A Probabilistic Approach to Default Reasoning*

Miodrag Rašković
Učiteljski fakultet, Narodnog fronta 43
11000 Beograd, Srbija i Crna Gora

Zoran Ognjanović, Zoran Marković
Matematički Institut, Kneza Mihaila 35
11000 Beograd, Srbija i Crna Gora
zorano@mi.sanu.ac.yu, zoranm@mi.sanu.ac.yu

Abstract

A logic is defined which in addition to propositional calculus contains several types of probabilistic operators which are applied only to propositional formulas. For every $s \in S$, where S is the unit interval of a recursive nonarchimedean field, an unary operator $P_{\geq s}(\alpha)$ and binary operators $CP_{=s}(\alpha, \beta)$ and $CP_{\geq s}(\alpha, \beta)$ (with the intended meaning "the probability of α is at least s", "the conditional probability of α given β is s, and "the conditional probability of α given β is at least s", respectively) are introduced. Since S is a non-archimedean field, we can also introduce a binary operator $CP_{\approx 1}(\alpha, \beta)$ with the intended meaning "probabilities of $\alpha \wedge \beta$ and β are infinitely close". Possible-world semantics with a probability measure on an algebra of subsets of the set of all possible worlds is provided. A simple set of axioms is given but some of the rules of inference are infinitary. As a result we can prove the strong completeness theorem for our logic. Formulas of the form $CP_{\approx 1}(\alpha, \beta)$ can be used to model default statements. We discuss some properties of the corresponding default entailment. We show that, if the language of defaults is considered, for every finite default base our default consequence relation coincides with the system P. If we allow arbitrary default bases, our system is more expressive than P. Finally, we analyze properties of the default consequence relation when we consider the full logic with negated defaults, imprecise observation etc.

Introduction

Different approaches to logics in which reasoning about probability is possible have been proposed. In some of them probabilities with finite ranges are allowed only (Đorđević, Rašković, & Ognjanović 2004; Ognjanović & Rašković 2000; Rašković 1993). In the others arbitrary probabilities are considered. In (Fagin, Halpern, & Megiddo 1990; Fagin & Halpern 1994) finite axiomatizations for various probabilistic logics were proposed. In that case the compactness theorem does not hold for the logics. Since the compactness theorem follows easily from the extended completeness theorem ('every consistent set of formulas is satisfiable'), one cannot hope for the extended completeness having a finitary axiomatic system. On the other hand, in (Ognjanović & Rašković 1999; 2000) an infinitary rule (from $\beta \rightarrow P_{\geq s-1/k}\alpha$, for every k, infer $\beta \rightarrow P_{\geq s}\alpha$, where $P_{\geq s}\alpha$ means probability of α is at least s) was added to axiomatic systems and the corresponding extended completeness theorems were proven.

In (Ognjanović & Rašković 2000) two kinds of probabilistic logics were distinguished: in the first logic higher order probabilities were allowed, while in the second one probabilities of classical formulas were considered only. In the later case the propositional connectives can be applied to probabilistic formulas, but mixing of classical propositional and new probabilistic formulas and iteration of probabilistic operators are not allowed. In the sequel, we concentrate on the later kind of logics. The new logic, denoted LPP^S (L for logic, the first P for propositional, the second P for probability, while S denotes a set which will be described later), is similar to the logic LPP_2 from (Ognjanović & Rašković 2000), with the above mentioned infinitary rule from LPP_2 replaced by a new one. With this new infinitary rule it is possible to determine the range of probability measure syntactically. We note that a similar rule was given in (Alechina 1995). The main novelty here is that we also introduce conditional probabilities in the syntax, together with the appropriate simple axioms. Namely, in addition to unary probabilistic operators of the type $P_{\geq s}\alpha$, for every element s of a syntactically defined set S, we also introduce binary operators of the types $CP_{=s}(\alpha, \beta)$, $CP_{\geq s}(\alpha, \beta)$ and $CP_{\approx 1}(\alpha, \beta)$ with the intended meaning "the conditional probability of α given β is s", "at least s", and "infinitely close to 1", respectively. The other novelty is that we specify, already in the syntax, that the range of the probability function is non-standard, in the sense that it contains infinitesimals. Although non-standard analysis is unfamiliar to most people, we should point out that it is essentially the original approach of Leibnitz and Newton, reworked into a consistent mathematical theory in 60's by A. Robinson using model theory. There is an elementary presentation for freshman (Keisler 1986), where everything is reduced to starting with nonarchimedean field, instead of the usual archimedean field of real numbers. This means that there exist infinite numbers, e.g., a number K, such that $K > n$ and $0 < \frac{1}{K} < \frac{1}{n}$ for every natural number n. We call $\frac{1}{K}$ an infinitesimal. Everything else is obtained using freshman mathematics. Another approach, less elementary but still simple for anyone proficient in standard analysis, is presented in (Benferhat,

*This work is supported by Ministartstvo nauke i zaštite životne sredine Republike Srbije, through Matematički Institut Beograd, grant number 1379.

Saffiotti, & Smets 2000, Section 2.4). The rationale for introducing non-standard analysis is that this is the simplest way of making precise statement like "probabilities of $\alpha \wedge \beta$ and β are almost the same", which in turn represents a default entailment "if β then, usually α". In our system this will be represented by the operator $CP_{\approx 1}(\alpha, \beta)$.

In the seminal papers (Kraus, Lehmann, & Magidor 1990; Lehmann & Magidor 1992) a set of properties which form a core of default reasoning and the corresponding formal system denoted P (or KLM) are proposed, while the default consequence relation is described in terms of preferential and rational models. In recent years many semantics for default entailment have been introduced and proven to be characterized by P (Friedman & Halpern 2001). Some of them are more or less close to our approach. In (Adams 1975; Lehmann & Magidor 1992) default rules are interpreted in terms of high conditional probabilities. In this paper we give a sound and complete axiomatization for a logic which extends those systems. It allows us to syntactically describe the behavior of the defaults in a probabilistic framework. A qualitative approach to nonstandard probabilities which uses only the power of the most significant term in the polynomial $p_w(\epsilon)$ (denoting the probability of a possible world w, ϵ is an infinitesimal) is proposed in (Goldszmidt & Pearl 1996) as a tool to work with defaults. In (Benferhat, Saffiotti, & Smets 2000) belief functions with extreme values are used to give semantics to default rules. We will show that, thanks to our probabilistic semantics, the default entailment defined in our system is different from all those approaches. The language of our logic allows one to express negated defaults, imprecise observation etc. (similarly as in the framework of conditional logic (Burghess 1981; Friedman & Halpern 2001), but we do not have nested defaults) and to deal with problems that are even not definable in the usual default systems.

