entailment would not be proper, as premise i could be removed without affecting its validity.

2. Consider the inequality in Expression (3) for $Z = X_1Y_1 \cdots X_kY_k$. Obviously Z witnesses every $X_i \to Y_i$, so $W_Z = [k]$. Assume for contradiction that $X_0Y_0 \not\subseteq X_1Y_1 \cdots X_kY_k$. Then the inequality reads either $-\gamma \geq (1-\gamma) \cdot \sum_{i\in[k]} \lambda_i$ or $0 \geq (1-\gamma) \cdot \sum_{i\in[k]} \lambda_i$, and both cases are impossible since the right-side is strictly positive by the previous item and the fact that $\gamma < 1$. Therefore $X_0Y_0 \subseteq X_1Y_1 \cdots X_kY_k$.

3. Considering still the same inequality, we know now that it reads $1 - \gamma \geq (1-\gamma) \cdot \sum_{i\in[k]} \lambda_i$. From this we conclude that $\sum_{i\in[k]} \lambda_i \leq 1$ since $\gamma < 1$.

4, 5 and 6. Now consider the inequality in Expression (3) for $Z = X_0$. As the entailment is proper we have $Y_0 \not\subseteq X_0 = Z$ and therefore Z violates $X_0 \to Y_0$. So the inequality reads $-\gamma \geq (1-\gamma) \cdot \sum_{i \in W_Z} \lambda_i - \gamma \cdot \sum_{i \in V_Z} \lambda_i$. As $\lambda_i \geq 0$ we get $-\gamma \geq -\gamma \cdot \sum_{i \in V_Z} \lambda_i$ and therefore $\sum_{i \in V_Z} \lambda_i \geq 1$ since $\gamma > 0$. But also $\sum_{i \in [k]} \lambda_i \leq 1$ from which it follows that $V_Z = [k]$ since each λ_i is strictly positive. Thus Z violates every $X_i \to Y_i$, so $X_i \subseteq Z = X_0$ and $X_i Y_i \not\subseteq Z = X_0$ for every i. Also $\sum_{i \in [k]} \lambda_i = 1$ follows.

7. For every $i \in [k]$, consider the inequality in Expression (3) for $Z = X_0 Y_i$. We proved in item 4 that $X_i \subseteq X_0$. It follows that $X_i Y_i \subseteq X_0 Y_i = Z$ and thus $i \in W_Z$. Now assume for contradiction that $Y_0 \not\subseteq Z$. Then Z violates $X_0 \to Y_0$ and the inequality reads $-\gamma \geq (1-\gamma) \cdot \sum_{j \in W_Z} \lambda_j - \gamma \cdot \sum_{j \in V_Z} \lambda_j$. Since $i \in W_Z$ and $\lambda_j \geq 0$ for every $j \in [k]$, the right-hand side of this inequality is at least $(1-\gamma) \cdot \lambda_i - \gamma \cdot \sum_{j \in [k] \setminus \{i\}} \lambda_j = \lambda_i - \gamma \cdot \sum_{j \in [k]} \lambda_j$. But this is strictly bigger than $-\gamma$ since $\lambda_i > 0$ by item 1 and $\sum_{j \in [k]} \lambda_i \leq 1$ by item 3. This contradiction proves that the assumption $Y_0 \not\subseteq Z$ was wrong. Thus $Y_0 \subseteq Z = X_0 Y_i$. □

4 Low thresholds

As it turns out, if the confidence parameter γ is too low, then there cannot be any entailment as in Expression (1) that does not already follow from one of its premises. In such a case the characterization follows from known ones. This is what the next theorem states:

Theorem 6. *Let γ be a confidence parameter in $(0, 1)$ and let $X_0 \to Y_0, \ldots, X_k \to Y_k$ be a set of partial implications with $k \geq 1$. If $\gamma < 1/k$, then the following are equivalent:*

1. $X_1 \to Y_1, \ldots, X_k \to Y_k \models_\gamma X_0 \to Y_0$,
2. $X_i \to Y_i \models_\gamma X_0 \to Y_0$ *for some* $i \in [k]$,
3. *either* $Y_0 \subseteq X_0$, *or* $X_i \subseteq X_0$ *and* $X_0 Y_0 \subseteq X_i Y_i$ *for some* $i \in [k]$.

Proof. The equivalence between *2.* and *3.* follows from the characterization of entailments with one premise. We prove the equivalence between *1.* and *2.*, and for that we just need to argue the implication *1.* to *2.* since the other one is obvious. Assume *1.* and let $L \subseteq [k]$ be minimal under set inclusion so that $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$. If $|L| \leq 1$ we already have what we want. Assuming $|L| \geq 2$ we prove $\gamma \geq 1/k$; this will prove the theorem.

