Artificial Intelligence 197 (2013) 1–24

Contents lists available at SciVerse ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Inconsistency measures for probabilistic logics

Matthias Thimm

Institute for Web Science and Technologies, Universität Koblenz–Landau, Germany

ARTICLE INFO

Article history:
Received 23 February 2012
Received in revised form 27 July 2012
Accepted 4 February 2013
Available online 7 February 2013

Keywords:
Inconsistency measures
Inconsistency management
Probabilistic reasoning
Probabilistic conditional logic

ABSTRACT

Inconsistencies in knowledge bases are of major concern in knowledge representation and reasoning. In formalisms that employ model-based reasoning mechanisms inconsistencies render a knowledge base useless due to the non-existence of a model. In order to restore consistency an analysis and understanding of inconsistencies are mandatory. Recently, the field of inconsistency measurement has gained some attention for knowledge representation formalisms based on classical logic. An inconsistency measure is a tool that helps the knowledge engineer in obtaining insights into inconsistencies by assessing their severity. In this paper, we investigate inconsistency measurement in probabilistic conditional logic, a logic that incorporates uncertainty and focuses on the role of conditionals, i.e. *if–then* rules. We do so by extending inconsistency measures for classical logic to the probabilistic setting. Further, we propose novel inconsistency measures that are specifically tailored for the probabilistic case. These novel measures use distance measures to assess the distance of a knowledge base to a consistent one and therefore takes the crucial role of probabilities into account. We analyze the properties of the discussed measures and compare them using a series of rationality postulates.

1. Introduction

The field of knowledge representation and reasoning [4] is concerned with formal representations of knowledge and how these formalizations can be used for *reasoning*, i.e., how new information can be automatically inferred using a formal system. One of the big issues in knowledge representation is *accuracy*. Usually, the term *"knowledge"* is used to describe *strict* or *objective* information that is considered to be absolutely true in the given frame of reference, i.e. the real world. The counterpart, denoted by *"subjective knowledge"* or *"beliefs"*, is used to describe information that is assumed to be true by the individual under consideration. While strict knowledge describes—by definition—a *consistent* state, subjective knowledge might be flawed in several aspects. Besides being incorrect with respect to the real world, subjective knowledge can be *incomplete*, *uncertain*, or *inconsistent*. That is, for some piece of information I it might be unknown whether I is true or false (*incompleteness*), I might be believed only to a certain degree (*uncertainty*), or I might be in conflict with another piece of information I' (*inconsistency*). Note that inconsistency of two pieces of information I and I' implies that at least one of them is *incorrect*. However, even without the possibility to compare I and I' with the state of the real world, an inconsistency *can* be detected by a being capable of reasoning, which is not necessarily true for incorrect information in general. In this paper, we do not consider the general problem of incorrect information and always assume that represented pieces of information are subjective. However, as some terms like *knowledge base* have been established in the literature we adapt those conventions.

Within the field of knowledge representation and reasoning there are several subfields that deal with incomplete, uncertain, and/or inconsistent knowledge such as *default* [27] and *defeasible reasoning* [19], *argumentation* [2,26], or *possibilistic*

E-mail address: thimm@uni-koblenz.de.

0004-3702/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.artint.2013.02.001

and fuzzy reasoning [33]. Among the most established logical frameworks for dealing with uncertainty is *probability theory* [23,25]. There have been numerous works on combining probability theory with knowledge representation. For example, Bayesian networks and Markov nets allow for derivation of uncertain beliefs from other uncertain beliefs. Especially in application areas such as medical diagnosis, where the user has to rely crucially on the certainty of individual recommendations, reasoning using probabilistic models of knowledge serves well [24].

In this paper we employ *probabilistic conditional logic* [28] for representing uncertain knowledge. In probabilistic conditional logic, knowledge is represented using probabilistic conditionals $(\psi \mid \phi)[p]$ with the intuitive meaning "if ϕ is true then ψ is true with probability p". Probabilistic conditional logic has been studied extensively under several aspects, e.g. effective reasoning mechanisms [9], default reasoning [20], or extensions with first-order logic fragments [15,16]. Moreover, the field of information theory provides a nice solution to the problem of *incomplete information* in probabilistic conditional logic. Using the principle of maximum entropy [23] one can complete uncertain and incomplete information in order to gain new information that was unspecified before, see also [28,14]. The expert system SPIRIT [30] is a working system that employs reasoning based on the principle of maximum entropy. It has been applied to various fields of operations research such as project risk management [1] and portfolio selection [29]. Though reasoning with maximum entropy can deal with incomplete and uncertain information, it is not suitable for reasoning with *inconsistent* information. But inconsistency is a ubiquitous matter and human beings have to deal with it all the time. In knowledge engineering and expert system design it becomes most apparent when multiple experts try to build up a common knowledge base. However, the issue of extending reasoning with maximum entropy to inconsistent knowledge bases has been dealt with in the literature only little so far, cf. [31,8,6].

In this paper, we investigate inconsistencies in probabilistic conditional logic from an analytical perspective. One way to analyze inconsistencies is by measuring them. An *inconsistency measure* is a function that quantifies the severity of inconsistencies in knowledge bases. An inconsistency value of zero indicates no inconsistency (and therefore consistency) while the larger the inconsistency value, the more severe the inconsistency. Thus, an inconsistency measure can be seen as the counterpart to an *information measure* [5] for the case of inconsistent information. Recently, there has been a gain in attention to approaches for measuring inconsistency in classical logics, see e.g. [13,11]. In general, an inconsistency measure can be used to support the knowledge engineer in building a consistent knowledge base or repairing an inconsistent one. For example, Grant and Hunter [11] develop an approach for stepwise inconsistency resolution of inconsistent knowledge bases that makes use of inconsistency measures. In their approach, a knowledge base is repaired by e.g. deleting or weakening formulas. There, inconsistency measures serve as heuristics for selecting the right formula that has to be modified, i.e. by selecting that one that maximizes consistency gain. Inconsistency measures can also be used to determine which pieces of information are most responsible for producing the inconsistency. In [13,36] the Shapley value [32] is used to distribute the inconsistency value of a knowledge base among the individual formulas. In a setting where knowledge is merged from different sources this information can help in identifying the responsible contributors.

However, classical approaches for inconsistency measurement do not grasp the nuances of probabilistic knowledge and allow only for a very coarse assessment of the severity of inconsistencies. In particular, those approaches do not take the crucial role of probabilities into account and exhibit a discontinuous behavior in measuring inconsistency. That is, a slight modification of the probability of a conditional in a knowledge base may yield a discontinuous change in the value of the inconsistency. Consequently, we develop novel inconsistency measures that are more apt for the probabilistic setting. We do so by continuing and largely extending previous work [34–36]. In particular, the contributions of this paper are as follows. First, we propose and discuss a series of rationality postulates for inconsistency measures in probabilistic conditional logic. Many of those postulates are inspired by similar properties for the classical case—see e.g. [13]—and others specifically address demands arising from the use of a probabilistic logic, such as the demand for a continuous behavior with respect to changes in the knowledge base. Second, we extend several inconsistency measures that were proposed for the classical case to the more expressive framework of probabilistic conditional logic and investigate their properties with respect to the rationality postulates. Third, we pick up an extended logical formalization [21] of the inconsistency measure proposed in [34] for probabilistic conditional logic, generalize it, and define a family of inconsistency measures based on minimizing the distance of a knowledge base to a consistent one. We also propose a novel compound measure that solves an issue with the previous measure. We thoroughly investigate the properties of all measures with respect to the rationality postulates and discuss their advantages and disadvantages with the use of examples.

The rest of this paper is organized as follows. We continue in Section 2 with an overview on probabilistic conditional logic and introduce further notation. In Section 3 we approach the problem of inconsistency measurement in probabilistic conditional logic by developing a series of rationality postulates. We continue with an overview on the technical results of the paper in Section 4. We extend inconsistency measures for classical logic to the probabilistic setting in Section 5 and present novel inconsistency measures that are more apt for the probabilistic setting in Section 6. In Section 7 we review related work and in Section 8 we conclude with some final remarks. All proofs of technical results can be found in Appendix A.

2. Probabilistic conditional logic

Let At be a propositional signature, i.e. a finite set of propositional atoms. Let $\mathcal{L}(\mathsf{At})$ be the corresponding propositional language generated by the atoms in At and the connectives $\wedge$ (*and*), $\vee$ (*or*), and $\neg$ (*negation*). For $\phi, \psi \in \mathcal{L}(\mathsf{At})$ we abbreviate

$\phi \wedge \psi$ by $\phi\psi$ and $\neg\phi$ by $\bar{\phi}$. The symbols $\top$ and $\bot$ denote *tautology* and *contradiction*, respectively. We use *possible worlds*, i.e. syntactical representations of *truth assignments*, for interpreting formulas in $\mathcal{L}(\mathsf{At})$. A possible world ω is a complete conjunction, i.e. a conjunction that contains for each $a \in \mathsf{At}$ either a or $\neg a$. Let $\Omega(\mathsf{At})$ denote the set of all possible worlds. A possible world $\omega \in \Omega(\mathsf{At})$ *satisfies* an atom $a \in \mathsf{At}$, denoted by $\omega \models a$ if and only if a positively appears in ω. The entailment relation $\models$ is extended to arbitrary formulas in $\mathcal{L}(\mathsf{At})$ in the usual way. Formulas $\psi, \phi \in \mathcal{L}(\mathsf{At})$ are *equivalent*, denoted by $\phi \equiv \psi$, if and only if $\omega \models \phi$ whenever $\omega \models \psi$ for every $\omega \in \Omega(\mathsf{At})$.

The central notion of probabilistic conditional logic [28] is that of a *probabilistic conditional*.

Definition 1 (*Probabilistic conditional*). If $\phi, \psi \in \mathcal{L}(\mathsf{At})$ with $p \in [0, 1]$ then $(\psi \mid \phi)[p]$ is called a *probabilistic conditional*.

A probabilistic conditional $c = (\psi \mid \phi)[p]$ is meant to describe a probabilistic *if–then* rule, i.e., the informal interpretation of c is that "if ϕ is true then ψ is true with probability p". If $\phi \equiv \top$ we abbreviate $(\psi \mid \phi)[p]$ by $(\psi)[p]$. Further, for $c = (\psi \mid \phi)[p]$ we denote with $\mathsf{head}(c) = \psi$ the *head* of c, with $\mathsf{body}(c) = \phi$ the *body* of c, and with $\mathsf{prob}(c) = p$ the *probability* of c. Let $\mathcal{C}(\mathsf{At})$ denote the set of all probabilistic conditionals with respect to At.

Definition 2 (*Knowledge base*). A *knowledge base* $\mathcal{K}$ is a finite sequence of probabilistic conditionals, i.e. it holds that $\mathcal{K} = \langle c_1, \ldots, c_n \rangle$ for some $c_1, \ldots, c_n \in \mathcal{C}(\mathsf{At})$.

We impose an ordering on the conditionals in a knowledge base $\mathcal{K}$ only for technical convenience. The order can be arbitrary and has no further meaning other than to enumerate the conditionals of a knowledge base in an unambiguous way. For similar reasons we allow a knowledge base to contain the same probabilistic conditional more than once. We come back to the reasons for these design choices later. However, for all practical purposes a knowledge base can be used as a set of probabilistic conditionals, as it is usually defined for knowledge representation issues. In particular, for knowledge bases $\mathcal{K} = \langle c_1, \ldots, c_n \rangle$, $\mathcal{K}' = \langle c'_1, \ldots, c'_m \rangle$ and a probabilistic conditional c we define $c \in \mathcal{K}$ via $c \in \{c_1, \ldots, c_n\}$, $\mathcal{K} \subseteq \mathcal{K}'$ via $\{c_1, \ldots, c_n\} \subseteq \{c'_1, \ldots, c'_m\}$, and $\mathcal{K} = \mathcal{K}'$ via $\{c_1, \ldots, c_n\} = \{c'_1, \ldots, c'_m\}$. The union of knowledge bases is defined via concatenation.

Semantics are given to probabilistic conditionals by *probability functions* on $\Omega(\mathsf{At})$. Let $\mathcal{F}(\mathsf{At})$ denote the set of all probability functions $P : \Omega(\mathsf{At}) \to [0, 1]$. A probability function $P \in \mathcal{F}(\mathsf{At})$ is extended to formulas $\phi \in \mathcal{L}(\mathsf{At})$ via

$$P(\phi) = \sum_{\omega \in \Omega(\mathsf{At}),\, \omega \models \phi} P(\omega). \tag{1}$$

That is, the probability of a formula is the sum of the probabilities of the possible worlds that satisfy that formula. If $P \in \mathcal{F}(\mathsf{At})$ then P satisfies a probabilistic conditional $(\psi \mid \phi)[p]$, denoted by $P \models^{pr} (\psi \mid \phi)[p]$, if and only if $P(\psi\phi) = pP(\phi)$. Note that we do not define probabilistic satisfaction via $P(\psi \mid \phi) = P(\psi\phi)/P(\phi) = p$ in order to avoid a case differentiation for $P(\phi) = 0$, see [23] for further justification. Note, that $P \models^{pr} (\psi)[p]$ if and only if $P(\psi) = p$ as $(\psi)[p]$ is the abbreviation for $(\psi \mid \top)[p]$ and $P \models^{pr} (\psi \mid \top)[p]$, if and only if $P(\psi \wedge \top) = pP(\top)$ which is equivalent to $P(\psi) = p$. A probability function P satisfies a knowledge base $\mathcal{K}$ (or is a model of $\mathcal{K}$), denoted by $P \models^{pr} \mathcal{K}$, if and only if $P \models^{pr} c$ for every $c \in \mathcal{K}$. Let $\mathsf{Mod}(\mathcal{K})$ be the set of models of $\mathcal{K}$. If $\mathsf{Mod}(\mathcal{K}) \neq \emptyset$ then $\mathcal{K}$ is *consistent* and if $\mathsf{Mod}(\mathcal{K}) = \emptyset$ then $\mathcal{K}$ is *inconsistent*.

Example 1. Consider the knowledge base

$$\mathcal{K} = \big\langle (f \mid b)[0.9], (b \mid p)[1], (f \mid p)[0.01] \big\rangle$$

with the intuitive meaning that birds (b) usually (with probability 0.9) fly (f), that penguins (p) are always birds, and that penguins usually do not fly (only with probability 0.01). The knowledge base $\mathcal{K}$ is consistent as for e.g. $P \in \mathcal{F}(\mathsf{At})$ with

$$\begin{array}{llll} P(bfp) = 0.005 & P(bf\bar{p}) = 0.49 & P(b\bar{f}p) = 0.045 & P(b\bar{f}\bar{p}) = 0.01 \\ P(\bar{b}fp) = 0.0 & P(\bar{b}f\bar{p}) = 0.2 & P(\bar{b}\bar{f}p) = 0.0 & P(\bar{b}\bar{f}\bar{p}) = 0.25 \end{array}$$

it holds that $P \models^{pr} \mathcal{K}$ as e.g.

$$P(b) = P(bfp) + P(bf\bar{p}) + P(b\bar{f}p) + P(b\bar{f}\bar{p}) = 0.55 \quad \text{and}$$
$$P(bf) = P(bfp) + P(bf\bar{p}) = 0.495$$

and therefore $P(f \mid b) = P(bf)/P(b) = 0.9$.

Example 2. The knowledge base $\{(a)[0.9], (a)[0.4]\}$ is inconsistent as there is no $P \in \mathcal{F}(\mathsf{At})$ with $P(a) = 0.9$ and $P(a) = 0.4$. Furthermore, observe that $\{(b \mid a)[0.8], (a)[0.6], (b)[0.4]\}$ is inconsistent as $P \models^{pr} \{(b \mid a)[0.8], (a)[0.6]\}$ implies $P(b) \geqslant 0.48$ which cannot simultaneously be satisfied with $P(b) = 0.4$.

A probabilistic conditional $(\psi \mid \phi)[p]$ is *normal* if and only if there are $\omega, \omega' \in \Omega(\mathsf{At})$ with $\omega \models \psi\phi$ and $\omega' \models \overline{\psi}\phi$. In other words, a probabilistic conditional c is normal if it is satisfiable but not tautological.

Example 3. The probabilistic conditionals $c_1 = (\top \mid a)[1]$ and $c_2 = (\bar{a} \mid a)[0.1]$ are *not* normal as c_1 is tautological (there is no $\omega \in \Omega(\mathsf{At})$ with $\omega \models \overline{\top}a$ as $\overline{\top}a \equiv \perp$) and c_2 is not satisfiable (there is no $\omega \in \Omega(\mathsf{At})$ with $\omega \models \bar{a}a$ as $\bar{a}a \equiv \perp$).

As a technical convenience, for the rest of this paper we consider only normal probabilistic conditionals, so let $\mathbb{K}$ be the set of all non-empty knowledge bases of $\mathcal{C}(\mathsf{At})$ that contain only normal probabilistic conditionals.

Proposition 1. *If $(\psi \mid \phi)[p]$ is normal then $(\psi \mid \phi)[x]$ is normal for every $x \in [0, 1]$.*

The proof of the above proposition is easy to see as the definition of normality does not depend on the probability of a conditional.

Knowledge bases $\mathcal{K}_1$, $\mathcal{K}_2$ are *extensionally equivalent*, denoted by $\mathcal{K}_1 \equiv^e \mathcal{K}_2$, if and only if $\mathsf{Mod}(\mathcal{K}_1) = \mathsf{Mod}(\mathcal{K}_2)$. Note that the notion of extensional equivalence does not distinguish between inconsistent knowledge bases, i.e. for inconsistent $\mathcal{K}_1$ and $\mathcal{K}_2$ it always holds that $\mathcal{K}_1 \equiv^e \mathcal{K}_2$. As we are interested particularly in inconsistent knowledge bases we require another means for comparing those. Knowledge bases $\mathcal{K}_1$, $\mathcal{K}_2$ are *semi-extensionally equivalent*, denoted by $\mathcal{K}_1 \equiv^s \mathcal{K}_2$, if and only if there is a bijection $\rho_{\mathcal{K}_1,\mathcal{K}_2} : \mathcal{K}_1 \to \mathcal{K}_2$ such that $c \equiv^e \rho_{\mathcal{K}_1,\mathcal{K}_2}(c)$ for every $c \in \mathcal{K}_1$. This means that two knowledge bases $\mathcal{K}_1$ and $\mathcal{K}_2$ are semi-extensionally equivalent if we find a mapping between the conditionals of both knowledge bases such that each conditional of $\mathcal{K}_1$ is extensionally equivalent to its image in $\mathcal{K}_2$. The following relationship is easy to see and given without proof.

Proposition 2. *It holds that $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ implies $\mathcal{K}_1 \equiv^e \mathcal{K}_2$.*

However, note that the other direction is not true in general.

Example 4. Consider the two knowledge bases $\mathcal{K}_1 = \langle (a)[0.7], (a)[0.4] \rangle$ and $\mathcal{K}_2 = \langle (b)[0.8], (b)[0.3] \rangle$. Both $\mathcal{K}_1$ and $\mathcal{K}_2$ are inconsistent and therefore $\mathcal{K}_1 \equiv^e \mathcal{K}_2$. But it holds that $\mathcal{K}_1 \not\equiv^s \mathcal{K}_2$ as both $(a)[0.7] \not\equiv^e (b)[0.8]$ and $(a)[0.7] \not\equiv^e (b)[0.3]$.

One way for reasoning with knowledge bases is by using model-based inductive reasoning techniques [23]. For example, reasoning based on the *principle of maximum entropy* selects among the models of a knowledge base $\mathcal{K}$ the one unique probability function with maximum entropy. Reasoning with this model satisfies several commonsense properties, see e.g. [23,14]. However, a necessary requirement for the application of model-based inductive reasoning techniques is the existence of at least one model of a knowledge base. In order to reason with inconsistent knowledge bases the inconsistency has to be resolved first. In the following, we discuss the topic of *inconsistency measurement* for probabilistic conditional logic as inconsistency measures can support the knowledge engineer in the task of resolving inconsistency.

