Proceedings ECAI'98, pp. 580-584, Wiley & sons 1998

Nonmonotonic reasoning in probabilistics

Gabriele Kern-Isberner[1]

Abstract. In probabilistics, reasoning at optimum entropy (ME-reasoning) has proved to be a most sound and consistent method for inference. This paper investigates its properties in the framework of nonmonotonic reasoning. In particular, we show that ME-reasoning satisfies *cumulativity* and *loop*, and we focus on its connection to conditionals. Finally, we introduce the notion of a *universal inference operation* to generalize nonmonotonic reasoning, thus providing a frame more appropriate for linking nonmonotonic reasoning with belief revision.

1 Introduction

The capability of revising knowledge and giving up conclusions in the light of conflicting evidence is one of the most outstanding features of commonsense reasoning. Though it seems to be practised in everyday life in a most natural and self-evident way, it challenges knowledge representation and inference procedures in AI because it clashes with the classical property of monotony. Therefore *defeasible reasoning*, as it is usually called, requires new formalisms to be realized adequately, and a lot of approaches to "nonmonotonic logics" have been made up to now. Makinson and others (cf. e.g. [8], [7]) set forth formal properties and axiom systems to judge and classify inference relations lacking monotony. Makinson's work also includes quite general inference procedures not being based on classical structures.

In particular, any *probabilistic logic*, i.e. any logic based on probabilistic reasoning, is non-classical with respect to knowledge representation as well as with respect to inference techniques. Usually, qualitative probabilistic approaches prove to be quite well-behaved, as e.g. Adams' $\epsilon-$semantics (cf. [1]; cf. [8], [7]). In contrast to this, quantified probabilistic reasoning appears to be much harder to be controlled and classified. Paris [10] set up a number of postulates for consistent reasoning with probabilistic beliefs, and it turns out that the only inference process that satisfies all of them is that being based on the *principles of maximum entropy* resp. *of minimum cross-entropy (ME-principles)*. Two other characterizations [13], [5] emphasize the outstanding position of ME-reasoning, too.

This paper investigates how infering at optimum entropy fits the formal framework for nonmonotonic reasoning of [8]. In particular, we show that any inference operation based on ME-reasoning is *cumulative* and satisfies *loop*. Moreover, it turns out to be *supraclassical* with respect to *standard probabilistic consequence* which obviously generalizes classical consequence within a probabilistic setting. We also focus on the relationships between nonmonotonic and conditional ME-reasoning. Once more, it becomes obvious that material implication and conditionals differ substantially. To make the differences clear, we extend the conditional probabilistic language we are working in so as to contain probabilistic formulas corresponding to material implication, too. We show that conditionalization in its usual sense relates to material implication, whereas the connections between nonmonotonic reasoning and conditionals are more complex.

Though the ME-methods genuinely manipulate knowledge of a numerical nature, ME-reasoning is not easily understood by observing the probabilities in change. ME-logic is not truth-functional, as fuzzy logic is, nor is its aim to raise or to lower probabilities, as in the framework of upper and lower probabilities, and there is no straightforward calculation algorithm, as for Bayesian networks. ME-infering rather makes use of the intensional structures of probabilistic knowledge (cf. [10],[13], [5]), so it seems to be better classified and appreciated by describing its formal properties as a nonmonotonic inference operation.

As a (full) revision operation, ME-infering exceeds the framework of nonmonotonic logics and actually should be considered under the somewhat wider point of view provided by *belief revision.* Though both formalisms are regarded as very closely related (cf. [9]) the crucial difference between them is that belief revision (also called *theory change*) substantially need a theory to base inferences on, whereas no such reference to prior knowledge occurs explicitly in nonmonotonic reasoning. Inspired by the possibilities of ME-reasoning, in particular by a property called *logical consistency* in [5] and here stated in theorem 2, we generalize the notion of an inference operation and introduce *universal inference operations* as a suitable counterpart to revision operators in nonmonotonic logics. We show that *logical consistency* may be considered as a strong version of *cumulativity* for universal inference operations. Using this more general framework not only allows a more accurate representation of belief revision via nonmonotonic formalisms, but gives also, vice versa, an important impetus to handle iterated revisions.

This paper makes use of the language $\mathcal{L}^{\sim}$ of probabilistic conditionals which we introduce in section 2, partly because of its intuitiveness, partly because the characterization of ME-reasoning in [5] is based on it. In fact, most of the results stated here will continue to hold if instead a more general probabilistic language (e.g. that of *linear probabilistic constraints*, cf. [10]) is used. In section 3 we define the *standard probabilistic consequence operation* as an appropriate general-

[1] Dept. of Computer Science, FernUniversitaet Hagen, P.O. Box 940, D-58084 Hagen, Germany; e-mail: Gabriele.Kern-Isberner@fernuni-hagen.de

ECAI 98. *13th European Conference on Artificial Intelligence*
Edited by Henri Prade
Published in 1998 by John Wiley & Sons, Ltd.

ization of classical logics within a probabilistic setting, and we fix the formal notations of *probabilistic inference operations*. Section 4 describes briefly the foundations of reasoning at optimum entropy, and in section 5 we investigate some basic properties of the corresponding ME-inference operation. The connections between ME-inference and conditional resp. material implication is the subject of section 6. Finally, section 7 presents a more general concept of nonmonotonic inference operations and a more comprehensive connection between the theories of nonmonotonic reasoning and of belief revision.