Let $\lambda = (\lambda_i : i \in L)$ be a solution to the inequalities in Expression (3) for $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$ as per Theorem 3. By the minimality of L, the entailment $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$ is proper. As γ is in the interval $(0, 1)$ and $|L| \geq 1$ (indeed ≥ 2), Lemma 5 applies to $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$ and says that $X_i \subseteq X_0$ for every $i \in L$, by part 4. Consequently, by the fact that $|L| \geq 2$, the minimality of L, and the characterization of entailment with at most one premise (Theorem 1), we have $X_0 Y_0 \not\subseteq X_i Y_i$ for every $i \in L$. Now, for fixed $i \in L$, let us look at the inequality in Expression (3) for $Z = X_i Y_i$. The above says that Z does not witness $X_0 \to Y_0$ so $w_Z(X_0 \to Y_0) \leq 0$. Of course Z witnesses $X_i \to Y_i$, so $w_Z(X_i \to Y_i) = 1 - \gamma$. Any other weight is at least $-\gamma$. Therefore, the inequality implies the following: $0 \geq \lambda_i \cdot (1-\gamma) - \gamma \cdot \sum_{j \in L \setminus \{i\}} \lambda_j = \lambda_i - \gamma \cdot \sum_{j \in L} \lambda_j$. By

Lemma 5, part 3, we have $\sum_{j\in L}\lambda_j \leq 1$. We conclude that $\lambda_i \leq \gamma$, and this holds for every $i \in L$. Adding over $i \in L$ we get $\sum_{i\in L}\lambda_i \leq \gamma \cdot |L|$, and the left-hand side is 1 by Lemma 5, part 6. Thus $\gamma \geq 1/|L| \geq 1/k$ and the theorem is proved. □

5 High thresholds

The goal of this section is to characterize entailments from k partial implications when the confidence parameter γ is large enough, and our proofs will show that $(k-1)/k$ is enough. Ideally, the characterization should make it easy to decide whether an entailment holds, or at least easier than solving the linear program given by Theorem 3. We come quite close to that. Before we get into the characterization, let us first discuss the key new concept on which it rests.

5.1 Enforcing homogeneity

We say that a set of partial implications $X_1 \to Y_1, \ldots, X_k \to Y_k$ enforces homogeneity if for every Z the following holds:

> if for all $i \in [k]$ either $X_i \not\subseteq Z$ or $X_iY_i \subseteq Z$ holds,
> then either $X_i \not\subseteq Z$ holds for all $i \in [k]$
> or $X_iY_i \subseteq Z$ holds for all $i \in [k]$.

In words, enforcing homogeneity means that every Z that does not violate any $X_i \to Y_i$, either witnesses them all, or does not cover any of them. Note that this definition does not depend on any confidence parameter. For economy of words, sometimes we refer to a set of partial implications that enforces homogeneity as being *nice*.

Note also that the empty set of partial implications vacuously enforces homogeneity; in fact, sets with less than two elements are trivially nice.

Homogeneity sounds like a very strong requirement. However, as the following lemma shows, it is at the heart of proper entailments.

Lemma 7. *Let $X_1 \to Y_1, \ldots, X_k \to Y_k$ be a set of partial implications with $k \geq 1$. If there exists a partial implication $X_0 \to Y_0$ and a confidence parameter γ in the interval $(0,1)$ for which the entailment $X_1 \to Y_1, \ldots, X_k \to Y_k \models_\gamma X_0 \to Y_0$ holds properly, then $X_1 \to Y_1, \ldots, X_k \to Y_k$ enforces homogeneity.*

Proof. Fix $X_0 \to Y_0$ and γ as in the statement of the lemma. We must show that if Z does not violate $X_i \to Y_i$ for any $i \in [k]$, then either Z witnesses all of them, or Z does not cover any of them. Fix Z that does not violate $X_i \to Y_i$ for any $i \in [k]$. In particular, for every $i \in [k]$, either Z does not cover $X_i \to Y_i$, or Z witnesses $X_i \to Y_i$. Thus $w_Z(X_i \to Y_i) \geq 0$ for every $i \in [k]$. If Z does not cover $X_j \to Y_j$ for any $j \in [k]$ we are done. Assume then that Z covers $X_j \to Y_j$ for some $j \in [k]$. Since it does not violate it, it witnesses it, which means that $w_Z(X_j \to Y_j) = 1 - \gamma$.

Now let us take a solution $\lambda = (\lambda_1, \ldots, \lambda_k)$ as promised by Theorem 3, and let us consider the inequality in Expression (3) for our fixed Z. This inequality reads $w_Z(X_0 \to Y_0) \geq \sum_{i \in [k]} \lambda_i \cdot w_Z(X_i \to Y_i)$. Since we proved that $w_Z(X_i \to Y_i) \geq 0$ for every $i \in [k]$, the right-hand side is at least $\lambda_j \cdot w_Z(X_j \to Y_j)$, which is $\lambda_j \cdot (1 - \gamma)$, for the j from the previous paragraph. Now, by Lemma 5.1 we have $\lambda_j > 0$ because the entailment is proper. Putting all this together we get $w_Z(X_0 \to Y_0) > 0$, so Z witnesses $X_0 \to Y_0$. Thus $X_0Y_0 \subseteq Z$. But we also know that $X_i \subseteq X_0$ for every $i \in [k]$ by Lemma 5.4. Thus $X_i \subseteq Z$ for every $i \in [k]$. Since Z does not violate $X_i \to Y_i$ for any $i \in [k]$, it must then be that Z witnesses $X_i \to Y_i$ for every $i \in [k]$. Precisely what we were trying to prove. □

The next lemma in this section characterizes *nicety*. For a partial implication $X \to Y$, let $X \Rightarrow Y$ denote its classical counterpart. Naturally, we write $Z \models X \Rightarrow Y$ if either $X \not\subseteq Z$ or $XY \subseteq Z$, i.e. if Z satisfies the implication classically. Also, in the context of classical implications, we use $\models$ to denote classical entailment.