The rest of the paper is organized as follows. We start with the syntax and semantics of our logic. Then we give a sound and complete axiomatic system. In the main part of the paper we describe in detail how our system can be used to model default reasoning and analyze properties of the corresponding default consequence relation. In the conclusion we summarize our results and mention some possible directions for further investigation. A sketch of the proof of the completeness theorem is presented in the appendix.

Syntax

Let S be the unit interval of a recursive nonarchimedean field containing all rational numbers. An example of such field is the Hardy field $Q[\epsilon]$. $Q[\epsilon]$ contains all rational functions of a fixed infinitesimal ϵ which belongs to a nonstandard elementary extension R^* of the standard real numbers (Robinson 1966). We use ϵ_1, ϵ_2, ... to denote infinitesimals from S. Note that every positive member of S is of the form

$$\epsilon^k \frac{\sum_{i=0}^{n} a_i \epsilon^i}{\sum_{i=0}^{m} b_i \epsilon^i},$$

where $a_0 \cdot b_0 \neq 0$, which means that for every infinitesimal $\epsilon_i \in Q[\epsilon]$, $\epsilon_i \leq c\epsilon^k$ for some positive rational number c and integer k. Note that there is no positive infinitesimal $\epsilon_i \in S$ such that for every positive integer k, $\epsilon_i \leq \epsilon^k$.

Let $\{s_0, s_1, \ldots\}$ be an enumeration of S. The language of the logic consists of:

- a denumerable set $\mathrm{Var} = \{p, q, r, \ldots\}$ of propositional letters,
- classical connectives $\neg$, and $\wedge$,
- a list of unary probabilistic operators $(P_{\geq s})_{s \in S}$,
- a list of binary probabilistic operators $(CP_{\geq s})_{s \in S}$,
- a list of binary probabilistic operators $(CP_{=s})_{s \in S}$ and
- a binary probabilistic operators $CP_{\approx 1}$.

The set For_C of classical propositional formulas is the smallest set X containing Var and closed under the formation rules: if α and β belong to X, then $\neg\alpha$ and $\alpha \wedge \beta$, are in X. Elements of For_C will be denoted by α, β, ... The set For_P^S of probabilistic propositional formulas is the smallest set Y containing all formulas of the forms:

- $P_{\geq s}\alpha$ for $\alpha \in For_C$, $s \in S$,
- $CP_{=s}(\alpha, \beta)$ for $\alpha, \beta \in For_C$, $s \in S$,
- $CP_{\geq s}(\alpha, \beta)$ for $\alpha, \beta \in For_C$, $s \in S$ and
- $CP_{\approx 1}(\alpha, \beta)$ for $\alpha, \beta \in For_C$.

and closed under the formation rules: if A and B belong to Y, then $\neg A$, and $A \wedge B$ are in Y. Formulas from For_P^S will be denoted by A, B, ... Note that we use the prefix notation $CP_{\geq s}(\alpha, \beta)$ (and similarly for $CP_{=s}(\alpha, \beta)$ and $CP_{\approx 1}(\alpha, \beta)$) rather than the corresponding infix notation $\alpha CP_{\geq s}\beta$ ($\alpha CP_{=s}\beta$, $\alpha CP_{\approx 1}\beta$).

As it can be seen, neither mixing of pure propositional formulas and probability formulas, nor nested probabilistic operators are allowed. For example, $\alpha \wedge P_{\geq s}\beta$ and $P_{\geq s}P_{\geq r}\alpha$ are not well defined formulas.

The other classical connectives ($\vee$, $\rightarrow$, $\leftrightarrow$) can be defined as usual, while we denote $\neg P_{\geq s}\alpha$ by $P_{<s}\alpha$ $P_{\geq 1-s}\neg\alpha$ by $P_{\leq s}\alpha$, $\neg P_{\leq s}\alpha$ by $P_{>s}\alpha$, $P_{\geq s}\alpha \wedge \neg P_{>s}\alpha$ by $P_{=s}\alpha$, $\neg P_{=s}\alpha$ by $P_{\neq s}\alpha$, $\neg CP_{\geq s}(\alpha, \beta)$ by $CP_{<s}(\alpha, \beta)$, $CP_{<s}(\alpha, \beta) \vee CP_{=s}(\alpha, \beta)$ by $CP_{\leq s}(\alpha, \beta)$, and $CP_{\geq s}(\alpha, \beta) \wedge \neg CP_{=s}(\alpha, \beta)$ by $CP_{>s}(\alpha, \beta)$.

Let $For^S = For_C \cup For_P^S$. φ, ψ, ... will be used to denote formulas from the set For^S. For $\alpha \in For_C$, and $A \in For_P^S$, we abbreviate both $\neg(\alpha \rightarrow \alpha)$ and $\neg(A \rightarrow A)$ by $\perp$ letting the context determine the meaning.

Semantics

The semantics for For^S will be based on the possible-world approach.

Definition 1 *An LPP^S-model is a structure $\langle W, H, \mu, v\rangle$ where:*

- *W is a nonempty set of elements called worlds,*
- *H is an algebra of subsets of W,*
- *$\mu : H \rightarrow S$ is a finitely additive probability measure, and*
- *$v : W \times \mathrm{Var} \rightarrow \{true, false\}$ is a valuation which associates with every world $w \in W$ a truth assignment $v(w)$ on the propositional letters.*

The valuation v is extended to a truth assignment on all classical propositional formula. Let M be an LPP^S model and $\alpha \in For_C$. The set $\{w : v(w)(\alpha) = true\}$ is denoted by $[\alpha]_M$.