2 Probabilistic logics

Suppose $\mathcal{V} = \{V_1, V_2, V_3, \ldots\}$ is a (fixed) finite set of propositional variables V_i which are assumed to be binary, and let $Dist(\mathcal{V})$ denote the set of all probability distributions P over $\mathcal{V}$. Let $\mathcal{L} = \mathcal{L}(\mathcal{V})$ be the propositional language with alphabet $\mathcal{V}$ and logical connectives $\vee, \wedge$ resp. juxtaposition and $\neg$. Elementary events ω over $\mathcal{V}$ symbolizes possible worlds, and thus each formula $\mathrm{A} \in \mathcal{L}$ corresponds to an event with probability $P(\mathrm{A}) = \sum_{\omega:\mathrm{A}(\omega)=1} P(\omega)$, where $\mathrm{A}(\omega) = 1$ resp. $\mathrm{A}(\omega) = 0$ means that A is true resp. false in the world represented by ω. $\mathcal{L}$ is extended to a probabilistic conditional language $\mathcal{L}^{\leadsto} = \mathcal{L}^{\leadsto}(\mathcal{V})$ by introducing a conditional operator $\leadsto$: A *probabilistic conditional* is an expression $\mathrm{A} \leadsto \mathrm{B}[x]$ with *antecedent* $\mathrm{A} \in \mathcal{L}$, *conclusion* $\mathrm{B} \in \mathcal{L}$ and probability $x \in [0,1]$. Probabilistic conditionals are interpreted via conditional probabilities: If P is a distribution, we write $P \models \mathrm{A} \leadsto \mathrm{B}[x]$ iff $P(\mathrm{A}) > 0$ and $P(\mathrm{B}|\mathrm{A}) = \dfrac{P(\mathrm{AB})}{P(\mathrm{A})} = x$. A *probabilistic fact* $\mathrm{A}[x]$ is regarded as equivalent to the conditional $\top \leadsto \mathrm{A}[x]$ with tautological antecedent, so $P \models \mathrm{A}[x]$ iff $P(\mathrm{A}) = x$.

Probabilistic facts and conditionals are also denoted by small Greek letters ϕ, ψ etc. So for a set $\mathcal{R} \subset \mathcal{L}^{\leadsto}$ of probabilistic conditionals, we write $P \models \mathcal{R}$ iff $P \models \phi$ for all $\phi \in \mathcal{R}$. *Models* of probabilistic conditionals are probability distributions that fulfill them, hence $Mod(\mathcal{R}) := \{Q | Q \in Dist(\mathcal{V}), Q \models \mathcal{R}\}$ for $\mathcal{R} \subset \mathcal{L}^{\leadsto}(\mathcal{V})$.

For $P \in Dist(\mathcal{V})$, let

$$Th(P) = \{\mathrm{A} \leadsto \mathrm{B}[x] \in \mathcal{L}^{\leadsto}(\mathcal{V}) | P \models \mathrm{A} \leadsto \mathrm{B}[x]\}$$

denote the set of all probabilistic conditionals which are valid in P. A set of conditionals $\mathcal{R} \subset \mathcal{L}^{\leadsto}(\mathcal{V})$ is *consistent* iff it has a probabilistic model, i.e. iff there is a $Q \in Dist(\mathcal{V})$ such that $Q \models \mathcal{R}$. For a (fixed) $P \in Dist(\mathcal{V})$, $\mathcal{R}$ is said to be *P-consistent* iff there is a distribution $Q \in Dist(\mathcal{V})$ with $Q \ll P$ (meaning Q is P-continuous, i.e. $P(\omega) = 0$ implies $Q(\omega) = 0$) and $Q \models \mathcal{R}$. P-consistency ensures a compatibility between P and $\mathcal{R}$. In particular, a P-consistent set $\mathcal{R} \subset \mathcal{L}^{\leadsto}$ does not contain any conditional $\mathrm{A} \leadsto \mathrm{B}[x]$ with $P(\mathrm{A}) = 0$.

Two distributions $P_1, P_2 \in Dist(\mathcal{V})$ are identical iff $P_1(\omega) = P_2(\omega)$ for all $\omega \in \Omega$, i.e. $P_1 \models \omega[x]$ iff $P_2 \models \omega[x]$. So there is a one-to-one correspondence between distributions P and their theories $Th(P)$, and we will sometimes identify both: $P \equiv Th(P)$.

