First Notes on Maximum Entropy Entailment for Quantified Implications

Francesco Kriegel

Institute of Theoretical Computer Science,
Technische Universität Dresden, Dresden, Germany
francesco.kriegel@tu-dresden.de
orcid.org/0000-0003-0219-0330

Abstract. Entropy is a measure for the *uninformativeness* or *randomness* of a data set, i.e., the higher the entropy is, the lower is the amount of information. In the field of propositional logic it has proven to constitute a suitable measure to be maximized when dealing with models of probabilistic propositional theories. More specifically, it was shown that the model of a probabilistic propositional theory with maximal entropy allows for the deduction of other formulae which are somehow expected by humans, i.e., allows for some kind of common sense reasoning.
In order to pull the technique of *maximum entropy entailment* to the field of Formal Concept Analysis, we define the notion of entropy of a formal context with respect to the frequency of its object intents, and then define maximum entropy entailment for quantified implication sets, i.e., for sets of partial implications where each implication has an assigned degree of confidence. Furthermore, then this entailment technique is utilized to define so-called *maximum entropy implicational bases (ME-bases)*, and a first general example of such a ME-base is provided.

Keywords: Maximum Entropy · Formal Context · Partial Implication · Formal Concept Analysis · Implicational Base · Uncertain Knowledge

1 Introduction

Entropy is a measure for describing the amount of *randomness* or *uninformativeness* of a particular system, and has variants in different fields, e.g., in *thermodynamics* (where it was defined first), in *statistical mechanics*, but also in *probability theory* (where it describes a lack of *predicatability*), and in *information theory* as well (where it is also called *Shannon entropy*, and describes the amount of information contained in a message sent through a channel). In particular, one of the first works on entropy in information theory was published by Shannon and Weaver in 1949, cf. [10]. Later, there were several researchers who adapted the notion of entropy to *probabilistic logic*. For example, a motivating and wide introduction can be found in the book *The Uncertain Reasoner's Companion - a Mathematical Perspective* [9] written by Paris. A thorough justification for reasoning under maximum entropy semantics can be found therein on pages 76 ff. Furthermore, there is ongoing and active research on maximum entropy semantics, in particular for the case of *probabilistic first-order-logic*, which is e.g. driven by Kern-Isberner, cf. [3]. Both Paris and Kern-Isberner show that reasoning under maximum entropy semantics somehow implements common sense reasoning, and yield conclusions which are expected by humans when dealing with probabilistic theories or data sets.

This paper shall give a short overview and some first notes on utilizing the measure of entropy in the field of *Formal Concept Analysis*, and on a possibility for defining maximum entropy semantics for sets of partial implications equipped with probabilities (which are called *quantified implication sets* herein). Additionally, this paper presents a definition of an implicational base under such a maximum entropy semantics (*ME-base*), and provides a first general example of such a base which is induced by Luxenburger's base, cf. [8]. However, there is a specific example of a data set showing that this first ME-base is neither non-redundant nor minimal, and thus leaves possibilities for future research.

2 Formal Concept Analysis

The field of *Formal Concept Analysis* was invented for at least two reasons: Firstly, it should formalize the philosophical notion of a *concept* – which is characterized by its *extent*, i.e., the objects it describes, and its *intent*, i.e., the properties it satisfies – in a formal and mathematical way, cf. [5, Page 58]. Secondly, it was meant to be a new approach to the field of *lattice theory*, and in particular each complete lattice is isomorphic to a concept lattice (the lattice that is induced by the set of all formal concepts of a formal context). In this section, all necessary definitions for understanding of the subsequent sections are presented. For a more sophisticated overview on Formal Context Analysis, the interested reader is rather referred to the standard book [5] by Ganter and Wille.

Definition 1 (Formal Context). *A* formal context *is a triple* $\mathbb{K} \coloneqq (G, M, I)$ *which consists of a set* G *of* objects*, a set* M *of* attributes*, and an* incidence relation $I \subseteq G \times M$*. In case* $(g, m) \in I$*, we say that the object* g *has the attribute* m *in* $\mathbb{K}$*, and we may also denote this infix as* $g\ I\ m$*. Furthermore,* $\mathbb{K}$ *induces two* derivation operators $\cdot^I \colon \wp(G) \to \wp(M)$ *and* $\cdot^I \colon \wp(M) \to \wp(G)$*, respectively, as follows:*

$$A^I \coloneqq \{\, m \in M \mid \forall\, g \in A \colon g\ I\ m \,\},$$
$$\textit{and} \quad B^I \coloneqq \{\, g \in G \mid \forall\, m \in B \colon g\ I\ m \,\}.$$

It is well-known [5] that both derivation operators form a so-called *Galois connection* between the powersets $\wp(G)$ and $\wp(M)$, i.e., the following statements hold true for all subsets $A, A_1, A_2 \subseteq G$ and $B, B_1, B_2 \subseteq M$:

1. $A \subseteq B^I \Leftrightarrow B \subseteq A^I \Leftrightarrow A \times B \subseteq I$
2. $A \subseteq A^{II}$
3. $A^I = A^{III}$
4. $A_1 \subseteq A_2 \Rightarrow A_2^I \subseteq A_1^I$
5. $B \subseteq B^{II}$
6. $B^I = B^{III}$
7. $B_1 \subseteq B_2 \Rightarrow B_2^I \subseteq B_1^I$

An attribute set $B \subseteq M$ with $B = B^{II}$ is called an *intent* of $\mathbb{K}$, and we shall denote the set of all intents of $\mathbb{K}$ as $\mathsf{Int}(\mathbb{K})$.