Lemma 8. *Let $X_1 \to Y_1, \ldots, X_k \to Y_k$ be a set of partial implications and let $U = X_1Y_1 \cdots X_kY_k$. Then, the following are equivalent:*

1. *$X_1 \to Y_1, \ldots, X_k \to Y_k$ enforces homogeneity,*
2. *$X_1 \Rightarrow Y_1, \ldots, X_k \Rightarrow Y_k \models X_i \Rightarrow U$, all $i \in [k]$.*

Proof. Assume $X_1 \to Y_1, \ldots, X_k \to Y_k$ enforces homogeneity. Let $Z \models X_i \Rightarrow Y_i$ for all $i \in [k]$. Then, by homogeneity, either $X_i \not\subseteq Z$ for all $i \in [k]$, and then it also holds $Z \models X_i \Rightarrow U$ for all $i \in [k]$, or $X_iY_i \subseteq Z$ for all $i \in [k]$ so that $U \subseteq Z$, and $Z \models X_i \Rightarrow U$ for all $i \in [k]$ as well. Therefore, $X_1 \Rightarrow Y_1, \ldots, X_k \Rightarrow Y_k$ entail every $X_i \Rightarrow U$.

Conversely, assume that $X_1 \Rightarrow Y_1, \ldots, X_k \Rightarrow Y_k$ entail every $X_i \Rightarrow U$ and let $Z \models X_i \Rightarrow Y_i$ for all $i \in [k]$, hence $Z \models X_i \Rightarrow U$ for all $i \in [k]$. Then either $U \subseteq Z$ and we are done, or, else, the only way to satisfy all these classical implications is by falsifying all the premises, so that $X_i \not\subseteq Z$ for all $i \in [k]$. Therefore we have proved that $X_1 \to Y_1, \ldots, X_k \to Y_k$ enforces homogeneity. □

This characterization is quite useful. Look at the set of three partial implications $B \to ACH, C \to AD, D \to AB$ on the attributes A, B, C, D, H. By the lemma this set enforces homogeneity, but any of its two-element subsets fails to do so. Note also that condition *2.* in the lemma can be decided efficiently by testing the unsatisfiability of all the propositional Horn formulas of the form $(X_1 \Rightarrow Y_1) \wedge \cdots \wedge (X_k \Rightarrow Y_k) \wedge X_j \wedge \neg A$ as j ranges over $[k]$ and A ranges over the attributes in U.

5.2 Main result for high threshold

We are ready to state and prove the characterization theorem for $\gamma \geq (k-1)/k$.

Theorem 9. *Let γ be a confidence parameter in $(0, 1)$ and let $X_0 \to Y_0, \ldots, X_k \to Y_k$ be a set of partial implications with $k \geq 1$. If $\gamma \geq (k-1)/k$, then the following are equivalent:*

1. $X_1 \to Y_1, \dots, X_k \to Y_k \models_\gamma X_0 \to Y_0$,

2. *there is a set* $L \subseteq [k]$ *such that* $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$ *holds properly,*

3. *either* $Y_0 \subseteq X_0$, *or there is a non-empty* $L \subseteq [k]$ *such that the following conditions hold:*

   (a) $\{X_i \to Y_i : i \in L\}$ *enforces homogeneity,*

   (b) $\bigcup_{i \in L} X_i \subseteq X_0 \subseteq \bigcup_{i \in L} X_i Y_i$,

   (c) $Y_0 \subseteq X_0 \cup \bigcap_{i \in L} Y_i$.

Proof. That *1.* implies *2.* is clear: the family of all sets $L \subseteq [k]$ for which the entailment $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$ holds is non-empty, as *1.* says that $[k]$ belongs to it. Since it is finite it has minimal elements. Let L be one of them.

From *2.* to *3.*, the index set L will be the same in both statements, unless $L = \emptyset$, in which case $Y_0 \subseteq X_0$ must hold and we are done. Assume then that L is not empty. Part *(a)* we get automatically from Lemma 7 since $\{X_i \to Y_i : i \in L\}$ properly entails $X_0 \to Y_0$ at γ, which is in the interval $(0, 1)$. Now we prove *(b)*. By Theorem 3, let $\lambda = (\lambda_i : i \in L)$ be a solution to the inequalities in Expression (3) for the entailment $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$. From the fact that this entailment is proper and the assumptions that $|L| \geq 1$ and $\gamma \in (0, 1)$, we are allowed to call Lemma 5. The first inclusion in *(b)* follows from that lemma, part 4. The second inclusion in *(b)* also follows from that lemma, part 2. Finally, for *(c)* we just refer to part 7 and straightforward distributivity.

For the implication from *3.* to *1.* we proceed as follows. If $Y_0 \subseteq X_0$ there is nothing to prove since then the entailment is trivial. Assume then that L is non-empty and satisfies *(a)*, *(b)*, and *(c)*. By Theorem 3 it suffices to show that the inequalities in Expression (3) for the entailment $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$ have a solution $\lambda = (\lambda_i : i \in L)$ with non-negative components. Let $\ell = |L|$ and set $\lambda_i = 1/\ell$ for $i \in L$. Recall that L is not empty so $\ell \geq 1$ and this is well-defined. For fixed Z, we prove that the inequality in Expression (3) for this Z is satisfied by these λ_i. In the following, let $X = \bigcup_{i \in L} X_i$ and $Y = \bigcap_{i \in L} Y_i$. We distinguish cases according to whether $X \subseteq Z$.