3 Probabilistic consequence and probabilistic inference

The inference operation Cn of classical logic satisfies the following characteristic conditions:

(i) *inclusion* or *reflexivity:* $X \subseteq Cn(X)$;
(ii) *idempotence:* $Cn(X) = Cn(Cn(X))$;
(iii) *cut:* $X \subseteq Y \subseteq Cn(X)$ implies $Cn(Y) \subseteq Cn(X)$;
(iv) *monotony:* $X \subseteq Y$ implies $Cn(X) \subseteq Cn(Y)$,

where $X, Y \subseteq \mathcal{L}$.

So Cn is a *consequence operation*, in the sense of Tarski. Its rigidity, however, appreciated e.g. in mathematics and establishing its fundamental meaning for logical deduction, makes it a poor candidate to represent commonsense reasoning – in particular, monotony prohibits defeasible conclusions. Nevertheless, many nonstandard logics have elements of classical logic built in in one way or another, and Makinson [8] and Kraus, Lehmann, Magidor [7] make use of it as a reference point to formulate their principles for nonmonotonic reasoning.

Within a probabilistic framework, the classical consequence operation has to be modified appropriately to provide a suitable tool for the description of formal inferences.

Semantically, Cn may be described by $Cn(X) = \{\mathrm{A} \in \mathcal{L} | X \models \mathrm{A}\}$, where $\models$ demands taking account of *every* model of X. So we define the *standard probabilistic consequence operation* $Cn_p : 2^{\mathcal{L}^{\leadsto}} \to 2^{\mathcal{L}^{\leadsto}}$ by virtue of

$$Cn_p(\mathcal{R}) = \{\phi \in \mathcal{L}^{\leadsto} | P \models \phi \text{ for all } P \in Mod(\mathcal{R})\},$$

with associated *standard probabilistic consequence relation* $\models^p \subset 2^{\mathcal{L}^{\leadsto}} \times \mathcal{L}^{\leadsto}$

$$\mathcal{R} \models^p \phi \text{ iff } \phi \in Cn_p(\mathcal{R}).$$

If $\mathcal{R} \subset \mathcal{L}^{\leadsto}$ is inconsistent, then $Mod(\mathcal{R}) = \emptyset$, hence $Cn_p(\mathcal{R}) = \mathcal{L}^{\leadsto}$ for inconsistent $\mathcal{R}$. Paris [10] presents a sequent calculus for $\models^p$.

Proposition 1 *Cn_p satisfies inclusion, idempotence, cut and monotony.*

Thus Cn_p takes the part of a classical logic within probabilistic logics. In general, however, probabilistic models are ill-famed for their extreme flexibility, so an inference operation based on *all* possible models is much too conservative and will yield little more but quite trivial results. In analogy to [8], we define *probabilistic inference operations* resp. *relations* to study nonclassical inferences:

Definition 1 *A* probabilistic inference operation *is an operation $C : 2^{\mathcal{L}^{\leadsto}} \to 2^{\mathcal{L}^{\leadsto}}$ on sets of probabilistic conditionals. Its associated* probabilistic inference relation *is given by $\mid\!\sim \; \subset 2^{\mathcal{L}^{\leadsto}} \times \mathcal{L}^{\leadsto}$, $\mathcal{R} \mid\!\sim \phi$ iff $\phi \in C(\mathcal{R})$. C is called* complete *iff it specifies for each set $\mathcal{R} \subset \mathcal{L}^{\leadsto}$ with $C(\mathcal{R}) \neq \emptyset, \mathcal{L}^{\leadsto}$ a unique distribution, i.e. iff there is a (unique) distribution $Q_{\mathcal{R}}$ such that $C(\mathcal{R}) = Th(Q_{\mathcal{R}})$.*

4 Infering at optimum entropy

The results of several papers (cf. [11], [13], [5]) show that making use of the principles of maximum resp. minimum (cross-) entropy cuts down the arbitrariness of probability in a reasonable way and pinpoints all available probabilistic knowledge in one distribution, the so-called *ME-distribution* (where ME abbreviates <u>M</u>aximum <u>E</u>ntropy or <u>M</u>inimum cross-<u>E</u>ntropy, respectively). It is even possible to take (fixed) prior knowledge P into account, so that ME-inferences based on P may be considered. For further information and references, confer any of the papers cited above; for an axiomatization based on probabilistic conditionals, cf. [5].

Let $*_e$ denote the *ME-adjustment operator*, i.e. $*_e$ assigns to each (prior) distribution P and each P-consistent set $\mathcal{R}$ of probabilistic conditionals the (unique) distribution $P_e = P *_e \mathcal{R}$ that has minimum cross-entropy to P among all models of $\mathcal{R}$. That is to say, P_e solves the minimization problem

$$\min_{Q \in Mod(\mathcal{R})} R(Q,P) = \sum_{\omega} q(\omega) \log \frac{q(\omega)}{p(\omega)}. \qquad (1)$$

If $\mathcal{R}$ is P-consistent, then P_e exists and is unique (cf. [2]).

