Probabilistic Entailment on First Order Languages

Soroush Rafiee Rad

1 Introduction

The treatment of inconsistencies is a long standing issue for mathematical logic. The process of reasoning in the classical logic has been devised with strong built-in consistency assumptions and it follows that the full force of classical entailment relation is too strong for reasoning with inconsistencies. Although limiting the scope of logical inference to only consistent domains fits well with the spirit of what one requires from reasoning in mathematical contexts, there are many aspects of reasoning where it does not. In particular, we have the case when the context of the reasoning is not assumed to represent some factual property of a structure nor objective facts concerning the real state of things but some not-necessarily-certain information or approximations regarding those facts.

There are different motivations for the development of logics that can accommodate inconsistencies and there have been several attempts in the literature to do so. The main difference between these motivations arise from the way that the inconsistent evidence is interpreted. One motivation stems from adopting the philosophical position of dialetheism, best advocated by Graham Priest for example. This position is characterised by submitting to the thesis that there are true contradictions. That is to accept that there are sentences which are true and false simultaneously, see for example (Priest, 1979, 1987, 1989). One approach to deal with inconsistencies in this view is to adopts a three valued logic with truth values $\{0, 1, \{0, 1\}\}$, for example, with truth value $\{0, 1\}$ for the sentences that are assumed to be both true and false.

Other motivations can arise from more pragmatic reasons which deal with reasoning in non-ideal contexts. Here the inconsistencies are interpreted as a property of the information and are taken to be anomalies that point out errors or shortcomings of the reasoners' information (or maybe communication channels). The approaches that arise from this latter motivation, primarily, try to deal with the inconsistent sets by reducing the inference to consistent reasoning. This is done either by defining the logical consequences of such sets on the basis of their maximal consistent subsets as is the case for da Costa's para-consistent logics, (da Costa, 1974, 1989, 1998), or by first revising the inconsistent sets to consistent ones. For example one might define the set of logical consequences of a possibly inconsistent set Γ as the union (or intersection) of the sets of logical consequences of its maximal consistent subsets. Or one might choose to apply some belief revision process to first arrive at a consistent information set Γ', as in AGM belief revision process for instance, (Alchourron et. al., 1985), and make the reasoning on the basis of this consistent set. The idea in an AGM-like belief revision process for example, is that upon receiving some inconsistent information ϕ, one will first retract the part of knowledge base that contradicts this new information and then expands the remaining knowledge set by adding ϕ. The assumption here, however, is that the new information is always more reliable than the old. An assumption which is counterintuitive in many aspects of reasoning. For example when the context of reasoning consists of statements derived from a not-completely-reliable sources or processes that are subject to errors. Even more pointed are cases where the context of reasoning consists of statements accumulated through different sources and processes which do not necessarily agree. This is indeed the case in almost all applications of reasoning outside some mathematical theory. As the information set expands by acquiring new information through possibly conflicting sources and processes, it may very well come to include conflicting and inconsistent evidence without any second order information that warrants discarding parts of these evidence in favour of others. This will void the possibility of using classical entailment (or other variations of it which still get trivialised in the presence of inconsistencies) as it validates any consequence from such an inconsistent set. In this sense having some inconsistency in a (possibly very large) set of evidence will render it completely useless for reasoning. There are many applications of reasoning, however, in which the inconsistencies should intuitively affect the reasoning only partially. As a very simple example, consider sentences ϕ and ψ that share no relation simple, function symbol or constants (hence have completely irrelevant informational content), then

$$\{\phi, \psi, \neg\phi\} \vDash \neg\psi$$

many instances of which are counterintuitive. For example, assume a case where ϕ is acquired from a source, say S_1, different from that of $\neg\phi$, say S_2 where both sources agree on ψ. Here the inconsistency of the information regarding ϕ may not provide any reason to affect the reasoning on the part of ψ. This motivates one to fashion inference processes that allow meaningful extraction of information from such sets of information. This is the motivation for what we shall pursue in these notes and the aspect of the literature we hope to contribute to. To this end wet will investigate how to accommodate inconsistencies of premisses from which one aims to make inference and will study a probabilistic entailment relation, on propositional languages introduced by Paris, (Paris, 2004), and further developed by Paris and Picado and Rosefield, (Paris et. al., 2008), and will extend their approach to first order languages.

The rest of this paper is organised as follows. In Section 2 we will set up our notation and preliminaries. In Section 3 we will investigate a revision process for reducing inconsistent sets of sentences to probabilistically consistent uncertain ones. We will investigate revision of sets of sentences in Section 3.1, and revision of inconsistent probabilistic assertions in Section 3.2 and prioritised sets of sentences or probabilistic assertions in Section 3.3. In Section 4 we will study a probabilistic entailment relation that allows meaningful inferences on possibly inconsistent sets. We shall give an analysis of this entailment relation in the first order logic in Section 4.3 and will then briefly point to a generalisation that allows us to limit the effect of inconsistency to only part of the reasoning in Section 5.

2 Preliminaries and Notation

Throughout these notes we will work with a first order language $\mathcal{L}$ with finitely many relation symbols, no function symbols and countably many constant symbols $a_1, a_2, a_3, \ldots$. Furthermore we assume that these individuals exhaust the universe. This means in particular that we have a name for every element in our universe. Thus a model is a structure M for the language $\mathcal{L}$ with domain $|M| = \{ a_i \,|\, i = 1, 2, \ldots\}$ where every constant symbol is interpreted as itself.Let $R\mathcal{L}, S\mathcal{L}$ denote the set of relation and the set of sentences of $\mathcal{L}$ respectively.

Definition 2.1 *We shall call $w : S\mathcal{L} \to [0\,,\,1]$ a probability function if for every $\phi, \psi, \exists x\psi(x) \in S\mathcal{L}$,*

- *P1. If $\models \phi$ then $w(\phi) = 1$.*
- *P2. $w(\phi \vee \psi) = w(\phi) + w(\psi) - w(\phi \wedge \psi)$.*
- *P3. $w(\exists x\psi(x)) = \lim_{n\to\infty} w(\bigvee_{i=1}^{n} \psi(a_i))$.*

Let $\mathcal{L}$ be a propositional language with propositional variables $p_1, p_2, \ldots, p_n$. By *atoms* of $\mathcal{L}$ we mean the set of sentences $\{ \alpha_i \,|\, i = 1, \ldots J\}$, $J = 2^n$ of the form

$$\pm p_1 \wedge \pm p_2 \wedge \ldots \wedge \pm p_n.$$

By disjunctive normal form theorem, for every sentence $\phi \in S\mathcal{L}$ there is unique set $\Gamma_\phi \subseteq \{ \alpha_i | i = 1, \ldots, J \}$ such that

$$\models \phi \leftrightarrow \bigvee_{\alpha_i \in \Gamma_\phi} \alpha_i.$$

It can be easily checked that $\Gamma_\phi = \{\, \alpha_j \,|\, \alpha_j \vDash \phi \,\}$.

Thus if $w : S\mathcal{L} \to [0\,,\,1]$ is a probability function then

$$w(\phi) = w(\bigvee_{\alpha_i \vDash \phi} \alpha_i) = \sum_{\alpha_i \vDash \phi} w(\alpha_i)$$

as the α_i's are mutually inconsistent. On the other hand since $\vDash \bigvee_{i=1}^{J} \alpha_i$ we have $\sum_{i=1}^{J} w(\alpha_i) = 1$. So the probability function w will be uniquely determined by its values on the α_i's, that is by the vector

$$< w(\alpha_1), ..., w(\alpha_J) > \in \mathbb{D}^{\mathcal{L}} \quad \text{where} \quad \mathbb{D}^{\mathcal{L}} = \{\, \vec{x} \in \mathbb{R}^J \,|\, \vec{x} \geq 0, \sum_{i=1}^{J} x_i = 1 \}.$$

Conversely if $\vec{a} \in \mathbb{D}^{\mathcal{L}}$ we can define a probability function $w' : S\mathcal{L} \to [0\,,\,1]$ such that $< w'(\alpha_1), ..., w'(\alpha_J) >= \vec{a}$ by setting

$$w'(\phi) = \sum_{\alpha_i \vDash \phi} a_i.$$

This gives a one to one correspondence between the probability functions on $\mathcal{L}$ and the points in $\mathbb{D}^{\mathcal{L}}$. In particular if a knowledge base K is taken to be a *satisfiable* set of linear constraints of the form

$$\sum_{j=1}^{n} a_{ij} w(\phi_j) = b_i, \quad i = 1, 2, ..., m$$

where $\phi_j \in S\mathcal{L}$, $a_{ij}, b_j \in \mathbb{R}$ and w is a probability function, then replacing each $w(\phi_j)$ in K with $\sum_{\alpha_i \vDash \phi_j} w(\alpha_i)$ and adding the equation $\sum_{i=1}^{J} w(\alpha_i) = 1$ we will get a new set of constraints given in terms of the probability of atoms

$$\sum_{j=1}^{J} a'_{ij} w(\alpha_j) = b_i, \quad i = 1, 2, ..., m$$

$$< w(\alpha_1), ..., w(\alpha_J) > A_K = \vec{b}_K.$$

The situation for first order languages is a bit more complicated. Here the atoms of the language are defined as the set of formulas

$$\bigwedge_{\substack{R\ j-ary \\ R \in R\mathcal{L}, j \in N^+}} \pm R(x_{i_1}, ..., x_{i_j}).$$

In the case of first order languages, what plays the role similar to the atoms for a propositional language, are the state descriptions.