First assume that $X \not\subseteq Z$. Then, by the first inclusion in *(b)*, $X_0 \not\subseteq Z$ so Z does not cover $X_0 \to Y_0$ and $w_Z(X_0 \to Y_0) = 0$. Also, there exists $j \in L$ such that $X_j \not\subseteq Z$. If $X_i Y_i \not\subseteq Z$ for every $i \in L$, then Z does not witness any $X_i \to Y_i$, so $w_Z(X_i \to Y_i) \leq 0$ for every $i \in L$. Whence $\sum_{i \in L} \lambda_i \cdot w_Z(X_i \to Y_i)$ is non-positive and then bounded by $w_Z(X_0 \to Y_0) = 0$ as required. Hence, suppose now that there exists $i \in L$ such that $X_i Y_i \subseteq Z$. We also have a $j \in L$ such that $X_j \not\subseteq Z$. Thus Z witnesses $X_i \to Y_i$ and fails to cover $X_j \to Y_j$, and both i and j are in L. As $\{X_i \to Y_i : i \in L\}$ enforces homogeneity, this means that Z violates $X_h \to Y_h$ for some $h \in L$. For that h we have $w_Z(X_h \to Y_h) = -\gamma$. The rest of weights are at most $1 - \gamma$ and therefore $\sum_{i \in L} \lambda_i \cdot w_Z(X_i \to Y_i)$ is bounded above by

$$-\frac{1}{\ell} \cdot \gamma + \frac{\ell - 1}{\ell} \cdot (1 - \gamma) = \frac{\ell - 1}{\ell} - \gamma.$$

Since $\ell \leq k$, this is at most $(k-1)/k - \gamma$. In turn, this is non-positive and then bounded by $w_Z(X_0 \to Y_0) = 0$ by the assumption that $\gamma \geq (k-1)/k$. This proves that the inequalities corresponding to these Z's are satisfied.

Assume now instead $X \subseteq Z$. In this case Z covers $X_i \to Y_i$ for every $i \in L$. Thus we split L into two sets, $L = V \cup W$, where V is the set of indices $i \in L$ such that Z violates $X_i \to Y_i$, and W is the set of indices $i \in L$ such that Z witnesses $X_i \to Y_i$. Of course $w_Z(X_i \to Y_i) = -\gamma$ for every $i \in V$ and $w_Z(X_i \to Y_i) = 1 - \gamma$ for every $i \in W$. We consider three subcases.

1. If $W = \emptyset$, then every $X_i \to Y_i$ with $i \in L$ is violated and then, using that the λ_i's add up to 1, $\sum_{i \in L} \lambda_i \cdot w_Z(X_i \to Y_i) = -\gamma \cdot \sum_{i \in L} \lambda_i = -\gamma \leq w_Z(X_0 \to Y_0)$; i.e. the inequality holds.

2. If $W = L$, then every $X_i \to Y_i$ with $i \in L$ is witnessed. Using *(b)* we get $X_0 \subseteq \bigcup_{i \in L} X_i Y_i \subseteq Z$, and the non-emptiness of L applied to *(c)* ensures the existence of some $i \in L$ for which $Y_0 \subseteq X_0 \cup Y \subseteq X_0 \cup Y_i \subseteq Z$. Thus $X_0 \to Y_0$ is also witnessed and $\sum_{i \in L} \lambda_i \cdot w_Z(X_i \to Y_i) = (1-\gamma) \cdot \sum_{i \in L} \lambda_i = 1 - \gamma = w_Z(X_0 \to Y_0)$; i.e. the inequality holds.

3. We consider now the general case where $W \neq \emptyset$ and $W \neq L$. The fact that $W \neq \emptyset$ ensures that there is some $i \in L$ such that $Y_i \subseteq Z$. Condition *(c)* then ensures that $Y_0 \subseteq X_0 \cup Y \subseteq X_0 \cup Y_i$ for this i. Altogether $X_0 \to Y_0$ is either witnessed or uncovered according to whether $X_0 \subseteq Z$. In both cases $w_Z(X_0 \to Y_0) \geq 0$. To complete the proof, let us split $\sum_{i \in L} \lambda_i \cdot w_Z(X_i \to Y_i)$ as follows:

$$\frac{1}{\ell} \cdot (1-\gamma) \cdot |W| - \frac{1}{\ell} \cdot \gamma \cdot (\ell - |W|).$$

The fact that $W \neq L$ implies $|W| \leq \ell - 1$. Therefore this is at most

$$\frac{1}{\ell} \cdot (|W| - \gamma \cdot \ell) \leq \frac{\ell - 1}{\ell} - \gamma \leq \frac{k-1}{k} - \gamma \leq 0 \leq w_Z(X_0 \to Y_0).$$

In the middle inequalities we used the fact that $\ell \leq k$ and the assumption that $\gamma \geq (k-1)/k$. We proved what we want; i.e. the inequality holds.

This closes the cycle of implications and the theorem is proved. □

5.3 Other properties of nicety

Enforcing homogeneity turned out to play a key role in the main result about the case of high confidence threshold. In this section we collect a few additional observations about it. The first one is quite trivial: sets of less than two partial implications are trivially nice. This does say, however, that every set of partial implications has some nice subset. The case $k = 2$ is a bit more interesting. Nicety corresponds exactly to the mysterious conditions in Theorem 2; cf. the discussion in Section 3.1.