For finite P-consistent sets $\mathcal{R} \subset \mathcal{L}^{\leadsto}$, $\mathcal{R} = \{\mathrm{A}_1 \leadsto \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \leadsto \mathrm{B}_n[x_n]\}$, P_e has a representation in the form

$$P_e(\omega) = \alpha_0 P(\omega) \prod_{\substack{1 \le i \le n \\ \mathrm{A}_i \mathrm{B}_i(\omega)=1}} \alpha_i^{1-x_i} \prod_{\substack{1 \le i \le n \\ \mathrm{A}_i \overline{\mathrm{B}_i}(\omega)=1}} \alpha_i^{-x_i}, \qquad (2)$$

where the α_i's are suitably chosen so as to make P_e fulfill $\mathcal{R}$, i.e. they are solutions of the nonlinear equations

$$\alpha_i = \frac{x_i}{1-x_i} \cdot \qquad (3)$$

$$\cdot \frac{\displaystyle\sum_{\omega:\mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1} P(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j \mathrm{B}_j(\omega)=1}} \alpha_j^{1-x_j} \prod_{\substack{j \neq i \\ \mathrm{A}_j \overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^{-x_j}}{\displaystyle\sum_{\omega:\mathrm{A}_i\mathrm{B}_i(\omega)=1} P(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j \mathrm{B}_j(\omega)=1}} \alpha_j^{1-x_j} \prod_{\substack{j \neq i \\ \mathrm{A}_j \overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^{-x_j}},$$

$$\text{with} \quad \alpha_i \begin{cases} > 0 & : \quad x_i \in (0,1) \\ = \infty & : \quad x_i = 1 \\ = 0 & : \quad x_i = 0 \end{cases} \quad , 1 \le i \le n,$$

using the conventions $\infty^0 = 1$, $\infty^{-1} = 0$ and $0^0 = 1$; α_0 simply arises as a normalizing factor (cf. [5]).

For fixed prior P define the *ME-inference operation* $C_P^e : 2^{\mathcal{L}^{\leadsto}} \to 2^{\mathcal{L}^{\leadsto}}$ by

$$C_P^e(\mathcal{R}) = \begin{cases} Th(P *_e \mathcal{R}) & \text{iff } \mathcal{R}\ P-\text{consistent} \\ \mathcal{L}^{\leadsto} & else \end{cases}$$

Thus C_P^e is a complete inference operation.

In [4], some derivation schemes of ME-inference are proved. In general, infering at maximum entropy may be realized by using the ME-system shell SPIRIT (cf. [12]).

5 Basic properties of the ME-inference operation

According to [11], [14], [5], the ME-adjustment operator $*_e$ satisfies the following two properties:

Theorem 2 *Let P be a (prior) distribution, and suppose $\mathcal{R}, \mathcal{S} \subseteq \mathcal{L}^{\leadsto}$ are two sets of probabilistic conditionals.*

(i) $P *_e \mathcal{R} = P \quad \Leftrightarrow \quad P \models \mathcal{R}$.
(ii) $(P *_e \mathcal{R}) *_e (\mathcal{R} \cup \mathcal{S}) = P *_e (\mathcal{R} \cup \mathcal{S})$.

Property (ii) above is named *logical consistency* in [5]. In that paper, it serves as a characterizing property of the ME-adjustment operator.

Proposition 3 *The ME-inference operation C_P^e satisfies*

(i) *inclusion:* $\mathcal{R} \subseteq C_P^e(\mathcal{R})$.
(ii) *idempotence:* $C_P^e(C_P^e(\mathcal{R})) = C_P^e(\mathcal{R})$.
(iii) *cumulativity:* $\mathcal{R} \subseteq \mathcal{S} \subseteq C_P^e(\mathcal{R})$ *implies* $C_P^e(\mathcal{R}) = C_P^e(\mathcal{S})$.

For a proof of proposition 3, cf. [5].

Cumulativity combines cut and *cautious monotony*, i.e. $\mathcal{R} \subseteq \mathcal{S} \subseteq C_P^e(\mathcal{R})$ implies $C_P^e(\mathcal{R}) \subseteq C_P^e(\mathcal{S})$. Given inclusion, cumulativity is equivalent to *reciprocity*, i.e. if $\mathcal{R} \subseteq C_P^e(\mathcal{S})$ and $\mathcal{S} \subseteq C_P^e(\mathcal{R})$ then $C_P^e(\mathcal{R}) = C_P^e(\mathcal{S})$ (cf. [8]).

In the works of Makinson [8] and Kraus, Lehmann and Magidor [7], cumulativity takes a central position as a very fundamental property of inference processes: cumulativity ensures the inference process to be stable, adding of derivable knowledge does not alter the set of nonmonotonic conclusions. As stated above (following [8]), cumulativity is a *pure* condition, without any reference to logical connectives in the underlying language and thus being applicable to probabilistic logics in a straightforward way. Besides the numerous postulates for nonmonotonic inference relations in [8] and [7] that are based on classical structures, there is another interesting pure condition called *loop*:

Theorem 4 *C_P^e satisfies the* loop-property*: If $\mathcal{R}_1, \ldots, \mathcal{R}_m \subseteq \mathcal{L}^{\leadsto}$ with $\mathcal{R}_{i+1} \subset C_P^e(\mathcal{R}_i) mod(m)$ then $C_P^e(\mathcal{R}_i) = C_P^e(\mathcal{R}_j)$ for all $i, j = 1, \ldots m$.*

ME-inference $\mid\!\sim_P^e$ is not transitive, but loop guarantees unambiguity for a sequence $\mathcal{R}_1 \mid\!\sim_P^e \mathcal{R}_2 \mid\!\sim_P^e \ldots \mid\!\sim_P^e \mathcal{R}_m \mid\!\sim_P^e \mathcal{R}_1$ of nonmonotonic derivations. Theorem 4 is proved by making use of the unique minimum property (1) of P_e.