3. Principles for inconsistency measurement

Inconsistency measurement is a research topic that has been mainly investigated in the field of classical theories only, see e.g. [11] for some recent work. In the following, we investigate inconsistency measurement for probabilistic conditional logic in a principled fashion but borrow some notation from classical inconsistency measurement like from [13]. An *inconsistency measure* $\mathcal{I}$ is a function that maps a (possibly inconsistent) knowledge base onto a non-negative real value, i.e., an inconsistency measure $\mathcal{I}$ is a function $\mathcal{I} : \mathbb{K} \to [0, \infty)$. The value $\mathcal{I}(\mathcal{K})$ for a knowledge base $\mathcal{K}$ is called the *inconsistency value* of $\mathcal{K}$ with respect to $\mathcal{I}$. In order to formalize the intuition behind inconsistency measures we give a list of principles that should be satisfied by any reasonable inconsistency measure. For that we need some further notation.

Definition 3 (*Minimal inconsistent set*). A set $\mathcal{M}$ of probabilistic conditionals is *minimal inconsistent* if $\mathcal{M}$ is inconsistent and every $\mathcal{M}' \subsetneq \mathcal{M}$ is consistent.

Let $\mathsf{MI}(\mathcal{K})$ be the set of the minimal inconsistent subsets of $\mathcal{K} \in \mathbb{K}$.

Example 5. Consider the knowledge base $\mathcal{K} = \langle (a)[0.3], (b)[0.5], (a \wedge b)[0.7] \rangle$. Then the set of minimal inconsistent subsets of $\mathcal{K}$ is given via

$$\mathsf{MI}(\mathcal{K}) = \big\{ \big\{ (a)[0.3], (a \wedge b)[0.7] \big\}, \big\{ (b)[0.5], (a \wedge b)[0.7] \big\} \big\}.$$

The notion of minimal inconsistent subsets captures those conditionals that are responsible for causing inconsistencies. Conditionals that do not take part in creating an inconsistency are *free*.

Definition 4 (*Free conditional*). A probabilistic conditional $c \in \mathcal{K}$ is *free* in $\mathcal{K}$ if and only if $c \notin \mathcal{M}$ for all $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$.

For a conditional or a knowledge base C let $\mathsf{At}(C)$ denote the set of atoms appearing in C.

Definition 5 (*Safe conditional*). A probabilistic conditional $c \in \mathcal{K}$ is *safe* in $\mathcal{K}$ if and only if $\mathsf{At}(c) \cap \mathsf{At}(\mathcal{K} \setminus \{c\}) = \emptyset$.

Note that the notion of safeness is due to Hunter and Konieczny [13]. The notion of a free conditional is clearly more general then the notion of a safe conditional.

Proposition 3. *If c is safe in $\mathcal{K}$ then c is free in $\mathcal{K}$.*

Consider now the following properties, cf. [13,34]. Let $\mathcal{K}$, $\mathcal{K}'$ be knowledge bases and c a probabilistic conditional.

Consistency $\mathcal{K}$ is consistent if and only if $\mathcal{I}(\mathcal{K}) = 0$.
Monotonicity $\mathcal{I}(\mathcal{K}) \leqslant \mathcal{I}(\mathcal{K} \cup \{c\})$.
Super-additivity If $\mathcal{K} \cap \mathcal{K}' = \emptyset$ then $\mathcal{I}(\mathcal{K} \cup \mathcal{K}') \geqslant \mathcal{I}(\mathcal{K}) + \mathcal{I}(\mathcal{K}')$.
Weak independence If $c \in \mathcal{K}$ is safe in $\mathcal{K}$ then $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K} \setminus \{c\})$.
Independence If $c \in \mathcal{K}$ is free in $\mathcal{K}$ then $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K} \setminus \{c\})$.
Penalty If $c \in \mathcal{K}$ is not free in $\mathcal{K}$ then $\mathcal{I}(\mathcal{K}) > \mathcal{I}(\mathcal{K} \setminus \{c\})$.
Irrelevance of syntax If $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ then $\mathcal{I}(\mathcal{K}_1) = \mathcal{I}(\mathcal{K}_2)$.
MI-separability If $\mathsf{MI}(\mathcal{K}_1 \cup \mathcal{K}_2) = \mathsf{MI}(\mathcal{K}_1) \cup \mathsf{MI}(\mathcal{K}_2)$ and $\mathsf{MI}(\mathcal{K}_1) \cap \mathsf{MI}(\mathcal{K}_2) = \emptyset$ then $\mathcal{I}(\mathcal{K}_1 \cup \mathcal{K}_2) = \mathcal{I}(\mathcal{K}_1) + \mathcal{I}(\mathcal{K}_2)$.

The property *consistency* demands that $\mathcal{I}(\mathcal{K})$ is minimal for consistent $\mathcal{K}$. The properties *monotonicity* and *super-additivity* demand that $\mathcal{I}$ is non-decreasing under the addition of new information. The properties *weak independence* and *independence* say that the inconsistency value should stay the same when adding “harmless” information. The property *penalty* is the counterpart of *independence* and demands that adding inconsistent information increases the inconsistency value. We define the property *irrelevance of syntax* in terms of the equivalence relation $\equiv^s$ as all inconsistent knowledge bases are equivalent with respect to $\equiv^e$. For an inconsistency measure $\mathcal{I}$, imposing *irrelevance of syntax* to hold in terms of $\equiv^e$ would yield $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K}')$ for every two inconsistent knowledge bases $\mathcal{K}$, $\mathcal{K}'$. The property *MI-separability*—which has been adapted from [13]—states that determining the value of $\mathcal{I}(\mathcal{K}_1 \cup \mathcal{K}_2)$ can be split into determining the values of $\mathcal{I}(\mathcal{K}_1)$ and $\mathcal{I}(\mathcal{K}_2)$ if the minimal inconsistent subsets of $\mathcal{K}_1 \cup \mathcal{K}_2$ are partitioned by $\mathcal{K}_1$ and $\mathcal{K}_2$.

The above properties do not take the crucial role of probabilities into account. In order to account for those we need some further notation. Let $\mathcal{K}$ be a knowledge base. For $\vec{x} \in [0,1]^{|\mathcal{K}|}$ we denote by $\mathcal{K}[\vec{x}]$ the knowledge base that is obtained from $\mathcal{K}$ by replacing the probabilities of the conditionals in $\mathcal{K}$ by the values in $\vec{x}$, respectively. More precisely, if $\mathcal{K} = \langle(\psi_1 \mid \phi_1)[p_1], \dots, (\psi_n \mid \phi_n)[p_n]\rangle$ then $\mathcal{K}[\vec{x}] = \langle(\psi_1 \mid \phi_1)[x_1], \dots, (\psi_n \mid \phi_n)[x_n]\rangle$ for $\vec{x} = \langle x_1, \dots, x_n\rangle \in [0,1]^n$. Similarly, for a single probabilistic conditional $c = (\psi \mid \phi)[p]$ and $x \in [0,1]$ we abbreviate $c[x] = (\psi \mid \phi)[x]$.

Definition 6 (*Characteristic function*). Let $\mathcal{K} \in \mathbb{K}$ be a knowledge base. The function $\Lambda_{\mathcal{K}} : [0,1]^{|\mathcal{K}|} \to \mathbb{K}$ with $\Lambda_{\mathcal{K}}(\vec{x}) = \mathcal{K}[\vec{x}]$ is called the *characteristic function* of $\mathcal{K}$.

Due to Proposition 1 the function $\Lambda_{\mathcal{K}}$ is well-defined. The above definition is also the justification for imposing an order on the probabilistic conditionals of a knowledge base.

Definition 7 (*Characteristic inconsistency function*). Let $\mathcal{I}$ be an inconsistency measure and let $\mathcal{K} \in \mathbb{K}$ be a knowledge base. The function

$$\theta_{\mathcal{I},\mathcal{K}} : [0,1]^{|\mathcal{K}|} \to [0,\infty)$$

with $\theta_{\mathcal{I},\mathcal{K}} = \mathcal{I} \circ \Lambda_{\mathcal{K}}$ is called the *characteristic inconsistency function* of $\mathcal{I}$ and $\mathcal{K}$.

The following property *continuity* describes our main demand for continuous inconsistency measurement, i.e., a “slight” change in the knowledge base should not result in a “vast” change of the inconsistency value.

Continuity $\theta_{\mathcal{I},\mathcal{K}}$ is continuous.

The above property demands a certain *smoothness* of the behavior of $\mathcal{I}$. Given a fixed set of probabilistic conditionals this property demands that changes in the *quantitative* part of the conditionals trigger a continuous change in the inconsistency value. Note that we require the *qualitative* part of the conditionals, i.e. premises and conclusions of the conditionals, to be fixed. This makes this property not applicable for the classical setting. In the probabilistic setting satisfaction of this property is helpful for the knowledge engineer in restoring consistency. Observe that for every knowledge base $\mathcal{K} \in \mathbb{K}$ there is always a $\vec{x} \in [0,1]^{|\mathcal{K}|}$ such that $\mathcal{K}[\vec{x}]$ is consistent, cf. [34]. While in the classical setting, consistency of knowledge bases can only be restored by either removing or weakening formulas, in the probabilistic setting every knowledge base can also be made consistent by changing probabilities, see [8] for a heuristic approach utilizing this observation. Given that we have a continuous inconsistency measure the search for a "close" consistent solution can be better guided, see [36] for approaches that utilize continuous inconsistency measures in order to implement a search procedure similar to gradient descent search in optimization [3].

Some relationships between the above properties are as follows.

Proposition 4. *Let $\mathcal{I}$ be an inconsistency measure and let $\mathcal{K}$, $\mathcal{K}'$ be some knowledge bases.*

1. *If $\mathcal{I}$ satisfies* super-additivity *then $\mathcal{I}$ satisfies* monotonicity.
2. *If $\mathcal{I}$ satisfies* independence *then $\mathcal{I}$ satisfies* weak independence.
3. *If $\mathcal{I}$ satisfies* MI-separability *then $\mathcal{I}$ satisfies* independence.
4. *$\mathcal{K} \subseteq \mathcal{K}'$ implies $\mathsf{MI}(\mathcal{K}) \subseteq \mathsf{MI}(\mathcal{K}')$.*
5. *If $\mathcal{I}$ satisfies* independence *then $\mathsf{MI}(\mathcal{K}) = \mathsf{MI}(\mathcal{K}')$ implies $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K}')$.*
6. *If $\mathcal{I}$ satisfies* independence *and* penalty *then $\mathsf{MI}(\mathcal{K}) \subsetneq \mathsf{MI}(\mathcal{K}')$ implies $\mathcal{I}(\mathcal{K}) < \mathcal{I}(\mathcal{K}')$.*

In [13] two further properties are discussed for classical inconsistency measurement: *normalization* and *dominance*. The property *normalization* can be phrased as follows (note that the term *normalization* is not the be confused with our notion of *normal* conditionals).

Normalization $\mathcal{I}(\mathcal{K}) \in [0,1]$.

The above property states that inconsistency values should be bounded from above by one. On the one hand, this demand makes perfect sense as this allows for comparing inconsistency values of different knowledge bases in a unified way. On the other hand, this demand is—in general—in conflict with the demand for *super-additivity* as the following example shows.

Example 6. Let $i, k \in \mathbb{N}$ with $k \leqslant i$. Consider the conditionals

$$c_1^k = (a_k)[0.6] \qquad c_2^k = (a_k)[0.4]$$

on a propositional signature $\mathsf{At}_i = \{a_1, \ldots, a_i\}$. Obviously, the knowledge base $\langle c_1^1, c_2^1 \rangle$ on At_1 is inconsistent and therefore some inconsistency measure $\mathcal{I}$ satisfying *consistency* assigns some non-zero inconsistency value to $\langle c_1^1, c_2^1 \rangle$, i.e. $\mathcal{I}(\langle c_1^1, c_2^1 \rangle) = x > 0$. Furthermore, any knowledge base $\langle c_1^i, c_2^i \rangle$ on At_i is inconsistent as well and should be assigned the same inconsistency value, i.e. $\mathcal{I}(\langle c_1^i, c_2^i \rangle) = x$. It follows that, if $\mathcal{I}$ satisfies *super-additivity* and does not take the size of signature of a knowledge base into account then there is a natural number $n \in \mathbb{N}$ such that for $\mathcal{K}_n = \langle c_1^1, c_2^1, \ldots, c_1^n, c_2^n \rangle$ it holds that

$$\mathcal{I}(\mathcal{K}_n) \geqslant \mathcal{I}\big(\langle c_1^1, c_2^1 \rangle\big) + \cdots + \mathcal{I}\big(\langle c_1^n, c_2^n \rangle\big) \geqslant nx > 1.$$

Thus, $\mathcal{I}$ cannot satisfy *normalization*.

The previous example showed that an inconsistency measure that does not take (the size of) the signature into account cannot satisfy *consistency*, *super-additivity*, and *normalization* at the same time. Furthermore, taking (the size of) the signature into account may become unintuitive. As for the case of Example 6, in order to allow $\mathcal{I}$ to satisfy *consistency*, *super-additivity*, and *normalization* it has to hold that for $\mathcal{K} = \langle c_1^1, c_2^1 \rangle$ defined on At_1 and $\mathcal{K}' = \langle c_1^1, c_2^1 \rangle$ defined on At_2 it follows that $\mathcal{I}(\mathcal{K}) \neq \mathcal{I}(\mathcal{K}')$. As $\mathcal{K} = \mathcal{K}'$ this result may be unintuitive. However, one has to note that for $\mathcal{K}$ the whole language is affected by the inconsistency while for $\mathcal{K}'$ only "half" of the language is affected. In particular, for the proposition $a_2 \in \mathsf{At}_2$ there is no conditional $c \in \mathcal{K}'$ such that $c \in \mathcal{M}$ for some $\mathcal{M} \in \mathsf{MI}(\mathcal{K}')$ and $a_2 \in \mathsf{At}(c)$. Provided that we employ a paraconsistent reasoning mechanism for probabilistic knowledge—like the one proposed in [6]—information about a_2 can consistently be derived, maybe only by inferring that there is no information on a_2, i.e., by deriving the probability 0.5 for a_2. This observation distinguishes $\mathcal{K}'$ from $\mathcal{K}$ as for the latter a_2 does not belong to the signature and therefore no information is derivable for a_2 at all. Although this distinction is marginal, observe that there is a difference in inferring that we have no information on a_2 and that inference on a_2 is not defined.

We now turn to the property *dominance* [13] which can be phrased as follows. Let $c_1 \models^{pr} c_2$ be defined via $\mathsf{Mod}(\{c_1\}) \subseteq \mathsf{Mod}(\{c_2\})$ for conditionals c_1, c_2.

Dominance If $c_1 \models^{pr} c_2$ then $\mathcal{I}(\mathcal{K} \cup \{c_1\}) \geqslant \mathcal{I}(\mathcal{K} \cup \{c_2\})$.

The motivation of the property *dominance* in the classical setting is that logically stronger formulas have the potential to bring more conflicts [13]. In the context of probabilistic conditional logic this property is vacuous as entailment by probabilistic conditionals is trivial.

Table 1
Comparison of inconsistency measures.

Property	$\mathcal{I}_0$	$\mathcal{I}_{\mathsf{MI}}$	$\mathcal{I}^C_{\mathsf{MI}}$	$\mathcal{I}_\eta$	$\mathcal{I}_p$	$\mathcal{I}^{\mathcal{I}_p}_\Sigma$	$\mathcal{I}^h_\mu$
Consistency	X	X	X	X	X	X	X
Monotonicity	X	X	X	X	X	X	X
Super-additivity	–	X	X	–	$p=1$	X	–
Irrelevance of syntax	X	X	X	X	X	X	X
Weak independence	X	X	X	X	X	X	X
Independence	X	X	X	X	X	X	?
MI-separability	–	X	X	–	$(p=1)$	X	–
Penalty	–	X	X	–	–	X	–
Normalization	X	–	–	X	–	–	X
Continuity	–	–	–	–	X	X	X

Proposition 5. *Let $c_1 = (\psi_1 \mid \phi_1)[p_1]$ and $c_2 = (\psi_2 \mid \phi_2)[p_2]$ be normal. If $c_1 \models^{pr} c_2$ then $c_2 \equiv^e c_1$.*

Applying this observation to the property *dominance* we obtain

Dominance If $c_1 \equiv^e c_2$ then $\mathcal{I}(\mathcal{K} \cup \{c_1\}) = \mathcal{I}(\mathcal{K} \cup \{c_2\})$,

which is a weakening of the property *irrelevance of syntax*. For this reason, we will not consider the property *dominance* in what follows.

4. Overview of results

In the following sections we investigate different inconsistency measures with respect to the properties defined above. We review inconsistency measures for classical logics and adapt them to the probabilistic case in Section 5. In particular, we investigate the drastic inconsistency measure $\mathcal{I}_0$, the MI inconsistency measure $\mathcal{I}_{\mathsf{MI}}$, the MI^C inconsistency measure $\mathcal{I}^C_{\mathsf{MI}}$, and the η-inconsistency measure $\mathcal{I}_\eta$. Afterwards, we develop novel inconsistency measures for the probabilistic case in Section 6. More specifically, we develop the family of d-inconsistency measures $\mathcal{I}_D$ that are based on distance measures D and the family of Σ-inconsistency measures $\mathcal{I}^{\mathcal{I}}_\Sigma$ that utilize other inconsistency measures. Finally, in Section 7 we also investigate another inconsistency measure $\mathcal{I}^h_\mu$ from related work [6].

Table 1 summarizes the properties of the inconsistency measures discussed in this paper. Note that we only show the properties of the p-norm distance inconsistency measure $\mathcal{I}_p$ as a particularly good representative for d-inconsistency measures $\mathcal{I}_D$. The properties of other d-inconsistency measures may vary, cf. Theorem 2. For the same reasons we only show the properties of the Σ-inconsistency measure instantiated with the p-norm distance inconsistency measure. In Table 1 the entry "X" means that the inconsistency measure satisfies the given property, the entry $p = 1$ means that the property is satisfied if the condition is satisfied, an entry in parentheses means that satisfaction of the property is conjectured, and a question mark means that it is unclear whether the property is satisfied.

In the following, we continue with providing the formal definitions of the inconsistency measures and the elaboration of the technical results.

5. Classical inconsistency measures

We start with a survey on existing approaches to inconsistency measurement for classical logic and adapt those to the probabilistic case. In particular, we have a look at the drastic inconsistency measure, the MI inconsistency measure, the MI^C inconsistency measure, and the η-inconsistency measure, see e.g. [12,17] for the classical definitions. What these approaches have in common, due to their origin, is that they concentrate on the qualitative part of inconsistency rather than the quantitative part, i.e. the probabilities.

5.1. Drastic inconsistency measure

The simplest approach to define an inconsistency measure is by just differentiating whether a knowledge base is consistent or inconsistent.

Definition 8 (*Drastic inconsistency measure*). Let $\mathcal{I}_0 : \mathbb{K} \to [0, \infty)$ be the function defined as

$$\mathcal{I}_0(\mathcal{K}) = \begin{cases} 0 & \text{if } \mathcal{K} \text{ is consistent} \\ 1 & \text{if } \mathcal{K} \text{ is inconsistent} \end{cases}$$

for $\mathcal{K} \in \mathbb{K}$. The function $\mathcal{I}_0$ is called the *drastic inconsistency measure*.

The drastic inconsistency measure allows only for a binary decision on inconsistencies and does not quantify the severity of inconsistencies. Although being a very simple inconsistency measure, $\mathcal{I}_0$ still satisfies several basic properties as the next proposition shows.

Proposition 6. *The function* $\mathcal{I}_0$ *satisfies* consistency, irrelevance of syntax, monotonicity, weak independence, independence, *and* normalization.

Notice, that $\mathcal{I}_0$ satisfies neither *super-additivity*, *penalty*, *MI-separability*, nor *continuity*.