Let $\mathbb{K} = (G, M, I)$ be a finite formal context. For an attribute set $X \subseteq M$, its *frequency* in $\mathbb{K}$ is defined as

$$\mathsf{freq}_{\mathbb{K}}(X) \coloneqq \frac{|\{\, g \in G \mid \{g\}^I = X \,\}|}{|G|},$$

and its *support* is given by

$$\mathsf{supp}_{\mathbb{K}}(X) := \frac{|X^I|}{|G|}.$$

It is easy to verify that $\mathsf{freq}_{\mathbb{K}}$ is a discrete probability measure on $\wp(M)$, i.e., $\sum_{X\subseteq M} \mathsf{freq}_{\mathbb{K}}(X) = 1$.

Definition 2 (Implication). *An* implication *over M is of the form $X \to Y$ where $X, Y \subseteq M$, and we call X the* premise*, and Y the* conclusion.[1] *It is* valid *in a formal context $\mathbb{K} := (G, M, I)$, denoted as $\mathbb{K} \models X \to Y$, if each object that possesses all attributes in X also has all attributes in Y, i.e., if $X^I \subseteq Y^I$. Furthermore, the* confidence *of $X \to Y$ in $\mathbb{K}$ is defined as*

$$\mathsf{conf}_{\mathbb{K}}(X \to Y) := \frac{|(X \cup Y)^I|}{|X^I|},$$

i.e., as the fraction of the number of objects satisfying both premise and conclusion compared to the number of objects satisfying only the premise.

An implication set *$\mathcal{L}$ is a set of implications. Then, $\mathcal{L}$ is* valid *in $\mathbb{K}$ if all implications in $\mathcal{L}$ are valid in $\mathbb{K}$, and we symbolize this by $\mathbb{K} \models \mathcal{L}$. If an implication $X \to Y$ is valid in all formal contexts in which $\mathcal{L}$ is valid, then we say that $\mathcal{L}$* entails *$X \to Y$ and denote this as $\mathcal{L} \models X \to Y$.*

The confidence of an implication measures the degree of validity in a given formal context. It is readily verified that an implication $X \to Y$ is valid in a formal context $\mathbb{K}$ if, and only if, its confidence in $\mathbb{K}$ equals 1. (Then $X^I \cap Y^I = (X \cup Y)^I = X^I$, and so $X^I \subseteq Y^I$.)

In [6], the *canonical base* of a formal context $\mathbb{K}$ was introduced as the implication set

$$\mathsf{Can}(\mathbb{K}) := \{\, P \to P^{II} \mid P \in \mathsf{PsInt}(\mathbb{K}) \,\},$$

where $\mathsf{PsInt}(\mathbb{K})$ consists of all *pseudo-intents* of $\mathbb{K}$, i.e., those sets $P \subseteq M$ that are no intents ($P \neq P^{II}$), but contain all intents Q^{II} for pseudo-intents $Q \subsetneq P$. In particular, the above mentioned canonical base is an *implicational base* for $\mathbb{K}$, i.e., for all implications $X \to Y$, it holds true that $X \to Y$ is valid in $\mathbb{K}$ if, and only if, $\mathsf{Can}(\mathbb{K})$ entails $X \to Y$. Furthermore, it is in fact a *minimal* implicational base, i.e., there is no implicational base of $\mathbb{K}$ containing fewer implications.

We continue by defining three different notions of equivalence of formal contexts, one for each of the measures freq, supp, and conf, that are introduced above. As a not very surprising result, it is afterwards shown that these three kinds of equivalence are indeed the same, i.e., two formal contexts are equivalent with respect to one of the measures if, and only if, they are equivalent with respect to all of the three measures.

Definition 3 (Equivalence of Formal Contexts). *Let $\mathbb{K}_1$ and $\mathbb{K}_2$ be two formal contexts with a common attribute set M.*

1. *$\mathbb{K}_1$ and $\mathbb{K}_2$ are* frequency-equivalent *if $\mathsf{freq}_{\mathbb{K}_1}(X) = \mathsf{freq}_{\mathbb{K}_2}(X)$ for all $X \subseteq M$.*

[1] In the field of machine learning, an implication is also called *association rule*, a premise is called *antedecent*, and a conclusion is called *consequent*.

2. $\mathbb{K}_1$ *and* $\mathbb{K}_2$ *are* support-equivalent *if* $\mathsf{supp}_{\mathbb{K}_1}(X) = \mathsf{supp}_{\mathbb{K}_2}(X)$ *for all* $X \subseteq M$.
3. $\mathbb{K}_1$ *and* $\mathbb{K}_2$ *are* confidence-equivalent *if* $\mathsf{conf}_{\mathbb{K}_1}(X \to Y) = \mathsf{conf}_{\mathbb{K}_2}(X \to Y)$ *for all* $X, Y \subseteq M$.

Lemma 4. *Let* $\mathbb{K}_1$ *and* $\mathbb{K}_2$ *be two formal contexts with a common attribute set* M*. Then the following statements are equivalent:*

1. $\mathbb{K}_1$ *and* $\mathbb{K}_2$ *are* frequency-equivalent.
2. $\mathbb{K}_1$ *and* $\mathbb{K}_2$ *are* support-equivalent.
3. $\mathbb{K}_1$ *and* $\mathbb{K}_2$ *are* confidence-equivalent.