In the rest of this section, we will investigate interrelationships between ME-inference and standard probabilistic consequence.

Proposition 5 *C_P^e is supraclassical, i.e. $Cn_p(\mathcal{R}) \subseteq C_P^e(\mathcal{R})$ for all $\mathcal{R} \subseteq \mathcal{L}^{\leadsto}$.*

The proof of this proposition is straightforward because $P *_e \mathcal{R}$ either is a model of $\mathcal{R}$, or $C_P^e(\mathcal{R}) = \mathcal{L}^{\leadsto}$.

Supraclassicality implies further important properties: Because C_P^e is cumulative and supraclassical, it also satisfies *full absorption*:

$$Cn_p C_P^e = C_P^e = C_P^e Cn_p.$$

So in particular (cf. [8]), C_P^e fulfills *right weakening*:

$$\mathcal{R} \subseteq C_P^e(\mathcal{S}) \text{ implies } Cn_p(\mathcal{R}) \subseteq C_P^e(\mathcal{S})$$

and *left logical equivalence*:

$$\text{if } Cn_p(\mathcal{R}) = Cn_p(\mathcal{S}) \text{ then } C_P^e(\mathcal{R}) = C_P^e(\mathcal{S}).$$

Therefore ME-inference respects standard deductional structures. But ME-inference fails *distribution*

$$C_P^e(\mathcal{R}) \cap C_P^e(\mathcal{S}) \subseteq C_P^e(Cn_p(\mathcal{R}) \cap Cn_p(\mathcal{S})),$$

as can easily be seen when considering the case $P = P_0(A, B, C)$ (= uniform distribution over three propositional variables A, B and C), $\mathcal{R} = \{a \leadsto b[x_1], a \leadsto c[x_2]\}$ and $\mathcal{S} = \{a \leadsto b[x_2], a \leadsto c[x_1]\}$ with $x_1 \neq x_2$. Then $a \leadsto bc[x_1 x_2] \in C_P^e(\mathcal{R}) \cap C_P^e(\mathcal{S})$ (cf. [4]), but $a \leadsto bc[x_1 x_2] \notin C_P^e(Cn_p(\mathcal{R}) \cap Cn_p(\mathcal{S}))$.

6 ME-Logic and conditionals

In [5] it is shown that ME-reasoning may be characterized as a sound and consistent method for handling probabilistic conditionals. Thus the notion of conditionals is fundamental to ME-logic, and the property of *conditionalization*

$$\text{If } \mathcal{R} \cup \{\phi\} \mathrel{|\!\sim} \psi \text{ then } \mathcal{R} \mathrel{|\!\sim} \phi \rightarrow \psi, \qquad (4)$$

as stated in standard, non-probabilistic terms in [8], is of special interest. Here $\rightarrow$ is meant to represent material implication which has no straightforward counterpart in probabilistic logics. It is clearly different from the operator $\rightsquigarrow$ used for probabilistic conditionals, not only for syntactical differences. The satisfaction relation between a distribution P and a conditional $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$ actually involves an adjustment process, namely shifting P to $P_A = P(\cdot|A)$, whereas material implication should be satisfied *in* P.

In the sequel, we will investigate conditionalization with respect to a suitably generalized material implication $\mathrm{A} \rightarrow \mathrm{B}[x]$ in probabilistics as well as with respect to the proper probabilistic conditionals of the form $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$.

So for the moment, let us extend our language $\mathcal{L}^{\rightsquigarrow}$ so as to also contain all formulas of the type $\mathrm{A} \rightarrow \mathrm{B}[x]$:

$$\mathcal{L}^{\rightsquigarrow,\rightarrow} = \mathcal{L}^{\rightsquigarrow} \cup \{\mathrm{A} \rightarrow \mathrm{B}[x] | \mathrm{A}, \mathrm{B} \in \mathcal{L}, x \in [0,1]\},$$

and we modify C^e_P slightly to take its values in $2^{\mathcal{L}^{\rightsquigarrow,\rightarrow}}$: $C^e_P : 2^{\mathcal{L}^{\rightsquigarrow}} \rightarrow 2^{\mathcal{L}^{\rightsquigarrow,\rightarrow}}$. In correspondence to classical material implication, and in compatibility with the use of the conditional operator $\rightsquigarrow$, we define the semantics of $\mathrm{A} \rightarrow \mathrm{B}[x]$ in a probabilistic setting as

$$P \models \mathrm{A} \rightarrow \mathrm{B}[x] \quad \text{iff} \quad P \models \mathrm{A}[1] \text{ implies } P \models \mathrm{B}[x]$$