Definition 2.2 *Let $\mathcal{L}$ be a first order language with the set of relation symbols $R\mathcal{L}$ and let $\mathcal{L}^{(k)}$ be a sub-language of $\mathcal{L}$ with only finitely many constant symbols $a_1, ..., a_k$. The* state descriptions *of $\mathcal{L}^{(k)}$ are the sentences $\Theta_1^{(k)}, ..., \Theta_{n_k}^{(k)}$ which enumerate all the sentences of the form*

$$\bigwedge_{\substack{i_1,...,i_j \leq k \\ R\ j-ary \\ R \in R\mathcal{L}, j \in N^+}} \pm R(a_{i_1}, ..., a_{i_j}).$$

The following theorem, due to Gaifman, provides a similar result, to that we had above, for the case of a first order language $\mathcal{L}$. Let $QFS\mathcal{L}$ be the set of quantifier free sentences of $\mathcal{L}$:

Theorem 2.3 *Let $v : QFS\mathcal{L} \to [0, 1]$ satisfy P1 and P2 for $\phi, \psi \in QFS\mathcal{L}$. Then v has a unique extension $w : S\mathcal{L} \to [0, 1]$ that satisfies P1, P2 and P3. In particular if $w : S\mathcal{L} \to [0, 1]$ satisfies P1, P2 and P3 then w is uniquely determined by its restriction to $QFS\mathcal{L}$.*

For $\phi \in QFS\mathcal{L}$ let k be an upper bound on the i such that a_i appears in ϕ. Then ϕ can be thought of as being from the propositional language $\mathcal{L}^{(k)}$ with propositional variables $R(a_{i_1}, ..., a_{i_j})$ for $i_1, ..., i_j \leq k$, $R \in R\mathcal{L}$ and $R\ j - ary$. Then the sentences $\Theta_i^{(k)}$ will be the atoms of $\mathcal{L}^{(K)}$ and

$$\phi \leftrightarrow \bigvee_{\Theta_i^{(k)} \models \phi} \Theta_i^{(k)} \quad so \quad w(\phi) = \sum_{\Theta_i^{(k)} \models \phi} w(\Theta_i^{(k)}).$$

Thus to determine the value $w(\phi)$ we only need to determine the values $w(\Theta_i^{(k)})$ and to require

- $w(\Theta_i^{(k)}) \geq 0$ and $\sum_{i=1}^{n_k} w(\Theta_i^{(k)}) = 1$.
- $w(\Theta_i^{(k)}) = \sum_{\Theta_j^{(k+1)} \models \Theta_i^{(k)}} w(\Theta_j^{(k+1)})$.

to ensure that w satisfies P1 and P2. Using this we will limit ourselves to only dealing with $QFS\mathcal{L}$.

3 Probabilistic Reduction of Inconsistent Theories

3.1 Probabilistic Reduction of Inconsistent Sets of Sentences

Consider a consistent set of sentences $B = \{\phi_1, \ldots, \phi_n\}$ and let θ be such that $B \cup \{\theta\}$ is inconsistent. In the setting we shall present here, this inconsistency will be characterised as uncertainty and will thus results in moving to some probabilistically consistent reduction of $B \cup \{\theta\}$, B', i.e., a set B' consisting of jointly satisfiable probabilistic statements of the form $w(\phi) = p$ for $\phi \in B \cup \{\theta\}$.

Definition 3.1 *For a set of sentences $\Gamma \subset S\mathcal{L}$, the maximal consistency of Γ, denoted by $mc(\Gamma)$ is defined as*

$$mc(\Gamma) = max\{\eta \,|\, \Gamma \textit{ is } \eta \textit{ consistent}\} =$$

$$max\{\eta \,|\, \textit{there is a probability function } w \textit{ on } S\mathcal{L} \textit{ such that } w(\phi) \geq \eta \textit{ for all } \phi \in \Gamma\}$$

Lemma 3.2 *Let $\Gamma = \{\phi_1, \ldots, \phi_n\} \subset S\mathcal{L}$ with $mc(\Gamma) = \eta$. Then there is a fixed subset of Γ, say Γ_1 such that for every probability function w on $S\mathcal{L}$, if $w(\phi) \geq \eta$ for all $\phi \in \Gamma$ then $w(\phi) = \eta$ for all $\phi \in \Gamma_1$.*

Proof Suppose not, then for every $\psi \in \Gamma$ there is a probability function w_ψ (not necessarily distinct) such that $w_\psi(\phi) \geq \eta$ for all $\phi \in \Gamma$ and $w_\psi(\psi) > \eta$. Let

$$w = 1/n \sum_{\psi \in \Gamma} w_\psi$$

then for every $\phi \in \Gamma$ we have

$$w(\phi) = 1/n \sum_{\psi \in \Gamma} w_\psi(\phi) > \eta$$

since every $w_\psi(\phi) \geq \eta$, $\psi \neq \phi$ and $w_\phi(\phi) > \eta$. This is a contradiction with $mc(\Gamma) = \eta$. □

Let $\Gamma \subset S\mathcal{L}$ and let $mc(\Gamma) = \eta_1$ and let Γ_1 as in Lemma 3.2. Set

$$\eta_2 = max\{\eta \,|\, w(\psi) \geq \eta \text{ for } \psi \in \Gamma - \Gamma_1$$
$$\text{where } w \text{ is a probability function such that } w(\phi) \geq \eta_1 \text{ for } \phi \in \Gamma\}.$$

With The same argument as in Lemma 3.2, one can show that there is a fixed subset $\Gamma_2 \subset \Gamma - \Gamma_1$ such that $w(\theta) = \eta_2$ for $\theta \in \Gamma_2$ and $w(\theta) \geq \eta_2$ for $\theta \in \Gamma - (\Gamma_1 \cup \Gamma_2)$ for every probability function w such that $w(\phi) \geq \eta_1$ for $\phi \in \Gamma$ (so $w(\phi) = \eta_1$ for $\phi \in \Gamma_1$) and $w(\psi) \geq \eta_2$ for $\psi \in \Gamma - \Gamma_1$. Following the same process finitely many times one will be left a partition $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \ldots \cup \Gamma_m$ and values $\eta_1, \ldots, \eta_m$. Then set

$$\vec{mc}(\Gamma) = < \delta_1, \ldots, \delta_n >, \text{ where } \delta_j = \eta_k \iff \phi_j \in \Gamma_k$$

Intuitively the values given in $\vec{mc}(\Gamma)$ are the highest probabilities that can be assigned to the sentences in Γ consistently. In the sense that there is no probability function that can assign a probability higher than η_1 to all the sentences in Γ_1 simultaneously and same for η_2 and Γ_2 and so on. In other words if we take $\vec{1} = < 1, \ldots, 1 >$ as an n-vector representing the assignment of reliability 1 to all sentences $\phi_1, \ldots, \phi_n$ (which will be inconsistent if Γ is) then for any probability function w if we set $\vec{w} = < w(\phi_1), \ldots, w(\phi_n) >$, we have

$$d(\vec{1}, \vec{mc}(\Gamma)) \leq d(\vec{1}, \vec{w})$$

thus accounting for $\vec{mc}(\Gamma)$ being the closest we can consistently get to the assumption that all sentences in our knowledge set Γ are correct.

Definition 3.3 *Let $B = \{\phi_1, \ldots, \phi_n\} \subset S\mathcal{L}$ be consistent set of sentences and $\phi_{n+1} \in S\mathcal{L}$ be such that $B \cup \{\phi_{n+1}\} \vDash \bot$, then the revision of B by ϕ_{n+1} is defined as*

$$B' = \{w(\phi_1) = p_1, \ldots, w(\phi_n) = p_n, w(\phi_{n+1}) = p_{n+1}\}$$

where

$$< p_1, \ldots, p_n, p_{n+1} > = \vec{mc}(\{\phi_1, \ldots, \phi_n, \phi_{n+1}\}).$$

Definition 3.3 is to intended to capture the idea that the revised assignments of probabilities to the sentences $\phi_1, \dots, \phi_n, \phi_{n+1}$ remain as close as possible to 1, that is to assign the highest reliability to the information that is consistently possible.

3.2 Reduction of Probabilistic Theories

Using the revision process described above, one will move, in the presence of inconsistencies, from a set of sentences to one consisting of probabilistic assertion on those sentences. To use this as a process for iterated revision one needs to define the revision process also on the sets of probabilistic assertions as above. The latter will be more general and include the categorical sets by identifying a set $\{\phi_1, \dots, \phi_n\}$ with the set $\{w(\phi_1) = 1, \dots, w(\phi_n) = 1\}$.