Definition 2 *An LPP^S-model M is measurable if $[\alpha]_M$ is measurable for every formula $\alpha \in For_C$ (i.e., $[\alpha]_M \in H$). An LPP^S-model M is neat if only the empty set has the zero probability.*

In this paper we focus on the class $LPP^S_{Meas,Neat}$ of all neat and measurable LPP^S-models.

Definition 3 *The satisfiability relation $\models \subset LPP^S_{Meas,Neat} \times For^S$ is defined by the following conditions for every $LPP^S_{Meas,Neat}$-model M:*

1. *if $\alpha \in For_C$, $M \models \alpha$ if $(\forall w \in W) v(w)(\alpha) = true$,*
2. *$M \models P_{\geq s}\alpha$ if $\mu([\alpha]_M) \geq s$,*
3. *$M \models CP_{\geq s}(\alpha, \beta)$ if either $\mu([\beta]_M) = 0$ or $\mu([\beta]_M) > 0$ and $\frac{\mu([\alpha\wedge\beta]_M)}{\mu([\beta]_M)} \geq s$,*
4. *$M \models CP_{=s}(\alpha, \beta)$ if either $\mu([\beta]_M) = 0$ and $s = 1$ or $\mu([\beta]_M) > 0$ and $\frac{\mu([\alpha\wedge\beta]_M)}{\mu([\beta]_M)} = s$,*
5. *$M \models CP_{\approx 1}(\alpha, \beta)$ if either $\mu([\beta]_M) = 0$ or $\mu([\beta]_M) > 0$ and for every positive integer n, $\frac{\mu([\alpha\wedge\beta]_M)}{\mu([\beta]_M)} \geq 1 - \frac{1}{n}$.*
6. *if $A \in For^S_P$, $M \models \neg A$ if $M \not\models A$,*
7. *if $A, B \in For^S_P$, $M \models A \wedge B$ if $M \models A$ and $M \models B$.*

Note that the condition 5 is equivalent to saying that the conditional probability equals $1 - \epsilon_i$ for some infinitesimal $\epsilon_i \in S$.

A formula $\varphi \in For^S$ is satisfiable if there is an $LPP^S_{Meas,Neat}$-model M such that $M \models \varphi$; φ is valid if for every $LPP^S_{Meas,Neat}$-model M, $M \models \varphi$; a set of formulas is satisfiable if there is a model in which every formula from the set is satisfiable. A formula $\varphi \in For^S$ is a semantical consequence of a set of formulas T ($T \models \varphi$) if φ holds in every $LP_{Meas,Neat}$-model in which all formulas from T are satisfied.

A sound and complete axiomatization

The set of all valid formulas can be characterized by the following set of axiom schemata:

1. all For_C-instances of classical propositional tautologies
2. all For^S_P-instances of classical propositional tautologies
3. $P_{\geq 0}\alpha$
4. $P_{\leq s}\alpha \rightarrow P_{<r}\alpha$, $r > s$
5. $P_{<s}\alpha \rightarrow P_{\leq s}\alpha$
6. $P_{\geq 1}(\alpha \leftrightarrow \beta) \rightarrow (P_{=s}\alpha \rightarrow P_{=s}\beta)$
7. $(P_{=s}\alpha \wedge P_{=r}\beta \wedge P_{\geq 1}\neg(\alpha \wedge \beta)) \rightarrow P_{=\min(1,s+r)}(\alpha \vee \beta)$
8. $CP_{=r}(\alpha, \beta) \rightarrow \neg CP_{=t}(\alpha, \beta)$, $r \neq t$
9. $P_{=0}\beta \rightarrow CP_{=1}(\alpha, \beta)$
10. $(P_{=r}\beta \wedge P_{=s}(\alpha \wedge \beta)) \rightarrow CP_{=s/r}(\alpha, \beta)$, $r \neq 0$
11. $CP_{=r}(\alpha, \beta) \rightarrow \neg CP_{\geq t}(\alpha, \beta)$, $r < t$
12. $CP_{=r}(\alpha, \beta) \rightarrow CP_{\geq t}(\alpha, \beta)$, $r \geq t$
13. $CP_{=r}(\alpha, \beta) \rightarrow (P_{=tr}(\alpha \wedge \beta) \leftrightarrow P_{=t}\beta)$, $t \neq 0$
14. $CP_{\approx 1}(\alpha, \beta) \rightarrow CP_{\geq r}(\alpha, \beta)$, for every rational $r \in [0, 1)$
15. $CP_{=1}(\alpha, \beta) \rightarrow CP_{\approx 1}(\alpha, \beta)$

and inference rules:

1. From φ and $\varphi \rightarrow \psi$ infer ψ.
2. If $\alpha \in For_C$, from α infer $P_{\geq 1}\alpha$.
3. From $A \rightarrow P_{\neq s}\alpha$, for every $s \in S$, infer $A \rightarrow \perp$.
4. From $A \rightarrow (P_{=tr}(\alpha\wedge\beta) \leftrightarrow P_{=r}\beta)$, for every $r \in S\backslash\{0\}$, infer $A \rightarrow CP_{=t}(\alpha, \beta)$.
5. From $A \rightarrow CP_{>r}(\alpha \wedge \beta)$, for every rational $r \neq 1$, infer $A \rightarrow CP_{\approx 1}(\alpha, \beta)$.