Proof. 1. $\Rightarrow$ 2. We may express $\mathsf{supp}_{\mathbb{K}}$ in terms of $\mathsf{freq}_{\mathbb{K}}$ as follows:

$$\mathsf{supp}_{\mathbb{K}}(X) = \sum_{X \subseteq Y} \mathsf{freq}_{\mathbb{K}}(Y).$$

It follows that frequency-equivalence implies support-equivalence.
2. $\Rightarrow$ 3. Obviously, $\mathsf{conf}_{\mathbb{K}}$ can be expressed by means of $\mathsf{supp}_{\mathbb{K}}$ by

$$\mathsf{conf}_{\mathbb{K}}(X \to Y) = \frac{\mathsf{supp}_{\mathbb{K}}(X \cup Y)}{\mathsf{supp}_{\mathbb{K}}(X)}.$$

Thus, support-equivalence implies confidence-equivalence.
3. $\Rightarrow$ 1. Note that for arbitrary formal contexts $\mathbb{K} := (G, M, I)$,

$$\begin{aligned}
\mathsf{conf}_{\mathbb{K}}(X \to Y) &= \frac{|(X \cup Y)^I|}{|X^I|} \\
&= \frac{|\{\, g \in G \mid X \cup Y \subseteq \{g\}^I \,\}|}{|\{\, g \in G \mid X \subseteq \{g\}^I \,\}|} \\
&= \frac{\sum_{X \cup Y \subseteq Z} |\{\, g \in G \mid Z = \{g\}^I \,\}|}{\sum_{X \subseteq Z} |\{\, g \in G \mid Z = \{g\}^I \,\}|} \\
&= \frac{\sum_{X \cup Y \subseteq Z} \mathsf{freq}_{\mathbb{K}}(Z)}{\sum_{X \subseteq Z} \mathsf{freq}_{\mathbb{K}}(Z)}.
\end{aligned}$$

Thus, it follows that

$$\mathsf{conf}_{\mathbb{K}}(\emptyset \to X) = \sum_{X \subseteq Z} \mathsf{freq}_{\mathbb{K}}(Z),$$

and hence

$$\mathsf{freq}_{\mathbb{K}}(X) = \mathsf{conf}_{\mathbb{K}}(\emptyset \to X) - \sum_{X \subsetneq Z} \mathsf{freq}_{\mathbb{K}}(Z)$$

yields a recursion with $\mathsf{freq}_{\mathbb{K}}(M) = \mathsf{conf}_{\mathbb{K}}(\emptyset \to M)$. Consequently, confidence-equivalence implies frequency-equivalence. $\square$

3 Quantified Implication Sets

In [8], Luxenburger introduced the notion of a *stochastic context* as a means to capture statistical knowledge between classes (of objects) and attributes. Formally, a stochastic context is a triple (G, M, i) consisting of a set G of *classes*, a set M of *attributes*, and an *incidence function* $i\colon G \times M \to [0,1]$. Then, the value $p = i(g, m)$ indicates that the (statistical) probability of an object in the class g having the attribute m is p. For introducing semantics, the notion of a *realizer* was defined: A formal context $(H, G \cup M, J)$ *realizes* a stochastic context (G, M, i) if, for all classes $g \in G$ and all attributes $m \in M$, the implication $g \to m$ has a confidence $i(g, m)$. Then, an implication is *certainly valid* (*possibly valid*) in (G, M, i) if it is valid in all (some) realizers of (G, M, i).

It turns out that this type of knowledge representation is too restricted. On the one hand, it is not possible to denote unknown incidence values,[2] and on the other hand, we may not express dependencies between attributes or sets of attributes. As a solution, we could consider i as a partial function with type $G \times \wp(M) \dashrightarrow [0,1]$, but then we would not be able to express relations between the different classes, e.g., when it is known that two classes are disjoint, or overlap to a certain degree. Generalizing further, we could utilize an incidence function $i\colon \wp(G \cup M) \times \wp(G \cup M) \dashrightarrow [0,1]$, which leads us directly to sets of implications that are annotated with a value between $[0,1]$. Then, we do not have to distinguish between the classes and attributes, and simply say that a (generalized) stochastic context over M is a partial function $\mathcal{L}\colon \mathsf{Imp}(M) \dashrightarrow [0,1]$.

Definition 5 (Quantified Implication). *A* quantified implication *over M is of the form $\mathsf{d} \bowtie p.\, X \to Y$ where $\bowtie\, \in \{<, \leq, =, \neq, \geq, >\}$, $p \in [0,1]$, and $X, Y \subseteq M$. It is* valid *in a formal context $\mathbb{K} := (G, M, I)$, symbolized as $\mathbb{K} \models \mathsf{d} \bowtie p.\, X \to Y$, if $\mathsf{conf}_{\mathbb{K}}(X \to Y) \bowtie p$. The set of all quantified implications over M is denoted as $\mathsf{dImp}(M)$.*

A quantified implication set *is a set of quantified implications. A formal context $\mathbb{K} := (G, M, I)$ is a* realizer *of a quantified implication set $\mathcal{L}$, and we shall denote this as $\mathbb{K} \models \mathcal{L}$, if all implications in $\mathcal{L}$ are valid in $\mathbb{K}$.*

In the following text, we will only consider quantified implications of the form $\mathsf{d} = p.\, X \to Y$. Consequently, a quantified implication set $\mathcal{L}$ may also be viewed as a partial mapping $\mathcal{L}\colon \mathsf{Imp}(M) \dashrightarrow [0,1]$ where $\mathcal{L}(X \to Y) := p$ if, and only if, $\mathsf{d} = p.\, X \to Y \in \mathcal{L}$. Of course, this requires that there are no two different quantified implications $\mathsf{d} = p.\, X \to Y$ and $\mathsf{d} = q.\, X \to Y$ where $p \neq q$. We will not distinguish between both notational variants, and use that one which is better readable. Furthermore, by $\mathsf{dom}(\mathcal{L})$ we denote the *domain* of a quantified implication set $\mathcal{L}\colon \mathsf{Imp}(M) \dashrightarrow [0,1]$, i.e., the set of all implications to which $\mathcal{L}$ assigns a probability.