For a set of probabilistic statements, $\Gamma = \{w(\phi_1) = p_1, \dots, w(\phi_n) = p_n\}$, we say that Γ is inconsistent when there is no probability function W with $W(\phi_i) = p_i$ for $i = 1, \dots, n$. In other words when w can not be extended to a probability function. Notice that in revising the $B = \{\phi_1, \dots, \phi_n\}$, with a sentence ϕ_{n+1}, the notion of maximal consistency of $B \cup \{\phi_{n+1}\}$ represent an attempt to consistently assign probabilities to these sentences while remaining as close as possible to their prior probabilities (namely, 1). The approach when dealing with inconsistent sets of probabilistic assertions is going to be the same. We shall try to consistently revise the probability assignments while remaining as close as possible to the prior probabilities, which might not necessarily be 1 any more. To this end we first generalise the notion of maximal consistency given above.

Definition 3.4 *Let $\Gamma = \{w(\phi_1) = p_1, \dots, w(\phi_n) = p_n\}$ be a (possibly inconsistent) set of probabilistic sentences. The minimal change consistency of Γ, $\vec{mcc}(\Gamma)$, is defined as the n-vector*

$$\vec{q} \in \{< a_1, \dots, a_n > \mid \textit{there is a probability function } W \textit{ on } S\mathcal{L} \textit{ with } W(\phi_i) = a_i\}$$

for which $d(\vec{q}, \vec{p})$ is minimal, where $\vec{p} =< p_1, \ldots, p_n >$ and d is the Euclidean distance.

Notice that for consistent $\Gamma = \{w(\phi_1) = p_1, \ldots, w(\phi_n) = p_n\}$, the $\vec{mcc}(\Gamma) =< p_1, \ldots, p_n >$. The process of revising a set of probabilistic assertions $B = \{w(\phi_1) = p_1, \ldots, w(\phi_n) = p_n\}$ with the statement $w(\phi_{n+1}) = p_{n+1}$ is the same as revising a on-probabilistic set of set of sentences but with $\vec{mcc}(B \cup \{w(\phi_{n+1}) = p_{n+1}\})$ instead of $\vec{mc}(B \cup \{\phi_{n+1}\})$.

Definition 3.5 *Let $B = \{w(\phi_1) = p_1, \ldots, w(\phi_n) = p_n\}$, where $\{\phi_1, \ldots, \phi_n\} \subset S\mathcal{L}$ and $\phi_{n+1} \in S\mathcal{L}$ be such that $B \cup \{w(\phi_{n+1}) = p_{n+1}\}$ is probabilistically inconsistent[1], then the revision of B by $w(\phi_{n+1}) = p_{n+1}$ is defined as*

$$B' = \{w(\phi_1) = q_1, \ldots, w(\phi_n) = q_n, w(\phi_{n+1}) = q_{n+1}\}$$

where

$$\vec{q} = \vec{mcc}(B \cup \{w(\phi_{n+1}) = p_{n+1}\}).$$

3.3 Revising Prioritised Sets of Sentences

One can immediately notice that in the revision process described above all the sentences are given the same priority. This can be readily relaxed. One can modify the distance used in the definition of $\vec{mcc}$ to account for a higher degree of reliability or trust in one or some of the proibabilistic assertions that are to revised. To this end we can for example take

$$d(\vec{q}, \vec{p}) := \sqrt{d_i(q_i - p_i)^2}$$

and define $\vec{mcc}(B)$, as the n-vectors

$$\vec{q} \in \{< a_1, \ldots, a_n > | \text{there is a probability function } W \text{ on } S\mathcal{L} \text{ with } W(\phi_j) = a_j\}$$

[1]that is there is no probability function that can simultaneously assign these values to the sentences in $\phi_1, \ldots, \phi_{n+1}$.

for which $d(\vec{q},\vec{p})$ is minimal. And as before let the revision of B by $w(\phi_{n+1}) = p_{n+1}$ be

$$B' = \{w(\phi_1) = q_1, \ldots, w(\phi_n) = q_n, w(\phi_{in+1}) = q_{n+1}\}$$

where

$$\vec{q} = \vec{mcc}(B \cup \{w(\phi_{n+1}) = p_{n+1}\}).$$

One can achieve the same results by taking a more detailed approach using some notion of ordinal ranking. To see this take the language $\mathcal{L}^{(k)}$ to have the same relation symbols as $\mathcal{L}$, say $R_1, \ldots, R_t$ but with the domain restricted to $\{a_1, \ldots, a_k\}$. If k is the largest such that a_k appears in ϕ_i, $i = 1, \ldots, n+1$, then the ϕ_i can be viewed as sentences in the propositional language with propositional variables

$$R_i(a_{j_1}, \ldots, a_{j_{s_i}})$$

with $1 \leq i \leq t$, $i_1, \ldots, i_{s_i} \in \{a_1, \ldots, a_k\}$ and s_i being the arity of R_i. Then the atoms of this language are the sentences of the form

$$\bigwedge_{\substack{j_1,\ldots,j_{s_i} \leq k \\ R\ s_i-ary \\ R_i \in R\mathcal{L}, j \in N^+}} \pm R_i(a_{j_1}, ..., a_{j_{s_i}}).$$

and given an ordinal ranking on these atoms in a way that contradictions are given rank 0, and the more plausible atoms get assigned a higher ordinal, one can take the coefficients d_i above as the highest rank such that there is an atom of that rank consistent with ϕ_i. On other contextual consideration one might choose to have the coefficients d_i to represent the reliability of the source or the process from which the information is acquired.

4 Probabilistic Entailment

4.1 The ${}^{\eta}\triangleright_{\zeta}$ Entailment

In this section we will first investigate a probabilistic entailment relation introduced by Paris (Paris, 2004) for propositional languages and further developed by Paris,

Picado and Rosefield (Paris et. al., 2008), and will then present analogous results to those in (Paris, 2004; Paris et. al., 2008) for first order languages and later on we shall study a generalisation of this entailment relation to multiple thresholds.
As will be clear shortly, the probabilistic entailment we study provides a spectrum of consequence relations, each at a different degree of reliability, which facilitate our goal in deriving meaningful inferences from an inconsistent set. As we shall see in details, this is in line with our initial thesis to identify an inconsistent theory with an uncertain theory which we shall represent as a probabilistic one. The inferences from such a theory will inevitably be probabilistic and we shall regard the entailment relation as preserving the reliability or "acceptability" of the consequences given that of the premises as opposed to preserving the categorical truth as is the case for the classical consequence relation. The "acceptability" in inferences here, will be represented with a probabilistic threshold which, we shall assume, can be set from the contextual considerations.

Definition 4.1 *Let $\Gamma \subset S\mathcal{L}$, $\psi \in S\mathcal{L}$ and $\eta, \zeta \in [0,1]$. Following (Paris et. al., 2008), we define*

$$\Gamma^{\eta} \triangleright_{\zeta} \psi \iff \textit{for all probability functions } w \textit{ on } \mathcal{L}, \textit{ if } w(\Gamma) \geq \eta \textit{ then } w(\psi) \geq \zeta$$

The idea here is that as long as one is in the position to assign to each of the sentences in Γ a probability of at least η, one is also in the position to assign a probability of at least ζ to the sentence ψ. The intuition for defining such a probabilistic entailment is more evident when $\eta = \zeta$ are interpreted as the thresholds for acceptance. In this situation the entailment relation $\Gamma^{\eta} \triangleright_{\eta} \psi$ can be read as: as long as we are prepared to *accept* all the sentences in Γ we are bound to *accept* ψ. There are situations, however, where the context of reasoning justifies different threshold for the assumptions and conclusion.
An important feature of this entailment relation is the observation that for the right value of η this is a para-consistent entailment relation. To see this notice for

example that

$$\{\phi, \neg\phi, \psi\} {}^{1/2}\!\not\triangleright_{1/2} \neg\psi$$

for ϕ and ψ syntactically disjoint (i.e., when they do not share any relation or constant symbols), since one can find a probability function w for which, $w(\phi) = w(\neg\phi) = 1/2$ and $w(\psi) = 1$ (and thus $w(\neg\psi) = 0$). This does however depend for each Γon the value of η. For $\eta > 1/2$, for example, ${}^{\eta} \triangleright_{\zeta}$ will be trivialised on the set $\{\phi, \neg\phi, \psi\}$ for any ζ since there would be no probability function that can assign a probability higher than 1/2 to all the sentences in this set. To be more precise, the entailment relation ${}^{\eta} \triangleright_{\zeta}$ is para-consistent on the set of sentences Γ for all $\eta \leq mc(\Gamma)$. Thus for the rest of this section we shall restrict ourselves to $\eta \in [0, mc(\Gamma)]$ whenever we make a reference to $\Gamma^{\eta} \triangleright_{\zeta}$.

4.2 Some Properties of ${}^{\eta} \triangleright_{\zeta}$

Proposition 4.2 *For any* $\Gamma \subset S\mathcal{L}$ *and* $\psi \in S\mathcal{L}$,
(i) $\Gamma^{\eta} \triangleright_0 \psi$.
(ii) For $\zeta > 0$, $\Gamma^1 \triangleright_{\zeta} \psi \iff \Gamma \vDash \psi$.
(iii) For $\eta > mc(\Gamma)$, $\Gamma^{\eta} \triangleright_1 \psi$.
(iv) For $\zeta > 0$, $\Gamma^0 \triangleright_{\zeta} \psi \iff \vDash \psi$.