We denote this axiomatic system by Ax_{LPP^S}. Let us briefly discuss it. Axiom 3 says that every formula is satisfied in a set of worlds of the probability at least 0. By substituting $\neg\alpha$ for α in Axiom 3, the formula $P_{\leq 1}\alpha$ ($= P_{\geq 0}\neg\alpha$) is obtained. This formula means that every formula is satisfied in a set of worlds of the probability at most 1. Let us denote it by 3'. Axiom 6 means that the equivalent formulas must have the same probability. Axiom 7 corresponds to the property of the finite additivity of probability. It says that, if the sets of worlds that satisfy α and β are disjoint, then the probability of the set of worlds that satisfy $\alpha \vee \beta$ is the sum of the probabilities of the former two sets. Axiom 13 and Rule 4 express the standard definition of conditional probability, while the axioms 14 and 15 and Rule 5 describe the relationship between the standard conditional probability and the conditional probability infinitesimally close to 1. From Axiom 3' and Rule 2 we obtain another inference rule: from α infer $P_{=1}\alpha$. The rules 3 – 5 are infinitary. Rule 3 guarantees that the probability of a formula belongs to the set S. Rule 4 corresponds to the standard meaning of the conditional probability, and Rule 5 syntactically defines the notion "infinitesimally close to 1". We should point out that, although infinitary rules might seem undesirable, especially to a computer scientist, similar types of logics with infinitary rules were proved to be decidable (Ognjanović & Rašković 2000). On the other hand, since the compactness theorem does not hold for our logic (there exists countably infinite set of formulas that is unsatisfiable although every finite subset is satisfiable: for instance, consider $\{\neg P_{=0}\alpha\} \cup \{P_{<\epsilon^n}\alpha : n \text{ is a positive integer}\}$) involving infinitary rules in the axiomatic system is the only way to obtain the extended completeness, as it is explained in the introduction.

A formula φ is deducible from a set T of formulas (denoted $T \vdash_{Ax_{LPP^S}} \alpha$) if there is an at most denumerable sequence (called proof) of formulas $\varphi_0, \varphi_1, \ldots, \varphi$, such that every φ_i is an axiom or a formula from the set T, or it is derived from the preceding formulas by an inference rule. A formula φ is a theorem ($\vdash \varphi$) if it is deducible from the empty set. A set T of formulas is consistent if there are at least a formula from For_C, and at least a formula from For^S_P that are not deducible from T. Now, following the ideas from (Ognjanović & Rašković 1999;

2000) we can prove the extended completeness theorem for the class of $LPP^S_{Meas,Neat}$-models:

Theorem 1 (Extended completeness theorem) *A set T of formulas is consistent if and only if T has an $LPP^S_{Meas,Neat}$-model.*

A sketch of the proof is provided in the appendix. In the usual manner we can derive the following corollary:

Corollary 1 *For every set T of formulas and every formula φ, $T \models \varphi$ if and only if $T \vdash \varphi$.*

Modeling default reasoning

The central notion in the field of default reasoning is the notion of default rules. A default rule, which can be seen as a sentence of the form 'if α, then generally β', can be written as[1] $\alpha \rightarrowtail \beta$. A default base Δ is a set of default rules. Default reasoning is described in terms of the corresponding consequence relation $\mid\!\sim$, i.e., we are interested in determining the set of defaults that are the consequences of a default base. Then, if α is a description of our knowledge and $\Delta \mid\!\sim \alpha \rightarrowtail \beta$, we (plausibly) conclude that β is the case. There are a number of papers which describe $\mid\!\sim$ in terms of classes of models and the corresponding satisfiability relations $\models$ such that $\Delta \mid\!\sim \alpha \rightarrowtail \beta$ if for every model M satisfying Δ, $M \models \alpha \rightarrowtail \beta$. In (Kraus, Lehmann, & Magidor 1990; Lehmann & Magidor 1992) a set of properties which form a core of default reasoning, and the corresponding formal system denoted P are proposed. The system P is based on the following axiom and rules ($\models$ denotes classical validity):

- $\alpha \rightarrowtail \alpha$ (Reflexivity)
- from $\models \alpha \leftrightarrow \alpha'$ and $\alpha \rightarrowtail \beta$, infer $\alpha' \rightarrowtail \beta$ (Left logical equivalence)
- from $\models \beta \rightarrow \beta'$ and $\alpha \rightarrowtail \beta$, infer $\alpha \rightarrowtail \beta'$ (Right weakening)
- from $\alpha \rightarrowtail \beta$ and $\alpha \rightarrowtail \gamma$, infer $\alpha \rightarrowtail \beta \wedge \gamma$ (And)
- from $\alpha \rightarrowtail \gamma$ and $\beta \rightarrowtail \gamma$, infer $\alpha \vee \beta \rightarrowtail \gamma$ (Or)
- from $\alpha \rightarrowtail \beta$ and $\alpha \rightarrowtail \gamma$, infer $\alpha \wedge \beta \rightarrowtail \gamma$ (Cautious monotonicity).

Then, for a default base Δ, $\Delta \vdash_P \alpha \rightarrowtail \beta$ if $\alpha \rightarrowtail \beta$ is deducible from Δ using the above axiom and rules. Default consequence relation is also described in terms of preferential models. A preferential model is a structure $M = \langle W, l, < \rangle$, where W is a set of states, l is a valuation which associates with each state a truth assignment on propositional letters, and $<$ is a strict partial order on W such that every set definable by formulas ($[\alpha] = \{s \in W : s \models \alpha\}$) satisfies the following condition: for every $t \in [\alpha]$ either there is minimal $u \in [\alpha]$ such $u < t$ or t is itself minimal in $[\alpha]$. A default $\alpha \rightarrowtail \beta$ holds in a preferential model M if for every s minimal in $[\alpha]$, $s \models \beta$. It is proved in (Kraus, Lehmann, & Magidor 1990; Lehmann & Magidor 1992) that the system P is sound and complete with respect to the class of all preferential models:

Theorem 2 *$\Delta \mid\!\sim \alpha \rightarrowtail \beta$ with respect to the class of all preferential models if and only if $\Delta \vdash_P \alpha \rightarrowtail \beta$.*