Clearly, quantified implication sets generalize the stochastic contexts of Luxenburger, since if (G, M, i) is a stochastic context, then $\mathcal{L}\colon (g \to m) \mapsto i(g, m)$ is a quantified implication set over $G \cup M$, and both have the same realizers.

As an alternative approach to defining semantics for the quantified implications, it is also possible to use probabilistic formal contexts which were introduced in [7]. In particular, we could define that a probabilistic formal context $\mathbb{K} := (G, M, W, I, \mathbb{P})$ *realizes* a quantified implication $\mathsf{d} \bowtie p.\, X \to Y$ if in $\mathbb{K}$ it holds true that $\mathbb{P}(X \to Y) \bowtie p$. However, we do not want to follow this approach here.

[2] Of course, this may be easily solved by regarding i as a partial function

As a next step, we define the notion of entailment for quantified implication sets. For this purpose, let $\mathcal{L} \cup \{\mathsf{d} = p.\, X \to Y\} \subseteq \mathsf{dImp}(M)$. Then $\mathcal{L}$ *certainly entails* $\mathsf{d} = p.\, X \to Y$ if each realizer of $\mathcal{L}$ is a realizer of $\mathsf{d} = p.\, X \to Y$, too, and we shall denote this by $\mathcal{L} \models_{\forall} \mathsf{d} = p.\, X \to Y$. Dually, $\mathcal{L}$ *possibly entails* $\mathsf{d} = p.\, X \to Y$ if there is a common realizer of $\mathcal{L}$ and $\mathsf{d} = p.\, X \to Y$, or if $\mathcal{L}$ is *inconsistent*, i.e., if $\mathcal{L}$ has no realizers at all, and this is symbolized as $\mathcal{L} \models_{\exists} \mathsf{d} = p.\, X \to Y$. The certain entailment $\models_{\forall}$ was investigated by Borchmann in [4], where a suitable system of linear equations must be solved in order to decide an entailment question $\mathcal{L} \models_{\forall} \mathsf{d} = p.\, X \to Y$. Furthermore, Borchmann showed that this entailment can be solved in polynomial space. In the next section, we will define the notion of *maximum entropy entailment* $\models_{\mathsf{ME}}$, which is stronger than $\models_{\exists}$, but weaker than $\models_{\forall}$.

4 Maximum Entropy Entailment

In [9], Paris motivates and introduces an inference process called *maximum entropy reasoning* for probabilistic propositional logic. Of course, there is a strong correspondence between propositional logic and Formal Concept Analysis, and we want to use this fact to apply the maximum entropy reasoning to Formal Concept Analysis.

Definition 6 (Entropy). *Let $\mathbb{K} := (G, M, I)$ be a formal context. Then, the* entropy *of $\mathbb{K}$ is defined as*[3]

$$\mathbf{H}(\mathbb{K}) := -\sum_{X \subseteq M} \mathsf{freq}_{\mathbb{K}}(X) \cdot \mathsf{log}(\mathsf{freq}_{\mathbb{K}}(X)).$$

For a quantified implication set $\mathcal{L}$, a maximum entropy realizer (ME-realizer) *of $\mathcal{L}$ is a realizer of $\mathcal{L}$ that has maximal entropy among all realizers of $\mathcal{L}$. It can be shown that modulo frequency-equivalence, maximum entropy realizers are unique, and hence we shall denote* the *maximum entropy realizer of $\mathcal{L}$ by $\mathcal{L}^*$. The* maximum entropy confidence (ME-confidence) *of an implication $X \to Y$ with respect to $\mathcal{L}$ is defined as the confidence of $X \to Y$ in the ME-realizer $\mathcal{L}^*$, i.e., $\mathsf{conf}_{\mathcal{L}}(X \to Y) := \mathsf{conf}_{\mathcal{L}^*}(X \to Y)$.*

Furthermore, for a quantified implication set $\mathcal{L} \cup \{\mathsf{d} \bowtie p.\, X \to Y\}$, we say that $\mathcal{L}$ entails $\mathsf{d} \bowtie p.\, X \to Y$ with respect to maximum entropy reasoning *(or abbreviated:* ME-entails*) if $\mathsf{d} \bowtie p.\, X \to Y$ is valid in the ME-realizer $\mathcal{L}^*$, or if $\mathcal{L}$ is inconsistent, i.e., if $\mathcal{L}$ has no realizers at all. We shall denote this as $\mathcal{L} \models_{\mathsf{ME}} \mathsf{d} \bowtie p.\, X \to Y$.*

For the case of one event and its converse, Figure 1 shows a plot of the corresponding entropy function, while for the case of two events, Figure 2 presents a suitable plot.