Proof Parts (i) and (iii) are immediate from the definition. Notice that classical valuations on $\mathcal{L}$ are themselves probability functions. Thus for consistent Γ, $\Gamma^1 \triangleright_{\zeta} \psi$ implies that $v(\psi) \geq \zeta$ for all valuations v for which $v(\Gamma) = 1$. Since $\zeta > 0$ this implies that $v(\psi) = 1$ and thus $\Gamma \vDash \psi$. If Γ is inconsistent then (ii) follows trivially. Conversely suppose $\Gamma \vDash \psi$ and $w(\Gamma) = 1$. Let β_i, $1 \leq i \leq m$, enumerate sentences of the form

$$\bigwedge_{i=1}^{n} \phi_i^{\epsilon_i}$$

where $\Gamma = \{\phi_1, \ldots, \phi_n\}$, $\epsilon_i \in \{0, 1\}$ and $\phi_i^1 = \phi_i$ and $\phi_i^0 = \neg\phi_i$. Then for any β_i such that $w(\beta_i) > 0$ we have $\beta_i \vDash \phi_i$ for all $1 \leq i \leq n$ since otherwise we will have

$$w(\phi_i) = \sum_{\beta_j \vDash \phi_i} w(\beta_j) < 1.$$

So $\beta_i \vDash \bigwedge \Gamma$ and since $\bigwedge \Gamma \vDash \psi$,

$$\zeta \leq 1 = \sum_{\beta_j \vDash \bigwedge \Gamma} = w(\bigwedge \Gamma) \leq w(\psi)$$

as required. For (iv), if $\nvDash \psi$ then there is a valuation v for which $v(\psi) = 0$. Since v is also a probability function and $v(\Gamma) \geq 0$, $\Gamma^0 \triangleright_\zeta$ will fail for any $\zeta > 0$. Conversely if $\Gamma^0 \triangleright_\zeta \psi$ fails then there si a probability function w for which $w(\psi) < \zeta \leq 1$ and thus $\nvDash \psi$. □

Proposition 4.3 *Assume that $\Gamma^\eta \triangleright_\zeta \psi$. Then*
(i) If $\tau \geq \eta$ and $\nu \leq \zeta$, then $\Gamma^\tau \triangleright_\nu \psi$.
(ii) if $\tau \geq 0$ and $\eta + \tau, \zeta + \tau \leq 1$, then $\Gamma^{\eta+\tau} \triangleright_{\zeta+\tau} \psi$

We will first prove the following lemma:

Lemma 4.4 *Take $\phi_1, \ldots, \phi_n \in S\mathcal{L}$, and let β_i enumerate the sentences*

$$\bigwedge_{i=1}^{n} \phi_i^{\epsilon_i}$$

as before and let $v(\beta_i)$ be such that $\sum_{i=1}^{2^n} v(\beta_i) = 1$. The there is a probability function, w on $S\mathcal{L}$ for which

$$w(\beta_i) = v(\beta_i).$$

Proof It is only enough to define w on $QFS\mathcal{L}$, the quantifier free sentences of $\mathcal{L}$. Choose any probability function u on $S\mathcal{L}$ such that $u(\beta_i) \neq 0$ for $i = 1, \ldots, 2^n$ and for $\psi \in QFS\mathcal{L}$, define

$$w(\psi) = \sum_{i=1}^{2^n} v(\beta_i) u(\psi | \beta_i).$$

□

Proof of Proposition (4.3). (i) is immediate from the definition. For (ii) suppose that $\Gamma^{\eta+\tau} \rhd_{\zeta+\tau} \psi$ failed. Thus there is a probability function w for which $w(\Gamma) \geq \eta+\tau$ but $w(\psi) < \zeta + \tau$. If $w(\psi) < \zeta$ we will have that $\Gamma^{\eta} \rhd_{\zeta} \psi$ fails. Otherwise let $\gamma \geq 0$ be such that

$$\gamma < \zeta < \gamma + (\zeta + \tau - w(\psi)).$$

Let β_i enumerate all the sentences of the form

$$\bigwedge_{i=1}^{n} \phi_i^{\epsilon_i} \wedge \psi^{\epsilon_{n+1}}.$$

Pick a β_i such that $w(\beta_i) > 0$ and $\beta_i \not\models \psi$ (such a β_i exists otherwise we should have $w(\psi) = 1$ and $\Gamma^{\eta+\tau} \rhd_{\zeta+\tau} \psi$ will hold). Define

$$v(\beta_k) = \begin{cases} w(\beta_k).(\gamma/w(\psi)) & \text{if } \beta_k \models \psi, \\ w(\beta_k) & \text{if } \beta_k \Vdash\!\!\!\!/\; \psi, \beta_k \neq \beta_i, \\ w(beta_i) + w(\psi) - \gamma & \text{if } \beta_k = \beta_i \end{cases}$$

so $\sum_{k=1}^{2^{n+1}} v(\beta_k) = 1$. Using Lemma (4.4), we can find a probability function w' on $S\mathcal{L}$ such that $w'(\beta_i) = v(\beta_i)$ for $i = 1, \ldots, 2^n$. Then we have:

$$w'(\psi) = \sum_{\beta_i \models \psi} w'(\beta_i) = \sum_{\beta_i \models \psi} w(\beta_i).\gamma/w(\psi) = \gamma$$

and for $\phi \in \Gamma$ we have

$$w(\phi) - w'(\phi) \leq \sum_{\beta_i \models \phi \wedge \psi} w(\beta_i)(1 - \gamma/w(\psi)) \leq w(\psi) - \gamma$$

since all other β_k increase in probability under w',

$$w'(\phi) \geq \eta + tau - (w(\psi) - \gamma) > \eta.$$

So we have $w'(\phi_i) > \eta$ while $w'(\psi) = \gamma < \zeta$ which contradicts $\Gamma^{\eta} \rhd_{\zeta} \psi$. $\square$

Proposition 4.5 *If* $\lim_{n\to\infty} \eta_n = \eta$ *and* $\lim_{n\to\infty} \zeta_n = \zeta$ *with* η_n *increasing and* $\Gamma^{\eta_n} \rhd_{\zeta_n} \psi$ *for all* n*, then* $\Gamma^{\eta} \rhd_{\zeta} \psi$.

Proof See (Paris et. al., 2008)

The next result shows that the entailment relation ${}^{\eta}\triangleright_{\zeta}$ does not depend on the choice of language. More precisely, let $\mathcal{L}_1, \mathcal{L}_2$ be finite first order languages and such that $\Gamma \subset S\mathcal{L}_1 \cap S\mathcal{L}_2$ and $\psi \in S\mathcal{L}_1 \cap S\mathcal{L}_2$, then $w_1(\psi) \geq \zeta$ for every probability function w_1 on $S\mathcal{L}$ such that $w_1(\Gamma) \geq \eta$ if and only if $w_2(\psi) \geq \zeta$ for every probability function w_2 on $S\mathcal{L}$ such that $w_2(\Gamma) \geq \eta$.

Proposition 4.6 *The relation* ${}^{\eta}\triangleright_{\zeta}$ *is language invariant.*

Proof Let $\Gamma \subset S\mathcal{L}$ and $\psi \in S\mathcal{L}$ such that $\Gamma^{\eta}\triangleright_{\zeta}\psi$ in the context of the language $\mathcal{L}$, i.e., for every probability function w on $S\mathcal{L}$ if $w(\Gamma) \geq \eta$ then $w(\psi) \geq \zeta$. It is enough to show that if $\mathcal{L}'$ is a language such that $\mathcal{L} \subset \mathcal{L}'$ then for every probability function w' on $S\mathcal{L}'$, if $w'(\Gamma) \geq \eta$ then $w'(\psi) \geq \zeta$ and conversely.
For the forward direction assume that w' is a probability function on $S\mathcal{L}'$ such that $w'(\Gamma) \geq \eta$ but $w'(\psi) < \zeta$. Let w be the restriction of w' to $S\mathcal{L}$. Then w will be a probability function that agrees with w' on Γ and ψ and thus $\Gamma^{\eta}\triangleright_{\zeta}\psi$ will fail in the context of the language $\mathcal{L}$. Conversely let w be a probability function on $S\mathcal{L}$ such that $w(\Gamma) \geq \eta$ but $w(\psi) < \zeta$. Let $\Gamma = \{\phi_1, \ldots, \phi_n\}$ and as before let β_i enumerate the sentences of the form

$$\bigwedge_{i=1}^{n} \phi_i^{\epsilon_i} \wedge \psi^{\epsilon_{i+1}}$$

and we have that

$$w(\psi) = \sum_{\beta_i \models \psi} w(\beta_i) < \zeta.$$

Since $\mathcal{L} \subset \mathcal{L}'$, we have $\beta_i \in S\mathcal{L}'$ and since w is a probability function we have that $\sum_{i=1}^{2^{n+1}} w(\beta_i) = 1$. Using lemma 4.4, we can find a probability function w' on $S\mathcal{L}'$ with $w'(\beta_i) = w(\beta_i)$. With the notation of lemma 4.4, for $\phi \in Gamma$,

$$w'(\phi) = \sum_{i=1}^{2^{n+1}} w(\beta_i)u(\phi|\beta_i) = \sum_{\beta_i \models \phi} w(\beta_i) = w(\phi) \geq \eta$$

and

$$w'(\psi) = \sum_{i=1}^{2^{n+1}} w(\beta_i)u(\psi|\beta_i) = \sum_{\beta_i \models \psi} w(\beta_i) = w(\psi) < \zeta.$$