A special subclass of preferential models, the class of rational (or ranked) models, is also considered in (Lehmann & Magidor 1992). A preferential model is rational if the corresponding set of states can be partitioned into equivalence classes (or ranks) such that states in a class are mutually $<$-incomparable and the classes are totally ordered. However, although rational models satisfy the property called rational monotonicity (if $\alpha \rightarrowtail \beta$, and it is not $\alpha \rightarrowtail \neg\gamma$, then $\alpha \wedge \gamma \rightarrowtail \beta$) it is shown that the system P is sound and complete with respect to the class of all rational models as well (when the language of defaults is considered). It turns out that many other approaches to default reasoning (Adams 1975; Benferhat, Dubois, & Prade 1997; Benferhat, Saffiotti, & Smets 2000; Goldszmidt & Pearl 1996) are characterized by P. It is explained in (Friedman & Halpern 2001) that the classes of models of all those systems can be mapped into the so-called rich classes of qualitative plausibility structures such that the following (weak) completeness theorem holds:

Theorem 3 ((Friedman & Halpern 2001) Theorem 5.8) *A set T of qualitative plausibility structures is rich if and only if for all finite default base Δ and defaults $\alpha \rightarrowtail \beta$, $\Delta \models_T \alpha \rightarrowtail \beta$ implies $\Delta \vdash_P \alpha \rightarrowtail \beta$.*

It is also argued that that the finiteness constraint for Δ can be overcome (which leads to the strong completeness), but the proof is not given there.

In this section we describe how our system can be used to model default reasoning. As it is noted in the Introduction, the ideas of using probabilities and infinitesimals in default reasoning are not new (see, for example (Adams 1975; Benferhat, Saffiotti, & Smets 2000; Goldszmidt & Pearl 1996; Lehmann & Magidor 1992; Satoh 1990)). In (Lehmann & Magidor 1992) a family of nonstandard (R^*) probabilistic models characterizing the default consequence relation defined by the system P, is proposed. An R^*-probabilistic model is a triple $M = \langle W, H, \mu \rangle$, where W is a set of possible worlds (truth assignments to propositional letters), H is an algebra of subsets of W containing all sets definable by propositional formulas, and $\mu : H \rightarrow R^*$ is a finitely additive R^*-valued probability measure. A default $\alpha \rightarrowtail \beta$ holds in an R^*-probabilistic model if either the probability of α is 0 or the conditional probability of β given α is infinitesimally close to 1. Obviously, that class of models is a superclass of $LPP^S_{Meas,Neat}$ since in our approach the range of probabilities is a countable subset of the unit interval of R^*.

We can use $CP_{\approx 1}(\beta, \alpha)$ to syntactically describe the behavior of the default $\alpha \rightarrowtail \beta$. In the sequel, we will use $\alpha \rightarrowtail \beta$ both in the original context of the system P and to denote the corresponding translation $CP_{\approx 1}(\beta, \alpha)$. In the case of a finite default base our approach produces the same result as the other mentioned approaches, namely it is equivalent to P.

Theorem 4 *For every finite default base Δ and for every*

[1]Note that the other authors use different symbols ($\rightarrow$, $\mid\!\sim$, for example) to denote the 'default implication'. In our opinion those symbols can cause confusion, so we prefer to introduce a new symbol here.

default $\alpha \rightarrowtail \beta$

$$\Delta \vdash_P \alpha \rightarrowtail \beta \quad \textit{iff} \quad \Delta \vdash_{Ax_{LPP^S}} \alpha \rightarrowtail \beta.$$

Proof. First, since (the corresponding translation of) all axioms and rules (e.g. $(CP_{\approx 1}(\beta,\alpha) \wedge CP_{\approx 1}(\gamma,\alpha)) \rightarrow CP_{\approx 1}(\beta \wedge \gamma, \alpha)$ corresponds to And rule) of the system P are valid in the class of nonstandard probability models from (Lehmann & Magidor 1992), and $LPP^S_{Meas,Neat}$ is a subclass of that class, P is sound with respect to $LPP^S_{Meas,Neat}$.

On the other hand, following the ideas from (Lehmann & Magidor 1992, Lemma 4.9), we can show that for every finite default base Δ and for every default $\alpha \rightarrowtail \beta$, if $\Delta \not\vdash_P \alpha \rightarrowtail \beta$ then $\Delta \not\vdash_{Ax_{LPP^S}} \alpha \rightarrowtail \beta$. The key step in the proof is that there is a finite rational model $M = \langle W, l, < \rangle$ which satisfies Δ and does not satisfy $\alpha \rightarrowtail \beta$. M can be transformed to an $LPP^S_{Meas,Neat}$-model M' such that for every default d, $M \models d$ iff $M' \models d$. The transformation can be as follows. For an arbitrary infinitesimal $\epsilon' \in S$ a probability distribution μ on W can be defined so that:

- $\frac{\mu(w_{n+1})}{\mu(w_n)} = \epsilon'$, where w_n and w_{n+1} are the sets of all states of the rank n and $n+1$ respectively, and
- all states of the same rank have equal probabilities.

Since $M \models \Delta$ and $M \not\models \alpha \rightarrowtail \beta$, the same holds for M', and from the completeness of Ax_{LPP^S} we obtain that $\Delta \not\vdash_{Ax_{LPP^S}} \alpha \rightarrowtail \beta$. □

Finally, as it is noted in (Lehmann & Magidor 1992, Theorem B.12), using the similar arguments as above we can also prove a stronger version of Theorem 4:

Theorem 5 *If the language of a default base Δ is finite, then for every default $\alpha \rightarrowtail \beta$, $\Delta \vdash_P \alpha \rightarrowtail \beta$ if and only if $\Delta \vdash_{Ax_{LPP^S}} \alpha \rightarrowtail \beta$.*