It is easy to see that $\mathcal{L} \models_{\forall} \mathsf{d} \bowtie p.\, X \to Y$ implies $\mathcal{L} \models_{\mathsf{ME}} \mathsf{d} \bowtie p.\, X \to Y$, and the latter implies $\mathcal{L} \models_{\exists} \mathsf{d} \bowtie p.\, X \to Y$, i.e., *ME-entailment* $\models_{\mathsf{ME}}$ is weaker than certain entailment $\models_{\forall}$, but stronger than possible entailment $\models_{\exists}$. A number of reasons that support the claim of ME-entailment being more natural than the other entailment types are presented by Paris in [9], where it is shown that reasoning under maximum entropy semantics somehow captures human expectation and common sense reasoning.

[3] We use the logarithm with base 2 here, since we are dealing with data sets or informations, respectively, which are encoded as bits. However, using another base would not cause any problem, since this would only distort the entropy by a multiplicative factor.

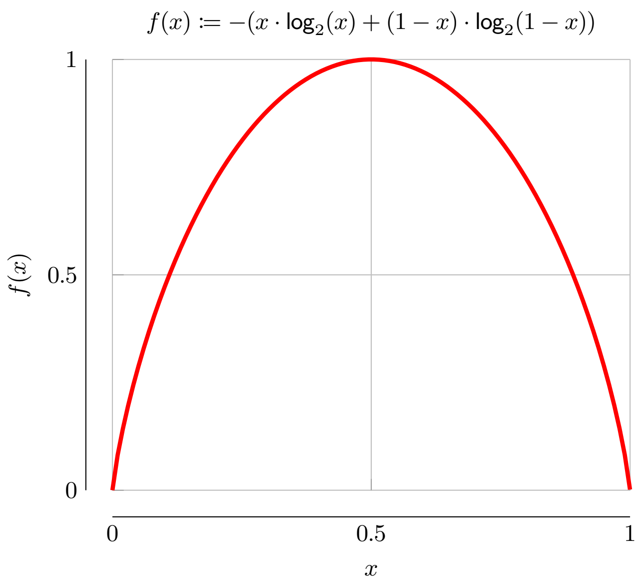

Fig. 1. A plot of the entropy function for one event.

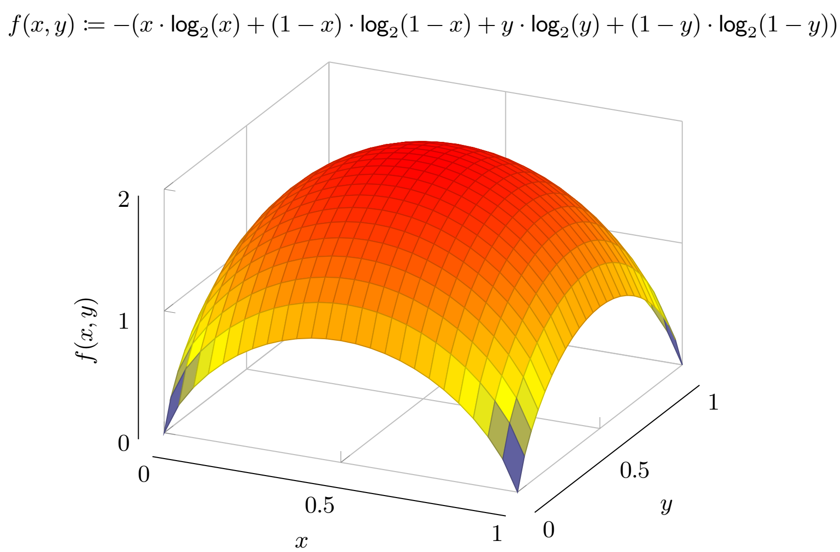

Fig. 2. A plot of the entropy function for two events.

In the remaining part of this section, the problem of deciding entailment under maximum entropy reasoning is considered. We hence assume that a quantified implication set $\mathcal{L}$ is given, and its ME-realizer is to be computed. For this purpose, we show how $\mathcal{L}$ induces a system of linear equations, for which the entropy function has to be maximized, i.e., we define an optimization problem the solution of which describes the ME-realizer $\mathcal{L}^*$ (modulo frequency-equivalence).

For each quantified implication $\mathsf{d} = p.\, X \to Y$ of $\mathcal{L}$, a realizer $\mathbb{K}$ of $\mathcal{L}$ must satisfy $\mathsf{conf}_{\mathbb{K}}(X \to Y) = p$, i.e.,

$$p = \frac{\mathsf{supp}_{\mathbb{K}}(X \cup Y)}{\mathsf{supp}_{\mathbb{K}}(X)} = \frac{\sum_{X \cup Y \subseteq Z} \frac{|\{\, g \in G \mid \{g\}^I = Z \,\}|}{|G|}}{\sum_{X \subseteq Z} \frac{|\{\, g \in G \mid \{g\}^I = Z \,\}|}{|G|}} = \frac{\sum_{X \cup Y \subseteq Z} \mathsf{freq}_{\mathbb{K}}(Z)}{\sum_{X \subseteq Z} \mathsf{freq}_{\mathbb{K}}(Z)}.$$

This is equivalent to the condition $\sum_{X \cup Y \subseteq Z} \mathsf{freq}_{\mathbb{K}}(Z) - p \cdot \sum_{X \subseteq Z} \mathsf{freq}_{\mathbb{K}}(Z) = 0$.