Hence $\Gamma^{\eta} \rhd_{\zeta} \psi$ fails in the context of language $\mathcal{L}'$. □

4.3 A Classical Analysis of $^{\eta} \rhd_{\zeta}$

Following in the footsteps of (Paris, 2004), we will now provide an analysis of the entailment relation $^{\eta} \rhd_{\zeta}$ in classical first order logic the intention behind this analysis becomes evident when we discuss its proof theory in the appendix. We start with the case where $\eta, \zeta > 0$ are rational and will generalise to the irrational η and ζ after introducing some more technicalities. Thus let $\eta = c/d$ and $\zeta = e/f$ for $c, d, e, f \in \mathbb{N}$ and assume

$$\phi_1, \ldots, \phi_n^{c/d} \rhd_{e/f} \psi. \tag{1}$$

As usual let $\beta_1, \ldots, \beta_m$ enumerate the sentences of the form

$$\bigwedge_{\phi_i}^{\epsilon_i} \wedge \psi^{\epsilon_{n+1}}.$$

Let $\vec{\phi}_i$ eb the m-vector with the jth coordinate 1 if and only if $\beta_j \models \phi_i$ and 0 otherwise (notice that if $\beta_j \not\models \phi_i$ then $\beta_j \models \neg\phi_i$) and define $\vec{\psi}$ the same way. Let

$$\mathbb{W}_m = \{< x_1, \ldots, x_m > | x_i \geq 0, \sum x_i = 1\}.$$

Notice that $\mathbb{W}_m$ is in one to one correspondence with the probability functions on $S\mathcal{L}$: using Lemma 4.4, every $\vec{v} \in \mathbb{W}_m$ can be extended to a probability function w on $S\mathcal{L}$ for which $w(\beta) =< \vec{v} >_i$ and for every probability function w, $< w(\beta_1), \ldots, w(\beta_m) >\in \mathbb{W}_m$ and and we have

$$w(\phi_i) = \sum_{\beta_j \models \phi_i} w(\beta_j) = \vec{\phi}_i \cdot < w(\beta_1), \ldots, w(\beta_m) > .$$

With this setting (1) will be equivalent to

$$\text{For all } \vec{w} \in \mathbb{W}_m, \text{ if } \vec{\phi}_i \cdot \vec{w} \geq c/d \text{ for } i = 1, \ldots, n, \text{ then } \vec{\psi} \cdot \vec{w} \geq e/f. \tag{2}$$

Let $\vec{1}$ be the m-vector with all coordinates 1, and set,

$$\underline{\vec{\phi_i}} = \vec{\phi_i} - (c/d)\vec{1}, \qquad \underline{\vec{\psi}} = \vec{\psi} - (e/f)\vec{1}$$

then (2) can be written as

$$\text{For all } \vec{w} \in \mathbb{W}_m, \text{ if } \underline{\vec{\phi_i}} \cdot \vec{w} \geq 0 \text{ for } i = 1, \ldots, n \text{ then } \underline{\vec{\psi}} \cdot \vec{w} \geq 0. \tag{3}$$

This means that $\underline{\vec{\psi}}$ is in the cone

$$\left\{ \sum_{i=1}^{n} a_i \underline{\vec{\phi_i}} + \sum_{j=1}^{m} b_j \vec{e_j} | 0 \leq a_i, b_j \in \mathbb{Q} \right\}$$

where $\vec{e}_j$ are the unit m-vectors. This means that for some $0 \leq a_i, b_j \in \mathbb{Q}$,

$$\underline{\vec{\psi}} = \sum_{i=1}^{m} a_i \underline{\phi_i} + \sum_{j=1}^{\vec{m}} b_j \vec{e_j}. \tag{4}$$

If we take M to be the product of the denominators of these a_i's, write the a_i's as N_i/M with $M, Ni \in \mathbb{N}$, remove the rightmost expression and multiply both sides by dM, we can rewrite (4) as

$$\sum_{i=1}^{n} N_i (df\vec{\phi_i} - cd\vec{1}) \leq M(df\vec{\phi} - de\vec{1}) \tag{5}$$

Setting $\vec{\neg\psi} = \vec{1} - \vec{\psi}$, (5) will be equivalent to

$$Mdf\vec{\neg\psi} + \sum_{i=1}^{n} dfN_i\vec{\phi_i} \leq [Md(f-e) + cf\sum_{i=1}^{n} m_i]\vec{1} \tag{6}$$

Conversely if (6) hlds for some $M, N_1, \ldots, N_n \geq 0$ then this process can be reversed to get back (2).

Let $\xi_1, \ldots, \xi_N \in \{\phi_1, \ldots, \phi_n\}$ be such that the sentence ϕ_i appears exactly dfN_i many times among $\xi_1, \ldots, \xi_N$ (so $N = df \sum_i N_i$). If $\beta_k \vDash \neg\psi$, by (**??**), the k-th coordinate of ξ_j is non-zero for at most $-deM + cf\sum_i N_i = (cN - d^2eM)/d$ many j. So

$$\bigvee_{\substack{J \subset \{1,\ldots,N\} \\ |J| > (cN - d^2eM)/d}} \bigwedge_{j \in J} \xi_j \vDash \psi \tag{7}$$

On the other hand, if $\beta_k \vDash \psi$ then k-th coordinate of ξ_j is non-zero for at most $(cN - d^2M(f-e))/d$ many j. So

$$\bigvee_{\substack{J\subset\{1,\dots,N\}\\ |J|>(cN-d^2M(f-e))/d}} \bigwedge_{j\in J} \xi_j \vDash \bot. \tag{8}$$

Now set,

$$Z = 1 + (cN - d^2eM)/d,$$

$$T = 1 + (cN - d^2M(f-e))/d$$

So

$$Td(f-e) = fcN - edZ + df$$

and we have $T < Z$. From (7) and (**??**),

$$\bigvee_{\substack{J\subset\{1,\dots,N\}\\ |J|=T}} \bigwedge_{j\in J} \xi_j \vDash \psi, \tag{9}$$

$$\bigvee_{\substack{J\subset\{1,\dots,N\}\\ |J|=Z}} \bigwedge_{j\in J} \xi_j \vDash \bot \tag{10}$$

$$Td(f-e) = fcN - edZ + df \mathrm{and} T < Z. \tag{11}$$

Conversely, if for some $\xi_1,\dots,\xi_N$ and some $Z,T \in \mathbb{N}$ (9), (10) hold and T and Z are related as in (11) then for any β_i, if $\beta_i \vDash \neg\psi$ then $\beta_i \vDash \xi_j$ for at most $T-1$ many j. the same way if $\beta_i \vDash \psi$ there are at most $Z-1$ many such j. So

$$\sum_{j=1}^{N} \vec{\xi}_j \le (T-1)\vec{1} + (Z-T)\vec{\psi}. \tag{12}$$

Now let $\vec{w} \in \mathbb{W}_m$ and $\vec{\xi}_j \cdot \vec{w} \ge c/d$ for $j = 1,\dots,N$. If we multiply both sides of (12) with $\vec{w}$ we get

$$(Z-T)\vec{\psi}\cdot\vec{w} \ge (c/d)N - T + 1.$$

But from (11),

$$\frac{(c/d)N - t + 1}{Z-T} = e/f$$

so $\vec{\psi}\cdot\vec{w} \ge e/f$. Thus if (9), (10) and (11) hold then

$$\xi_1, \dots, \xi_N^{c/d} \triangleright_{e/f} \psi$$

And conversely if

$$\phi_1, \dots, \phi_n^{c/d} \triangleright_{e/f} \psi$$

then there is a θ and sentences $\xi_1, \dots, \xi_N \in \Gamma$ (possibly with repeats) such that $\theta \vDash \psi$ and for some $T, Z \in \mathbb{N}$, (9), (10) and (11) hold.

Theorem 4.7 *For $\eta, \zeta \in (0,1]$ and $\phi_1, \dots, \phi_n \in S\mathcal{L}$,*

$$\phi_1, \dots, \phi_n{}^{\eta} \triangleright_{\zeta} \psi \iff \exists \xi_1, \dots, \xi_N \in \{\phi_1, \dots, \phi_n\}, \text{ and } T, Z \in \mathbb{N} \text{ with}$$

$$T(1-\zeta) \leq \eta N - \zeta Z + 1, \quad T < Z \text{ and}$$

$$\bigvee_{\substack{J \subset \{1,\dots,N\} \\ |J|=T}} \bigwedge_{j \in J} \xi_j \vDash \psi, \quad \bigvee_{\substack{J \subset \{1,\dots,N\} \\ |J|=Z}} \bigwedge_{j \in J} \xi_j \vDash \bot$$

Proof The preceding analysis gives the proof for the case of rational η and ζ. First consider the case where η is irrational but $\zeta \in \mathbb{Q}$ and assume that $\Gamma^{\eta} \triangleright_{\zeta} \psi$. Before preceding to show the result for the case where either η or ζ are irrational we need to introduce some notation and technicalities.