($P \in Dist(\mathcal{V})$). Note that $P \models \mathrm{A} \rightarrow \mathrm{B}[x]$ is not equivalent to $P \models (\neg\mathrm{A} \vee \mathrm{B})[x]$, i.e. to $P(\neg\mathrm{A} \vee \mathrm{B}) = x$. Rather we have $\models^p \mathrm{A} \rightarrow \mathrm{B}[x]$ iff $Mod(\mathrm{A}[1]) \subseteq Mod(\mathrm{B}[x])$, and in particular, $\models^p \mathrm{A} \rightarrow \mathrm{B}[x]$ implies $\mathrel{|\!\sim}_P \mathrm{A} \rightsquigarrow \mathrm{B}[x]$ for all $P \in Dist(\mathcal{V})$.

Using this notion of material implication, ME-inference satisfies conditionalization in the sense of (4):

Proposition 6 *Suppose $P \in Dist(\mathcal{V})$. Whenever $\mathrm{B}[x] \in C^e_P(\mathcal{R} \cup \{\mathrm{A}[1]\})$ then $\mathrm{A} \rightarrow \mathrm{B}[x] \in C^e_P(\mathcal{R})$.*

Proof: Suppose $\mathrm{B}[x] \in C^e_P(\mathcal{R} \cup \{\mathrm{A}[1]\})$, i.e. $P *_e (\mathcal{R} \cup \{\mathrm{A}[1]\}) \models \mathrm{B}[x]$. If $P *_e \mathcal{R} \models \mathrm{A}[1]$, then $P *_e (\mathcal{R} \cup \{\mathrm{A}[1]\}) = P *_e \mathcal{R}$, according to theorem 2. Therefore $P *_e \mathcal{R} \models \mathrm{B}[x]$. Hence $P *_e \mathcal{R} \models \mathrm{A} \rightarrow \mathrm{B}[x]$, which was to be proven. □

But note that proposition 6 is false for probabilistic conditionals: in general, $\mathrm{B}[x] \in C^e_P(\mathcal{R} \cup \{\mathrm{A}[1]\})$ does not imply $\mathrm{A} \rightsquigarrow \mathrm{B}[x] \in C^e_P(\mathcal{R})$. Instead, we have the following connections between ME-inference and conditional implication:

Proposition 7 *For any $P \in Dist(\mathcal{V})$, $\mathcal{R} \subset \mathcal{L}^{\rightsquigarrow}$, $\mathrm{A} \rightsquigarrow \mathrm{B}[x] \in \mathcal{L}^{\rightsquigarrow}$, we have*

(i) $(P *_e \{\mathrm{A}[1]\}) = P_\mathrm{A} = P(\cdot|\mathrm{A})$

(ii) $P *_e (\mathcal{R} \cup \{\mathrm{A}[1]\}) = P_\mathrm{A} *_e \mathcal{R}$, *i.e.* $C^e_P(\mathcal{R} \cup \{\mathrm{A}[1]\}) = C^e_{P_\mathrm{A}}(\mathcal{R})$

(iii) $(P *_e \mathcal{R}) *_e \{\mathrm{A}[1]\} \models \mathrm{B}[x]$ *implies* $P *_e \mathcal{R} \models \mathrm{A} \rightsquigarrow \mathrm{B}[x]$.

Proof: (i) shows that ME-infering generalizes Bayesian conditioning. It may easily be proved by using (2) and (3) but can also be found in the works of other authors (cf. e.g. [10]). The proof of (ii) is formal but straightforward from (2) and (3). (iii) follows from (i) and the definitions. □

Proposition 7 shows that $P *_e (\mathcal{R} \cup \{\mathrm{A}[1]\})$ and $(P *_e \mathcal{R}) *_e \{\mathrm{A}[1]\}$ differ in general, and that only the last of these distributions reveals a behavior that is compatible to conditional implication.

Proposition 7 (ii) links up two inference operations based on different probabilistic theories in a very special case. The next section is dedicated to the studying of similar relationships in a more general context.

7 Generalized nonmonotonic reasoning and belief revision

Nonmonotonic reasoning is closely related to belief revision (cf. e.g. [9]). The correspondences between axioms of belief change and properties of nonmonotonic inference operations, however, are usually elaborated only in the case that revisions are based on a fixed theory (cf. [9]), and very few work has been done to incorporate iterated belief revision in that framework.

Within the context of ME-inference, iterated adjustments arise quite naturally and are dealt with in a consistent manner (cf. theorem 2 and proposition 7). The crucial point here is that the ME-operator $*_e$ actually is a full revision operator taking *two* entries, namely a distribution P on its left and a (compatible and consistent) set of conditionals on its right. Nonmonotonic reasoning and belief revision usually focus on its right entry, while considering its left entry – i.e. the theory inferences are based on – to be fixed.