Hence, we formulate a matrix equation

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b},$$

where $\mathbf{A}\colon\ \mathsf{dom}(\mathcal{L}) \cup \{*\} \times \wp(M) \to [0,1]$, $\mathbf{x}\colon\ \wp(M) \to [0,1]$, and $\mathbf{b}\colon\ \mathsf{dom}(\mathcal{L}) \cup \{*\} \to [0,1]$, such that

$$\mathbf{A}(X \to Y, Z) \coloneqq \begin{cases} 1 - p & \text{if } X \cup Y \subseteq Z, \\ -p & \text{else if } X \subseteq Z, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$
$$\mathbf{A}(*, Z) \coloneqq 1$$

and $\mathbf{b}(X \to Y) \coloneqq 0$ for all $X \to Y \in \mathsf{dom}(\mathcal{L})$, and $\mathbf{b}(*) \coloneqq 1$. Then a solution $\mathbf{x}$ induces a class of formal contexts $\mathbb{K}$ such that for all $Z \subseteq M$, $\mathsf{freq}_{\mathbb{K}}(Z) = \mathbf{x}(Z)$ – we just have to determine the solution $\mathbf{x}$, for which the entropy $\mathbf{H}(\mathbf{x}) \coloneqq -\sum_{X \subseteq M} \mathbf{x}(X) \cdot \mathsf{log}(\mathbf{x}(X))$ is maximal.

5 Maximum Entropy Bases

For the axiomatization of data sets containing assertional knowledge, implicational bases were intensively used to describe the deducible implications in a sound and complete manner. If only valid implications are of interest, then the *canonical base* of Guigues and Duquenne in [6] describes a minimal implicational base, i.e., it is of smallest size among all sound and complete implication sets for the input formal context. There are also some investigations on *direct bases*, for which entailment can be checked in one pass, i.e., each implication in a direct base must only be considered once when deciding entailment, cf., Adaricheva, Nation, and Rand [1]. Furthermore, several different bases for partial implications, i.e., for the implications in a formal context with a confidence less than 1, were introduced, and an overview as well as a comparison on different types are presented by Balcázar in [2], among which the most popular base is of course Luxenburger's base [8]. However, in this section a new notion of a base of partial implications of a formal context shall be defined, which utilizes the maximum entropy entailment from the preceeding sections.

Definition 7 (Maximum Entropy Implicational Base). *Let* $\mathbb{K}$ *be a formal context with attribute set* M*. Then, a quantified implication set* $\mathcal{L}$ *over* M *is a* maximum entropy implicational base (ME-base) *of* $\mathbb{K}$ *if for all* $X \to Y \in \mathsf{Imp}(M)$*,*

$$\mathsf{conf}_{\mathbb{K}}(X \to Y) = \mathsf{conf}_{\mathcal{L}}(X \to Y).$$

A ME-base $\mathcal{L}$ *is* non-redundant*, if none of its values can be removed, i.e., for all* $X \to Y \in \mathsf{dom}(\mathcal{L})$ *it holds that* $\mathcal{L}(X \to Y) \neq \mathsf{conf}_{\mathcal{L}\setminus\{X\to Y\}}(X \to Y)$*. Furthermore, a ME-base* $\mathcal{L}$ *of* $\mathbb{K}$ *is* minimal *if there is no ME-base with a smaller domain, i.e., if* $|\mathsf{dom}(\mathcal{L})| \leq |\mathsf{dom}(\mathcal{L}')|$ *for all ME-bases* $\mathcal{L}'$ *of* $\mathbb{K}$*.*[4]

Lemma 8. *Let* $\mathbb{K}$ *be a formal context, and* $\mathcal{L}$ *be a quantified implication set. Then the following statements are equivalent:*

1. $\mathcal{L}$ *is a maximum entropy implicational base of* $\mathbb{K}$*.*
2. $\mathcal{L}^*$ *and* $\mathbb{K}$ *are confidence-equivalent.*
3. $\mathcal{L}^*$ *and* $\mathbb{K}$ *are support-equivalent.*
4. $\mathcal{L}^*$ *and* $\mathbb{K}$ *are frequency-equivalent.*

Proof. By the definition of realizers of quantified implication sets, we have $\mathcal{L}(X \to Y) = \mathsf{conf}_{\mathcal{L}^*}(X \to Y)$ for all implications $X \to Y \in \mathsf{dom}(\mathcal{L})$. Hence, $\mathcal{L}$ is a maximum entropy implicational base of $\mathbb{K}$ if, and only if, $\mathcal{L}^*$ and $\mathbb{K}$ are confidence-equivalent. The remaining equivalences are immediate consequences of Lemma 4 ☐

As a corollary we infer that, since all realizers of the quantified implication set $\mathsf{conf}_{\mathbb{K}}$ are frequency-equivalent to $\mathbb{K}$, $\mathsf{conf}_{\mathbb{K}}$ is a maximal entropy base of $\mathbb{K}$. Furthermore, it follows that $\mathcal{L}$ with $\mathcal{L}(\emptyset \to X) := \mathsf{conf}_{\mathbb{K}}(\emptyset \to X)$ is a maximum entropy base of $\mathbb{K}$.

For an implication $X \to Y \in \mathsf{Imp}(M)$, it holds that $\mathsf{conf}_{\mathbb{K}}(X \to Y) = \mathsf{conf}_{\mathbb{K}}(X^{II} \to Y^{II})$.