Definition 4.8 *For $\Gamma \subset S\mathcal{L}, \psi \in S\mathcal{L}$ and $\eta \in [0,1]$, let*

$$\zeta_{\Gamma,\eta}^{\psi} = sup\{\zeta \in [0,1] \,|\, \Gamma^{\eta} \triangleright_{\zeta} \psi\}.$$

Using Propositions 4.2 and 4.5, it is easy to check that this is well defined and that there is a probability function w for which $w(\Gamma) \geq \eta$ and $w(\psi) = \zeta_{\Gamma,\eta}^{\psi}$.
We will first show that for all x in some non-empty neighborhood $(\eta - \epsilon, \eta + \epsilon)$ we have $\zeta_{\Gamma,x}^{\psi} = q_1 x + q_2$ for some $q_1, q_2 \in \mathbb{Q}$. To show this we will first argue that the set of points $(x, \zeta_{\Gamma,x}^{\psi})$ is convex and then we will show that the function $\zeta_{\Gamma,x}^{\psi}$ is continuous on $[0, mc(\Gamma)]$ and so it should be made up of straight lines $y = q_i 1x + q_i 2$ on this interval. By taking ϵ small enough we will end up on a single

one of such straight lines in the interval $(\eta - \epsilon, \eta + \epsilon)$.

First notice that by Proposition 4.3, if $x_1 \le x_2$ then $\zeta^{\psi}_{\Gamma,x_1} \le \zeta^{\psi}_{\Gamma,x_2}$. Thus $\zeta^{\psi}_{\Gamma,x}$ in creasing in x. Second notice that for $\eta_1 < \eta_2 \le mc(\Gamma)$ and $0 < \delta < 1$ we can find probability functions w_1 and w_2 such that $w_1(\Gamma) \ge \eta_1$ and $w_1(\psi) = \zeta^{\psi}_{\Gamma,\eta_1}$ and similarly $w_2(\Gamma) \ge \eta_2$ and $w_2(\psi) = \zeta^{\psi}_{\Gamma,\eta_2}$. Take $w = \delta w_1 + (1-\delta)w_2$ and we will have

$$w(\phi) = \delta w_1(\phi) + (1-\delta)w_2(\phi) \ge \delta\eta_1 + (1-\delta)\eta_2$$

$$w(\psi) = \delta w_1(\psi) + (1-\delta)w_2(\psi) = \delta\zeta^{\psi}_{\Gamma,\eta_1} + (1-\delta)\zeta^{\psi}_{\Gamma,\eta_2}$$

Thus we have

$$\zeta^{\psi}_{\Gamma,(\delta\eta_1+(1-\delta)\eta_2)} \le \delta\zeta^{\psi}_{\Gamma,\eta_1} + (1-\delta)\zeta^{\psi}_{\Gamma,\eta_2}$$

This shows that on $[0, mc(\Gamma)]$, $\zeta^{\psi}_{\Gamma,x}$ is both increasing and convex as a function on x. Thus to show its continuity it would be enough to show that $\lim_{x\to mc(\Gamma)} \zeta^{\psi}_{\Gamma,x} = \zeta^{\psi}_{\Gamma,mc(\Gamma)}$. Using Proposition 4.5, we have

$$\lim_{x\to mc(\Gamma)} \zeta^{\psi}_{\Gamma,x} \le \zeta^{\psi}_{\Gamma,mc(\Gamma)}. \tag{13}$$

If $\lim_{x\to mc(\Gamma)} \zeta^{\psi}_{\Gamma,x} < \zeta^{\psi}_{\Gamma,mc(\Gamma)}$ then we can take

$$\lim_{x\to mc(\Gamma)} \zeta^{\psi}_{\Gamma,} < t < \zeta^{\psi}_{\Gamma,mc(\Gamma)}$$

we can then find an increasing sequence r_n converging to $mc(\Gamma)$ and probability functions w_n for which $w_n(\Gamma) \ge r_n$ and $w_n(\psi) < t$. For $\Gamma = \{\phi_1, \dots, \phi_n\}$ and as usual let β_i, $i = 1, \dots, m$ enumerate the sentences

$$\bigwedge_i^n \phi_i^{\epsilon_i} \wedge \psi^{\epsilon_{n+1}}$$

and consider the vector

$$\vec{w}_j =< w_j(\beta_1), \dots, w_j(\beta_m) >$$

Since $w_j(\beta_1)$ is a bounded sequence it has a convergent subsequence, say $w_{1_j}(\beta_1)$, converging to say, $w(\beta_1)$. Let

$$\vec{w}_j^1 =< w_j^1(\beta_1), \dots, w_j^1(\beta_m) >$$

be a subsequence of $\vec{w}_j$ such that w_j^1 is a subsequence of w_{1_j} (so $w_j^1(\beta_1)$ converges to $w(\beta_1)$). The same way we have $w_1^j(\beta_2)$ is a bounded sequence and so has a convergent subsequence, say $w_{2_j}(\beta_2)$, converging to say $w(\beta_2)$ and let

$$\vec{w}_j^2 =< w_j^2(\beta_1), \dots, w_j^2(\beta_m) >$$

be a subsequence of $\vec{w}_j^1$ for which $w_j^2(\beta_2)$ is a subsequence of $w_{2_j}(\beta_2)$ and so converges to $w(\beta_2)$. By the same method we will eventually construct a convergent subsequence of $\vec{w}_j$, namely $\vec{w}_j^m$, that converges to

$$\vec{w} =< w(\beta_1), \dots, w(\beta_m) > .$$

Using Lemma 4.4 we can extend this to a probability function w on $S\mathcal{L}$ and for all $\phi \in \Gamma$

$$w(\phi) = \sum_{\beta_k \vDash \phi} w(\beta_k) = \sum_{\beta_k \vDash \phi} \lim_{j\to\infty} w_j^m(\beta_k) = \lim_{j\to\infty} \sum_{\beta_k \vDash \phi} w_j(\beta_k) = \lim_{j\to\infty} w_j(\phi) \geq \lim_{j\to\infty} r_j = r$$

while

$$w(\psi) = \sum_{\beta_k \vDash \psi} w(\beta_k) = \sum_{\beta_k \vDash \psi} \lim_{j\to\infty} w_j^m(\beta_k) = \lim_{j\to\infty} \sum_{\beta_k \vDash \psi} w_j(\beta_k) = \lim_{j\to\infty} w_j(\psi) < \lim_{j\to\infty} t = t < \zeta^{\psi}_{\Gamma, mc(\Gamma)}$$

which is a contradiction. Thus the strict inequality can hold in (13) and we have

$$\lim_{x \to mc(\Gamma)} \zeta^{\psi}_{\Gamma,x} = \zeta^{\psi}_{\Gamma,mc(\Gamma)}$$

as required. It only remains to show that the set of pints $(x, \zeta^{\psi}_{\Gamma,x})$ is convex. Take $\Psi(x,y)$ to be a formula in the language $\mathcal{R} =< \mathbb{R}, +, \leq, 0, 1 >$ such that for $\eta, \zeta \in [0,1]$, $\mathcal{R} \vDash \Psi(\eta, \zeta) \iff \zeta^{\psi}_{\Gamma,\eta} = \zeta$. Then since $\mathcal{R}$ admits quantifier elimination and is an elementary extension of $\mathcal{Q} =< \mathbb{Q}, +, \leq, 0, 1 >$ we can suppose that $\Psi(x, y)$ is of the form

$$\bigvee_{i=1}^{s} \bigwedge_{j=1}^{u_s} (m_{ij} y * n_{ij} x + k_{ij})$$

for some $m_{ij}, n_{ij}, k_{ij} \in \mathbb{Z}$, where $*$ is either $<$ or $\leq$. The set of pairs (x, y) for which $\mathcal{R} \vDash \bigwedge_{j=1}^{u_s} (m_{ij} y * n_{ij} x + k_{ij})$ is convex. Since $\zeta_{\Gamma,x}^{\psi}$ is a continuous and convex function of x it must be a straight line $y = q_i 1x + qi2$ with coefficients $q_i 1, q_i 2 \in \mathbb{Q}$ with x ranging over some proper interval (which we can take to be closed since $\zeta_{\Gamma,x}^{\psi}$ is continuous).