Hitherto we shared this point of view. In this section, however, we will present a wider framework for probabilistic nonmonotonic reasoning by explicitly taking account of the prior distribution P.

Definition 2 *A* universal probabilistic inference operation $\mathbf{C}$ *assigns a complete probabilistic inference operation C_P to each distribution $P \in Dist(\mathcal{V})$: $\mathbf{C} : P \mapsto C_P$. $\mathbf{C}$ is* founded *iff for each distribution P and for any $\mathcal{R} \subset \mathcal{L}^{\rightsquigarrow}$, $P \models \mathcal{R}$ implies $C_P(\mathcal{R}) = Th(P)$. A universal probabilistic inference operation is said to be* reflexive (idempotent, cumulative) *iff all its involved inference operations have the corresponding property.*

If $\mathbf{C} : P \mapsto C_P$ is a universal probabilistic inference operation, C_P is complete for each $P \in Dist(\mathcal{V})$. That means, for each set $\mathcal{R} \subset \mathcal{L}^{\rightsquigarrow}$, $C_P(\mathcal{R})$ is either $\emptyset$ or $\mathcal{L}^{\rightsquigarrow}$, or it specifies a distribution: $C_P(\mathcal{R}) = Th(Q_{P,\mathcal{R}})$. Define the set of all such distributions by

$$Dist(C_P) = \{Q \in Dist(\mathcal{V}) \,|\, \exists \mathcal{R} \subset \mathcal{L}^{\rightsquigarrow} : C_P(\mathcal{R}) = Th(Q)\}.$$

The property of foundedness establishes a close and intuitive relationship between a distribution P and its associated inference operation C_P, distinguishing P as its stable starting point. In particular, if $\mathbf{C}$ is founded then $C_P(\emptyset) = Th(P)$ (this property is called *faithfulness* in [6]). As to the universal inference operation $\mathbf{C}$, foundedness (resp. faithfulness) ensures injectivity, as can easily be proved:

Proposition 8 *If $\mathbf{C}$ is founded, then it is injective.*

In one-dimensional, i.e. standard nonmonotonic reasoning, as it was developed in [8] and [7], cumulativity occupies a central and fundamental position, fixing the inferences of a *set* $\mathcal{S}$ that "lies between" another set $\mathcal{R}$ and its nonmonotonic consequences $C(\mathcal{R})$. To establish a similar well-behavedness of $\mathbf{C}$ with respect to *distributions*, i.e. to *theories*, we introduce suitable relations to compare them with each other.

Definition 3 *Let* $\mathbf{C}: P \mapsto C_P$ *be a universal inference operation. For each distribution* P, *define a relation* $\sqsubseteq_P$ *on* $Dist(C_P)$ *by setting* $P_1 \sqsubseteq_P P_2$ *iff there are sets* $\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \mathcal{L}^{\sim}$ *such that* $P_1 \equiv C_P(\mathcal{R}_1)$ *and* $P_2 \equiv C_P(\mathcal{R}_2)$.

For faithful universal inference operations, i.e. $C_P(\emptyset) = Th(P)$ for all $P \in Dist(\mathcal{V})$, P is a minimal element of $Dist(C_P)$ with respect to $\sqsubseteq_P$:

Proposition 9 *If* $\mathbf{C}$ *is faithful, then for all* $P \in Dist(\mathcal{V})$ *and for all* $P' \in Dist(C_P)$, $P \sqsubseteq_P P'$.

We will now generalize the notion of cumulativity for universal inference relations:

Definition 4 $\mathbf{C}$ *is called* strongly cumulative *iff for each* $P \in Dist(\mathcal{V})$ *and for any distributions* $P_1, P_2 \in Dist(C_P)$, $P \sqsubseteq_P P_1 \sqsubseteq_P P_2$ *implies: whenever* $\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \mathcal{L}^{\sim}$ *such that* $P_1 \equiv C_P(\mathcal{R}_1)$ *and* $P_2 \equiv C_P(\mathcal{R}_2)$, *we have* $P_2 \equiv C_P(\mathcal{R}_2) = C_{P_1}(\mathcal{R}_2)$.

Strong cumulativity describes a relationship between inference operations based on different distributions, resp. probabilistic theories, thus linking up the inference operations of $\mathbf{C}$.

The next proposition states that strong cumulativity actually generalizes cumulativity for an important class of universal inference operations:

Theorem 10 *If* $\mathbf{C}$ *is founded, then strong cumulativity implies cumulativity.*