Proposition 9. *Let* $\mathbb{K}$ *be a formal context, and define the quantified implication set* $\mathbb{K}^*\colon \mathsf{Imp}(M) \dashrightarrow [0,1]$ *where*

$$\begin{aligned} \mathbb{K}^*(P \to P^{II}) &:= 1 && \text{for all } P \in \mathsf{PsInt}(\mathbb{K}), \text{ and} \\ \mathbb{K}^*(X \to Y) &:= \mathsf{conf}_{\mathbb{K}}(X \to Y) && \text{for all } X, Y \in \mathsf{Int}(\mathbb{K}) \text{ with } X \subsetneq Y \\ &&& \text{where no } Z \in \mathsf{Int}(\mathbb{K}) \text{ with } X \subsetneq Z \subsetneq Y \text{ exists.} \end{aligned}$$

Then $\mathbb{K}^*$ *is a maximum entropy implicational base for* $\mathbb{K}$*.*

Proof. We utilize Lemma 8, i.e., we prove that $\mathbb{K}$ and $\mathbb{K}^{**}$ are confidence-equivalent. Thus, consider an arbitrary implication $X \to Y \in \mathsf{Imp}(M)$ – we need to show that the confidences of $X \to Y$ in $\mathbb{K}$ and $\mathbb{K}^{**}$ are equal. Beforehand, it is useful to give a characterization or construction of the ME-realizer $\mathbb{K}^{**}$.

Consider an implication $X \to Y$ that is valid in $\mathbb{K}$, i.e., has a confidence of 1. Then it follows from the canonical base. Furthermore, all implications $P \to P^{II}$ where P is a pseudo-intent of $\mathbb{K}$ are valid in $\mathbb{K}^{**}$. Consequently, also $X \to Y$ must be valid in $\mathbb{K}^{**}$, i.e., has a confidence of 1 in $\mathbb{K}^{**}$. Consequently, $\mathbb{K}$ and $\mathbb{K}^{**}$ are confidence-equivalent for all valid implications of $\mathbb{K}$.

Now assume that $\mathsf{conf}_{\mathbb{K}}(X \to Y) = p < 1$. We consider the following four cases:

[4] We denote by $\mathcal{L} \setminus \{X \to Y\}$ the quantified implication set which assigns no probability to $X \to Y$, but otherwise coincides with $\mathcal{L}$.

1. $\{X, Y\} \subseteq \mathsf{Int}(\mathbb{K})$ with $X \subseteq Y$, and there is no $Z \in \mathsf{Int}(\mathbb{K})$ such that $X \subsetneq Z \subsetneq Y$.
2. $\{X, Y\} \subseteq \mathsf{Int}(\mathbb{K})$ with $X \subseteq Y$, and there is a non-empty subset $\{Z_1, \ldots, Z_n\} \subseteq \mathsf{Int}(\mathbb{K})$ such that $X \subsetneq Z_1 \subsetneq \ldots \subsetneq Z_n \subsetneq Y$, and for all $i \in \{0, \ldots, n\}$, there is no $Z \in \mathsf{Int}(\mathbb{K})$ with $Z_i \subsetneq Z \subsetneq Z_{i+1}$ where $Z_0 \coloneqq X$ and $Z_{n+1} \coloneqq Y$.
3. $\{X, Y\} \subseteq \mathsf{Int}(\mathbb{K})$ with $X \not\subseteq Y$.
4. $\{X, Y\} \not\subseteq \mathsf{Int}(\mathbb{K})$.

We continue by proving the claim for each of the cases above.

1. is true by definition of $\mathbb{K}^*$.
2. Note that for sets $X \subseteq Y \subseteq Z \subseteq M$, and arbitrary formal contexts (G, M, I), it holds true that $\mathsf{conf}_{(G,M,I)}(X \to Y) \cdot \mathsf{conf}_{(G,M,I)}(Y \to Z) = \mathsf{conf}_{(G,M,I)}(X \to Z)$. Then, by Case 1, each of the implications $Z_i \to Z_{i+1}$ where $i \in \{0, \ldots, n\}$ has the same confidence in $\mathbb{K}$ and $\mathbb{K}^{**}$. Consequently, we know that the following equations are valid:

$$\begin{aligned}
\mathsf{conf}_{\mathbb{K}}(X \to Y) &= \mathsf{conf}_{\mathbb{K}}(Z_0 \to Z_{n+1}) \\
&= \mathsf{conf}_{\mathbb{K}}(Z_0 \to Z_1) \cdot \ldots \cdot \mathsf{conf}_{\mathbb{K}}(Z_n \to Z_{n+1}) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(Z_0 \to Z_1) \cdot \ldots \cdot \mathsf{conf}_{\mathbb{K}^{**}}(Z_n \to Z_{n+1}) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(Z_0 \to Z_{n+1}) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(X \to Y).
\end{aligned}$$

3. The following equations are valid.

$$\begin{aligned}
\mathsf{conf}_{\mathbb{K}}(X \to Y) &= \mathsf{conf}_{\mathbb{K}}(X \to X \cup Y) \\
&= \mathsf{conf}_{\mathbb{K}}(X \to (X \cup Y)^{II}) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(X \to (X \cup Y)^{II}) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(X \to X \cup Y) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(X \to Y).
\end{aligned}$$

4. We know that both implications $X \to Y$ and $X^{II} \to Y^{II}$ have the same confidence in $\mathbb{K}$. Furthermore, $X \to Y$ is entailed by $\{X \to X^{II}, X^{II} \to Y^{II}, Y^{II} \to Y\}$, where $X \to X^{II}$ and $Y^{II} \to Y$ both have confidence 1 in $\mathbb{K}$, and hence also in $\mathbb{K}^{**}$. We prove that both implications $X \to Y$ and $X^{II} \to Y^{II}$ also have the same confidence in the ME-realizer $\mathbb{K}^{**}$.
Assume $X \subseteq Y$. Then we have that

$$\begin{aligned}
\mathsf{conf}_{\mathbb{K}^{**}}(X \to Y) &= \mathsf{conf}_{\mathbb{K}^{**}}(X \to Y) \cdot \mathsf{conf}_{\mathbb{K}^{**}}(Y \to Y^{II}) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(X \to Y^{II}) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(X \to X^{II}) \cdot \mathsf{conf}_{\mathbb{K}^{**}}(X^{II} \to Y^{II}) \\
&= \mathsf{conf}_{\mathbb{K}^{**}}(X^{II} \to Y^{II}).
\end{aligned}$$