Lemma 10. *A set of two partial implications $X_1 \to Y_1, X_2 \to Y_2$ enforces homogeneity if and only of both $X_1 \subseteq X_2 Y_2$ and $X_2 \subseteq X_1 Y_1$ hold.*

Proof. Assume $X_1 \not\subseteq X_2Y_2$. Then $Z = X_2Y_2 \models X_1 \Rightarrow Y_1$ and $Z \models X_2 \Rightarrow Y_2$, but this does not happen homogenously. The same holds if $X_2 \not\subseteq X_1Y_1$ by symmetry. Conversely, if both inclusions hold, consider any Z such that $Z \models X_1 \Rightarrow Y_1$ and $Z \models X_2 \Rightarrow Y_2$. If $X_1 \not\subseteq Z$, then $X_2Y_2 \not\subseteq Z$ either, hence $X_2 \not\subseteq Z$ is the only way to satisfy the second implication; by symmetry, we obtain $X_1 \not\subseteq Z$ if and only if $X_2 \not\subseteq Z$. Thus homogeneity holds. □

Finally, a recurrent situation concerns sets of partial implications with a common left-hand side. The next lemma says that every such set is nice.

Lemma 11. *Every set of partial implications of the form $X \to Y_1, \ldots, X \to Y_k$ enforces homogeneity.*

Proof. This is a direct application of Lemma 8. □

6 Intervening thresholds

The rest of the values of γ require ad hoc consideration in terms of the actual partial implications involved. We start by defining what will end up being the *critical* confidence threshold for a given entailment.

6.1 Critical threshold

Let $\Sigma = \{X_1 \to Y_1, \ldots, X_k \to Y_k\}$ be a set of partial implications with $k \geq 1$ and all its attributes in $[n]$, and let $X \subseteq [n]$. Define:

$$\gamma^* = \gamma^*(\Sigma, X) := \inf_{\lambda} \max_{Z} \frac{\sum_{i \in W_Z} \lambda_i}{\sum_{i \in V_Z \cup W_Z} \lambda_i} \tag{4}$$

where

1. Z ranges over all subsets of $[n]$ with $X \not\subseteq Z$,
2. $V_Z = \{i \in [k] : Z \text{ violates } X_i \to Y_i\}$,
3. $W_Z = \{i \in [k] : Z \text{ witnesses } X_i \to Y_i\}$,
4. λ ranges over vectors $(\lambda_1, \ldots, \lambda_k)$ of non-negative reals such that $\sum_{i \in [k]} \lambda_i = 1$,

and, by convention any occurrence of $0/0$ in the definition of γ^* is taken as 0, and a vacuous maximum is taken as 0. Note that this last case occurs only if $X = \emptyset$ since otherwise there is always the possibility of taking $Z = \emptyset$. Note also that since all λ_i are non-negative, the only way the denominator can be zero is by making the numerator also zero. It should be pointed out that the convention about $0/0$ is *not* an attempt to repair a discontinuity; in general, the discontinuities of the rational functions inside the max are not repairable. A final comment on the definition is that we required $k \geq 1$. This ensures that the inf is not vacuous, which in turn implies $0 \leq \gamma^* \leq 1$: the lower bound is obvious, and for the upper bound just take $\lambda_i = 1/k$ for every $i \in [k]$, which is well-defined when $k \geq 1$.

Observe that γ^* is defined for a set of partial inequalities and a single set X of attributes. Typically X will be the left-hand side of another partial inequality $X_0 \to Y_0$, but $\gamma^*(\Sigma, X_0)$ is explicitly defined not to depend on Y_0. For later reference let us also point out that, with the notation V_Z and W_Z from above, the inequalities in Expression (3) for an entailment $X_1 \to Y_1, \ldots, X_k \to Y_k \models_\gamma X_0 \to Y_0$ can be written as $w_Z(X_0 \to Y_0) \geq (1-\gamma) \cdot \sum_{i \in W_Z} \lambda_i - \gamma \cdot \sum_{i \in V_Z} \lambda_i$. It is not the first time we use this sort of notation.

6.2 Characterization for all thresholds

The main result of this section is a characterization theorem in the style of Theorem 9 that captures all possible confidence parameters.

Theorem 12. *Let γ be a confidence parameter in $(0,1)$ and let $X_0 \to Y_0, \ldots, X_k \to Y_k$ be a set of partial implications with $k \geq 1$. The following are equivalent:*

1. $X_1 \to Y_1, \ldots, X_k \to Y_k \models_\gamma X_0 \to Y_0$,
2. *there is a set $L \subseteq [k]$ such that $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$ holds properly,*
3. *either $Y_0 \subseteq X_0$, or there is a non-empty $L \subseteq [k]$ such that the following conditions hold:*
   - *(a) $\{X_i \to Y_i : i \in L\}$ enforces homogeneity,*
   - *(b) $\bigcup_{i \in L} X_i \subseteq X_0 \subseteq \bigcup_{i \in L} X_i Y_i$,*
   - *(c) $Y_0 \subseteq X_0 \cup \bigcap_{i \in L} Y_i$,*
   - *(d) $\gamma \geq \gamma^*(\{X_i \to Y_i : i \in L\}, X_0)$.*

Proof. That *1.* implies *2.* is clear, as in Theorem 9. From *2.* to *3.*, we may assume that L is non-empty as in Theorem 9. Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ be a vector of non-negative reals that satisfy the inequalities in Expression (3) as per Theorem 3. Then properties *(a)*, *(b)*, and *(c)* just follow from Lemma 5 in the same way as in Theorem 3. It remains to argue *(d)*. To see this first note that for every Z such that $X_0 \not\subseteq Z$ we have $w_Z(X_0 \to Y_0) = 0$ and therefore the inequality in Expression (3) for this Z reads as $0 \geq (1-\gamma) \cdot \sum_{i \in W_Z} \lambda_i - \gamma \cdot \sum_{i \in V_Z}$. Rearranging we get $\gamma \geq \left(\sum_{i \in W_Z} \lambda_i\right) / \left(\sum_{i \in V_Z \cup W_Z} \lambda_i\right)$, where $0/0$ is interpreted as 0. In particular, the maximum of the right-hand side over all Z such that $X_0 \not\subseteq Z$ is bounded by γ, and thus γ^* is also bounded by γ. Note that this also covers the case of empty X_0 since in that case the max in γ^* is vacuous, which we conveyed to define as 0.