Example 7. Consider the knowledge bases $\mathcal{K}_1 = \langle c_1, c_2 \rangle$ and $\mathcal{K}_2 = \langle c_3, c_4 \rangle$ given via

$$c_1 = (a)[0.4] \qquad c_2 = (a)[0.6] \qquad c_3 = (b)[0.4] \qquad c_4 = (b)[0.6].$$

It follows that $\mathcal{I}_0(\mathcal{K}_1) = \mathcal{I}_0(\mathcal{K}_2) = 1$ but

$$\mathcal{I}_0(\mathcal{K}_1 \cup \mathcal{K}_2) = 1 \neq \mathcal{I}_0(\mathcal{K}_1) + \mathcal{I}_0(\mathcal{K}_2),$$

therefore violating both *super-additivity* and *MI-separability*. Furthermore, c_4 is not a free conditional in $\mathcal{K}_1 \cup \mathcal{K}_2$ but $\mathcal{I}_0(\mathcal{K}_1 \cup \mathcal{K}_2 \setminus \{c_4\}) = \mathcal{I}_0(\mathcal{K}_1 \cup \mathcal{K}_2)$ violating *penalty*. Also, $\mathcal{I}_0$ fails to satisfy *continuity* as $\operatorname{Im} \mathcal{I}_0 = \{0, 1\}$ ($\operatorname{Im} f$ denotes the image of the function f).

One thing to note is that $\mathcal{I}_0$ is the upper bound for any inconsistency measure that satisfies *consistency* and *normalization*, i.e., if $\mathcal{I}$ satisfies *consistency* and *normalization* then $\mathcal{I}(\mathcal{K}) \leqslant \mathcal{I}_0(\mathcal{K})$ for every $\mathcal{K} \in \mathbb{K}$ [36].

5.2. MI inconsistency measure

The next inconsistency measure quantifies inconsistency by the number of minimal inconsistent subsets of a knowledge base.

Definition 9 (*MI inconsistency measure*). Let $\mathcal{I}_{\mathsf{MI}} : \mathbb{K} \to [0, \infty)$ be the function defined as

$$\mathcal{I}_{\mathsf{MI}}(\mathcal{K}) = \left|\mathsf{MI}(\mathcal{K})\right|$$

for $\mathcal{K} \in \mathbb{K}$. The function $\mathcal{I}_{\mathsf{MI}}$ is called the *MI inconsistency measure*.

The definition of the MI inconsistency measure is motivated by the intuition that the more minimal inconsistent subsets the greater the inconsistency.

Proposition 7. *The function* $\mathcal{I}_{\mathsf{MI}}$ *satisfies* consistency, monotonicity, super-additivity, weak independence, independence, irrelevance of syntax, MI-separability, *and* penalty.

Notice, that $\mathcal{I}_{\mathsf{MI}}$ satisfies neither *normalization* nor *continuity*.

Example 8. Consider again $\mathcal{K}_1$ and $\mathcal{K}_2$ from Example 7. It holds that $\mathcal{I}_{\mathsf{MI}}(\mathcal{K}_1 \cup \mathcal{K}_2) = 2$ violating *normalization*. Also, $\mathcal{I}_{\mathsf{MI}}$ fails to satisfy *continuity* as $\operatorname{Im} \mathcal{I}_{\mathsf{MI}} = \mathbb{N}_0$ (the non-negative natural numbers).

For a further discussion of the MI inconsistency measure we refer to [36].

5.3. MI^C inconsistency measure

Only considering the number of minimal inconsistent subsets may be too simple for assessing inconsistencies in general. Another indicator for the severity of inconsistencies is the size of minimal inconsistent subsets. A large minimal inconsistent subset means that the inconsistency is distributed over a large number of conditionals. The more conditionals involved in an inconsistency the less severe the inconsistency is. Furthermore, a small minimal inconsistent subset means that the participating conditionals strongly represent contradictory information. Consider the following example for classical logic that can be found in e.g. [12].

Example 9. In a lottery there are n lottery tickets and only one of them is the winning ticket. If w_i denotes the proposition that ticket i will win the lottery then the (classical) formula $\phi = w_1 \vee \cdots \vee w_n$ can be regarded as true. Furthermore, the *belief* of each ticket buyer i is that he will not win the lottery, i.e., the formula $\phi_i = \neg w_i$ is regarded to be true for each $i = 1, \ldots, n$. Obviously the set $\{\phi, \phi_1, \ldots, \phi_n\}$ is inconsistent as ϕ demands that one ticket has to win and, hence, one ticket owner k is wrong in assuming $\neg w_k$. However, with increasing number of available tickets the inconsistency becomes negligible and each ticket owner is justified in believing that he will not win.

Although the previous example has been formulated for classical logic the argument stands for probabilistic logics as well.

The following inconsistency measure is inspired by [12] and aims at differentiating between minimal inconsistent sets of different size.

Definition 10 (MI^C *inconsistency measure*). Let $\mathcal{I}^C_{\mathsf{MI}} : \mathbb{K} \to [0, \infty)$ be the function defined as

$$\mathcal{I}^C_{\mathsf{MI}}(\mathcal{K}) = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \frac{1}{|\mathcal{M}|}$$

for $\mathcal{K} \in \mathbb{K}$. The function $\mathcal{I}^C_{\mathsf{MI}}$ is called the MI^C *inconsistency measure*.

Note that $\mathcal{I}^C_{\mathsf{MI}}(\mathcal{K}) = 0$ if $\mathsf{MI}(\mathcal{K}) = \emptyset$.

The MI^C inconsistency measure sums over the reciprocal of the sizes of all minimal inconsistent subsets. In that way, a large minimal inconsistent subset contributes less to the inconsistency value than a small minimal inconsistent subset. As the MI inconsistency measure the MI^C inconsistency measure behaves well with respect to many desirable properties.

Proposition 8. *The function $\mathcal{I}^C_{\mathsf{MI}}$ satisfies* consistency, monotonicity, super-additivity, weak independence, independence, irrelevance of syntax, MI-separability, *and* penalty.

Note that $\mathcal{I}^C_{\mathsf{MI}}$ satisfies neither *normalization* nor *continuity*.

Example 10. Consider the knowledge base $\mathcal{K} = \langle c_1, \ldots, c_6 \rangle$ given via

$$\begin{aligned} &c_1 = (a)[0.4] \qquad c_2 = (a)[0.6] \qquad c_3 = (b)[0.4] \\ &c_4 = (b)[0.6] \qquad c_5 = (c)[0.4] \qquad c_6 = (c)[0.6]. \end{aligned}$$

It follows that $\mathcal{I}^C_{\mathsf{MI}}(\mathcal{K}) = 1.5$ thus violating *normalization*. $\mathcal{I}^C_{\mathsf{MI}}$ also fails to satisfy *continuity* as $\mathsf{Im}\,\mathcal{I}^C_{\mathsf{MI}} = \mathbb{Q}^+_0$ (the non-negative rational numbers).

For a further discussion of the MI^C inconsistency measure we refer to [36].

5.4. η-Inconsistency measure

The work [17] employs probability theory itself to measure inconsistency in classical theories by considering probability functions on classical interpretations. Those ideas can be extended for measuring inconsistency in probabilistic logics by considering probability functions on probability functions. Let $\hat{P} : \mathcal{F}(\mathsf{At}) \to [0, 1]$ be a probability function on $\mathcal{F}(\mathsf{At})$ such that $\hat{P}(P) > 0$ only for finitely many $P \in \mathcal{F}(\mathsf{At})$. Let $\mathcal{F}^2(\mathsf{At})$ be the set of those probability functions. Then define

$$\hat{P}(c) = \sum_{P \in \mathcal{F}(\mathsf{At}),\ P \models^{pr} c} \hat{P}(P) \tag{2}$$

for a conditional c. This means that the probability (in terms of $\hat{P}$) of a conditional is the sum of the probabilities of probability functions that satisfy c. Note that this definition is similar in spirit to the definition of the probability of formulas in (1). The main difference is that in (1) formulas of the object level are propositional formulas and in (2) formulas of the object level are probabilistic conditionals. Note also that by restricting $\hat{P}$ to assign a non-zero value only to finitely many $P \in \mathcal{F}(\mathsf{At})$, the sum in (2) is well-defined.

Now consider the following definition of the η-inconsistency measure.

Definition 11 (η-*Inconsistency measure*). Let $\mathcal{I}_\eta : \mathbb{K} \to [0, \infty)$ be the function defined as

$$\mathcal{I}_\eta(\mathcal{K}) = 1 - \max\bigl\{\eta \bigm| \exists \hat{P} \in \mathcal{F}^2(\mathsf{At}){:}\ \forall c \in \mathcal{K}{:}\ \hat{P}(c) \geqslant \eta\bigr\}$$

for $\mathcal{K} \in \mathbb{K}$. The function $\mathcal{I}_\eta$ is called the η-*inconsistency measure*.

The idea of the η-inconsistency measure is that it looks for the largest probability that can be consistently assigned to the conditionals of a knowledge base and defines the inconsistency value inversely proportional to this probability.

Example 11. Let $\mathcal{K}$ be a knowledge base with $\mathcal{K} = \langle (b \mid a)[0.9], (a)[0.9], (b)[0.1] \rangle$. Note that $\mathcal{K}$ is inconsistent. As $A_1 = \{(b \mid a)[0.9], (a)[0.9]\}$ is consistent, let $P_1 \in \mathcal{F}(\mathsf{At})$ be a probability function with $P_1 \models^{pr} A_1$. Similarly, let

$A_2 = \{(b \mid a)[0.9], (b)[0.1]\}$, $A_3 = \{(a)[0.9], (b)[0.1]\}$ and $P_2, P_3 \in \mathcal{F}(\mathsf{At})$ such that $P_2 \models^{pr} A_2$ and $P_3 \models^{pr} A_3$. Then define $\hat{P} \in \mathcal{F}^2(\mathsf{At})$ via

$$\hat{P}(P_1) = \hat{P}(P_2) = \hat{P}(P_3) = 1/3$$
$$\hat{P}(P) = 0 \quad \text{for } P \in \mathcal{F}(\mathsf{At}) \setminus \{P_1, P_2, P_3\}.$$

It follows

$$\hat{P}\big((b \mid a)[0.9]\big) = \sum_{P \in \mathcal{F}(\mathsf{At}),\, P \models^{pr} c} \hat{P}(P) = \hat{P}(P_1) + \hat{P}(P_2) = 2/3$$

and similarly $\hat{P}((a)[0.9]) = \hat{P}((b)[0.1]) = 2/3$. It is also easy to see that there is no $\hat{P}' \in \mathcal{F}^2(\mathsf{At})$ such that $\hat{P}'(c) > 2/3$ for all $c \in \mathcal{K}$. Therefore, it follows $\mathcal{I}_\eta(\mathcal{K}) = 1 - 2/3 = 1/3$.

Several properties for the η-inconsistency measure can be directly derived from properties of its classical counterpart. For example, the following proposition is a direct extension of Theorem 2.12 in [18].

Proposition 9. *If* $\mathsf{MI}(\mathcal{K}) = \{\mathcal{K}\}$ *then* $\mathcal{I}_\eta(\mathcal{K}) = 1/|\mathcal{K}|$.

As for the properties proposed in the previous section consider the following proposition.

Proposition 10. *The function* $\mathcal{I}_\eta$ *satisfies* consistency, monotonicity, weak independence, independence, irrelevance of syntax *and* normalization.

Note that $\mathcal{I}_\eta$ does not satisfy *super-additivity*, *penalty*, *MI-separability* and *continuity*.

Example 12. Consider again the knowledge bases $\mathcal{K}_1 = \langle c_1, c_2\rangle$ and $\mathcal{K}_2 = \langle c_3, c_4\rangle$ from Example 7 given via

$$c_1 = (a)[0.4] \qquad c_2 = (a)[0.6] \qquad c_3 = (b)[0.4] \qquad c_4 = (b)[0.6].$$

By Proposition 9 it follows that $\mathcal{I}_\eta(\mathcal{K}_1) = \mathcal{I}_\eta(\mathcal{K}_2) = 1/2$ but

$$\mathcal{I}_\eta(\mathcal{K}_1 \cup \mathcal{K}_2) = 1/2$$

as well, therefore violating both *super-additivity* and *MI-separability* (observe that there are P_1, P_2 such that $P_1 \models^{pr} c_1, c_3$ and $P_2 \models c_2, c_4$). Consider now the knowledge base $\mathcal{K}_3 = \langle (a)[0.4], (a)[0.6], (\neg a)[0.4]\rangle$. Note that $(\neg a)[0.4] \in \mathcal{K}_3$ is not a free conditional in $\mathcal{K}_3$. However, it holds that

$$\mathcal{I}_\eta(\mathcal{K}_3) = \mathcal{I}_\eta\big(\mathcal{K}_3 \setminus \big\{(\neg a)[0.4]\big\}\big) = 1/2$$

as there are $P_1, P_2 \in \mathcal{F}(\mathsf{At})$ with $P_1 \models^{pr} (a)[0.4], (\neg a)[0.6]$ and $P_2 \models^{pr} (a)[0.6]$. Consider now the knowledge base $\mathcal{K}_x = \langle (a)[0.2], (a)[x]\rangle$. It holds that $\mathcal{I}_\eta(\mathcal{K}_x) = 1/2$ for $x \neq 0.2$ and $\mathcal{I}_\eta(\mathcal{K}_x) = 0$ for $x = 0.2$. Therefore, $\mathcal{I}_\eta$ fails to satisfy *continuity*.

5.5. Classical inconsistency measures and continuity

The inconsistency measures discussed above were initially developed for inconsistency measurement in classical theories and therefore allow only for a "discrete" measurement. Hence, all of the above discussed inconsistency measures do not satisfy *continuity*. But satisfaction of *continuity* is crucial for an inconsistency measure in probabilistic logics in order to assess inconsistencies in a meaningful manner, cf. also Section 3.

Example 13. Consider the knowledge base $\mathcal{K} = \langle c_1, c_2, c_3\rangle$ given via

$$c_1 = (b \mid a)[1] \qquad c_2 = (a)[1] \qquad c_3 = (b)[0].$$

The knowledge base $\mathcal{K}$ is inconsistent and the set of minimal inconsistent subsets is given by $\mathsf{MI}(\mathcal{K}) = \{\{c_1, c_2, c_3\}\}$. It follows that

$$\mathcal{I}_0(\mathcal{K}) = 1 \qquad \mathcal{I}_{\mathsf{MI}}(\mathcal{K}) = 1 \qquad \mathcal{I}^C_{\mathsf{MI}}(\mathcal{K}) = \frac{1}{3} \qquad \mathcal{I}_\eta(\mathcal{K}) = \frac{1}{3}.$$

Consider a modification $\mathcal{K}' = \langle c'_1, c'_2, c'_3\rangle$ of $\mathcal{K}$ given via

$$c'_1 = (b \mid a)[1] \qquad c'_2 = (a)[1] \qquad c'_3 = (b)[0.999].$$

The knowledge base $\mathcal{K}'$ is still inconsistent and it holds that $\mathcal{I}(\mathcal{K}') = \mathcal{I}(\mathcal{K})$ for $\mathcal{I} \in \{\mathcal{I}_0, \mathcal{I}_{\mathrm{MI}}, \mathcal{I}^{\mathrm{C}}_{\mathrm{MI}}, \mathcal{I}_\eta\}$. Now consider the knowledge base $\mathcal{K}'' = \langle c''_1, c''_2, c''_3\rangle$ given via

$$c''_1 = (b \mid a)[1] \qquad c'_2 = (a)[1] \qquad c''_3 = (b)[1].$$

The knowledge base $\mathcal{K}''$ is consistent and it follows that $\mathcal{I}_0(\mathcal{K}'') = \mathcal{I}_{\mathrm{MI}}(\mathcal{K}'') = \mathcal{I}^{\mathrm{C}}_{\mathrm{MI}}(\mathcal{K}'') = \mathcal{I}_\eta(\mathcal{K}'') = 0$. By comparing $\mathcal{K}'$ and $\mathcal{K}''$ one can discover only a minor difference of the modeled knowledge. Whereas in $\mathcal{K}''$ the proposition b is assigned a probability of 1 in $\mathcal{K}'$ it is assigned a probability of 0.999. From a practical point of view this difference may be of no relevance. Still, a knowledge engineer may not grasp the harmlessness of the inconsistency in $\mathcal{K}'$ as $\mathcal{K}$ has the same degree of inconsistency with respect to those classical measures.

The above example motivates the need for a more graded approach to measure the inconsistencies in $\mathcal{K}$, $\mathcal{K}'$, and $\mathcal{K}''$. This measure should assign $\mathcal{K}'$ a much smaller inconsistency value than to $\mathcal{K}$ in order to distinguish their severities. In the next section, we continue with an investigation of inconsistency measures that take the probabilities of conditionals into account and therefore satisfy those needs.

6. Inconsistency measures based on distance minimization

As can be seen in Example 13 the probabilities of conditionals play a crucial role in creating inconsistencies. In order to respect this role we propose a family of inconsistency measures that is based on the distance to consistency. To this end we employ the notion of a distance measure.

6.1. The d-inconsistency measure

The obvious difference between classical knowledge bases—i.e. sets of classical formulas—and probabilistic knowledge bases is that the latter are parametrized by probabilities. Therefore, given a knowledge base of a fixed qualitative structure the different instantiations of probabilities can be represented within the vector space $[0,1]^{|\mathcal{K}|}$. In a vector space, the traditional means of measuring differences are *distance measures*.

Definition 12 (*Distance measure*). Let $n \in \mathbb{N}^+$. A function $d_n : \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty)$ is called a *distance measure* if it satisfies the following properties:

1. $d_n(\vec{x}, \vec{y}) = 0$ if and only if $\vec{x} = \vec{y}$ (*reflexivity*).
2. $d_n(\vec{x}, \vec{y}) = d_n(\vec{y}, \vec{x})$ (*symmetry*).
3. $d_n(\vec{x}, \vec{y}) \leqslant d_n(\vec{x}, \vec{z}) + d_n(\vec{z}, \vec{y})$ (*triangle inequality*).

For $n \in \mathbb{N}^+$ let $\mathcal{D}_n$ denote the set of all distance measures $d_n : \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty)$. Let $\mathcal{D} = \bigcup_{n\in\mathbb{N}^+} \mathcal{D}_n$.

The simplest form of a distance measure is the *drastic distance measure* d^0_n defined as $d^0_n(\vec{x}, \vec{y}) = 0$ for $\vec{x} = \vec{y}$ and $d^0_n(\vec{x}, \vec{y}) = 1$ for $\vec{x} \neq \vec{y}$ (for $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $n \in \mathbb{N}^+$). A more interesting distance measure is the p-norm distance.

Definition 13 (p*-Norm distance*). Let $n, p \in \mathbb{N}^+$. The function $d^p_n : \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty)$ defined via

$$d^p_n(\vec{x}, \vec{y}) = \sqrt[p]{|x_1 - y_1|^p + \cdots + |x_n - y_n|^p}$$

for $\vec{x} = \langle x_1, \ldots, x_n\rangle$, $\vec{y} = \langle y_1, \ldots, y_n\rangle \in \mathbb{R}^n$ is called the p*-norm distance*.

Special cases of the p-norm distance include the *Manhattan distance* (for $p = 1$) and the *Euclidean distance* (for $p = 2$).

In order to deal with vector spaces of different dimensions we also consider *distance generators* which map a dimension $n \in \mathbb{N}^+$ to a corresponding distance function.

Definition 14 (*Distance generator*). A distance generator D is a function $D : \mathbb{N}^+ \to \mathcal{D}$ such that $D(n) \in \mathcal{D}_n$ for all $n \in \mathbb{N}^+$. Let D be a distance generator.

1. D is *monotonously generating* if

$$D(n)\big(\langle x_1, \ldots, x_n\rangle, \langle y_1, \ldots, y_n\rangle\big) \leqslant D(n+1)\big(\langle x_1, \ldots, x_{n+1}\rangle, \langle y_1, \ldots, y_{n+1}\rangle\big) \tag{3}$$

for every $n \in \mathbb{N}^+$ and $x_1, \ldots, x_{n+1}, y_1, \ldots, y_{n+1} \in \mathbb{R}$.