Returning to our proof of Theorem 4.7, take η irrational and ζ rational and assume

$$\phi_1, \dots, \phi_n \,^{\eta} \triangleright_{\zeta} \; \psi$$

By the discussion above, $\zeta_{\Gamma,x}^{\psi} = q_1 x + q_2$ for all x in some non-empty interval $(\eta - \epsilon, \eta + \epsilon)$. Since $q_1 \eta + q_2$ is irrational we should have $q_1 \eta + q_2 > \zeta$ (notice that $\zeta_{\Gamma,x}^{\psi} = q_1 x + q_2$ is the maximum on all $zeta$ for which $\phi_1, \dots, \phi_n \,^{\eta} \triangleright_{\zeta} \; \psi$ and the equlity cannot hold) so there are $r_1, r_2 \in \mathbb{Q}$ such that $r_1 < \eta, r_2 > \zeta$ and $q_1 r_1 + q_2 > r_2$. Taking r_1 withing the ϵ of η then $\zeta_{\Gamma,x}^{\psi} > r_2$ so from the first case for rational thresholds we can find $\xi_1, \dots, \xi_N \in \Gamma$ and Z, T such that $T(1 - r_2) \leq r_1 N - r_2 Z + 1, T < Z$

$$\bigvee_{\substack{J \subset \{1,\dots,N\} \\ |J| = T}} \bigwedge_{j \in J} \xi_j \vDash \psi, \qquad \bigvee_{\substack{J \subset \{1,\dots,N\} \\ |J| = Z}} \bigwedge_{j \in J} \xi_j \vDash \bot \tag{14}$$

and by taking r_1 and r_2 close enough to η and ζ we will have

$$T(1 - \zeta) \leq \eta N - \zeta Z + 1 \tag{15}$$

as required by Theorem 4.7. Conversely if we have $\xi_1, \dots, \xi_N \in \Gamma$ and Z, T satisfying (14) and (15), there must be $r_1 < \eta$ and $r_2 > \zeta$ such that

$$T(1 - r_2) \leq r_1 N - r_2 Z + 1$$

and by the rational case above we should have $\phi_1, \dots, \phi_n \,^{r_1} \triangleright_{r_2} \; \psi$ and thus by Proposition 4.3 we have

$$\phi_1, \dots, \phi_n \,^{\eta} \triangleright_{\zeta} \; \psi.$$

The third case where $\eta \in \mathbb{Q}$ and $\zeta \notin \mathbb{Q}$ is proved similarly. For the last case where η, ζ are both irrational assume $\Gamma^\eta \triangleright_\zeta \psi$. First notice that if $\zeta^\psi_{\Gamma,x} > \zeta$ then we can take a rational $r_2 > \zeta$ and close enough to ζ and proceed as above so we will assume that $\zeta^\psi_{\Gamma,x} = q_1\eta + q_2 = \zeta$. The by the discussion above we have that $\zeta^\psi_{\Gamma,x} = q_1\eta + q_2$ in some non-empty interval $(\eta+\epsilon, \eta-\epsilon)$, and we can choose $r_1 \in \mathbb{Q}$ in this interval and set $r_2 = q_1r_1 + q_2$ and by the first case for rational thresholds we have $\xi_1, \ldots, \xi_N \in \Gamma$ and $Z < T$ with $T(1-r_2) \leq r_1N - r_2Z + 1$. We notice that we should have equality here otherwise we could increase r_2 while keeping r_1 fixed and show that $\zeta^\psi_{\Gamma,r_1} > r_2$ which contradicts the choice of r_2. These Z and T work for r_1 arbitrarily close to η (and $r_2 = q_1r_1 + q_2$) and so by taking the limit one can readily check that the same Z and T will satisfy the required inequality also for η and ζ. In the other direction, suppose we have $\xi_1, \ldots \xi_N$ for which (13) hold and Z, T that satisfy the required inequalities. Then for rational r_1 close to η and $r_2 \leq \frac{(r_1N-T+1)}{(Z-T)}$ close to ζ these same $\xi_1, \ldots, \xi_N$, Z and T will give $\Gamma^{r_1} \triangleright_{r_2} \psi$. since r_1 and r_2 can be made arbitrarily close to η and ζ respectively we can use Proposition 4.5 to get $\Gamma^\eta \triangleright_\zeta \psi$.
□

5 Generalising to Multiple Thresholds; $\vec{\eta} \triangleright_\zeta$

The intuition behind the probabilistic entailment that we studied in the previous sections is to view it as extending the (classical) relation between the *truth* of the premises and conclusions of an entailment to a relation between their *reliability*. As mentioned before this relation will only make sense if we restrict ourselves to $\eta \in [0, mc(\Gamma)]$ and in particular when dealing with inconsistent sets, we will be interested in the case where $\eta = mc(\Gamma)$. This intuitively means that we are interested to investigate the probabilistic inferences from a set Γ if we are ready to accept it with the highest reliability that is consistently possible. That reliability will of course be 1 for a consistent Γ in which case the entailment relation will coincide with the classical one. This however might be too coarse a view in many cases. One such case, for example, is when the statements in Γ are accumulated from different sources and their reliability is inevitably bound by the reliability of the corresponding source. Consider a set Γ where some statements in Γ are proved analytically and some are driven from experiments with a certain degree of reliability or error margin. The relation ${}^{\eta}\triangleright_{\zeta}$, fails to distinguish between this difference of reliability amongst statements in Γ and the threshold η is assigned to all these statements indiscriminatingly. Although this entailment relation provides suitable grounds for deriving probabilistic inferences from an inconsistent set of statements, as we shall elaborate more in the next section, it fails to limit the effect of such inconsistency to only the part of the information that is relevant to it. This is because, the threshold η is assigned to the set Γ as a whole and the presence of inconsistencies in Γ will change the maximal consistency for it as a whole. With this idea in mind, one can set out to generalise this entailment relation to a more fine graded relation that allows distinguishing between different statements in the set of premisses. To this end, we can consider a generalisation of the relation ${}^{\eta}\triangleright_{\zeta}$ investigated in the previous section. The idea here is that the entailment relation between the set Γ and a sentence ψ is to account not only for the relation between the reliability of ψ and that of Γ as a whole but between ψ and the individual sentences in Γ.

Definition 5.1 *Let* $\Gamma = \{\phi_1, \ldots, \phi_n\} \subset S\mathcal{L}$, $\psi \in S\mathcal{L}$ *and* $\vec{\eta} \in [0,1]^n, \zeta \in [0,1]$. *Define*

$$\Gamma^{\vec{\eta}} \triangleright_{\zeta} \psi \iff \textit{for all probability functions } w \textit{ on } \mathcal{L},$$
$$\textit{if } w(\phi_i) \geq <\vec{\eta}>_i \textit{ for } i = 1, \ldots, n \textit{ then } w(\psi) \geq \zeta.$$

Generalising the entailment relation to multiple thresholds allows us to first make inferences from an inconsistent set while avoiding trivialisation and secondly to limit the effect of inconsistencies to the parts of the reasoning relevant to them.

Given a set of sentences $\Gamma \subset S\mathcal{L}$, let $\eta = mc(\Gamma)$ and define

$$\Gamma \approx_\zeta \psi \iff \Gamma^\eta \rhd_\zeta \psi.$$

Intuitively we have $\Gamma \approx_\zeta \psi$ if assuming the highest reliability for the sentences of Γ, ψ will be at least as reliable as ζ. This gives, for each Γ, a spectrum of inference relations $\approx_\zeta$ for $\zeta \in [0,1]$ each at a different degree of reliability. Notice that if we denote the set of consequences of Γ at reliability degree ζ by C_Γ^ζ then for $\zeta \leq \delta$ we have

$$C_\Gamma^\delta \subseteq C_\Gamma^\zeta.$$

This does address our first goal to make valid nontrivial inferences from an inconsistent set. To address the second goal we shall move to the fine graded version of the entailment relation; Given a set of sentences $\Gamma \subset S\mathcal{L}$, with $\vec{mcc}(\Gamma) = \vec{\eta}$, define

$$\Gamma \approx_\zeta \psi \iff \Gamma^{\vec{\eta}} \rhd_\zeta \psi.$$

Again, we have a spectrum of entailment relations from the set Γ each at a different degree of reliability in $[0,1]$. To see how this allows limiting the effect of inconsistencies consider the following case; Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be disjoint languages with $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ and let $\Gamma_1 \subset S\mathcal{L}_1$ and $\Gamma_2 \subset S\mathcal{L}_2$ so $\Gamma = \Gamma_1 \cup \Gamma_2 \subset S\mathcal{L}$. Let $\Gamma_1 = \{\phi_1, \dots, \phi_n\}$ be inconsistent with $\vec{mcc}_{\Gamma_1} =< \eta_1, \dots, \eta_n >$ and assume that $\Gamma_2 = \{\psi_1, \dots, \psi_m\}$ is consistent and so $\vec{mcc}_{\Gamma_2} =< \delta_1, \dots, \delta_m >=< 1, \dots, 1 >$. Then taking

$$\Gamma = \{\phi_1, \dots, \phi_n, \psi_1, \dots, \psi_m\}$$

in this fixed order, we have

$$\vec{mcc}(\Gamma) =< \eta_1, \dots, \eta_n, 1, \dots, 1 > .$$

Now for $\theta \in S\mathcal{L}_2 \subset S\mathcal{L}$ we have

$$\Gamma \approx_\zeta \theta \iff \Gamma_2 \models \theta$$

thus reducing the inference on sentences of $\mathcal{L}_2$ where the relevant knowledge is consistent to the classical inference, hence limiting the pathological effect of the inconsistency only to inferences on sentences $\mathcal{L}_1$ where the knowledge is inconsistent.