Theorem 5 cannot be generalized to an arbitrary default base Δ, as it is illustrated by the following example. It is proved in (Lehmann & Magidor 1992, Lemma 2.7) that the infinite set of defaults $T = \{p_i \rightarrowtail p_{i+1}, p_{i+1} \rightarrowtail \neg p_i\}$, where p_i's are propositional letters for every integer $i \geq 0$, has only non well-founded preferential models (a preferential model containing an infinite descending chain of states) in which $p_0 \not\rightarrowtail \bot$, i.e., p_0 is consistent. It means that $T \not\vdash_P p_0 \rightarrowtail \bot$. On the other hand, $T \vdash_{Ax_{LPP^S}} p_0 \rightarrowtail \bot$ since the following holds. Let an $LPP^S_{Meas,Neat}$-model $M = \langle W, H, \mu, v \rangle$ satisfy the set T. If $\mu([p_i]) = 0$, for some $i > 0$, then it must be $\mu([p_0]) = 0$, and $M \models p_0 \rightarrowtail \bot$. Thus, suppose that $\mu([p_i]) \neq 0$, for every $i > 0$. Then:

$$\frac{\mu([p_i \wedge p_{i+1}])}{\mu([p_i])} \approx 1 \text{ and } \frac{\mu([\neg p_i \wedge p_{i+1}])}{\mu([p_{i+1}])} \approx 1,$$

i.e.,

$$\frac{\mu([p_i \wedge p_{i+1}])}{\mu([p_i])} = 1 - \epsilon_1 \text{ and } \frac{\mu([\neg p_i \wedge p_{i+1}])}{\mu([p_{i+1}])} = 1 - \epsilon_2,$$

for some infinitesimals ϵ_1 and ϵ_2, and for every $i \geq 0$. A simple calculation shows that

$$\mu([p_i \wedge p_{i+1}]) = \epsilon_2 \mu([\neg p_i \wedge p_{i+1}]),$$

$$\mu([p_i \wedge \neg p_{i+1}]) = \epsilon_1 \mu([p_i \wedge p_{i+1}])$$

and

$$\mu([p_i]) \leq \epsilon_0 \mu([p_{i+1}]),$$

for every $i \geq 0$, where $\epsilon_0 \leq \{\epsilon_1, \epsilon_2\}$. Since, for some c and k, $\epsilon_0 \leq c\epsilon^k$, it follows that for every $i > 0$,

$$0 \leq \mu([p_0]) \leq \epsilon^i.$$

Since $\mu([p_0]) \in S$ and there is no positive element of S with such property, it follows that

$$\mu([p_0]) = 0,\ [p_0] = \emptyset \text{ and } M \models p_0 \rightarrowtail \bot.$$

Since M is an arbitrary $LPP^S_{Meas,Neat}$-model, $T \vdash_{Ax_{LPP^S}} p_0 \rightarrowtail \bot$. Note that the above explanation why $\mu([p_0]) = 0$ does not hold in the case when the range of the probability is the unit interval of R^* because R^* is ω_1-saturated (which means that the intersection of any countable decreasing sequence of nonempty internal sets must be nonempty). As a consequence, thanks to the restricted ranges of probabilities that are allowed in $LPP^S_{Meas,Neat}$-class of models, our system goes beyond the system P, when we consider infinite default bases.

There are some weaknesses of the system P. The most notable are that it suffers from the problems of irrelevance and inheritance blocking from classes to exceptional subclasses (Benferhat, Saffiotti, & Smets 2000; Goldszmidt & Pearl 1996). To overcome these problems the following approach is usually taken: considering a default base, one determines a subset of the corresponding class of models and reasons about the behavior of defaults in that subclass only. For example, the system Z (Goldszmidt & Pearl 1996) is based on the class which contains only one model of the considered default base. That model is distinguished because the corresponding ranking function is minimal. In the system Z the irrelevance problem is correctly addressed. Similarly, the system Z* from (Goldszmidt & Pearl 1996) solves the drawback of the inheritance blocking. Since ranking functions that are used in those systems are qualitative approximations of nonstandard probabilities, it is not hard to see, having in mind the procedure for the probability calculation mentioned in the proof of Theorem 4, that those systems can be described in the semantical framework given by our approach.

All the above remarks do not take into account that the language of our system is rich enough not only to express formulas that represents defaults but also to describe more: probabilities of formulas, negations of defaults, combinations of defaults with the other (probabilistic) formulas etc. Let us now considerer some situations where these possibilities allow us to obtain more conclusions than in the framework of the language of defaults.

For example, the translation of rational monotonicity, $((\alpha \rightarrowtail \beta) \wedge \neg(\alpha \rightarrowtail \neg\gamma)) \rightarrow ((\alpha \wedge \gamma) \rightarrowtail \beta)$, is $LPP^S_{Meas,Neat}$-valid since rational monotonicity is satisfied in every R^*-probabilistic model, and $LPP^S_{Meas,Neat}$ is a subclass of that class of models. The same holds for the formula $\neg(true \rightarrowtail false)$ corresponding to another property called normality in (Friedman & Halpern 2001). In (Benferhat, Saffiotti, & Smets 2000) the following example is presented. Let the default base consist of the following two defaults $s \rightarrowtail b$ and $s \rightarrowtail t$, where s, b and t means swedes, blond and tall, respectively. Because of the inheritance blocking problem, in some systems (for example in P) it is not possible to conclude that swedes who are not tall are blond ($(s \wedge \neg t) \rightarrowtail b$). Since our system and P coincide if the default base is finite, the same holds in our framework. In fact, there are some $LPP^{S}_{Meas,Neat}$-models in which the previous formula is not satisfied. Avoiding a discussion of intuitive acceptability of the above conclusion, we point out that by adding an additional assumption to the default base we can entail that conclusion too. The assumption says that the probability of $s \wedge \neg t \wedge \neg b$ is at least an order of magnitude lower than the probability of $s \wedge \neg t$, which means that there are significantly more short swedes than short swedes that are not blond. It can be expressed in our logic as the set $\{P_{\leq r}(s \wedge \neg t) \rightarrow P_{\leq r/n}(s \wedge \neg t \wedge \neg b) :$ for every $r \in S, n > 0\}$. An easy calculation which respects that assumption shows that $(s \wedge \neg t) \rightarrowtail b$ follows from the new knowledge base.