2. D is *super-additively generating* if
$$D(n)\big(\langle x_1,\dots,x_n\rangle,\langle y_1,\dots,y_n\rangle\big)+D(m)\big(\langle x_{n+1},\dots,x_{n+m}\rangle,\langle y_{n+1},\dots,y_{n+m}\rangle\big)$$
$$\leqslant D(n+m)\big(\langle x_1,\dots,x_{n+m}\rangle,\langle y_1,\dots,y_{n+m}\rangle\big) \tag{4}$$
for every $n,m\in\mathbb{N}^+$ and $x_1,\dots,x_{n+m},y_1,\dots,y_{n+m}\in\mathbb{R}$.
3. D is *symmetric generating* if
$$D(n)\big(\langle x_1,\dots,x_n\rangle,\langle y_1,\dots,y_n\rangle\big)=D(n)\big(\langle x_1,\dots,1-x_i,\dots,x_n\rangle,\langle y_1,\dots,1-y_i,\dots,y_n\rangle\big) \tag{5}$$
for every $i=1,\dots,n$ and $n\in\mathbb{N}^+$.
4. D is *continuously generating* if $D(n)$ is continuous for every $n\in\mathbb{N}^+$.

Although distance generators may be defined quite arbitrarily we consider the *drastic distance generator* D^0 defined via $D^0(n)=d_n^0$ (for every $n\in\mathbb{N}^+$) and the *p-norm distance generator* D^p defined via $D^p(n)=d_n^p$ (for every $n,p\in\mathbb{N}^+$).

Coming back to the issue of measuring inconsistency one can define the "severity of inconsistency" in a knowledge base by the minimal distance of the knowledge base to a consistent one. As we are able to identify knowledge bases of the same qualitative structure in a vector space, we can employ distance measures for measuring inconsistency.

Definition 15 (*d-Inconsistency measure*). Let D be a distance generator. Then the function $\mathcal{I}_D:\mathbb{K}\to[0,\infty)$ defined via
$$\mathcal{I}_D(\mathcal{K})=\min\big\{D\big(|\mathcal{K}|\big)(\vec{x},\vec{y})\;\big|\;\mathcal{K}=\mathcal{K}[\vec{x}]\text{ and }\mathcal{K}[\vec{y}]\text{ is consistent}\big\} \tag{6}$$
for $\mathcal{K}\in\mathbb{K}$ is called the *d-inconsistency measure*.

The idea behind the d-inconsistency measure is that we look for a consistent knowledge base that both (1) has the same qualitative structure as the input knowledge base and (2) is as close as possible to the input knowledge base (that there always exists such a knowledge base see the proof of Theorem 1 below). That is, if the input knowledge base is $\mathcal{K}[\vec{x}]$ we look at all $\vec{y}\in[0,1]^{|\mathcal{K}|}$ such that $\mathcal{K}[\vec{y}]$ is consistent and $\vec{x}$ and $\vec{y}$ are as close as possible with respect to the distance measure $D(|\mathcal{K}|)$.

As we are mainly working with the p-norm distance we abbreviate $\mathcal{I}_{D^p}$ simply by $\mathcal{I}_p$. Looking at Eq. (6) it is not obvious that $\mathcal{I}_D$ is well-defined as e.g. the minimum might not be defined on the set of distance values. However, as the following theorem shows the d-inconsistency measure is well-defined for every reasonable distance measure.

Theorem 1. *If D is continuously generating the function $\mathcal{I}_D$ is well-defined.*

As the p-norm distance is a continuous function it also follows that $\mathcal{I}_p$ is well-defined for every $p\in\mathbb{N}$. In [34,36] the measure $\mathcal{I}_1$ has been investigated in a preliminary fashion while [21,35] contain some first discussions of the general p-norm distance inconsistency measure. In particular, in [21] it has been shown that for every $p,p'\in\mathbb{N}^+$ with $p\neq p'$ the two measures $\mathcal{I}_p$ and $\mathcal{I}_{p'}$ are not equivalent, i.e., there are knowledge bases $\mathcal{K}_1$ and $\mathcal{K}_2$ such that $\mathcal{I}_p(\mathcal{K}_1)>\mathcal{I}_p(\mathcal{K}_2)$ but $\mathcal{I}_{p'}(\mathcal{K}_1)<\mathcal{I}_{p'}(\mathcal{K}_2)$.

Note that D being continuously generating is a sufficient but not necessary requirement for $\mathcal{I}_D$ being well-defined. Consider also the following observation.

Proposition 11. *The function $\mathcal{I}_{D^0}$ is well-defined and it holds that $\mathcal{I}_{D^0}=\mathcal{I}_0$.*

Before we investigate the formal properties of the above measure we first have a look at an example.

Example 14. We continue Example 13 with the knowledge base $\mathcal{K}=\langle c_1,c_2,c_3\rangle$ given via
$$c_1=(b\mid a)[1]\qquad c_2=(a)[1]\qquad c_3=(b)[0]$$
and consider the p-norm distance inconsistency measure. In particular, observe that for $\mathcal{K}^*=\langle(b\mid a)[1],(a)[0.5],(b)[0.5]\rangle$ it holds that
$$\mathcal{K}^*\in\arg\min\big\{d_3^p(\vec{x},\vec{y})\;\big|\;\mathcal{K}=\mathcal{K}[\vec{x}]\text{ and }\mathcal{K}[\vec{y}]\text{ is consistent}\big\}$$
for all $p\in\mathbb{N}^+$. That is, $\mathcal{K}^*$ is a consistent knowledge base that has minimal p-norm distance to $\mathcal{K}$ for all $p\in\mathbb{N}^+$. In particular, it holds that $\mathcal{I}_p(\mathcal{K})=\sqrt[p]{2\cdot 0.5^p}$. For example, it holds that
$$\mathcal{I}_1(\mathcal{K})=1\quad\text{and}\quad\mathcal{I}_2(\mathcal{K})\approx 0.707.$$
Furthermore, it holds that $\mathcal{I}_1(\mathcal{K}')=0.001$ and $\mathcal{I}_2(\mathcal{K}')\approx 0.00071$, and clearly $\mathcal{I}_1(\mathcal{K}'')=\mathcal{I}_2(\mathcal{K}'')=0$.

As with respect to the properties proposed in the previous section consider the following results.

Theorem 2. *Let D be a distance generator such that $\mathcal{I}_D$ is well-defined.*

1. *The function $\mathcal{I}_D$ satisfies* consistency.
2. *If D is monotonously generating then $\mathcal{I}_D$ satisfies* monotonicity.
3. *If D is super-additively generating then $\mathcal{I}_D$ satisfies* super-additivity.
4. *If D is symmetric generating then $\mathcal{I}_D$ satisfies* irrelevance of syntax.
5. *If D is continuously generating then $\mathcal{I}_D$ satisfies* continuity.

As for the specific case of the p-norm distance inconsistency measure consider the following theorem.

Theorem 3. *Let $p \in \mathbb{N}^+$.*

1. *The function $\mathcal{I}_p$ satisfies* consistency, monotonicity, weak independence, independence, irrelevance of syntax, *and* continuity.
2. *If $p = 1$ then $\mathcal{I}_p$ satisfies* super-additivity.

The property *MI-separability* is not, in general, satisfied by $\mathcal{I}_D$ as the following example shows.

Example 15. Let $\mathcal{K} = \langle (a)[0.3], (a)[0.7], (b)[0.3], (b)[0.7] \rangle$. It is easy to see that

$$\mathcal{I}_1(\mathcal{K}) = 0.4 + 0.4 = 0.8$$
$$\mathcal{I}_2(\mathcal{K}) = \sqrt{0.2^2 + 0.2^2 + 0.2^2 + 0.2^2} = 0.4.$$

It also holds that

$$\mathcal{I}_1\big(\langle (a)[0.3], (a)[0.7] \rangle\big) = \mathcal{I}_1\big(\langle (b)[0.3], (b)[0.7] \rangle\big) = 0.4 \quad \text{and}$$
$$\mathcal{I}_2\big(\langle (a)[0.3], (a)[0.7] \rangle\big) = \mathcal{I}_2\big(\langle (b)[0.3], (b)[0.7] \rangle\big) = \sqrt{0.2^2 + 0.2^2} \approx 0.283.$$

For $p = 1$ it follows that

$$\mathcal{I}_1(\mathcal{K}) = \mathcal{I}_1\big(\langle (a)[0.3], (a)[0.7] \rangle\big) + \mathcal{I}_1\big(\langle (b)[0.3], (b)[0.7] \rangle\big),$$

therefore satisfying *MI-separability*. However, for $p = 2$ it follows that

$$\mathcal{I}_2(\mathcal{K}) < \mathcal{I}_2\big(\langle (a)[0.3], (a)[0.7] \rangle\big) + \mathcal{I}_2\big(\langle (b)[0.3], (b)[0.7] \rangle\big)$$

violating *MI-separability*—and also *super-additivity*—as $\langle (a)[0.3], (a)[0.7] \rangle$ and $\langle (b)[0.3], (b)[0.7] \rangle$ partition the set of minimal inconsistent subsets of $\mathcal{K}$.

As the above example suggests *MI-separability* seems to be satisfied for $\mathcal{I}_p$ with $p = 1$. However, neither a counterexample nor a formal proof has been found yet.

Conjecture 1. *If $p = 1$ then $\mathcal{I}_p$ satisfies* MI-separability.

Observe that $\mathcal{I}_p$ does not satisfy *penalty* which has been mistakenly claimed in [34]. Consider the following counterexample.

Example 16. Consider the knowledge base $\mathcal{K} = \langle (a)[0.7], (a)[0.3] \rangle$ and the probabilistic conditional $(a)[0.5]$. Then $(a)[0.5]$ is not free in $\mathcal{K}' = \mathcal{K} \cup \{(a)[0.5]\}$ as $\{(a)[0.3], (a)[0.5]\} \in \mathsf{MI}(\mathcal{K}')$. However, it holds that $\mathcal{I}_1(\mathcal{K}) = \mathcal{I}_1(\mathcal{K}') = 0.4$—as $\langle (a)[0.5], (a)[0.5] \rangle$ has minimal distance to $\mathcal{K}$ and $\langle (a)[0.5], (a)[0.5], (a)[0.5] \rangle$ has minimal distance to $\mathcal{K}'$—which violates *penalty*.

As for *normalization* consider the following counterexample.

Example 17. Consider the knowledge base $\mathcal{K} = \langle (a)[0], (a)[1], (b)[0], (b)[1] \rangle$. It is easy to see that $\mathcal{I}_1(\mathcal{K}) = 2$ violating *normalization*.

However, a normalized variant of $\mathcal{I}_p$ can be defined by exploiting $\mathcal{I}_p(\mathcal{K}) \leqslant |\mathcal{K}|$ for all $\mathcal{K} \in \mathbb{K}$, cf. [36].

6.2. The Σ-inconsistency measure

The main drawback of the inconsistency measure discussed above is that it does not satisfy *penalty*. However, this issue can be solved by the following compound measure.

Definition 16. Let $\mathcal{K}$ be a knowledge base and let $\mathcal{I}$ be an inconsistency measure. Then define the *Σ-inconsistency measure* $\mathcal{I}_\Sigma^{\mathcal{I}}(\mathcal{K})$ of $\mathcal{K}$ and $\mathcal{I}$ via

$$\mathcal{I}_\Sigma^{\mathcal{I}}(\mathcal{K}) = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \mathcal{I}(\mathcal{M}).$$

The Σ-inconsistency measure is defined as the sum of the inconsistency values of all minimal inconsistent subsets of the knowledge base under consideration. The following property is easy to see and given without proof.

Proposition 12. *Let $\mathcal{I}$ be an inconsistency measure. If $\mathsf{MI}(\mathcal{K}) = \{\mathcal{K}\}$ then $\mathcal{I}_\Sigma^{\mathcal{I}}(\mathcal{K}) = \mathcal{I}(\mathcal{K})$.*

The above proposition states that $\mathcal{I}_\Sigma^{\mathcal{I}}(\mathcal{K})$ is the same as $\mathcal{I}(\mathcal{K})$ if $\mathcal{K}$ is minimally inconsistent. For general knowledge bases consider the following example.

Example 18. We continue Example 16 and consider the knowledge base $\mathcal{K} = \langle (a)[0.7], (a)[0.3] \rangle$ and the probabilistic conditional $(a)[0.5]$. Observe that

$$\mathcal{I}_\Sigma^{\mathcal{I}_1}(\mathcal{K}) = \mathcal{I}_1(\mathcal{K}) = 0.4$$
$$\mathcal{I}_\Sigma^{\mathcal{I}_1}\big(\mathcal{K} \cup \{(a)[0.5]\}\big) = \mathcal{I}_1(\mathcal{K}) + \mathcal{I}_1\big(\langle (a)[0.7], (a)[0.5] \rangle\big) + \mathcal{I}_1\big(\langle (a)[0.3], (a)[0.5] \rangle\big) = 0.4 + 0.2 + 0.2 = 0.8.$$

Therefore, the addition of the non-free conditional $(a)[0.5]$ to $\mathcal{K}$ has been penalized by $\mathcal{I}_\Sigma^{\mathcal{I}_1}$.

As hinted above, the Σ-inconsistency measure $\mathcal{I}_\Sigma^{\mathcal{I}}(\mathcal{K})$ behaves well with respect to the property *penalty*, provided that the inner measure is a reasonable inconsistency measure. Consider the following theorem.

Theorem 4. *Let $\mathcal{I}$ be an inconsistency measure.*

1. *$\mathcal{I}_\Sigma^{\mathcal{I}}$ satisfies* monotonicity, super-additivity, weak independence, independence, *and* MI-separability.
2. *If $\mathcal{I}$ satisfies* consistency *then $\mathcal{I}_\Sigma^{\mathcal{I}}$ satisfies* consistency *and* penalty.
3. *If $\mathcal{I}$ satisfies* irrelevance of syntax *then $\mathcal{I}_\Sigma^{\mathcal{I}}$ satisfies* irrelevance of syntax.
4. *If $\mathcal{I}$ satisfies* continuity *then $\mathcal{I}_\Sigma^{\mathcal{I}}$ satisfies* continuity.

The following corollary is a direct application of Theorems 3 and 4.

Corollary 1. *If $p \in \mathbb{N}^+$ then $\mathcal{I}_\Sigma^{\mathcal{I}_p}$ satisfies* consistency, monotonicity, super-additivity, weak independence, independence, MI-separability, penalty, irrelevance of syntax, *and* continuity.

As one can see, the Σ-inconsistency measure performs well with respect to all properties except *normalization*.

7. Related work

The problem of measuring inconsistency in probabilistic knowledge bases is relatively novel and has—to our knowledge—only been addressed before in [31,6] and [21]. We have a more closer look on the works [21] and [6] below.

Further related work is concerned with measuring inconsistency in classical theories, see e.g. the works by Hunter et al. [12,10,13]. While [12,13] deal with measuring inconsistency in propositional logic, the work [10] considers first-order logic. Those works also take a principled approach to measuring inconsistency and many of our properties have been adapted from e.g. [12]. Furthermore, the inconsistency measures presented in Section 6 are straightforward translations of inconsistency measures from those works. However, Hunter et al. are working with classical theories and as such do not have to deal with probabilities as a means for knowledge representation. In order to adhere for the presence of probabilities we introduced continuous inconsistency measures which have no correspondent in the classical setting.

Besides the inconsistency measures discussed here another form of measuring inconsistency can be realized using *culpability measures* [6,34], also used under the term *inconsistency values* in [13]. A culpability measure does not assign a degree of inconsistency to the whole knowledge base but to each individual element of the knowledge base. The interpretation of culpability measures is that they assign a degree of "blame" for creating an inconsistency to an element. In [13] such a measure has been defined in terms of some ordinary inconsistency measure and the *Shapley value*, a well-known solution for solving coalition games in game theory [32]. This approach can also be applied for inconsistency measures for probabilistic logics as has been done for the measure $\mathcal{I}_1$ in [34]. Furthermore, in [6] the measure $\mathcal{I}^h_\mu$ (see below) has also been extended to a culpability measure.

We go on by taking a closer look on the works by Muiño [21] and Daniel [6], for some analysis on [31] see [36].

7.1. Infinitesimal inconsistency values

The research presented in this paper is complementary to the work in [21]. The paper [21] also discusses the $\mathcal{I}_p$ measure but focuses on (1) the problem of infinitesimal inconsistency values and (2) the application of $\mathcal{I}_p$ on the medical knowledge base CADIAG-2. In particular, it is not investigated how $\mathcal{I}_p$ behaves with respect to the principles above.

The problem of infinitesimal inconsistency values appears when one defines probabilistic satisfaction via $P \models^{pr}_{\mathrm{alt}} (\psi \mid \phi)[p]$ if and only if

$$P(\psi \mid \phi) = p \quad \text{and} \quad P(\phi) > 0.$$

A knowledge base $\mathcal{K}$ is $\models^{pr}_{\mathrm{alt}}$-consistent if there exists $P \in \mathcal{F}(\mathsf{At})$ with $P \models^{pr}_{\mathrm{alt}} \mathcal{K}$. Using our notation, the inconsistency measure $\mathcal{I}'_p$ from [21] can be defined via

$$\mathcal{I}'_p(\mathcal{K}) = \min\bigl\{D^p\bigl(|\mathcal{K}|\bigr)(\vec{x}, \vec{y}) \bigm| \mathcal{K} = \mathcal{K}[\vec{x}] \text{ and } \mathcal{K}[\vec{y}] \text{ is } \models^{pr}_{\mathrm{alt}}\text{-consistent}\bigr\}. \tag{7}$$

Theorem 1 does not apply for $\mathcal{I}'_p$ as the set $\{\vec{y} \mid \mathcal{K}[\vec{y}]$ is $\models^{pr}_{\mathrm{alt}}$-consistent$\}$ is not closed. Therefore, the minimum in (7) is not always defined.

Example 19. Consider the knowledge base $\mathcal{K} = \langle (a)[0], (b \mid a)[0.7]\rangle$. Note that $\mathcal{K}$ is consistent (using our notion of probabilistic satisfaction) but not $\models^{pr}_{\mathrm{alt}}$-consistent as there is no probability function P with $P \models^{pr}_{\mathrm{alt}} (a)[0]$ and $P \models^{pr}_{\mathrm{alt}} (b \mid a)[0.7]$. However, one can easily construct a sequence of probability functions $P_1, P_2, \ldots$ such that $P_i(a) > 0$ and $P_i(b \mid a) = 0.7$ for $i \in \mathbb{N}$ and

$$\lim_{n\to\infty} P(a) = 0.$$

In [21] knowledge bases like $\mathcal{K}$ above are assigned an *infinitesimal* inconsistency value. The motivation for introducing infinitesimal inconsistency values stems from the application of $\mathcal{I}_p$ on the medical knowledge base, a collection of expert rules relating symptoms and diseases. In [21] it is shown that CADIAG-2 has an infinitesimal inconsistency value.

7.2. Candidacy degrees of best candidates

Among others, one contribution of [6] is an inconsistency measure on knowledge bases of probabilistic constraints. In particular, the work [6] focuses on linear probabilistic knowledge bases but also considers generalizations such as polynomial probabilistic knowledge bases. However, in order to compare it to our work we simplify several notations and present the inconsistency measure $\mathcal{I}^h_\mu$ of [6] only for probabilistic conditional logic.