Now we prove that *3.* implies *1.* Assuming *(a)* through *(d)*, it is enough to find a solution to the inequalities in Expression (3) for the entailment $\{X_i \to Y_i : i \in L\} \models_\gamma X_0 \to Y_0$. What we show is that for every positive real $\epsilon > 0$ there is a vector $\lambda = (\lambda_i : i \in L)$ with non-negative real components such that the following inequality holds uniformly for every $Z \subseteq [n]$:

$$\sum_{i \in L} \lambda_i \cdot w_Z(X_i \to Y_i) \leq w_Z(X_0 \to Y_0) + \epsilon. \tag{5}$$

By basic real analysis this will be enough (it is worth pointing out that a more direct *continuity* argument to replace inf by min would not work here; as stated earlier, the discontinuities of the rational functions at 0/0 are, in general, not repairable). Fix then a positive real $\epsilon > 0$ and let $\lambda = (\lambda_i : i \in L)$ be such that the max in the definition of γ^* is at most $\gamma^* + \epsilon$. For fixed Z, we prove Expression (5) by cases:

1. First assume that $X_0Y_0 \subseteq Z$. Then, Z witnesses $X_0 \to Y_0$ and $w_Z(X_0 \to Y_0) = 1 - \gamma$. The left-hand side in Expression (5) can be written as $(1-\gamma) \cdot \sum_{i \in W_Z} \lambda_i - \gamma \cdot \sum_{i \in V_Z} \lambda_i$. Using $\lambda_i \geq 0$ and $\sum_{i \in L} \lambda_i = 1$ this is at most $(1-\gamma) \cdot \sum_{i \in L} \lambda_i = (1-\gamma) = w_Z(X_0 \to Y_0)$, which in turn is at most the right-hand side in Expression (5); i.e. the inequality holds.

2. From now on, we assume that $X_0Y_0 \not\subseteq Z$. For this case assume additionally that $X_0 \subseteq Z$. In particular $Y_0 \not\subseteq Z$ and Z violates $X_0 \to Y_0$, so $w_Z(X_0 \to Y_0) = -\gamma$. By *(b)* we have $X_i \subseteq X_0$, whereas, by *(c)* we know that $Y_0 \subseteq X_0Y_i$ for every $i \in L$. Since $X_0 \subseteq Z$ and $Y_0 \not\subseteq Z$, this means that $X_i \subseteq Z$ but $Y_i \not\subseteq Z$ for every $i \in L$. It follows that Z violates $X_i \to Y_i$ and $w_Z(X_i \to Y_i) = -\gamma$ for every $i \in L$. Using $\sum_{i \in L} \lambda_i = 1$, the left-hand side in Expression (5) is $-\gamma \cdot \sum_{i \in L} \lambda_i = -\gamma = w_Z(X_0 \to Y_0)$, which is at most the right-hand side in Expression (5); i.e. the inequality holds.

3. Given the previous cases, we can assume now $X_0 \not\subseteq Z$, so Z does not cover $X_0 \to Y_0$ and $w_Z(X_0 \to Y_0) = 0$. The choice of $(\lambda_i : i \in L)$ implies that the ratio inside the max in the definition of γ^* is at most $\gamma^* + \epsilon$ for our Z; since we are in the case $X_0 \not\subseteq Z$, the ratio for our Z is in the max. By *(d)* it is also at most $\gamma + \epsilon$. It follows that $(\gamma+\epsilon) \cdot \sum_{i \in V_Z \cup W_Z} \lambda_i \geq \sum_{i \in W_Z} \lambda_i$ by non-negativity of the λ_i. Rearranging we get $(1-\gamma) \cdot \sum_{i \in W_Z} \lambda_i - \gamma \cdot \sum_{i \in V_Z} \lambda_i \leq \epsilon \cdot \sum_{i \in V_Z \cup W_Z} \lambda_i$. Since $\lambda_i \geq 0$ and $\sum_{i \in L} \lambda_i \leq 1$, the right-hand side is at most ϵ, which is precisely $w_Z(X_0 \to Y_0) + \epsilon$ since Z does not cover $X_0 \to Y_0$ and $w_Z(X_0 \to Y_0) = 0$. This is the right-hand side in Expression (5); i.e. the inequality holds.

This closes the cycle of implications and the proof. ∎

6.3 An interesting example

In view of the characterization theorems obtained so far, one may wonder if the critical γ of any entailment among partial implications is of the form $(k-1)/k$. This was certainly the case for $k = 1$ and $k = 2$, and Theorems 9 and 12 may sound as hints that this could be the case. In this section we refute this for $k = 3$ in a strong way: we compute γ^* for a specific entailment for $k = 3$ to find out that it is the unique real solution of the equation

$$1 - \gamma + (1-\gamma)^2/\gamma + (1-\gamma)^3/\gamma^2 = 1. \tag{6}$$

Numerically [13], the unique real solution is

$$\gamma_c \approx 0.56984\ldots.$$

Consider the following 5-attribute entailment for a generic confidence parameter γ:

$$B \to ACH,\ C \to AD,\ D \to AB\ \models_\gamma\ BCDH \to A.$$

Let us compute its $\gamma^*(\Sigma, X)$ where Σ is the left-hand side, and $X = BCDH$. In other words, we want to determine a triple $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ that minimizes