Thus reasoning with conflicting evidence in our approach amounts to first identifying the maximal (probabilistic) consistency of the evidence. This will give the reliability of the evidence which in turn identifies the relevant evaluation functions. In the case of classical consequence relation, logical consequences of a set of sentences are those that get value 1 from all relevant evaluation functions. In the same manner the set of logical consequences of a set Γ in our setting are those that receive a probability higher than a certain threshold by all the relevant evaluation functions: the probability functions that satisfy the reliabilities for the evidence given in the $\vec{mcc}(\Gamma)$.

6 conclusion

Our approach to deal with inconsistencies is motivated by reasoning in non-ideal contexts and lies on the assumption that the inconsistent evidence do not point out the inconsistencies of the reality under investigation but point to an inconsistent valuation of facts. Receiving contradictory information should thus affect such valuations. In this view, receiving some piece of information ϕ while having $\neg\phi$ in our knowledge base has the effect of changing the valuation of ϕ (and thus $\neg\phi$). In case of categorical knowledge (with truth values of zero or one), this means moving from categorical belief in ϕ and $\neg\phi$ to some uncertain valuation of them and in case of probabilistic knowledge this would entail re-evaluation of the probabilities.
Our approach is based on two assumptions,

- the inconsistencies are identified with the uncertainty that they induce in the information set
- the information is assumed to be as reliable as possibly allowed by the consistency considerations.

Thus receiving inconsistent information will change the context of reasoning from a categorical one to an uncertain one, which we shall represent by means of probabilities. One can also hope to do so in a way that allows us to limit the pathological effect of inconsistencies to the part of the reasoning relevant to it. To make this clear, suppose as above that one is left, after receiving $\neg\phi$, with the inconsistent knowledge $\{\phi, \psi, \neg\phi\}$ where again ϕ is acquired from source S_1 and $\neg\phi$ from source S_2 while both sources agree on ψ. This inconsistency is accommodated by changing the categorical belief in ϕ and $\neg\phi$ to uncertain one by assignment of probabilities with the probabilities of ϕ and $\neg\phi$ adding up to 1 but without changing the valuation of ψ as it is irrelevant to the inconsistency.
How the change in the information set induced by the inconsistency is carried out, depends on one's approach to the weighting of the new information with respect to the old information. For example, if we take the new information to be infinitely more reliable than the old, we will end up with the same retraction and expansion process as in the AGM. But as we have seen, one can also devise the change in a manner that allows a wider range of epistemic attitudes towards the new information in comparison to the old. Since the inconsistencies will reduce our categorical knowledge to probabilistic one, any inference based on such knowledge will essentially be probabilistic. Our main goal was thus study an entailment relation that allows meaningful inference from such probabilistic knowledge bases. The idea was to investigate a consequence relation that generalises the classical consequence relation from a relation that preserves the truth to one that preserves, or more precisely ensures, some degree of reliability. To this end wet investigated how to accommodate inconsistencies of premisses from which one aims to make inference and studied a probabilistic entailment relation.
It is also worth mentioning that one can choose a different route altogether and deal with the inconsistent evidence by adopting a richer language in which the source of information is also coded in the information. Thus, for example, ϕ received from source S_1 is replaced by $(\phi)^1$ to the effect that "according to S_1, ϕ".

In this approach receiving ϕ^1 (according to S_1, ϕ) and $(\neg\phi)^2$ (according to S_2, $\neg\phi$) pose no contradiction any more while contradictory information from the same source has the effect of reducing the reliability of the source. The evaluation of information is carried out by weighting them with the reliability of the sources. As it would be immediately clear however, this approach will be equivalent to ours. The simplest case we will discuss corresponds to receiving information from equally reliable sources. The case of prioritised evidence corresponds to receiving information from sources with different reliabilities. Our approach, however, has the advantage of avoiding unnecessary complication of the language.
Of course our notion of "closeness" when revising the inconsistent theories into probabilistically consistent one can be subject to debate. The use of Euclidean distance was motivated by trying to choose the closest values for all sentences simultaneously. It would be interesting to investigate if other notions of "closeness" can improve this approach.

7 Appendix

7.1 Proof Theory

We say the the ϕ proves with thresholds η and ζ the sentence ψ, $\phi^\eta \vdash_\zeta \psi$ if and only if there is finite sequence of sentence $\phi_0, \ldots, \phi_m$ where $\phi_0 = \phi, \phi_m = \psi$ and every sentence in the sequence is either an axiom or follows from privious sentences in the sequence by some rule of inference.

The rules of inference here are

(i) *Right Weakening:* $\frac{\phi_1,\ldots,\phi_n|\psi,\psi\vDash\theta}{\phi_1,\ldots,\phi_n|\theta}$

(ii) *Monotonicity:* $\frac{\phi_1,\ldots,\phi_n|\psi}{\theta_1,\ldots,\theta_m|\psi}$ where $\{\phi_1, \ldots, \phi_n\} \subset \{\theta_1, \ldots, \theta_m\}$

(iii)(η,ζ)*-rule:* $\frac{\bigvee_{\substack{J\subset\{1,\dots,N\}\\|J|=Z}}\bigwedge_{j\in J}\xi_j\vDash\bot}{\xi_1,\dots,\xi_N|\bigvee_{\substack{J\subset\{1,\dots,N\}\\|J|=T}}\bigwedge_{j\in J}\xi_j}$ where $T(1-\zeta)\leq\eta N-\zeta Z+1$ and $T<Z$.

Although this seems very weak and some what unsatisfactory proof system, mainly because of the complicated and unnatural structure of the (η,ζ) rule, it is important to notice, as pointed out in (Paris, 2004), that given the fact that our entailment relation is based on probabilities and essentially has roots in measure theory it is quite interesting and surprising that one can get this far without introducing any of those concepts here. Nevertheless we shall give the following theorem for the sake of completeness.

Theorem 7.1 *Let* $\eta,\zeta\in(0,1]$. *Then for* $\phi_1,\dots,\phi_n,\psi\in S\mathcal{L}$,

$$\phi_1,\dots,\phi_n^\eta\rhd_\zeta\psi\iff\phi_1,\dots,\phi_n^\eta\vdash_\zeta$$

Proof By the discussion in the previous section if $\phi_1,\dots,\phi_n^\eta\rhd_\zeta\psi$ then we can find $\xi_1,\dots,\xi_N$ from $\phi_1,\dots,\phi_n$ and Z,T such that the hypothesis of the (η,ζ)-rule are satisfied and

$$\bigvee_{\substack{J\subset\{1,\dots,N\}\\|J|=T}}\bigwedge_{j\in J}\xi_j\vDash\psi.$$

$$\xi_1,\dots,\xi_N|\bigvee_{\substack{J\subset\{1,\dots,N\}\\|J|=T}}\bigwedge_{j\in J}\xi_j\ \text{ by }(\eta,\zeta)\text{-rule}$$

$$\xi_1,\dots,\xi_N|\psi\ \text{ by right weakening}$$

$$\phi_1,\dots,\phi_n|\psi\ \text{ by monotonicity since }\{\xi_1,\dots,\xi_N\}\subset\{\phi_1,\dots,\phi_n\}$$

In the other direction, the discussion in the last section provides the argument for the soundness of the (η,ζ)-rule and the soundness of other rules are immediate.

References

Alchourrn, C. E., Gadenfors, P., and Makinson, D. (1985). "On the logic of theory change: Partial meet contraction and revision functions", *Journal of Symbolic Logic* 50: 510-530.

da Costa, N.C.A. (1974). On the Theory of Inconsistent Formal Systems, *Notre Dame Journal of Formal Logic* 15(4): 497-510.

da Costa, N.C.A. and Subrahmanian, V.S. (1989). "Paraconsistent logic as a formalism for reasoning about inconsistent knowledge bases", *Artificial Intelligence in Medicine* 1: 167-174.

da Costa, N.C.A. (1998), "Paraconsistent logic", *Stanisaw Jakowski Memorial Symposium* 29–35.

Paris, J.B. (2004). "Deriving information from inconsistent knowledge bases: A completeness theorem", *Logic Journal of the IGPL* 12: 345–353.

Paris, J.B., Picado-Muino, D. and Rosefield, M. (2008). "Information from inconsistent knowledge: A Probability Logic approach", Van-Nam Huynh et al (eds.) *Interval/Probabilistic Uncertainty and Non-Classical Logics, Advances in Soft Computing* 46: 291–307.

Paris, J.B. and Vencovska, A. (1989). "Maximum Entropy And Inductive Inference", J. Skilling (ed.) *Maximum Entropy and Bayesian Methods* 397–403.

Priest, G. (1979). "Logic of paradox", *Journal of Philosophical Logic* 8: 219241.

Priest, G. (1987), *In Contradiction*, Nijhoff.

Priest, G., Routley, R. and Norman, J. (1989). *Paraconsistent Logic.* Philosophia Verlag .