The central notion of [6] is the *candidacy function*. A candidacy function is similar to a fuzzy set as it assigns a degree of membership of a probability function belonging to the models of a knowledge base. More specifically, a candidacy function $\mathfrak{C}$ is a function $\mathfrak{C} : \mathcal{F}(\mathsf{At}) \to [0, 1]$. A uniquely determined candidacy function $\mathfrak{C}_\mathcal{K}$ can be assigned to a (consistent or inconsistent) knowledge base $\mathcal{K}$ as follows. For a probability function $P \in \mathcal{F}(\mathsf{At})$ and a set of probability functions $S \subseteq \mathcal{F}(\mathsf{At})$ let $d^E(P, S)$ denote the distance of P to S with respect to the Euclidean norm, i.e., $d^E(P, S)$ is defined via

$$d^E(P, S) = \inf\left\{\sqrt{\sum_{\omega\in\Omega(\mathsf{At})} \bigl(P(\omega) - P'(\omega)\bigr)^2} \;\middle|\; P' \in S\right\}.$$

Let $h : \mathbb{R}^+ \to (0, 1]$ be a strictly decreasing, positive, and continuous log-concave function with $h(0) = 1$. Then the candidacy function $\mathfrak{C}^h_\mathcal{K}$ for a knowledge base $\mathcal{K}$ is defined as

$$\mathfrak{C}^h_\mathcal{K}(P) = \prod_{c\in\mathcal{K}} h\Bigl(\sqrt{2^{|\mathsf{At}|}}\, d^E\bigl(P, \mathsf{Mod}\bigl(\{c\}\bigr)\bigr)\Bigr)$$

for every $P \in \mathcal{F}(\mathsf{At})$. Note that the definition of the candidacy function $\mathfrak{C}^h_\mathcal{K}$ depends on the size of the signature At. The intuition behind this definition is that a probability function P that is near to the models of each probabilistic conditional in $\mathcal{K}$ gets a high candidacy degree in $\mathfrak{C}^h_{\mathcal{K}}(P)$. It is easy to see that it holds that $\mathfrak{C}^h_{\mathcal{K}}(P) = 1$ if and only if $P \models^{pr} \mathcal{K}$. Using the candidacy function $\mathfrak{C}^h_{\mathcal{K}}$ the inconsistency measure $\mathcal{I}^h_\mu$ can be defined via

$$\mathcal{I}^h_\mu(\mathcal{K}) = 1 - \max_{P \in \mathcal{F}(\mathsf{At})} \mathfrak{C}^h_{\mathcal{K}}(P)$$

for a knowledge base $\mathcal{K}$. In [6] it is shown that $\mathcal{I}^h_\mu$ satisfies (among others) the following properties.

Proposition 13. *(See [6].) $\mathcal{I}^h_\mu$ satisfies* consistency, monotonicity, continuity, *and* normalization.

Furthermore, it can also be shown that the function $\mathcal{I}^h_\mu$ satisfies the following properties.

Theorem 5. *$\mathcal{I}^h_\mu$ satisfies* irrelevance of syntax *and* weak independence.

In Example 6 we talked about the issue of an inconsistency measure satisfying all three of *consistency*, *super-additivity*, and *normalization*. We showed that an inconsistency measure that does not take the cardinality of the signature into account cannot satisfy all these properties at once. As one can see above, the function $\mathcal{I}^h_\mu$ takes the cardinality of the signature into account and it may be possible that $\mathcal{I}^h_\mu$ satisfies *super-additivity*. However, this is not the case as the following example shows.

Example 20. Let $\mathsf{At} = \{a_1, a_2\}$ be a propositional signature and let $\mathcal{K}_1 = \langle c_1, c_2\rangle$ and $\mathcal{K}_2 = \langle c_3, c_4\rangle$ be knowledge bases with

$$c_1 = (a_1)[1] \qquad c_2 = (a_1)[0] \qquad c_3 = (a_2)[1] \qquad c_4 = (a_2)[0]$$

and let $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2$. Note that both $\mathcal{K}_1$ and $\mathcal{K}_2$ are inconsistent and $\mathcal{K}_1 \cap \mathcal{K}_2 = \emptyset$. As $\mathcal{I}^h_\mu$ is defined on the semantic level and does not take the names of propositions into account it follows that $\mathcal{I}^h_\mu(\mathcal{K}_1) = \mathcal{I}^h_\mu(\mathcal{K}_2)$. As the situations in $\mathcal{K}_1$ and $\mathcal{K}_2$ are symmetric and $\mathcal{K}_i$ is symmetric with respect to c_1 and c_2 and with respect to c_3 and c_4 there are probability functions P_i with $\mathcal{I}^h_\mu(\mathcal{K}_i) = 1 - C^h_{\mathcal{K}_i}(P_i)$ for $i = 1, 2$ and

$$d^E\big(P_1, \mathsf{Mod}\big(\{c_1\}\big)\big) = d^E\big(P_1, \mathsf{Mod}\big(\{c_2\}\big)\big) = d^E\big(P_2, \mathsf{Mod}\big(\{c_3\}\big)\big) = d^E\big(P_2, \mathsf{Mod}\big(\{c_4\}\big)\big).$$

Let $x = d^E(P_1, \mathsf{Mod}(\{c_1\}))$ and let $h^* : \mathbb{R}^+ \to (0, 1]$ be a strictly decreasing, positive, and continuous log-concave function with $h^*(0) = 1$ and $h^*(\sqrt{2^{|\mathsf{At}|}}x) = 0.5$. Then it follows $C^{h^*}_{\mathcal{K}_1}(P_1) = 0.25$ and $\mathcal{I}^{h^*}_\mu(\mathcal{K}_1) = 0.75$. In order to satisfy *super-additivity* $\mathcal{I}^{h^*}_\mu$ must satisfy

$$\mathcal{I}^{h^*}_\mu(\mathcal{K}) \geqslant \mathcal{I}^{h^*}_\mu(\mathcal{K}_1) + \mathcal{I}^{h^*}_\mu(\mathcal{K}_2) = 1.5$$

which is a contradiction since $\mathcal{I}^{h^*}_\mu$ satisfies *normalization*.

The above example is also a counterexample for *MI-separability* as $\mathcal{K}_1$ and $\mathcal{K}_2$ partition the set of minimal inconsistent subsets. Furthermore, it can be easily seen that $\mathcal{I}^h_\mu$ also fails to satisfy *penalty* for similar reasons as $\mathcal{I}_d$ fails to satisfy *penalty*. For the knowledge base $\mathcal{K} = \langle (b \mid a)[1], (a)[1], (b)[0]\rangle$ let P' be such that

$$\max_{P \in \mathcal{F}(\mathsf{At})} \mathfrak{C}^h_{\mathcal{K}}(P) = \mathfrak{C}^h_{\mathcal{K}}\big(P'\big). \tag{8}$$

In other words, P' is a probability function that has the maximal candidacy degree with respect to $\mathcal{K}$. As $\mathcal{K}$ is inconsistent, it follows that P' fails to satisfy at least one of the probabilistic conditionals of $\mathcal{K}$. Assume that it holds that $P' \not\models^{pr} (b \mid a)[1]$ which implies $P'(a) > 0$. Consider the knowledge base $\mathcal{K}' = \mathcal{K} \cup \{c'\}$ with $c' = (b \mid a)[P'(b \mid a)]$. As $\mathcal{I}^h_\mu$ satisfies *monotonicity* it follows $\mathcal{I}^h_\mu(\mathcal{K}') \geqslant \mathcal{I}^h_\mu(\mathcal{K})$ and due to $h(\sqrt{2^{|\mathsf{At}|}} d^E(P', \mathsf{Mod}(\{c'\}))) = 1$, as $d^E(P', \mathsf{Mod}(\{c'\})) = 0$, it follows that P' also satisfies

$$\max_{P \in \mathcal{F}(\mathsf{At})} \mathfrak{C}^h_{\mathcal{K}'}(P) = \mathfrak{C}^h_{\mathcal{K}'}\big(P'\big).$$

Therefore, P' has also maximal candidacy degree with respect to $\mathcal{K}'$ which is clear as we only added information consistent with P' (otherwise P' would have violated (8)). It follows $\mathcal{I}^h_\mu(\mathcal{K}') \leqslant \mathcal{I}^h_\mu(\mathcal{K})$ and as $\{(b \mid a)[1], (a)[1], c'\}$ is a minimal inconsistent subset of $\mathcal{K}'$ this contradicts *penalty*. Similar observations can be made when $P' \not\models^{pr} (a)[1]$ or $P' \not\models^{pr} (b)[0]$.

In [6] it is shown that $\mathcal{I}^h_\mu$ satisfies several other properties that cannot be related directly to our properties of Section 3, see [36] for a discussion. It is also still an open issue whether $\mathcal{I}^h_\mu$ satisfies *independence*.

8. Summary and discussion

Analyzing inconsistencies is of major concern in the area of knowledge representation as consistency is a necessary prerequisite for many knowledge representation formalisms. In particular, the task of inference bases mostly on the consistency of the underlying information. In this paper, we investigated inconsistency measures for probabilistic conditional logic. For that, we developed a series of rationality postulates for inconsistency measures motivated by both inconsistency measurement for classical logics and the peculiarities of probabilistic knowledge representation. We adapted several classical inconsistency measures and showed that they lack a particularly important property for the probabilistic domain, namely, a continuous behavior with respect to modifications of the knowledge base. Consequently, we investigated inconsistency measures based on distance measures and showed that these measures are more apt for the probabilistic domain. We compared these measures with related work, in particular with the approach of [6].

In this paper, we used probabilistic conditional logic for knowledge representation which suffices for many application areas that need to represent rule-like information. However, probabilistic conditional logic is not capable of expressing general linear relationships such as "a is twice as probable as b" or polynomial relationships such as "a and b are probabilistically independent". Furthermore, using point probabilities can be seen as too restricting as well and one may want to represent conditionals of the form $(\psi \mid \phi)[u,l]$ with the intended meaning that $P(\psi \mid \phi) \in [u,l]$, cf. [20]. The motivation of using the simple framework of probabilistic conditional logic here merely stems from reasons of presentation rather than inadequacy of the ideas to more complex frameworks. In [36] the inconsistency measure $\mathcal{I}_1$ has also been defined for the more general frameworks of linear probabilistic knowledge bases and probabilistic conditional logic with intervals. Generalizing those ideas to the discussion from this paper is straightforward.

The focus of the discussion in this paper was mainly on properties of inconsistency measures and not on their algorithmic computation and complexity. However, Eq. (6) already induces a straightforward method to compute the value $\mathcal{I}_D(\mathcal{K})$ for a specific knowledge base $\mathcal{K}$ by representing (6) as an optimization problem, see [34,21] for formalizations. Note that these optimization problems are, in general, non-convex. Furthermore, a subproblem of determining the inconsistency value for a knowledge base $\mathcal{K}$ is checking consistency of probabilistic conditional knowledge bases which is an NP-hard problem [23]. Therefore, determining $\mathcal{I}_D(\mathcal{K})$ is in general a hard task and future work comprises of investigating scalable approaches. Some first steps have already been conducted in [36] by approximating $\mathcal{I}_D(\mathcal{K})$ by "similar" convex optimization problems. Future work comprises of developing optimized algorithms by utilizing e.g. more sophisticated methods for probabilistic consistency checking [7]. The basic approach for computing $\mathcal{I}_D(\mathcal{K})$ using non-convex optimization methods has also been implemented in the Tweety library for artificial intelligence.[1]

Appendix A. Proofs of technical results

Proposition 3. *If c is safe in $\mathcal{K}$ then c is free in $\mathcal{K}$.*

Proof. Assume that c is not free in $\mathcal{K} \cup \{c\}$. Then there is a set $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$ with $c \in \mathcal{M}$. As $\mathcal{M} \setminus \{c\}$ is consistent and $\mathsf{At}(\mathcal{M} \setminus \{c\}) \cap \mathsf{At}(\{c\}) = \emptyset$ (as c is safe in $\mathcal{K}$) let P_1 be a probability function in $\mathcal{F}(\mathsf{At} \setminus \mathsf{At}(\{c\}))$ with $P_1 \models^{pr} \mathcal{M} \setminus \{c\}$. As c is normal let P_2 be a probability function in $\mathcal{F}(\mathsf{At}(\{c\}))$ with $P_2 \models^{pr} c$. Let $\omega \in \Omega(\mathsf{At})$ and define $\omega_{\mathcal{A}}$ with $\mathcal{A} \subseteq \mathsf{At}$ to be the projection of ω on $\mathcal{A}$, i.e. $\omega_{\mathcal{A}} = \bigwedge\{a \mid a \in \mathcal{A}, \omega \models a\} \cup \{\neg a \mid a \in \mathcal{A}, \omega \models \neg a\}$. Define a probability function P in $\mathcal{F}(\mathsf{At})$ via

$$P(\omega) = P_1(\omega_{\mathsf{At}\setminus\mathsf{At}(\{c\})}) \cdot P_2(\omega_{\mathsf{At}(\{c\})})$$

for all $\omega \in \Omega(\mathsf{At})$. Note that $f : \Omega(\mathsf{At}) \to \Omega(\mathsf{At} \setminus \mathsf{At}(\{c\})) \times \Omega(\mathsf{At}(\{c\}))$ with

$$f(\omega) = (\omega_{\mathsf{At}\setminus\mathsf{At}(\{c\})}, \omega_{\mathsf{At}(\{c\})})$$

is a bijection. It follows that P is indeed a probability function as

$$\begin{aligned}
\sum_{\omega\in\Omega(\mathsf{At})} P(\omega) &= \sum_{\omega\in\Omega(\mathsf{At})} P_1(\omega_{\mathsf{At}\setminus\mathsf{At}(\{c\})}) \cdot P_2(\omega_{\mathsf{At}(\{c\})}) \\
&= \sum_{(\omega_1,\omega_2)\in\Omega(\mathsf{At}\setminus\mathsf{At}(\{r\}))\times\Omega(\mathsf{At}(\{c\}))} P_1(\omega_1)P_2(\omega_2) \\
&= \sum_{\omega_1\in\Omega(\mathsf{At}\setminus\mathsf{At}(\{c\}))} \sum_{\omega_2\in\Omega(\mathsf{At}(\{c\}))} P_1(\omega_1)P_2(\omega_2) \\
&= \sum_{\omega_1\in\Omega(\mathsf{At}\setminus\mathsf{At}(\{c\}))} \left(P_1(\omega_1) \cdot \sum_{\omega_2\in\Omega(\mathsf{At}(\{c\}))} P_2(\omega_2)\right) \\
&= 1.
\end{aligned}$$

[1] http://sourceforge.net/projects/tweety/.

Furthermore, for $\omega \in \Omega(\mathsf{At} \setminus \mathsf{At}(\{c\}))$ it holds that

$$P(\omega) = \sum_{\omega' \in \Omega(\mathsf{At}(\{c\}))} P(\omega \wedge \omega') = \sum_{\omega' \in \Omega(\mathsf{At}(\{c\}))} P_1(\omega) P_2(\omega') = P_1(\omega) \sum_{\omega' \in \Omega(\mathsf{At}(\{c\}))} P_2(\omega') = P_1(\omega)$$

and similarly $P(\omega') = P_2(\omega')$. It follows that $P \models^{pr} \mathcal{M} \setminus \{c\}$ and $P \models^{pr} c$ contradicting the assumption that $\mathcal{M}$ is a minimal inconsistent subset. $\square$

Proposition 4. *Let $\mathcal{I}$ be an inconsistency measure and let $\mathcal{K}, \mathcal{K}'$ be some knowledge bases.*

1. *If $\mathcal{I}$ satisfies* super-additivity *then $\mathcal{I}$ satisfies* monotonicity.
2. *If $\mathcal{I}$ satisfies* independence *then $\mathcal{I}$ satisfies* weak independence.
3. *If $\mathcal{I}$ satisfies* MI-separability *then $\mathcal{I}$ satisfies* independence.
4. *$\mathcal{K} \subseteq \mathcal{K}'$ implies $\mathsf{MI}(\mathcal{K}) \subseteq \mathsf{MI}(\mathcal{K}')$.*
5. *If $\mathcal{I}$ satisfies* independence *then $\mathsf{MI}(\mathcal{K}) = \mathsf{MI}(\mathcal{K}')$ implies $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K}')$.*
6. *If $\mathcal{I}$ satisfies* independence *and* penalty *then $\mathsf{MI}(\mathcal{K}) \subsetneq \mathsf{MI}(\mathcal{K}')$ implies $\mathcal{I}(\mathcal{K}) < \mathcal{I}(\mathcal{K}')$.*

Proof.

1. Let $\mathcal{I}$ satisfy *super-additivity*. If $c \in \mathcal{K}$ then $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K} \cup \{c\})$. If $c \notin \mathcal{K}$ then $\mathcal{I}(\mathcal{K} \cup \{c\}) \geqslant \mathcal{I}(\mathcal{K}) + \mathcal{I}(\{c\}) \geqslant \mathcal{I}(\mathcal{K})$ due to *super-additivity*.
2. Let $\mathcal{I}$ satisfy *independence* and let c be safe in $\mathcal{K}$. By Proposition 3, c is also free in $\mathcal{K}$ and it follows $\mathcal{I}(\mathcal{K} \setminus \{c\}) = \mathcal{I}(\mathcal{K})$ by *independence* and, hence, $\mathcal{I}$ satisfies *weak independence*.
3. Let $\mathcal{I}$ satisfy *MI-separability* and let c be free in $\mathcal{K}$. Observe that $\mathsf{MI}(\{c\}) = \emptyset$ as c is normal. Then it also holds that $\mathsf{MI}(\mathcal{K}) = \mathsf{MI}(\mathcal{K} \setminus \{c\}) = \mathsf{MI}(\mathcal{K} \setminus \{c\}) \cup \mathsf{MI}(\{c\})$ and $\mathsf{MI}(\mathcal{K} \setminus \{c\}) \cap \mathsf{MI}(\{c\}) = \emptyset$. By *MI-separability* it follows that $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K} \setminus \{c\}) + \mathcal{I}(\{c\}) = \mathcal{I}(\mathcal{K} \setminus \{c\})$.
4. Let $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$ be a minimal inconsistent subset of $\mathcal{K}$. Then it holds that $\mathcal{M} \subseteq \mathcal{K} \subseteq \mathcal{K}'$. Suppose $\mathcal{M} \notin \mathsf{MI}(\mathcal{K}')$ which is equivalent to stating that either $\mathcal{M}$ is not minimal or not inconsistent. Both cases contradict the assumption $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$.
5. Let $\mathcal{K}'' = \bigcup_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \mathcal{M}$. It holds that $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K}'')$ due to the facts that $\mathcal{K} \setminus \mathcal{K}''$ only contains free conditionals of $\mathcal{K}$ and that $\mathcal{I}$ satisfies *independence*. As the same is true for $\mathcal{K}'$ it follows $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K}')$.
6. Let $\mathcal{K}'' = \bigcup_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \mathcal{M}$. It holds that $\mathcal{I}(\mathcal{K}) = \mathcal{I}(\mathcal{K}'')$ due to the facts that $\mathcal{K} \setminus \mathcal{K}''$ only contains free conditionals of $\mathcal{K}$ and that $\mathcal{I}$ satisfies *independence*. As $\mathcal{K}'' \subsetneq \mathcal{K}'$ due to $\mathsf{MI}(\mathcal{K}) \subsetneq \mathsf{MI}(\mathcal{K}')$ and $\mathcal{K}' \setminus \mathcal{K}$ contains at least one conditional c that is not free in $\mathcal{K}'$—otherwise it would be $\mathsf{MI}(\mathcal{K}) = \mathsf{MI}(\mathcal{K}')$—it follows $\mathcal{I}(\mathcal{K}') > \mathcal{I}(\mathcal{K}'') = \mathcal{I}(\mathcal{K})$ as $\mathcal{I}$ satisfies *penalty*. $\square$

Proposition 5. *Let $c_1 = (\psi_1 \mid \phi_1)[p_1]$ and $c_2 = (\psi_2 \mid \phi_2)[p_2]$ be normal. If $c_1 \models^{pr} c_2$ then $c_2 \equiv^e c_1$.*

Proof. Observe that the set of models $\mathsf{Mod}(\{c\})$ of a probabilistic conditional $c = (\psi \mid \phi)[p]$ can be described via $\mathsf{Mod}(\{c\}) = \{P \in \mathcal{F}(\mathsf{At}) \mid P(\phi) = x, P(\psi\phi) = px, x \in [0,1]\}$. That is, $\mathsf{Mod}(\{c\})$ is the intersection of $\mathcal{F}(\mathsf{At})$ with a hyperplane, i.e. a linear space of dimension $2^{|\mathsf{At}|} - 1$ (note that $\mathcal{F}(\mathsf{At})$ can be embedded in a space of dimension $2^{|\mathsf{At}|}$, one dimension for each probability of an interpretation), see also [23]. In general, there are three different possible relationships between any two hyperplanes in a space of dimension $2^{|\mathsf{At}|}$: they may either be parallel, intersect in a linear space of dimension $2^{|\mathsf{At}|} - 2$, or coincide. For example, two planes in a 3-dimensional space are either parallel, intersect in a line, or are the same. Then $c_1 \models^{pr} c_2$ implies that the hyperplanes corresponding to c_1 and c_2 coincide, otherwise there would be a model of c_1 that is not a model of c_2. It follows $\mathsf{Mod}(\{c_1\}) = \mathsf{Mod}(\{c_2\})$ and the claim. $\square$

Proposition 6. *The function $\mathcal{I}_0$ satisfies* consistency, irrelevance of syntax, monotonicity, weak independence, independence, *and* normalization.