Conclusion

In this paper we consider a language, a class of probabilistic model and a sound and complete axiomatic system (at a price of introducing infinitary deduction rules). In the formalization most parts of field theory are moved to the meta theory, so the axioms are rather simple. Our system allow us to model default reasoning. The corresponding entailment is characterized by the following:

- if we consider the language of defaults and finite default bases, the entailment coincides with the one in the system P,
- if we consider the language of defaults and arbitrary default bases, more conclusions can be obtained in our system than in the system P,
- when we consider our full language, we can express rational monotonicity, normality and the other properties that can not be formulated in the usual systems for default reasoning,
- it is not sensitive to the syntactical form which represents the available knowledge (for example, duplications of rules in the knowledge base).

There are many possible directions for further investigations. First of all, the question of decidability of our logic naturally arises. We believe that the ideas from (Ognjanović & Rašković 2000) can help us in obtaining axiomatization of the logic with higher order conditional probabilities and the corresponding first order logic. It would be interesting to compare such a logic and the random world approach from (Grove, Halpern, & Koller 1996). Finally, although some approaches which avoid the problems of irrelevance and inheritance blocking may be (for someone) not completely intuitively acceptable, it is clear that deeper comparison of our probabilistic framework and those systems can help us to better understand default entailment proposed in this paper.

References

Adams, E. W. 1975. *The logic of Conditional*. Dotrecht: Reidel.

Alechina, N. 1995. Logic with probabilistic operators. In *Proc. of the ACCOLADE '94*, 121–138.

Benferhat, S.; Dubois, D.; and Prade, H. 1997. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence* (92):259–276.

Benferhat, S.; Saffiotti, A.; and Smets, P. 2000. Belief functions and default reasoning. *Artificial Intelligence* (122):1–69.

Burghess, J. 1981. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of formal Logic* 22(1):76–84.

Đorđević, R.; Rašković, M.; and Ognjanović, Z. 2004. Completeness theorem for propositional probabilistic models whose measures have only finite ranges. *to appear in Archive for Mathematical Logic*.

Fagin, R., and Halpern, J. 1994. Reasoning about knowledge and probability. *Journal of the ACM* 41(2):340 – 367.

Fagin, R.; Halpern, J.; and Megiddo, N. 1990. A logic for reasoning about probabilities. *Information and Computation* 87(1-2):78–128.

Friedman, N., and Halpern, J. 2001. Plausibility measures and default reasoning. *Journal of the ACM* 48(6):648 – 685.

Goldszmidt, M., and Pearl, J. 1996. Qualitative probabilities for default reasoning, belief revision and causal modeling. *Artificial Intelligence* 84(1-2):57 – 112.

Grove, F. B. A. J.; Halpern, J.; and Koller, D. 1996. From statistical knowledge bases to degrees of belief. *Artificial Intelligence* 87(1-2):75–143.

Keisler, J. 1986. *Elementary calculus. An infinitesimal approach. 2nd ed.* Boston, Massachusetts: Prindle, Weber & Schmidt.

Kraus, S.; Lehmann, D.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44:167 – 207.

Lehmann, D., and Magidor, M. 1992. What does a conditional knowledge base entail? *Artificial Intelligence* 55:1 – 60.

Marković, Z.; Ognjanović, Z.; and Rašković, M. 2003. A probabilistic extension of intuitionistic logic. *Mathematical Logic Quarterly* 49:415–424.

Ognjanović, Z., and Rašković, M. 1999. Some probability logics with new types of probability operators. *Journal of Logic and Computation* 9(2):181 – 195.

Ognjanović, Z., and Rašković, M. 2000. Some first-order probability logics. *Theoretical Computer Science* 247(1-2):191 – 212.

Rašković, M. 1993. Classical logic with some probability operators. *Publications de l'Institut Mathématique, Nouvelle Série, Beograd* (53(67)):1 – 3.

Robinson, A. 1966. *Non-standard analysis*. Amsterdam: North-Holland.

Satoh, K. 1990. A probabilistic interpretation for lazy non-monotonic reasoning. In *Proceedings of the Eighth American Conference on Artificial Intelligence*, 659 – 664.

Appendix. Proof of Theorem 1

Soundness of our system follows from the soundness of propositional classical logics, as well as from the properties of probabilistic measures. The arguments are of the type presented in the proof of Theorem 13 in (Marković, Ognjanović, & Rašković 2003). In the proof of the completeness theorem the following strategy is applied. We start with a form of the deduction theorem. In the next step we show how to extend a consistent set T of formulas to a maximal consistent set T^*. Finally, a canonical $LPP^S_{Meas,Neat}$-model M is constructed out of the formulas from the set T^* such that $M \models \varphi$ iff $\varphi \in T^*$.

Theorem 6 (Deduction theorem) *If T is a set of formulas and $T \cup \{\varphi\} \vdash \psi$, then $T \vdash \varphi \to \psi$, where either $\varphi, \psi \in For_C$ or $\varphi, \psi \in For^S_P$.*

Proof. We use the transfinite induction on the length of the proof of ψ from $T \cup \{\varphi\}$. For example, let $\psi = C \to \bot$ is obtained from $T \cup \{\varphi\}$ by an application of Rule 3, and $\varphi \in For^S_P$. Then:

$T, \varphi \vdash C \to P_{\neq s}\delta$, for every $s \in S$

$T \vdash \varphi \to (C \to P_{\neq s}\delta)$, for every $s \in S$, by the induction hypothesis

$T \vdash (\varphi \wedge C) \to P_{\neq s}\delta$, for every $s \in S$

$T \vdash (\varphi \wedge C) \to \bot$, by Rule 3

$T \vdash \varphi \to \psi$.

The other cases follow similarly. □

A consistent set T of formulas is said to be maximal consistent if the following holds:

- for every $\alpha \in For_C$, if $T \vdash \alpha$, then $\alpha \in T$ and $P_{\geq 1}\alpha \in T$, and
- for every $A \in For^S_P$, either $A \in T$ or $\neg A \in T$.

A set T is deductively closed if for every $\varphi \in For^S$, if $T \vdash \varphi$, then $\varphi \in T$.