$$\max_Z \frac{\sum_{i \in W_Z} \lambda_i}{\sum_{i \in V_Z \cup W_Z} \lambda_i}$$

as Z ranges over the sets that do not include $X = BCDH$, and subject to the constraints that $\lambda_1, \lambda_2, \lambda_3 \geq 0$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. There are $2^5 = 32$ possible Z's out of which two ($ABCDH$ and $BCDH$) contain X and therefore do not contribute to the maximum. Some others give value 0 to the ratio and therefore do not contribute to the maximum either. Note that if either $|Z| \leq 2$, or $|Z| = 3$ and $A \notin Z$, then $W_Z = \emptyset$, so the numerator is 0 and hence the ratio is also 0 (recall the convention that 0/0 is 0). Thus, the only sets Z that can contribute non-trivially to the maximum are those of cardinality 4 or 3 that contain the attribute A. There are four Z of the first type ($ABCD$, $ABCH$, $ABDH$ and $ACDH$) and six Z of the second type (ABC, ABD, ABH, ACD, ACH and ADH). The corresponding ratios are

$$\frac{\lambda_2 + \lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}, \frac{\lambda_1}{\lambda_1 + \lambda_2}, \frac{\lambda_3}{\lambda_1 + \lambda_3}, \frac{\lambda_2}{\lambda_2 + \lambda_3}, \frac{0}{\lambda_1 + \lambda_2}, \frac{\lambda_3}{\lambda_1 + \lambda_3}, \frac{0}{\lambda_1}, \frac{\lambda_2}{\lambda_2 + \lambda_3}, \frac{0}{\lambda_2}, \frac{0}{\lambda_3}.$$

Those with 0 numerator cannot contribute to the maximum so, removing those as well as duplicates, we are left with

$$\frac{\lambda_2 + \lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}, \frac{\lambda_1}{\lambda_1 + \lambda_2}, \frac{\lambda_3}{\lambda_1 + \lambda_3}, \frac{\lambda_2}{\lambda_2 + \lambda_3}.$$

Since all λ_i are non-negative, the first dominates the third and we are left with three ratios:

$$\frac{\lambda_2 + \lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}, \frac{\lambda_1}{\lambda_1 + \lambda_2}, \frac{\lambda_2}{\lambda_2 + \lambda_3}. \tag{7}$$

We claim that a λ_c that satisfies the constraints and minimizes the maximum of the three terms in (7) is

$$\begin{aligned} \lambda_{c,1} &= 1 - \gamma_c \\ \lambda_{c,2} &= (1 - \gamma_c)^2 / \gamma_c \\ \lambda_{c,3} &= (1 - \gamma_c)^3 / \gamma_c^2 \end{aligned}$$

where γ_c is the unique real solution of the equation in Expression (6). Clearly this choice of λ_c satisfies the constraints of non-negativity, and they add up to one precisely because their sum is the left-hand side in Expression (6). By plugging in, note also that this λ_c makes all three terms in (7) equal to γ_c; that is,

$$\frac{\lambda_{c,2} + \lambda_{c,3}}{\lambda_{c,1} + \lambda_{c,2} + \lambda_{c,3}} = \frac{\lambda_{c,1}}{\lambda_{c,1} + \lambda_{c,2}} = \frac{\lambda_{c,2}}{\lambda_{c,2} + \lambda_{c,3}} = \gamma_c. \tag{8}$$

For later reference, let us note that the left-hand side of (6) is a strictly decreasing function of γ in the interval $(0, 1)$ (e.g. differentiate it, or just plot it) and therefore

$$1 - \gamma_0 + (1 - \gamma_0)^2/\gamma_0 + (1 - \gamma_0)^3/\gamma_0^2 > 1 \qquad (9)$$

whenever $0 < \gamma_0 < \gamma_c$.

In order to see that λ_c minimizes the maximum of the three terms in (7) suppose for contradiction that λ satisfies the constraints and achieves a smaller maximum, say $0 < \gamma_0 < \gamma_c$. Since γ_0 is the maximum of the three terms in (7) we have

$$\begin{aligned} \gamma_0 &\geq (\lambda_2 + \lambda_3)/(\lambda_1 + \lambda_2 + \lambda_3) \\ \gamma_0 &\geq \lambda_1/(\lambda_1 + \lambda_2) \\ \gamma_0 &\geq \lambda_2/(\lambda_2 + \lambda_3). \end{aligned}$$

Using $\lambda_1, \lambda_2, \lambda_3 \geq 0$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$, and rearranging, we get

$$\begin{aligned} \lambda_1 &\geq 1 - \gamma_0 \\ \lambda_2 &\geq \lambda_1 \cdot (1 - \gamma_0)/\gamma_0 \geq (1 - \gamma_0)^2/\gamma_0 \\ \lambda_3 &\geq \lambda_2 \cdot (1 - \gamma_0)/\gamma_0 \geq (1 - \gamma_0)^3/\gamma_0^2. \end{aligned}$$

Adding all three inequalities we get

$$\lambda_1 + \lambda_2 + \lambda_3 \geq 1 - \gamma_0 + (1 - \gamma_0)^2/\gamma_0 + (1 - \gamma_0)^3/\gamma_0^2.$$

But this is a contradiction: the left-hand side is 1 since λ satisfies the constraints, and the right-hand side is strictly bigger than 1 by (9). This proves the claim.