Proof. We only show that $\mathcal{I}_0$ satisfies *consistency*, *irrelevance of syntax*, *monotonicity*, *independence*, and *normalization* as *weak independence* follows from *independence* due to Proposition 4.

Consistency A knowledge base $\mathcal{K}$ is consistent if and only if $\mathcal{I}_0(\mathcal{K}) = 0$ by definition.

Irrelevance of syntax From $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ follows $\mathcal{K}_1 \equiv^e \mathcal{K}_2$ by Proposition 2. Therefore, $\mathcal{K}_1$ is inconsistent if and only if $\mathcal{K}_2$ is inconsistent. It follows $\mathcal{I}_0(\mathcal{K}_1) = \mathcal{I}_0(\mathcal{K}_2)$.

Monotonicity If $\mathcal{K}$ is inconsistent so is any superset of $\mathcal{K}$. It follows $\mathcal{I}_0(\mathcal{K}) = 1 = \mathcal{I}_0(\mathcal{K} \cup \{c\})$. If $\mathcal{K}$ is consistent then $\mathcal{I}_0(\mathcal{K} \cup \{c\}) \geqslant \mathcal{I}_0(\mathcal{K}) = 0$ by definition.

Independence Assume that $\mathcal{K}$ is consistent and c is free in $\mathcal{K} \cup \{c\}$. If $\mathcal{K} \cup \{c\}$ would be inconsistent then for every minimal inconsistent subset $\mathcal{M}$ of $\mathcal{K} \cup \{c\}$ it holds that $c \notin \mathcal{M}$. Hence, $\mathcal{M}$ is also a minimal inconsistent subset of $\mathcal{K}$ rendering $\mathcal{K}$ inconsistent. As $\mathcal{K}$ is consistent it follows that $\mathcal{K} \cup \{c\}$ is consistent and therefore $\mathcal{I}_0(\mathcal{K} \cup \{c\}) = 0 = \mathcal{I}_0(\mathcal{K})$. If $\mathcal{K}$ is inconsistent so is any superset of $\mathcal{K}$ and hence $\mathcal{I}_0(\mathcal{K} \cup \{c\}) = 1 = \mathcal{I}_0(\mathcal{K})$.

Normalization For every $\mathcal{K}$ it holds that either $\mathcal{I}_0(\mathcal{K})=0$ or $\mathcal{I}_0(\mathcal{K})=1$ and therefore $\mathcal{I}_0(\mathcal{K})\in[0,1]$. □

Proposition 7. *The function* $\mathcal{I}_{\mathsf{MI}}$ *satisfies* consistency, monotonicity, super-additivity, weak independence, independence, irrelevance of syntax, MI-separability, *and* penalty.

Proof. We only show that $\mathcal{I}_{\mathsf{MI}}$ satisfies *consistency*, *super-additivity*, *irrelevance of syntax*, MI-*separability*, and *penalty*, as *monotonicity* follows from *super-additivity*, *weak independence* follows from *independence*, and *independence* follows from MI-*separability*, cf. Proposition 4.

Consistency If $\mathcal{K}$ is consistent it follows that $\mathsf{MI}(\mathcal{K})=\emptyset$ and therefore $\mathcal{I}_{\mathsf{MI}}(\mathcal{K})=0$. If $\mathcal{K}$ is inconsistent then $\mathsf{MI}(\mathcal{K})\neq\emptyset$ and $\mathcal{I}_{\mathsf{MI}}(\mathcal{K})>0$.

Super-additivity Let $\mathcal{K}\cap\mathcal{K}'=\emptyset$. Due to Proposition 4 it holds that $\mathsf{MI}(\mathcal{K})\subseteq\mathsf{MI}(\mathcal{K}\cup\mathcal{K}')$ and $\mathsf{MI}(\mathcal{K}')\subseteq\mathsf{MI}(\mathcal{K}\cup\mathcal{K}')$. Due to $\mathcal{K}\cap\mathcal{K}'=\emptyset$ it follows that $\mathsf{MI}(\mathcal{K})\cap\mathsf{MI}(\mathcal{K}')=\emptyset$ and therefore $\mathcal{I}_{\mathsf{MI}}(\mathcal{K}\cup\mathcal{K}')=|\mathsf{MI}(\mathcal{K}\cup\mathcal{K}')|\geqslant|\mathsf{MI}(\mathcal{K})\cup\mathsf{MI}(\mathcal{K}')|=|\mathsf{MI}(\mathcal{K})|+|\mathsf{MI}(\mathcal{K}')|=\mathcal{I}_{\mathsf{MI}}(\mathcal{K}')+\mathcal{I}_{\mathsf{MI}}(\mathcal{K}')$.

Irrelevance of syntax Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be knowledge bases with $\mathcal{K}_1\equiv^s\mathcal{K}_2$ and let $\rho_{\mathcal{K}_1,\mathcal{K}_2}:\mathcal{K}_1\to\mathcal{K}_2$ be a bijection with $c\equiv^e\rho_{\mathcal{K}_1,\mathcal{K}_2}(c)$ for all $c\in\mathcal{K}_1$. Let $\mathcal{C}\subseteq\mathcal{K}_1$ and let

$$\rho_{\mathcal{K}_1,\mathcal{K}_2}(\mathcal{C})=\bigl\{\rho_{\mathcal{K}_1,\mathcal{K}_2}(c)\bigm| c\in\mathcal{C}\bigr\}. \tag{A.1}$$

As $\mathsf{Mod}(c)=\mathsf{Mod}(\rho_{\mathcal{K}_1,\mathcal{K}_2}(c))$ for every $c\in\mathcal{K}_1$ and due to the fact that $\rho_{\mathcal{K}_1,\mathcal{K}_2}$ is a bijection it follows that $\mathcal{M}$ is a minimal inconsistent subset of $\mathcal{K}_1$ if and only if $\rho_{\mathcal{K}_1,\mathcal{K}_2}(\mathcal{M})$ is a minimal inconsistent subset of $\mathcal{K}_2$. Hence, it follows $\mathcal{I}_{\mathsf{MI}}(\mathcal{K}_1)=\mathcal{I}_{\mathsf{MI}}(\mathcal{K}_2)$.

MI-separability Let $\mathcal{K}_1,\mathcal{K}_2$ be knowledge bases with $\mathsf{MI}(\mathcal{K}_1\cup\mathcal{K}_2)=\mathsf{MI}(\mathcal{K}_1)\cup\mathsf{MI}(\mathcal{K}_2)$ and $\mathsf{MI}(\mathcal{K}_1)\cap\mathsf{MI}(\mathcal{K}_2)=\emptyset$. It follows directly that $\mathcal{I}_{\mathsf{MI}}(\mathcal{K}_1\cup\mathcal{K}_2)=|\mathsf{MI}(\mathcal{K}_1\cup\mathcal{K}_2)|=|\mathsf{MI}(\mathcal{K}_1)|+|\mathsf{MI}(\mathcal{K}_2)|=\mathcal{I}_{\mathsf{MI}}(\mathcal{K}_1)+\mathcal{I}_{\mathsf{MI}}(\mathcal{K}_2)$.

Penalty Let $c\notin\mathcal{K}$ be a conditional that is not free in $\mathcal{K}\cup\{c\}$. By the facts that $\mathsf{MI}(\mathcal{K})\subseteq\mathsf{MI}(\mathcal{K}\cup\{c\})$ and that there is an $\mathcal{M}\in\mathsf{MI}(\mathcal{K}\cup\{c\})$ with $c\in\mathcal{M}$ it follows that $|\mathsf{MI}(\mathcal{K})|<|\mathsf{MI}(\mathcal{K}\cup\{c\})|$ and therefore $\mathcal{I}_{\mathsf{MI}}(\mathcal{K})<\mathcal{I}_{\mathsf{MI}}(\mathcal{K}\cup\{c\})$. □

Proposition 8. *The function* $\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}$ *satisfies* consistency, monotonicity, super-additivity, weak independence, independence, irrelevance of syntax, MI-separability, *and* penalty.

Proof. We only show that $\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}$ satisfies *consistency*, *super-additivity*, *irrelevance of syntax*, MI-*separability*, and *penalty* as *monotonicity* follows from *super-additivity*, *weak independence* follows from *independence*, and *independence* follows from MI-*separability*, cf. Proposition 4.

Consistency If $\mathcal{K}$ is consistent it follows that $\mathsf{MI}(\mathcal{K})=\emptyset$ and therefore $\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K})=0$ (the empty sum). If $\mathcal{K}$ is inconsistent then $\mathsf{MI}(\mathcal{K})\neq\emptyset$ with $\mathcal{M}\in\mathsf{MI}(\mathcal{K})$ and $|\mathcal{M}|>0$. It follows that $\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K})>0$.

Super-additivity Let $\mathcal{K}\cap\mathcal{K}'=\emptyset$. Due to Proposition 4 it holds that $\mathsf{MI}(\mathcal{K})\subseteq\mathsf{MI}(\mathcal{K}\cup\mathcal{K}')$ and $\mathsf{MI}(\mathcal{K}')\subseteq\mathsf{MI}(\mathcal{K}\cup\mathcal{K}')$. Due to $\mathcal{K}\cap\mathcal{K}'=\emptyset$ it follows that $\mathsf{MI}(\mathcal{K})\cap\mathsf{MI}(\mathcal{K}')=\emptyset$ and therefore

$$\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}\bigl(\mathcal{K}\cup\mathcal{K}'\bigr)=\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K}\cup\mathcal{K}')}\frac{1}{|\mathcal{M}|}\geqslant\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K})}\frac{1}{|\mathcal{M}|}+\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K}')}\frac{1}{|\mathcal{M}|}=\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K})+\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}\bigl(\mathcal{K}'\bigr).$$

Irrelevance of syntax Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be knowledge bases with $\mathcal{K}_1\equiv^s\mathcal{K}_2$ and let $\rho_{\mathcal{K}_1,\mathcal{K}_2}:\mathcal{K}_1\to\mathcal{K}_2$ be a bijection with $c\equiv^e\rho_{\mathcal{K}_1,\mathcal{K}_2}(c)$ for all $c\in\mathcal{K}_1$. In the proof of Proposition 7 it has already been shown that $\mathcal{M}$ is a minimal inconsistent subset of $\mathcal{K}_1$ if and only if $\rho_{\mathcal{K}_1,\mathcal{K}_2}(\mathcal{M})$ is a minimal inconsistent subset of $\mathcal{K}_2$, cf. the definition of $\rho_{\mathcal{K}_1,\mathcal{K}_2}(\mathcal{M})$ in Eq. (A.1). As $\rho_{\mathcal{K}_1,\mathcal{K}_2}$ is a bijection it also follows that $|\mathcal{M}|=|\rho_{\mathcal{K}_1,\mathcal{K}_2}(\mathcal{M})|$ and hence

$$\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K}_2)=\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K}_2)}\frac{1}{|\mathcal{M}|}=\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K}_1)}\frac{1}{|\rho_{\mathcal{K}_1,\mathcal{K}_2}(\mathcal{M})|}=\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K}_1)}\frac{1}{|\mathcal{M}|}=\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K}_1).$$

MI-separability Let $\mathcal{K}_1,\mathcal{K}_2$ be knowledge bases with $\mathsf{MI}(\mathcal{K}_1\cup\mathcal{K}_2)=\mathsf{MI}(\mathcal{K}_1)\cup\mathsf{MI}(\mathcal{K}_2)$ and $\mathsf{MI}(\mathcal{K}_1)\cap\mathsf{MI}(\mathcal{K}_2)=\emptyset$. It follows directly that

$$\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K}_1\cup\mathcal{K}_2)=\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K}_1\cup\mathcal{K}_2)}\frac{1}{|\mathcal{M}|}=\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K}_1)}\frac{1}{|\mathcal{M}|}+\sum_{\mathcal{M}\in\mathsf{MI}(\mathcal{K}_2)}\frac{1}{|\mathcal{M}|}=\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K}_1)+\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K}_2).$$

Penalty Let $c\notin\mathcal{K}$ be a conditional that is not free in $\mathcal{K}\cup\{c\}$. By the facts that $\mathsf{MI}(\mathcal{K})\subseteq\mathsf{MI}(\mathcal{K}\cup\{c\})$ and that there is an $\mathcal{M}\in\mathsf{MI}(\mathcal{K}\cup\{c\})$ with $c\in\mathcal{M}$ it follows that $|\mathsf{MI}(\mathcal{K})|<|\mathsf{MI}(\mathcal{K}\cup\{c\})|$ and therefore $\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K})<\mathcal{I}^{\mathsf{C}}_{\mathsf{MI}}(\mathcal{K}\cup\{c\})$. □

Proposition 9. *If* $\mathsf{MI}(\mathcal{K})=\{\mathcal{K}\}$ *then* $\mathcal{I}_\eta(\mathcal{K})=1/|\mathcal{K}|$.

Proof. Let $\mathcal{K} = \langle c_1, \dots, c_n \rangle$ and let $\mathcal{K}_1, \dots, \mathcal{K}_n$ be defined via $\mathcal{K}_i = \mathcal{K} \setminus \{c_i\}$ for $i = 1, \dots, n$. Each $\mathcal{K}_i$ for $i = 1, \dots, n$ is consistent as $\mathcal{K}$ is minimally inconsistent. Therefore, let $P_1, \dots, P_n$ be probability functions with $P_i \models^{pr} \mathcal{K}_i$ for $i = 1, \dots, n$. Define $\hat{P}$ through $\hat{P}(P_i) = 1/n$ and $\hat{P}(P) = 0$ for all $P \in \mathcal{F}(\mathsf{At})$ with $P \notin \{P_1, \dots, P_n\}$. Note that every c_i is contained in every $\mathcal{K}_j$ with $j \neq i$. Therefore, all probability functions P_j with $j \neq i$ satisfy c_i and it follows

$$\hat{P}(c_i) = \hat{P}(P_1) + \dots + \hat{P}(P_{i-1}) + \hat{P}(P_{i+1}) + \dots + \hat{P}(P_n) = \frac{n-1}{n} = 1 - \frac{1}{n}.$$

It follows that $\hat{P}(c_i) = 1 - 1/n$ for every $i = 1, \dots, n$ and, hence, $\mathcal{I}_\eta(\mathcal{K}) \geqslant (1 - 1/n)$. It is also easy to see that there can be no $\hat{P}'$ with $\hat{P}'(c_i) > 1 - 1/n$ for all $i = 1, \dots, n$, see [18] for details. It follows $\mathcal{I}_\eta(\mathcal{K}) = (1 - 1/n)$. □

Proposition 10. *The function $\mathcal{I}_\eta$ satisfies* consistency, monotonicity, weak independence, independence, irrelevance of syntax *and* normalization.

Proof. We only show that $\mathcal{I}_\eta$ satisfies *consistency*, *monotonicity*, *independence*, *irrelevance of syntax* and *normalization* as *weak independence* follows from *independence*, cf. Proposition 4.

Consistency Let $\mathcal{K}$ be consistent and P be a probability function with $P \models^{pr} \mathcal{K}$. Define $\hat{P}_P$ via $\hat{P}_P(P) = 1$ and $\hat{P}_P(P') = 0$ for all $P' \in \mathcal{F}(\mathsf{At})$ with $P' \neq P$. It follows that $\hat{P}_P(c) = 1$ for every $c \in \mathcal{K}$ and due to *normalization* it follows $\mathcal{I}_\eta(\mathcal{K}) = 1 - 1 = 0$. If $\mathcal{K}$ is inconsistent there can be no $\hat{P}$ with $\hat{P}(c) = 1$ for every $c \in \mathcal{K}$ because otherwise every P with $\hat{P}(P) > 0$ would obey $P \models^{pr} \mathcal{K}$. Therefore $\max\{\eta \mid \exists \hat{P} : \forall c \in \mathcal{K}\colon\ \hat{P}(c) \geqslant \eta\} < 1$ and $\mathcal{I}_\eta(\mathcal{K}) > 0$.

Monotonicity Let $\mathcal{K}$ be a knowledge base, c a conditional and $\mathcal{K}' = \mathcal{K} \cup \{c\}$. Let $\hat{P} \in \mathcal{F}^2(\mathsf{At})$ be a probability function and $\eta' \in [0, 1]$ be such that

$$\mathcal{I}_\eta\big(\mathcal{K}'\big) = 1 - \eta' \quad \text{and} \quad \eta' = \max\big\{\eta \bigm| \forall c \in \mathcal{K}'\colon\ \hat{P}(c) \geqslant \eta\big\}.$$

In particular, it holds that $\hat{P}(c) \geqslant \eta'$ for all $c \in \mathcal{K}$ and therefore

$$\mathcal{I}_\eta(\mathcal{K}) = 1 - \max\big\{\eta \bigm| \forall c \in \mathcal{K}\colon\ \hat{P}(c) \geqslant \eta\big\} \leqslant 1 - \eta' = \mathcal{I}_\eta\big(\mathcal{K}'\big).$$

Independence Let $\mathcal{K}$ be a knowledge base and let $c \in \mathcal{K}$ be free in $\mathcal{K}$. Due to monotonicity it follows $\mathcal{I}_\eta(\mathcal{K}) \geqslant \mathcal{I}_\eta(\mathcal{K} \setminus \{c\})$. The proof of $\mathcal{I}_\eta(\mathcal{K}) \leqslant \mathcal{I}_\eta(\mathcal{K} \setminus \{c\})$ is analogous to the proof of Corollary 2.20 in [18].

Irrelevance of syntax Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be knowledge bases with $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ and let $\rho_{\mathcal{K}_1,\mathcal{K}_2} : \mathcal{K}_1 \to \mathcal{K}_2$ be a bijection with $c \equiv^e \rho_{\mathcal{K}_1,\mathcal{K}_2}(c)$ for all $c \in \mathcal{K}_1$. As $\mathsf{Mod}(c) = \mathsf{Mod}(\rho_{\mathcal{K}_1,\mathcal{K}_2}(c))$ for all $c \in \mathcal{K}_1$ it follows that $\hat{P}(c) = \hat{P}(\rho_{\mathcal{K}_1,\mathcal{K}_2}(c))$ for every $\hat{P} \in \mathcal{F}^2(\mathsf{At})$ and therefore $\mathcal{I}_\eta(\mathcal{K}_1) = \mathcal{I}_\eta(\mathcal{K}_2)$.