Proof: Suppose $\mathbf{C}$ to be founded and strongly cumulative. Let $P \in Dist(\mathcal{V})$, and let $\mathcal{R}, \mathcal{S} \subset \mathcal{L}^{\sim}$ with $\mathcal{R} \subset \mathcal{S} \subset C_P(\mathcal{R})$. To prove cumulativity for C_P, we have to show $C_P(\mathcal{R}) = C_P(\mathcal{S})$. Set $P_1 \equiv C_P(\mathcal{R}), P_2 \equiv C_P(\mathcal{S})$. Because of $\mathcal{R} \subset \mathcal{S}$, we have $P_1 \sqsubseteq_P P_2$. Strong cumulativity now implies $P_2 \equiv C_P(\mathcal{S}) = C_{P_1}(\mathcal{S})$. Moreover, we presupposed $\mathcal{S} \subset C_P(\mathcal{R})$, therefore $P_1 \models \mathcal{S}$. Because of foundedness, $C_{P_1}(\mathcal{S}) \equiv P_1$. Thus $P_1 = P_2$, i.e. $C_P(\mathcal{R}) = C_P(\mathcal{S})$. □

To conclude this section, let us return to the framework of belief revision. Universal inference operations $\mathbf{C}: P \mapsto C_P$ are a proper counterpart of revision operators $*$ in nonmonotonic reasoning by virtue of $P * \mathcal{R} = C_P(\mathcal{R})$ for $P \in Dist(\mathcal{V}), \mathcal{R} \subseteq \mathcal{L}^{\sim}$. Using this notation, foundedness means $P * \mathcal{R} = P$ for $P \models \mathcal{R}$, and strong cumulativity is equivalent to $P * (\mathcal{R} \cup \mathcal{S}) = (P * \mathcal{R}) * (\mathcal{R} \cup \mathcal{S})$. Thus both properties may serve to control iterated revisions.

Theorem 2 shows that ME-inference defines a founded and strongly cumulative universal operation $\mathbf{C}^e: P \mapsto C_P^e$.

Following the intention of the present paper, this section only deals with inferences within a probabilistic framework. But actually, all of the definitions given above use probabilistic structures in quite a formal manner. So analogous results may be obtained as well for other nonmonotonic theories taking the role of belief states. It is worth noticing that strong cumulativity, introduced here as a generalization of cumulativity for binary probabilistic revision operators and satisfied by ME-inference, was quite recently proposed as a postulate for iterated belief change in general (cf. [3]).

8 Conclusion

This papers centers round ME-reasoning as a nonmonotonic inference operation. Based on the work of Makinson [8], we prove its well-behavedness and point out interesting connections to conditional reasoning. As a final highlight, we extend the framework of nonmonotonic reasoning so as to be more adequate for representing its connections to belief revision. Though this is done in a purely probabilistic setting, analogous results may be proved for universal inference operations based on non-probabilistic theories.

REFERENCES

[1] E.W. Adams, *The Logic of Conditionals*, D. Reidel, Dordrecht, 1975.

[2] I. Csiszár, 'I-divergence geometry of probability distributions and minimization problems', *Ann. Prob.*, **3**, 146–158, (1975).

[3] A. Darwiche and J. Pearl, 'On the logic of iterated belief revision', *Artificial Intelligence*, **89**, 1–29, (1997).

[4] G. Kern-Isberner, 'A logically sound method for uncertain reasoning with quantified conditionals', in *Proceedings First International Conference on Qualitative and Quantitative Practical Reasoning, ECSQARU-FAPR'97*, pp. 365 – 379, Berlin Heidelberg New York, (1997). Springer.

[5] G. Kern-Isberner, 'Characterizing the principle of minimum cross-entropy within a conditional-logical framework', *Artificial Intelligence*, **98**, 169–208, (1998).

[6] G. Kern-Isberner. Revising and updating probabilistic beliefs, 1998. submitted.

[7] S. Kraus, D. Lehmann, and M. Magidor, 'Nonmonotonic reasoning, preferential models and cumulative logics', *Artificial Intelligence*, **44**, 167–207, (1990).

[8] D. Makinson, 'General patterns in nonmonotonic reasoning', in *Handbook of Logic in Artificial Intelligence and Logic Programming*, eds., D.M. Gabbay, C.H. Hogger, and J.A. Robinson, volume 3, 35–110, Oxford University Press, (1994).

[9] D. Makinson and P. Gärdenfors, 'Relations between the logic of theory change and nonmonotonic logic', in *Proceedings Workshop The Logic of Theory Change, Konstanz, Germany, 1989*, pp. 185–205, Berlin Heidelberg New York, (1991). Springer.

[10] J.B. Paris, *The uncertain reasoner's companion – A mathematical perspective*, Cambridge University Press, 1994.

[11] J.B. Paris and A. Vencovská, 'A note on the inevitability of maximum entropy', *International Journal of Approximate Reasoning*, **14**, 183–223, (1990).

[12] W. Rödder and C.-H. Meyer, 'Coherent knowledge processing at maximum entropy by SPIRIT', in *Proceedings 12th Conference on Uncertainty in Artificial Intelligence*, eds., E. Horvitz and F. Jensen, pp. 470–476, San Francisco, Ca., (1996). Morgan Kaufmann.

[13] J.E. Shore and R.W. Johnson, 'Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy', *IEEE Transactions on Information Theory*, **IT-26**, 26–37, (1980).

[14] J.E. Shore and R.W. Johnson, 'Properties of cross-entropy minimization', *IEEE Transactions on Information Theory*, **IT-27**, 472–482, (1981).