Finally, this example also shows that for γ midway through $1/k$ and $(k-1)/k$, the vector solution to the inequalities in Expression (3) could be very non-uniform. In this example with $\gamma = \gamma_c$, the solution is $\lambda_c \approx (0.43016, 0.32472, 0.24512)$. In contrast, for $\gamma \geq (k-1)/k$, the proof of Theorem 9 shows that it is always possible to take $\lambda_i = 1/|L|$ for $i \in L$ and $\lambda_i = 0$ for $i \in [k] \setminus L$. In this case, the vector $(\lambda_1, \lambda_2, \lambda_3) = (1/3, 1/3, 1/3)$ works for $\gamma \geq 2/3$, but fails otherwise. To see that it fails when $\gamma < 2/3$, take the inequality for $Z = ABCD$ in Expression (3).

By the way, it is easy to check that conditions *(a)*, *(b)* and *(c)* hold for this example, thus Theorem 12 says that $\gamma_c \approx 0.56984$ is the smallest confidence at which the entailment holds.

7 Closing remarks

Our study gives a useful handle on entailments among partial or probabilistic implications. The very last comment of the previous section is a good illustration of its power. However, there are a few questions that arose and were not fully answered by our work.

For the forthcoming discussion, let us take $\gamma = (k-1)/k$ for concreteness. The linear programming characterization in Theorem 3 gives an algorithm to decide if entailment holds that is polynomial in k, the number of premises, but exponential in n, the number of attributes. This is due to the dimensions of the matrix that defines the dual LP: this is a $2^n \times k$ matrix of rational numbers in the order of $1/k$ (for our fixed $\gamma = (k-1)/k$). On the other hand, the characterization theorem in Theorem 9 reverses the situation: there the algorithm is polynomial in n but exponential in k. In order to see this, first note that condition *(a)* can be solved by running $O(nk)$ Horn satisfiability tests of size $O(nk)$ each, as discussed at the end of Section 5.1. Second, conditions *(b)* and *(c)* are really straightforward to check if the sets are given as bit-vectors, say. So far we spent time polynomial in both n and k in checking the conditions of the characterization. The exponential in k blow-up comes, however, from the need to *pass* to a subset $L \subseteq [k]$, as potentially there are 2^k many of those sets to check. It does show, however, that the general problem in the case of $\gamma \geq (k-1)/k$ is in NP. This does not seem to follow from the linear programming characterization by itself, let alone the definition of entailment. But is it NP-hard? Or is there an algorithm that is polynomial in both k and n? One comment worth making is that an efficient *separation oracle* for the exponentially many constraints in the LP of Theorem 3 might well exist, from which a polynomial-time algorithm would follow from the ellipsoid method.

It is tempting to think that the search over subsets of $[k]$ can be avoided when we start with a proper entailment. And indeed, this is correct. However, we do not know if this gives a characterization of proper entailment. In other words, we do not know if conditions *(a)*, *(b)* and *(c)*, by themselves, guarantee proper entailment. The proof of the direction *3.* to *1.* in Theorem 9 does not seem to give this, and we suspect that it does not. If they did, we would get an algorithm to check for proper entailment that is polynomial in both n and k.

From a wider and less theoretical prespective, it would be very interesting to find real-life situations in problems of data analysis, say, in which partial implications abound, but many are redundant. In such situations, our characterization and algorithmic results could perhaps be useful for detecting and removing such redundancies, thus producing outputs of better quality for the final user. This was one of the original motivations for the work in [9], and our continuation here.

References

[1] L. De Raedt, *Logical and relational learning.* Springer, 2008.

[2] L. Geng and H. J. Hamilton, "Interestingness measures for data mining: A survey," *ACM Comput. Surv.*, vol. 38, no. 3, 2006.

[3] M. Luxenburger, "Implications partielles dans un contexte," *Mathématiques et Sciences Humaines*, vol. 29, pp. 35–55, 1991.

[4] R. Agrawal, T. Imielinski, and A. N. Swami, "Mining association rules between sets of items in large databases," in *SIGMOD Conference*, P. Buneman and S. Jajodia, Eds. ACM Press, 1993, pp. 207–216.

[5] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo, "Fast discovery of association rules," in *Advances in Knowledge Discovery and Data Mining.* AAAI/MIT Press, 1996, pp. 307–328.

[6] A. Ceglar and J. F. Roddick, "Association mining," *ACM Comput. Surv.*, vol. 38, no. 2, 2006.

[7] C. C. Aggarwal and P. S. Yu, "A new approach to online generation of association rules," *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 4, pp. 527–540, 2001.

[8] M. Kryszkiewicz, "Representative association rules," in *Proc. of the 2nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, ser. Lecture Notes in Artificial Intelligence, X. Wu, K. Ramamohanarao, and K. B. Korb, Eds., vol. 1394. Springer-Verlag, 1998, pp. 198–209.

[9] J. L. Balcázar, "Redundancy, deduction schemes, and minimum-size bases for association rules," *Logical Methods in Computer Science*, vol. 6, no. 2:3, pp. 1–33, 2010.

[10] J. B. Paris and R. Simmonds, "O is not enough," *The Review of Symbolic Logic*, vol. 2, pp. 298–309, 6 2009.

[11] H. Karloff, *Linear Programming*, 1st ed. Birkhäuser, 1991.

[12] D. Angluin, M. Frazier, and L. Pitt, "Learning conjunctions of Horn clauses," *Machine Learning*, vol. 9, pp. 147–164, 1992.

[13] Wolfram Alpha. Wolfram Alpha LCC, 15h30m, January 15, 2015.