Normalization For every $\hat{P} : \mathcal{F}(\mathsf{At}) \to [0, 1]$ and probabilistic conditional c it holds that $\hat{P}(c) \in [0, 1]$ as $\hat{P}$ is a probability function. It follows that $\max\{\eta \mid \exists \hat{P}\colon\ \forall c \in \mathcal{K}\colon\ \hat{P}(c) \geqslant \eta\} \in [0, 1]$ and therefore $\mathcal{I}_\eta(\mathcal{K}) \in [0, 1]$. □

Theorem 1. *If D is continuously generating the function $\mathcal{I}_D$ is well-defined.*

Proof. Let $\mathcal{K} = \langle c_1, \dots, c_n \rangle$ be a knowledge base with $c_i = (\psi_i \mid \phi_i)[p_i]$ for $i = 1, \dots, n$ and $d = D(n)$. Consider the set $\mathcal{P}_\mathcal{K} \subseteq \mathcal{F}(\mathsf{At}) \times [0, 1]^n$ defined via

$$\mathcal{P}_\mathcal{K} = \big\{\big\langle P, \langle x_1, \dots, x_n \rangle\big\rangle \in \mathcal{F}(\mathsf{At}) \times [0, 1]^n \bigm| P \models^{pr} \Lambda_\mathcal{K}(x_1, \dots, x_n)\big\}.$$

We show now that $\mathcal{P}_\mathcal{K}$ is a closed set. Let $\langle P_i, \langle x_1^i, \dots, x_n^i \rangle\rangle \in \mathcal{P}_\mathcal{K}$ for $i \in \mathbb{N}$ be such that $\lim_{i\to\infty}(P_i, (x_1^i, \dots, x_n^i))$ exists and define

$$\big\langle Q, \langle y_1, \dots, y_n \rangle\big\rangle = \lim_{i\to\infty} \big\langle P_i, \big\langle x_1^i, \dots, x_n^i \big\rangle\big\rangle.$$

In particular, it holds that $\lim_{i\to\infty} P_i = Q$ with $Q \in \mathcal{F}(\mathsf{At})$.[2] For $j = 1, \dots, n$, if $Q(\phi_j) > 0$ then there is some $k \in \mathbb{N}$ such that for all $i > k$ it holds that $P_i(\phi_j) > 0$ as well. Therefore, for $i > k$ it holds that $P_i(\psi_j \mid \phi_j) = x_j^i$ and

$$Q(\psi_j \mid \phi_j) = \frac{Q(\psi_j \phi_j)}{Q(\phi_j)} = \frac{\lim_{i\to\infty} P_i(\psi_j \phi_j)}{\lim_{i\to\infty} P_i(\phi_j)} = \lim_{i\to\infty} \frac{P_i(\psi_j \phi_j)}{P_i(\phi_j)} = \lim_{i\to\infty} P_i(\psi_j \mid \phi_j) = \lim_{i\to\infty} x_j^i = y_j$$

which implies $Q \models^{pr} (\psi_j \mid \phi_j)[y_j]$. Furthermore, for $j = 1, \dots, n$, if $Q(\phi_j) = 0$ then trivially $Q \models^{pr} (\psi_j \mid \phi_j)[y_j]$ due to our definition of probabilistic satisfaction. It follows that $Q \models^{pr} \Lambda_\mathcal{K}(y_1, \dots, y_n)$ and therefore $Q \in \mathcal{P}_\mathcal{K}$, i.e., $\mathcal{P}_\mathcal{K}$ is closed. Consider now the projection $\rho : \mathcal{P}_\mathcal{K} \to [0, 1]^n$ defined via $\rho(\langle P, \langle x_1, \dots, x_n \rangle\rangle) = \langle x_1, \dots, x_n \rangle$ for $\langle P, \langle x_1, \dots, x_n \rangle\rangle \in \mathcal{P}_\mathcal{K}$. As

[2] Note that the set $\mathcal{F}(\mathsf{At})$ is a closed set, see e.g. [36].

$\mathcal{F}(\mathrm{At})$ is compact—see e.g. [36]—it follows that ρ is a *closed map*, cf. the Tube Lemma[3] [22]. Therefore, ρ maps closed sets to closed sets and it follows that

$$\rho(\mathcal{P}_{\mathcal{K}}) = \bigl\{\langle x_1,\dots,x_n\rangle \in [0,1]^n \bigm| \exists P\colon \bigl\langle P, \langle x_1,\dots,x_n\rangle\bigr\rangle \in \mathcal{P}_{\mathcal{K}}\bigr\} = \bigl\{\vec{x} \in [0,1]^n \bigm| \mathcal{K}[\vec{x}] \text{ is consistent}\bigr\}$$

is a closed set. Observe that we can write $\mathcal{I}_D(\mathcal{K})$ as

$$\mathcal{I}_D(\mathcal{K}) = \min\bigl\{d\bigl(\langle p_1,\dots,p_n\rangle, \langle x_1,\dots,x_n\rangle\bigr) \bigm| \langle x_1,\dots,x_n\rangle \in \rho(\mathcal{P}_{\mathcal{K}})\bigr\}.$$

As $\rho(\mathcal{P}_{\mathcal{K}})$ is a closed set—and also compact as it is bounded due to $\rho(\mathcal{P}_{\mathcal{K}}) \subseteq [0,1]^n$—and the mapping $\langle x_1,\dots,x_n\rangle \mapsto d(\langle p_1,\dots,p_n\rangle, \langle x_1,\dots,x_n\rangle)$ is continuous—as d is a continuous function and $d_1,\dots,d_n$ are constants—the set

$$N^d_{\mathcal{K}} = \bigl\{d\bigl(\langle p_1,\dots,p_n\rangle, \langle x_1,\dots,x_n\rangle\bigr) \bigm| \langle x_1,\dots,x_n\rangle \in \rho(\mathcal{P}_{\mathcal{K}})\bigr\}$$

is closed as well. Note that $\rho(\mathcal{P}_{\mathcal{K}})$ and therefore $N^d_{\mathcal{K}}$ are non-empty as for every $\mathcal{K}$ there is always an $\vec{x}$ such that $\mathcal{K}[\vec{x}]$ is consistent (take an arbitrary positive probability function P and define $x_i = P(\psi_i \mid \phi_i)$, see also [34]). It follows that $\mathcal{I}_D(\mathcal{K}) = \min N^d_{\mathcal{K}}$ is well-defined. $\square$

Proposition 11. *The function $\mathcal{I}_{D^0}$ is well-defined and it holds that $\mathcal{I}_{D^0} = \mathcal{I}_0$.*

Proof. Let $\mathcal{K} = \mathcal{K}[\vec{x}]$ for some $\vec{x} \in [0,1]^n$ be a consistent knowledge base and let $d^0 = D^0(n)$. Then clearly $\mathcal{I}_{D^0}(\mathcal{K}) = 0 = \mathcal{I}_0(\mathcal{K})$ as $d^0(\vec{x},\vec{x}) = 0$ is minimal and $\mathcal{K}[\vec{x}]$ is consistent. Let $\mathcal{K} = \mathcal{K}[\vec{x}]$ for some $\vec{x} \in [0,1]^n$ be an inconsistent knowledge base. As noted in the proof of Theorem 1 there is a $\vec{y} \in [0,1]^n$ such that $\mathcal{K}[\vec{y}]$ is consistent. It follows that $\mathcal{I}_{D^0}(\mathcal{K}) \leqslant d^0(\vec{x},\vec{y}) = 1$ and $\mathcal{I}_{D^0}(\mathcal{K}) > d^0(\vec{x},\vec{x}) = 0$ as $\mathcal{K}[\vec{x}]$ is inconsistent. Due to $\operatorname{Im} d^0 = \{0,1\}$ it follows $\mathcal{I}_{D^0}(\mathcal{K}) = 1 = \mathcal{I}_0(\mathcal{K})$ and therefore $\mathcal{I}_{D^0} = \mathcal{I}_0$. As $\mathcal{I}_0$ is well-defined so is $\mathcal{I}_{D^0}$. $\square$

Theorem 2. *Let D be a distance generator such that $\mathcal{I}_D$ is well-defined.*

1. *The function $\mathcal{I}_D$ satisfies* consistency.
2. *If D is monotonously generating then $\mathcal{I}_D$ satisfies* monotonicity.
3. *If D is super-additively generating then $\mathcal{I}_D$ satisfies* super-additivity.
4. *If D is symmetric generating then $\mathcal{I}_D$ satisfies* irrelevance of syntax.
5. *If D is continuously generating then $\mathcal{I}_D$ satisfies* continuity.

Proof. Let $\mathcal{K} = \langle c_1,\dots,c_n\rangle \in \mathbb{K}$ be a knowledge base with $c_i = (\psi_i \mid \phi_i)[p_i]$ for $i = 1,\dots,n$, $d = D(n)$, and define $\Theta_{\mathcal{K}} = \{\vec{y} \mid \mathcal{K}[\vec{y}] \text{ is consistent}\}$.

1. If $\mathcal{K} = \mathcal{K}[\vec{x}]$ is consistent then due to $d(\vec{x},\vec{x}) = 0$ and $d(\vec{x},\vec{y}) \geqslant 0$ for all $\vec{y} \in [0,1]^{|\mathcal{K}|}$ it follows $\mathcal{I}_D(\mathcal{K}) = 0$.
2. Without loss of generality we only show that $\mathcal{I}_D(\mathcal{K}) \geqslant \mathcal{I}_D(\mathcal{K} \setminus \{c_n\})$. First, note that if $\mathcal{K}' = \langle c_1,\dots,c_{n-1}\rangle[\langle y_1,\dots,y_{n-1}\rangle]$ is consistent there is a $y_n \in [0,1]$ such that $\langle c_1,\dots,c_n\rangle[\langle y_1,\dots,y_n\rangle]$ is consistent (by taking some model P of $\mathcal{K}'$ and defining $y_n = P(\psi_n \mid \phi_n)$; the latter is defined as c_n is normal). Furthermore, if $\langle c_1,\dots,c_n\rangle[\langle y_1,\dots,y_n\rangle]$ is consistent so is $\langle c_1,\dots,c_{n-1}\rangle[\langle y_1,\dots,y_{n-1}\rangle]$. It follows that $\langle y_1,\dots,y_{n-1}\rangle \in \Theta_{\mathcal{K}'}$ if and only if there is a $y_n \in [0,1]$ such that $\langle y_1,\dots,y_n\rangle \in \Theta_{\mathcal{K}}$. Let now $\mathcal{K} = \mathcal{K}[\vec{x}]$ for some $\vec{x} = \langle x_1,\dots,x_n\rangle \in [0,1]$ and $\langle y_1,\dots,y_{n-1}\rangle \in \Theta_{\mathcal{K}'}$. Then for every $y_n \in [0,1]$ such that $\langle y_1,\dots,y_n\rangle \in \Theta_{\mathcal{K}}$ it holds that
$$d\bigl(\langle x_1,\dots,x_{n-1}\rangle, \langle y_1,\dots,y_{n-1}\rangle\bigr) \leqslant d\bigl(\langle x_1,\dots,x_n\rangle, \langle y_1,\dots,y_n\rangle\bigr)$$
as D is monotonously generating. It follows that every element of $M_1 = \{d(\vec{x},\vec{y}) \mid \vec{y} \in \Theta_{\mathcal{K}}\}$ is greater or equal to an element in $M_2 = \{d(\langle x_1,\dots,x_{n-1}\rangle, \vec{y}) \mid \vec{y} \in \Theta_{\mathcal{K}'}\}$. Consequently, $\mathcal{I}_D(\mathcal{K}') = \min M_2 \leqslant \min M_1 = \mathcal{I}_D(\mathcal{K})$ proving *monotonicity*.
3. This proof is analogous to the proof of (2).
4. Let $\mathcal{K}_1 = \mathcal{K}_1[\vec{x_1}]$ and $\mathcal{K}_2 = \mathcal{K}_2[\vec{x_2}]$ be knowledge bases with $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ and $\vec{x_1} = \langle x^1_1,\dots,x^n_1\rangle$ and $\vec{x_2} = \langle x^1_2,\dots,x^n_2\rangle$ (with $|\mathcal{K}_1| = n = |\mathcal{K}_2|$). Let $\mathcal{K}'_1 = \mathcal{K}_1[\vec{y_1}]$ be consistent such that $\mathcal{I}_D(\mathcal{K}_1) = d(\vec{x_1},\vec{y_1})$ for some $\vec{y_1} = \langle y^1_1,\dots,y^n_1\rangle$. In Proposition 5 it has been shown that for normal $c = (\psi \mid \phi)[p]$ and $c' = (\psi' \mid \phi')[p']$ with $c \equiv^e c'$ it holds that either
(a) $\phi \equiv \phi'$ and $\psi \wedge \phi \equiv \psi' \wedge \phi'$ and $p = p'$ or
(b) $\phi \equiv \phi'$ and $\psi \wedge \phi \equiv \overline{\psi'} \wedge \phi'$ and $p = 1 - p'$.
Define now $\vec{y_2} = \langle y^1_2,\dots,y^n_2\rangle$ via $y^i_2 = y^i_1$ if $x^i_2 = x^i_1$ and via $y^i_2 = 1 - y^i_1$ if $x^i_2 = 1 - x^i_1$ (for $i = 1,\dots,n$). As D is symmetric generating, by iteratively applying (5) it follows that $d(\vec{x_2},\vec{y_2}) = d(\vec{x_1},\vec{y_1})$. Note also that, by construction,

[3] An equivalent formalization of the Tube Lemma is "If X is Hausdorff and Y is Hausdorff and compact then $p: X \times Y \to X$ with $p(x,y) = x$ is a closed map". Note, that all spaces above are Hausdorff as they are subsets of Euclidean spaces (or can be characterized as such).

$\mathcal{K}_2[\vec{y}_2]$ is consistent as $\mathcal{K}_2[\vec{y}_2] \equiv^s \mathcal{K}_1[\vec{y}_1]$. It follows that $\mathcal{I}_D(\mathcal{K}_2) \leqslant d(\vec{x}_2, \vec{y}_2) = d(\vec{x}_1, \vec{y}_1) = \mathcal{I}_D(\mathcal{K}_1)$. Similarly we obtain $\mathcal{I}_D(\mathcal{K}_1) \leqslant \mathcal{I}_D(\mathcal{K}_2)$ and therefore the claim.
5. Let $\vec{x} \in [0,1]^{|\mathcal{K}|}$. For every $\vec{y} \in \Theta_{\mathcal{K}}$ the mapping $\vec{x} \mapsto d(\vec{x}, \vec{y})$ is continuous as D is continuously generating. As the minimum of a set of continuous functions is continuous it follows that the mapping $\vec{x} \mapsto \mathcal{I}_d(\mathcal{K}[\vec{x}])$ is continuous as well. $\square$

Theorem 3. *Let $p \in \mathbb{N}^+$.*

1. *The function $\mathcal{I}_p$ satisfies* consistency, monotonicity, weak independence, independence, irrelevance of syntax, *and* continuity.
2. *If $p = 1$ then $\mathcal{I}_p$ satisfies* super-additivity.

Proof.

1. It has already been noted that D^p is continuously generating and therefore $\mathcal{I}_p$ is well-defined. By Theorem 2 it also follows that $\mathcal{I}_p$ satisfies *consistency* and *continuity*. We continue with showing that D^p is also monotonously and symmetric generating. Let $x_1, \ldots, x_{n+1}, y_1, \ldots, y_{n+1} \in \mathbb{R}^+$ for some $n \in \mathbb{N}^+$.

$$\begin{aligned} D^p(n)\big(\langle x_1, \ldots, x_n\rangle, \langle y_1, \ldots, y_n\rangle\big) &= \sqrt[p]{|x_1 - y_1|^p + \cdots + |x_n - y_n|^p} \\ &\leqslant \sqrt[p]{|x_1 - y_1|^p + \cdots + |x_n - y_n|^p + |x_{n+1} - y_{n+1}|^p} \\ &= D^p(n+1)\big(\langle x_1, \ldots, x_{n+1}\rangle, \langle y_1, \ldots, y_{n+1}\rangle\big) \end{aligned}$$

as $|x_{n+1} - y_{n+1}|^p \geqslant 0$ and the root function is monotonous. Let $i \in \{1, \ldots, n\}$.

$$\begin{aligned} D^p(n)\big(\langle x_1, \ldots, x_n\rangle, \langle y_1, \ldots, y_n\rangle\big) &= \sqrt[p]{|x_1 - y_1|^p + \cdots + |x_i - y_i|^p + \cdots + |x_n - y_n|^p} \\ &= \sqrt[p]{|x_1 - y_1|^p + \cdots + |1 - x_i - (1 - y_i)|^p + \cdots + |x_n - y_n|^p} \\ &= D^p(n)\big(\langle x_1, \ldots, 1 - x_i, \ldots, x_n\rangle, \langle y_1, \ldots, 1 - y_i, \ldots, y_n\rangle\big). \end{aligned}$$

By Theorem 2 it follows that $\mathcal{I}_p$ satisfies *monotonicity* and *irrelevance of syntax*. It remains to show that $\mathcal{I}_p$ satisfies *weak independence* and *independence*. Before proving *independence* we first show that from both $\mathcal{K} \cup \{(\psi \mid \phi)[p_1]\}$ and $\mathcal{K} \cup \{(\psi \mid \phi)[p_2]\}$ being consistent for some knowledge base $\mathcal{K}$ and $p_1 \leqslant p_2$ it follows that $\mathcal{K} \cup \{(\psi \mid \phi)[y]\}$ is consistent for every $y \in [0,1]$ that satisfies $p_1 \leqslant y \leqslant p_2$. Let $P_1 \models^{pr} \mathcal{K} \cup \{(\psi \mid \phi)[p_1]\}$ and let $P_2 \models^{pr} \mathcal{K} \cup \{(\psi \mid \phi)[p_2]\}$. If $P_1(\phi) = 0$ then clearly $P_1 \models^{pr} \mathcal{K} \cup \{(\psi \mid \phi)[y]\}$ for every $y \in [0,1]$ due to our definition of probabilistic satisfaction. If $P_2(\phi) = 0$ then $P_2 \models^{pr} \mathcal{K} \cup \{(\psi \mid \phi)[y]\}$ for every $y \in [0,1]$ accordingly. So assume $P_1(\phi) > 0$ and $P_2(\phi) > 0$. Let $\delta \in [0,1]$ and consider the probability function P_δ defined via $P_\delta(\omega) = \delta P_1(\omega) + (1 - \delta) P_2(\omega)$ for all $\omega \in \Omega(\mathsf{At})$. Then $P_\delta \models^{pr} \mathcal{K}$ for all $\delta \in [0,1]$ as the set of models of a knowledge base is a convex set, cf. [23]. Furthermore, note that $P_\delta(\phi) > 0$ for every $\delta \in [0,1]$ as both $P_1(\phi) > 0$ and $P_2(\phi) > 0$. Then $P_\delta(\psi \mid \phi)$ is continuous in δ and for every $y \in [p_1, p_2]$ there is a $\delta_y \in [0,1]$ such that $P_{\delta_y}(\psi \mid \phi) = y$. It follows that $P_{\delta_y} \models^{pr} \mathcal{K} \cup \{(\psi \mid \phi)[y]\}$ for every $y \in [p_1, p_2]$ and therefore $\mathcal{K} \cup \{(\psi \mid \phi)[y]\}$ is consistent for every $y \in [p_1, p_2]$.