Furthermore, let $X \not\subseteq Y$. The implication $(X \cup Y)^{II} \to X^{II} \cup Y^{II}$ is valid in $\mathbb{K}$, i.e., has confidence 1 both in $\mathbb{K}$ and in the maximum entropy realizer $\mathbb{K}^{**}$. We proceed as follows:

$$\begin{aligned}\mathsf{conf}_{\mathbb{K}^{**}}(X \to Y) &= \mathsf{conf}_{\mathbb{K}^{**}}(X \to X \cup Y)\\ &= \mathsf{conf}_{\mathbb{K}^{**}}(X^{II} \to (X \cup Y)^{II})\\ &= \mathsf{conf}_{\mathbb{K}^{**}}(X^{II} \to (X \cup Y)^{II}) \cdot \mathsf{conf}_{\mathbb{K}^{**}}((X \cup Y)^{II} \to X^{II} \cup Y^{II})\\ &= \mathsf{conf}_{\mathbb{K}^{**}}(X^{II} \to X^{II} \cup Y^{II})\\ &= \mathsf{conf}_{\mathbb{K}^{**}}(X^{II} \to Y^{II}).\end{aligned}$$

We have shown that $\mathbb{K}$ and $\mathbb{K}^{**}$ are confidence-equivalent. Consequently, both are also frequency-equivalent, and thus must have the same intents. (The set of object intents consists of all those attribute sets with a frequency > 0, hence both $\mathbb{K}$ and $\mathbb{K}^{**}$ have the same object intents.) □

It should be emphasized that the ME-base from the preceding Proposition 9 is not a minimal ME-base – which is based on the example of a Boolean formal context $(M, M, \subseteq)$ for some $M \coloneqq \wp(N)$: its Luxenburger base is

$$\{\, X \to X \cup \{m\} \mid X \subseteq M,\, m \notin X \,\}$$

where all contained implications possess a confidence of $\frac{1}{2}$. Of course, it can be easily verified that the empty implication set $\emptyset \subseteq \mathsf{dImp}(M)$ also constitutes a ME-base of $(M, M, \subseteq)$, since then $\mathsf{freq}_{\emptyset^*}$ is uniform in order to maximize the entropy, and hence each formal context being a ME-realizer of $\emptyset$ must be isomorphic to $(M, M, \subseteq)$ – in particular, it must be a subposition of some copies of the Boolean formal context $(M, M, \subseteq)$, modulo reordering of rows.

6 Conclusion

In this paper, the widely used measure of *entropy* is defined for the basic structure of a formal context in the field of *Formal Concept Analysis*. It is then utilized to define entailment for quantified implications under maximum entropy semantics, yielding a reasoning procedure which is more common for human thinking than the existing certain semantics and the possible semantics. Then some first steps towards implicational bases under maximum entropy semantics are presented. However, there is large room for future research. For example, possibilities of smaller ME-bases shall be investigated, and particularly and most importantly searching for a (canonical) minimal ME-base is a major goal. Of course, this would require a more sophisticated study of this idea.

Acknowledgements. The author gratefully thanks the anonymous reviewers for their constructive hints and helpful remarks.

References

[1] Kira V. Adaricheva, James B. Nation, and Robert Rand. "Ordered direct implicational basis of a finite closure system". In: *Discrete Applied Mathematics* 161.6 (2013), pp. 707–723.

[2] José L. Balcázar. "Minimum-Size Bases of Association Rules". In: *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML/PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part I*. Ed. by Walter Daelemans, Bart Goethals, and Katharina Morik. Vol. 5211. Lecture Notes in Computer Science. Springer, 2008, pp. 86–101.

[3] Christoph Beierle et al. "Extending and Completing Probabilistic Knowledge and Beliefs Without Bias". In: *KI* 29.3 (2015), pp. 255–262.

[4] Daniel Borchmann. "Deciding Entailment of Implications with Support and Confidence in Polynomial Space". In: *CoRR* abs/1201.5719 (2012).

[5] Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, 1999.

[6] Jean-Luc Guigues and Vincent Duquenne. "Famille minimale d'implications informatives résultant d'un tableau de données binaires". In: *Mathématiques et Sciences Humaines* 95 (1986), pp. 5–18.

[7] Francesco Kriegel. "Probabilistic Implicational Bases in FCA and Probabilistic Bases of GCIs in $\mathcal{EL}^{\bot}$". In: *Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications, Clermont-Ferrand, France, October 13-16, 2015*. Ed. by Sadok Ben Yahia and Jan Konecny. Vol. 1466. CEUR Workshop Proceedings. CEUR-WS.org, 2015, pp. 193–204.

[8] Michael Luxenburger. "Implikationen, Abhängigkeiten und Galois Abbildungen - Beiträge zur formalen Begriffsanalyse". PhD thesis. Technische Hochschule Darmstadt, 1993.

[9] Jeff B. Paris. *The Uncertain Reasoner's Companion - a Mathematical Perspective*. Cambridge Tracts in Theoretical Computer Science 39. Cambridge University Press, 1994.

[10] Claude Elwood Shannon and Warren Weaver. *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press, 1949.