Theorem 7 *Every consistent set T can be extended to a maximal consistent set.*

Proof. Let T be a consistent set, $Cn_C(T)$ the set of all classical formulas that are consequences of T, A_0, A_1, ... an enumeration of all formulas from For^S_P and α_0, α_1, ... an enumeration of all formulas from For_C. We define a sequence of sets T_i, $i = 0, 1, 2, \ldots$ such that:

1. $T_0 = T \cup Cn_C(T) \cup \{P_{\geq 1}\alpha : \alpha \in Cn_C(T)\}$
2. for every $i \geq 0$, if $T_{2i} \cup \{A_i\}$ is consistent, then $T_{2i+1} = T_{2i} \cup \{A_i\}$; otherwise, if A_i is of the form $A \to CP_{=s}(\alpha, \beta)$, then $T_{2i+1} = T_{2i} \cup \{\neg A_i, A \to \neg(P_{=st}(\alpha \wedge \beta) \leftrightarrow P_{=t}\beta)\}$, for some $t > 0$; otherwise, if A_i is of the form $A \to CP_{\approx 1}(\alpha, \beta)$, then $T_{2i+1} = T_{2i} \cup \{\neg A_i, A \to \neg CP_{>r}(\alpha, \beta)\}$, for some rational number $r \in [0, 1)$; otherwise, $T_{2i+1} = T_{2i} \cup \{\neg A_i\}$,
3. for every $i \geq 0$, $T_{2i+2} = T_{2i+1} \cup \{P_{=r}\alpha_i\}$, for some $r \in S$, so that T_{2i+2} is consistent,
4. for every $i \geq 0$, if T_i is enlarged by a formula of the form $P_{=0}\alpha$, add $\neg\alpha$ to $T_i \cup \{P_{=0}\alpha\}$ as well.

We have to show that every T_i is a consistent set. T_0 is consistent because it is a set of consequences of a consistent set. Suppose that T_{2i+1} is obtained by the step 2 of the above construction and that neither $T_{2i} \cup \{A_i\}$, nor $T_{2i} \cup \{\neg A_i\}$ are consistent. It follows by the deduction theorem that $T_{2i} \vdash A_i \wedge \neg A_i$, which is a contradiction. The proof proceeds by analyzing possible cases in the other steps of the construction. We continue by showing that the set $T^* = \cup_i T_i$ is a deductively closed set which does not contain all formulas, and, as a consequence, that T^* is consistent. For example, if a formula $\alpha \in For_C$, by the construction of T_0, α and $\neg\alpha$ cannot be simultaneously in T_0, and if $T^* \vdash \alpha$, then by the construction of T_0, $\alpha, P_{\geq 1}\alpha \in T^*$. Finally, according to the above definition of a maximal set, the construction guarantees that T^* is maximal. □

Now, using T^* we can define a tuple $M = \langle W, \{[\alpha]_M : \alpha \in For_C\}, \mu, v\rangle$, where:

- $W = \{w \models Cn_C(T)\}$ contains all the classical propositional interpretations that satisfy the set $Cn_C(T)$ of all classical consequences of the set T,
- $[\alpha]_M = \{w \in W : w \models \alpha\}$,
- for every world w and every propositional letter $p \in \text{Var}$, $v(w)(p) = true$ iff $w \models p$, and
- μ is defined on $\{[\alpha]_M : \alpha \in For_C\}$ by $\mu([\alpha]_M) = s$ iff $P_{=s}\alpha \in T^*$.

The next theorem states that M is an $LPP^S_{Meas,Neat}$-model.

Theorem 8 *Let $M = \langle W, \{[\alpha]_M : \alpha \in For_C\}, \mu, v\rangle$ be defined as above. Then, the following hold:*

1. *μ is a well-defined function.*
2. *$\{[\alpha]_M : \alpha \in For_C\}$ is an algebra of subsets of W.*
3. *μ is a finitely additive probability measure.*
4. *for every $\alpha \in For_C$, $\mu([\alpha]_M) = 0$ iff $[\alpha]_M = \emptyset$.*

Proof of Theorem 1. The ($\Leftarrow$)-direction follows from the soundness of the above axiomatic system. In order to prove the ($\Rightarrow$)-direction we construct the $LPP^S_{Meas,Neat}$-model M as above, and show that for every $\varphi \in For^S$, $M \models \varphi$ iff $\varphi \in T^*$. For example, let $\varphi = CP_{\approx 1}(\alpha, \beta)$. Suppose that $CP_{\approx 1}(\alpha, \beta) \in T^*$. If $\mu([\beta]_M) = 0$, it follows that $M \models CP_{\approx 1}(\alpha, \beta)$. Next, suppose that $\mu([\beta]_M) \neq 0$. From Axiom 14, we have that for every rational $r \in [0, 1)$ $CP_{\geq r}(\alpha, \beta) \in T^*$, and $CP_{=r}(\alpha, \beta) \notin T^*$. It means that for every rational $r \in [0, 1)$, $M \models CP_{\geq r}(\alpha, \beta)$, and $M \not\models CP_{=r}(\alpha, \beta)$, i.e. that for every positive integer n, $\frac{\mu([\alpha \wedge \beta]_M)}{\mu([\beta]_M)} \geq 1 - \frac{1}{n}$. It follows that $M \models CP_{\approx 1}(\alpha, \beta)$. Let $CP_{\approx 1}(\alpha, \beta) \notin T^*$. If $\mu([\beta]_M) = 0$, then $M \models CP_{=1}(\alpha, \beta)$, $CP_{=1}(\alpha, \beta) \in T^*$, and using Axiom 15 $CP_{\approx 1}(\alpha, \beta) \in T^*$, a contradiction. Thus, let $\mu([\beta]_M) \neq 0$. By the step 2 of the construction of T^*, there is some rational number $r \in [0, 1)$ such that $\neg CP_{>r}(\alpha, \beta) \in T^*$. It means that there is some rational number $r \in [0, 1)$ such that $M \not\models CP_{\geq r}(\alpha, \beta)$, and it does not hold that for every positive integer n, $\frac{\mu([\alpha \wedge \beta]_M)}{\mu([\beta]_M)} \geq 1 - \frac{1}{n}$. Thus, we have that $M \not\models CP_{\approx 1}(\alpha, \beta)$. □