   Let now $\mathcal{K} = \langle c_1, \ldots, c_n\rangle$ and $c_i = (\psi_i \mid \phi_i)[p_i]$ for $i = 1, \ldots, n$ be a knowledge base and let $c = (\psi \mid \phi)[p]$ be free in $\mathcal{K} \cup \{c\}$. Assume that $\mathcal{K}$ is also a minimal inconsistent set, i.e. $\mathsf{MI}(\mathcal{K}) = \{\mathcal{K}\}$. Let $\mathcal{I}_p(\mathcal{K}) = x$ and let $\langle x_1, \ldots, x_n\rangle \in [0,1]^n$ be such that $\Lambda_{\mathcal{K}}(x_1, \ldots, x_n)$ is consistent and $|p_1 - x_1| + \cdots + |p_n - x_n| = x$. Consider now $\mathcal{K}' = \langle (\psi_1 \mid \phi_1)[p_1], \ldots, (\psi_n \mid \phi_n)[p_n], (\psi \mid \phi)[p]\rangle$. It suffices to show that $\Lambda_{\mathcal{K}'}(x_1, \ldots, x_n, p)$ is consistent. Define $C_j = \mathcal{K} \setminus \{(\psi_j \mid \phi_j)[p_j]\}$ for every $j = 1, \ldots, n$. Then both C_j and $C_j \cup \{c\}$ are consistent. Let p_j be such that there is a P with $P \models^{pr} C_j \cup \{c\}$, $P \models^{pr} (\psi_j \mid \phi_j)[p'_j]$ and $|p_j - p'_j|$ is minimal. It follows that $|p_j - p'_j| \geqslant x$ (otherwise this would contradict $\mathcal{I}_p(\mathcal{K}) = x$). Assume w.l.o.g. $p'_j > p_j$. As $\{c, c_j\}$ is consistent as well (as c is free) it follows that $\{c, (\psi_j \mid \phi_j)[y]\}$ is consistent for every $y \in [p_j, p'_j]$ due to our elaboration above. As $|p_j - x_j| \leqslant x$ it follows $x_j \in [p_j, p'_j]$ as well (or $x_j \in [p'_j, p_j]$ if $p_j > p'_j$). Hence, $\{c, (\psi_j \mid \phi_j)[x_j]\}$ is consistent for every $j = 1, \ldots, n$. As $\Lambda_{\mathcal{K}}(x_1, \ldots, x_n)$ is consistent and c is consistent with every combination of conditionals in $\Lambda_{\mathcal{K}}(x_1, \ldots, x_n)$ it follows that $\Lambda_{\mathcal{K}'}(x_1, \ldots, x_n, p)$ is consistent. The above can be generalized if $\mathcal{K}$ contains multiple minimal inconsistent subsets by iteratively considering each minimal inconsistent subset of $\mathcal{K}$. By Proposition 4 it also follows that $\mathcal{I}_p$ satisfies *weak independence*.
2. Due to Theorem 2 it suffices to show that D^1 is super-additively generating. Let $n, m \in \mathbb{N}^+$ and $x_1, \ldots, x_{n+m}, y_1, \ldots, y_{n+m} \in \mathbb{R}$.

$$\begin{aligned} &D^1(n)\big(\langle x_1, \ldots, x_n\rangle, \langle y_1, \ldots, y_n\rangle\big) + D^1(m)\big(\langle x_{n+1}, \ldots, x_{n+m}\rangle, \langle y_{n+1}, \ldots, y_{n+m}\rangle\big) \\ &\quad = |x_1 - y_1| + \cdots + |x_n - y_n| + |x_{n+1} - y_{n+1}| + \cdots + |x_{n+m} - y_{n+m}| \\ &\quad = D^1(n+m)\big(\langle x_1, \ldots, x_{n+m}\rangle, \langle y_1, \ldots, y_{n+m}\rangle\big). \qquad \square \end{aligned}$$

Theorem 4. *Let $\mathcal{I}$ be an inconsistency measure.*

1. $\mathcal{I}_\Sigma^\mathcal{I}$ *satisfies* monotonicity, super-additivity, weak independence, independence, *and* MI-separability.
2. *If $\mathcal{I}$ satisfies* consistency *then* $\mathcal{I}_\Sigma^\mathcal{I}$ *satisfies* consistency *and* penalty.
3. *If $\mathcal{I}$ satisfies* irrelevance of syntax *then* $\mathcal{I}_\Sigma^\mathcal{I}$ *satisfies* irrelevance of syntax.
4. *If $\mathcal{I}$ satisfies* continuity *then* $\mathcal{I}_\Sigma^\mathcal{I}$ *satisfies* continuity.

Proof.

1. We first show that $\mathcal{I}_\Sigma^\mathcal{I}$ satisfies *super-additivity*. If $\mathcal{K} \cap \mathcal{K}' = \emptyset$ then it holds that $\mathsf{MI}(\mathcal{K}) \cap \mathsf{MI}(\mathcal{K}') = \emptyset$ as well. Due to Proposition 4 it follows that $\mathsf{MI}(\mathcal{K}) \cup \mathsf{MI}(\mathcal{K}') \subseteq \mathsf{MI}(\mathcal{K} \cup \mathcal{K}')$. It follows
$$\mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K} \cup \mathcal{K}') = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K} \cup \mathcal{K}')} \mathcal{I}(\mathcal{M}) \geqslant \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \mathcal{I}(\mathcal{M}) + \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K}')} \mathcal{I}(\mathcal{M}) = \mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K}) + \mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K}').$$
Due Proposition 4 it also follows that $\mathcal{I}_\Sigma^\mathcal{I}$ satisfies *monotonicity*. We now show that $\mathcal{I}_\Sigma^\mathcal{I}$ satisfies *MI-separability*. Let $\mathsf{MI}(\mathcal{K} \cup \mathcal{K}') = \mathsf{MI}(\mathcal{K}) \cup \mathsf{MI}(\mathcal{K}')$ and $\mathsf{MI}(\mathcal{K}) \cap \mathsf{MI}(\mathcal{K}') = \emptyset$. Then clearly
$$\mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K} \cup \mathcal{K}') = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K} \cup \mathcal{K}')} \mathcal{I}(\mathcal{M}) = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \mathcal{I}(\mathcal{M}) + \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K}')} \mathcal{I}(\mathcal{M}) = \mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K}) + \mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K}').$$
Due to Proposition 4 it also follows that $\mathcal{I}_\Sigma^\mathcal{I}$ satisfies *independence* and *weak independence*.
2. We first show that $\mathcal{I}_\Sigma^\mathcal{I}$ satisfies *consistency*. If $\mathcal{K}$ is consistent then $\mathsf{MI}(\mathcal{K}) = \emptyset$ and $\mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K}) = 0$. If $\mathcal{K}$ is inconsistent then there is an $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$ and as $\mathcal{I}$ satisfies *consistency* it follows that $\mathcal{I}(\mathcal{M}) > 0$. Hence, $\mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K}) > 0$ as well. We now show that $\mathcal{I}_\Sigma^\mathcal{I}$ satisfies *penalty*. Let $c \in \mathcal{K}$ be a probabilistic conditional that is not free in $\mathcal{K}$. Due to Proposition 4 it follows that $\mathsf{MI}(\mathcal{K} \setminus \{c\}) \subseteq \mathsf{MI}(\mathcal{K})$. As $c \notin \mathcal{K} \setminus \{c\}$ and there is at least one $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$ with $c \in \mathcal{M}$ it follows that $\mathsf{MI}(\mathcal{K} \setminus \{c\}) \subsetneq \mathsf{MI}(\mathcal{K})$. As $\mathcal{I}$ satisfies *consistency* it follows that $\mathcal{I}(\mathcal{M}) > 0$ and therefore $\mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K} \setminus \{c\}) < \mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K})$.
3. Let it hold that $\mathcal{K}_1 \equiv^s \mathcal{K}_2$. It follows that for every $\mathcal{M} \in \mathsf{MI}(\mathcal{K}_1)$ there is $\mathcal{M}' \in \mathsf{MI}(\mathcal{K}_2)$ with $\mathcal{M} \equiv^s \mathcal{M}'$, and vice versa. As $\mathcal{I}$ satisfies *irrelevance of syntax* it follows that $\mathcal{I}(\mathcal{M}) = \mathcal{I}(\mathcal{M}')$ for every $\mathcal{M} \in \mathsf{MI}(\mathcal{K}_1)$. Hence, it holds that $\mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K}_1) = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K}_1)} \mathcal{I}(\mathcal{M}') = \sum_{\mathcal{M}' \in \mathsf{MI}(\mathcal{K}_2)} \mathcal{I}(\mathcal{M}') = \mathcal{I}_\Sigma^\mathcal{I}(\mathcal{K}_2)$.
4. It is easy to see that $\theta_{\mathcal{I}_\Sigma^\mathcal{I},\mathcal{K}}$ is given via $\theta_{\mathcal{I}_\Sigma^\mathcal{I},\mathcal{K}} = \sum_{\mathcal{M} \in \mathsf{MI}(\mathcal{K})} \theta_{\mathcal{I},\mathcal{M}}$ (given an adequate ordering of the conditionals in $\mathcal{K}$). It follows directly, that $\theta_{\mathcal{I}_\Sigma^\mathcal{I},\mathcal{K}}$ is continuous if $\theta_{\mathcal{I},\mathcal{M}}$ is continuous for every $\mathcal{M} \in \mathsf{MI}(\mathcal{K})$, i.e., if $\mathcal{I}$ satisfies *continuity*. □

Theorem 5. $\mathcal{I}_\mu^h$ *satisfies* irrelevance of syntax *and* weak independence.

Proof.

Irrelevance of syntax Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be knowledge bases with $\mathcal{K}_1 \equiv^s \mathcal{K}_2$ and let $\rho_{\mathcal{K}_1,\mathcal{K}_2} : \mathcal{K}_1 \to \mathcal{K}_2$ be a bijection with $c \equiv^e \rho_{\mathcal{K}_1,\mathcal{K}_2}(c)$ for all $c \in \mathcal{K}_1$. Due to $\mathsf{Mod}(\{c\}) = \mathsf{Mod}(\{\rho_{\mathcal{K}_1,\mathcal{K}_2}(c)\})$ it follows that $d^E(P, \mathsf{Mod}(\{c\})) = d^E(P, \mathsf{Mod}(\{\rho_{\mathcal{K}_1,\mathcal{K}_2}(c)\}))$ and therefore $\mathfrak{C}^h_{\mathcal{K}_1}(P) = \mathfrak{C}^h_{\mathcal{K}_2}(P)$ for every $c \in \mathcal{K}_1$ and $P \in \mathcal{F}(\mathsf{At})$. It follows $\mathcal{I}_\mu^h(\mathcal{K}_1) = \mathcal{I}_\mu^h(\mathcal{K}_2)$.

Weak independence Let $\mathcal{K} = \langle c_1, \ldots, c_n \rangle$ be a knowledge base with $c_i = (\psi_i \mid \phi_i)[p_i]$ for $i = 1, \ldots, n$ and assume w.l.o.g. that c_n is safe in $\mathcal{K}$, i.e. $\mathsf{At}(c_n) \cap \mathsf{At}(\mathcal{K} \setminus \{c_n\}) = \emptyset$. For $\mathcal{B}$ with $\mathsf{At}(\mathcal{K}) \subseteq \mathcal{B}$ let
$$\hat{\Omega}_h^\mathcal{B}(\mathcal{K}) = \bigl\{ P \in \mathcal{F}(\mathcal{B}) \bigm| \mathcal{I}_\mu^h(\mathcal{K}) = 1 - \mathfrak{C}^h_\mathcal{K}(P) \bigr\}.$$
It suffices to show that there is a $P^* \in \hat{\Omega}_h^{\mathsf{At}(\mathcal{K})}(\mathcal{K} \setminus \{c_n\})$ with $P \models^{pr} c_n$ as this implies $\mathfrak{C}^h_\mathcal{K}(P^*) = \mathfrak{C}^h_{\mathcal{K} \setminus \{c_n\}}(P^*)$ (due to $h(0) = 1$) and therefore $\mathcal{I}_\mu^h(\mathcal{K}) \leqslant \mathcal{I}_\mu^h(\mathcal{K} \setminus \{c_n\})$. Together with monotonicity it follows $\mathcal{I}_\mu^h(\mathcal{K}) = \mathcal{I}_\mu^h(\mathcal{K} \setminus \{c_n\})$.
Let $\omega \in \Omega(\mathsf{At})$ and define $\omega_\mathcal{A} \in \Omega(\mathcal{A})$ with $\mathcal{A} \subseteq \mathsf{At}$ to be the projection of ω onto $\mathcal{A}$, e.g. for $\mathsf{At} = \{a, b, c\}$ and $\omega = a \wedge \neg b \wedge c$ it is $\omega_{\{a,b\}} = a \wedge \neg b$. Furthermore, if $P \in \mathcal{F}(\mathsf{At})$ let $P|_\mathcal{A} \in \mathcal{F}(\mathsf{At})$ denote the projection of P onto $\mathcal{A} \subseteq \mathsf{At}$, that is
$$P|_\mathcal{A}(\omega') = \sum_{\omega \in \Omega(\mathsf{At}),\, \omega \models \omega'} P(\omega)$$
for all $\omega' \in \Omega(\mathcal{A})$. In [6] it has been shown that $\mathfrak{C}^h_\mathcal{K}$ is *language invariant*, that is in particular, for every $P \in \mathcal{F}(\mathsf{At} \setminus \mathsf{At}(c_n))$ and every $P' \in \mathcal{F}(\mathsf{At})$ such that $P = P'|_{\mathsf{At} \setminus \mathsf{At}(c_n)}$ it holds that $\mathfrak{C}^h_{\mathcal{K} \setminus \{c_n\}}(P) = \mathfrak{C}^h_{\mathcal{K} \setminus \{c_n\}}(P')$. In other words, as no atom of $\mathsf{At}(c_n)$ is mentioned in $\mathcal{K} \setminus \{c_n\}$ it holds that $\mathfrak{C}^h_{\mathcal{K} \setminus \{c_n\}}(P')$ is the same as $\mathfrak{C}^h_{\mathcal{K} \setminus \{c_n\}}(P)$ if P is the projection of P' onto $\mathsf{At} \setminus \mathsf{At}(c_n)$. In particular, it follows
$$\hat{\Omega}_h^{\mathsf{At}(\mathcal{K})}(\mathcal{K} \setminus \{c_n\}) = \bigl\{ P \bigm| P|_{\mathsf{At}(\mathcal{K}) \setminus \mathsf{At}(c_n)} \in \hat{\Omega}_h^{\mathsf{At}(\mathcal{K}) \setminus \mathsf{At}(c_n)}(\mathcal{K} \setminus \{c_n\}) \bigr\}.$$

Let now $P'' \in \mathcal{F}(\mathsf{At}(c_n))$ with $P'' \models^{pr} c_n$ and $P' \in \hat{\Omega}_h^{\mathsf{At}(\mathcal{K})\setminus\mathsf{At}(c_n)}(\mathcal{K} \setminus \{c_n\})$. Define $P''' \in \mathcal{F}(\mathsf{At}(\mathcal{K}))$ via

$$P'''(\omega) = P''(\omega_{\mathsf{At}(c_n)})P'(\omega_{\mathsf{At}(\mathcal{K})\setminus\mathsf{At}(c_n)}).$$

By construction it holds that $P'''|_{\mathsf{At}(\mathcal{K})\setminus\mathsf{At}(c_n)} = P' \in \hat{\Omega}_h^{\mathsf{At}(\mathcal{K})\setminus\mathsf{At}(c_n)}(\mathcal{K} \setminus \{c_n\})$ and therefore $P''' \in \hat{\Omega}_h^{\mathsf{At}(\mathcal{K})}(\mathcal{K} \setminus \{c_n\})$. As $P''' \models^{pr} c_n$ the claim follows. □

References

[1] A. Ahuja, W. Rödder, Project risk management by a probabilistic expert system, in: Proceedings of the International Conference on Operations Research, 2002, pp. 329–334.
[2] T.J.M. Bench-Capon, P.E. Dunne, Argumentation in artificial intelligence, Artificial Intelligence 171 (10–15) (2007) 619–641.
[3] S.P. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[4] R.J. Brachman, H.J. Levesque, Knowledge Representation and Reasoning, The Morgan Kaufmann Series in Artificial Intelligence, Morgan Kaufmann Publishers, 2004.
[5] T.M. Cover, Elements of Information Theory, 2nd edition, Wiley–Interscience, New York, 2001.
[6] L. Daniel, Paraconsistent probabilistic reasoning, Ph.D. thesis, L'École Nationale Supérieure des Mines de Paris, 2009.
[7] M. Finger, G.D. Bona, Probabilistic satisfiability: Logic-based algorithms and phase transition, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11), 2011, pp. 528–533.
[8] M. Finthammer, G. Kern-Isberner, M. Ritterskamp, Resolving inconsistencies in probabilistic knowledge bases, in: KI 2007: Advances in Artificial Intelligence – 30th Annual German Conference on Artificial Intelligence, in: Lecture Notes in Computer Science, vol. 4667, Springer-Verlag, 2007, pp. 114–128.
[9] A.M. Frisch, P. Haddawy, Anytime deduction for probabilistic logic, Artificial Intelligence 69 (1–2) (1994) 93–122.
[10] J. Grant, A. Hunter, Analysing inconsistent first-order knowledge bases, Artificial Intelligence 172 (8–9) (2008) 1064–1093.
[11] J. Grant, A. Hunter, Measuring the good and bad in inconsistent information, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11), 2011, pp. 2632–2637.
[12] A. Hunter, S. Konieczny, Measuring inconsistency through minimal inconsistent sets, in: Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning, AAAI Press, 2008, pp. 358–366.
[13] A. Hunter, S. Konieczny, On the measure of conflicts: Shapley inconsistency values, Artificial Intelligence 174 (14) (2010) 1007–1026.
[14] G. Kern-Isberner, Conditionals in Nonmonotonic Reasoning and Belief Revision, Lecture Notes in Computer Science, vol. 2087, Springer-Verlag, 2001.
[15] G. Kern-Isberner, T. Lukasiewicz, Combining probabilistic logic programming with the power of maximum entropy, Artificial Intelligence 157 (1–2) (2004) 139–202.
[16] G. Kern-Isberner, M. Thimm, Novel semantical approaches to relational probabilistic conditionals, in: F. Lin, U. Sattler, M. Truszczyński (Eds.), Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR'10), AAAI Press, 2010, pp. 382–392.
[17] K.M. Knight, Measuring inconsistency, Journal of Philosophical Logic 31 (2001) 77–98.
[18] K.M. Knight, A theory of inconsistency, Ph.D. thesis, University of Manchester, 2002.
[19] H.E. Kyburg, R.P. Loui, G.N. Carlson (Eds.), Knowledge Representation and Defeasible Reasoning, Studies in Cognitive Systems, Springer-Verlag, 1990.
[20] T. Lukasiewicz, Probabilistic default reasoning with conditional constraints, Annals of Mathematics and Artificial Intelligence 34 (2002) 35–88.
[21] D.P. Muiño, Measuring and repairing inconsistency in probabilistic knowledge bases, International Journal of Approximate Reasoning 52 (6) (2011) 828–840.
[22] J. Munkres, Topology, 2nd edition, Prentice Hall, 1999.
[23] J.B. Paris, The Uncertain Reasoner's Companion – A Mathematical Perspective, Cambridge University Press, 2006.
[24] G. Parmigiani, Modeling in Medical Decision Making: A Bayesian Approach, Statistics in Practice, John Wiley and Sons, 2002.
[25] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, 1998.
[26] I. Rahwan, G.R. Simari (Eds.), Argumentation in Artificial Intelligence, Springer-Verlag, 2009.
[27] R. Reiter, A logic for default reasoning, Artificial Intelligence 13 (1–2) (1980) 81–132.
[28] W. Rödder, Conditional logic and the principle of entropy, Artificial Intelligence 117 (2000) 83–106.
[29] W. Rödder, I.R. Gartner, S. Rudolph, Entropy-driven portfolio selection – A downside and upside risk framework, Tech. rep., Diskussionsbeitrag der Fakultät für Wirtschaftswissenschaft, FernUniversität in Hagen, 2009.
[30] W. Rödder, C.-H. Meyer, Coherent knowledge processing at maximum entropy by SPIRIT, in: Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI'96), 1996, pp. 470–476.
[31] W. Rödder, L. Xu, Elimination of inconsistent knowledge in the probabilistic expert system-shell SPIRIT (in German), in: B. Fleischmann, R. Lasch, K. Derigs, W. Domschke, K. Riedler (Eds.), Operations Research Proceedings: Selected Papers of the Symposium on Operations Research 2000, Springer-Verlag, 2001, pp. 260–265.
[32] L.S. Shapley, A value for n-person games, in: H. Kuhn, A. Tucker (Eds.), Contributions to the Theory of Games II, in: Annals of Mathematics Studies, vol. 28, Princeton University Press, 1953, pp. 307–317.
[33] W. Siler, J.J. Buckley, Fuzzy Expert Systems and Fuzzy Reasoning, John Wiley and Sons, 2005.
[34] M. Thimm, Measuring inconsistency in probabilistic knowledge bases, in: J. Bilmes, A. Ng (Eds.), Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI'09), AUAI Press, 2009, pp. 530–537.
[35] M. Thimm, Analyzing inconsistencies in probabilistic conditional knowledge bases using continuous inconsistency measures, in: Proceedings of the Third Workshop on Dynamics of Knowledge and Belief (DKB'11), 2011, pp. 31–45.
[36] M. Thimm, Probabilistic reasoning with incomplete and inconsistent beliefs, Ph.D. thesis, Technische Universität Dortmund, Germany, 2011.