Revising and updating probabilistic beliefs

Gabriele Kern-Isberner
FernUniversität Hagen
Dept. of Computer Science, LG Prakt. Informatik VIII
P.O. Box 940, D-58084 Hagen, Germany
e-mail: gabriele.kern-isberner@fernuni-hagen.de

Abstract

A new perspective of probabilistic belief change is developed in this paper. Based on the ideas of AGM [1] and Katsuno and Mendelzon [8], the operations of revision and update are investigated within a probabilistic framework, where we assume representation of knowledge to be achieved by a nonmonotonic inference operation. We distinguish between revision as a knowledge adding process, and updating as a change-recording process. A number of axioms is set forth to describe each of the change operations adequately, and we derive representation results. Moreover, we elaborate and deepen the close relationship between nonmonotonic inference and belief change for probabilistic logics by introducing *universal inference operations* as an expressive counterpart to belief change operations. As an example, we present inference based on the techniques of optimum entropy as an adequate and powerful method to realize probabilistic belief change.

1 Introduction

Besides the representation of knowledge, the handling of its dynamics in the light of new information, generally termed *belief change*, is one of the fundamental problems in Artificial Intelligence. The great variety of approaches that has been set forth up to now, usually each method coming along with a descriptive axiom scheme (for a survey, cf. [7]), corresponds to the many different interpretations and names the term *change* has been given. Gärdenfors [6] identified three fundamental types of belief change, *revision, expansion* and *update*. Katsuno and Mendelzon [8] argued that the axioms of Gärdenfors for revision are only adequate to describe a change in knowledge about a *static* world, but not for recording changes in an *evolving* world. They called this latter type of change *update*, with *erasure* being its inverse operation (cf. [8]). *Conditioning* has been regarded as an adequate method for revising probabilistic beliefs (cf. e.g. [13], [6]), but Dubois and Prade [5] emphasize that actually, conditioning does not correspond to revision but rather to *focusing*.

Most of the work in describing and classifying belief change operations has been done for belief sets based on classical logics (cf. e.g. [7]). Probabilistic knowledge is much more difficult to deal with because of its too solid structure when being represented by a probability distribution (representing a complete knowledge base), or, even worse, because of its too weak structure when being represented only by a set of probabilistic constraints – usually, there will be a myriad of different distributions all satisfying the constraints specified.

We will open up probabilistic structures by distinguishing between *belief bases*, as explicitly stated knowledge, and *belief states*, i.e. distributions representing states of belief in equilibrium (cf. [6], [13]). The transition from belief bases to belief states is achieved by using probabilistic inference operations which are assumed to be nonmonotonic to yield nontrivial inferences. With this set of tools, we present axiom schemata and representation results for probabilistic revision and updating. Revision operates on belief bases, while updating is defined on belief states. To realize updating, we make use of the inference operation underlying knowledge representation. This establishes a strong connection between nonmonotonic logics and probabilistic belief change, similar to that in more classical approaches. The notion of a *universal inference operation*, generalizing usual inference operations and introduced here as an expressive counterpart for belief change operations, allows a more adequate comparison between characteristics of nonmonotonic logics and of belief change.

As an example for an adequate and powerful universal inference operation, we present reasoning based on the techniques of maximum entropy resp. minimum cross-entropy. And we show that in this case, revision is different from focusing, thus revealing a behavior of probabilistic inference analogous to that proved by Dubois and Prade in [5] for upper and lower probabilities.

This paper is organized as follows: Section 2 lists fundamental definitions and terms of probabilistic knowledge representation. In section 3, two ways of representing probabilistic knowledge, *belief bases* and *belief states* are motivated and introduced. Section 4 describes the transition from belief bases to belief states. We begin section 5 with a short review of related work and develop the appropriate ideas for revision and updating within a probabilistic framework. In the following two sections 6 and 7 we make both notions concrete by setting up axiom schemata, and we state representation results. Section 8 compares both belief change operations directly, and section 9 shows that within the framework of optimum entropy, focusing is different from revision but akin to updating. Section 10 concludes this paper.

2 Probabilistic logics

We consider probability distributions P over a finite set of binary variables $\mathcal{V}$. Let Ω denote the set of all elementary events ω (also called atoms or complete conjunctions), and let $\mathcal{L}$ be the propositional language over the alphabet $\mathcal{V}$ with logical connectives $\vee, \wedge$ resp. juxtaposition and $\neg$. Thus each propositional formula $\mathrm{A} \in \mathcal{L}$ corresponds to an event and has probability $P(\mathrm{A}) = \sum_{\omega:\mathrm{A}(\omega)=1} P(\omega)$, where $\mathrm{A}(\omega) = 1$ resp. $= 0$ means that A is true resp. false in the world described by ω.

A *probabilistic conditional* or *probabilistic rule* has the form $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$ with $\mathrm{A}, \mathrm{B} \in \mathcal{L}, x \in [0,1]$. Let $\mathcal{L}^*$ denote the set of all probabilistic conditionals over $\mathcal{L}$: $\mathcal{L}^* = \{\mathrm{A} \rightsquigarrow \mathrm{B}[x] | \mathrm{A}, \mathrm{B} \in \mathcal{L}, x \in [0,1]\}$. Their semantics is based on a distribution P via conditional probabilities: We write $P \models \mathrm{A} \rightsquigarrow \mathrm{B}[x]$ iff $P(\mathrm{A}) > 0$ and $P(\mathrm{B}|\mathrm{A}) = \dfrac{P(\mathrm{AB})}{P(\mathrm{A})} = x$. Let $Th(P)$ denote the set of all conditionals that may be derived from P: $Th(P) = \{\mathrm{A} \rightsquigarrow \mathrm{B}[x] | P \models \mathrm{A} \rightsquigarrow \mathrm{B}[x]\}$. *Probabilistic facts* are probabilistic rules $\top \rightsquigarrow \mathrm{B}$ with tautological antecedent. Thus the classical language $\mathcal{L}$ may be embedded into $\mathcal{L}^*$ by identifying the classical propositional formula B with the probabilistic conditional $\top \rightsquigarrow \mathrm{B}[1]$.

A set $\mathcal{R}$ of probabilistic conditionals is *consistent* iff it has a model, i.e. iff there is a distribution Q with $Q \models \mathcal{R}$. If P is a distribution, then $\mathcal{R}$ is called *P-consistent* iff there is a distribution Q 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 prior knowledge P and new information $\mathcal{R}$. In particular, a P-consistent set $\mathcal{R} \subset \mathcal{L}^*$ does not contain any conditional $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$ with $P(\mathrm{A}) = 0$. Two sets $\mathcal{R}_1, \mathcal{R}_2 \subset \mathcal{L}^*$ are *probabilistically equivalent* iff each conditional of one set may be derived from conditionals in the other set by using the axioms of probability theory (cf. [13]).

Probabilistic conditionals constitute quite a general and popular form of probabilistic constraints. Thus in this paper, we base probabilistic knowledge representation upon $\mathcal{L}^*$. Most of the results to be derived, however, will also hold for more general probabilistic languages.

3 Probabilistic belief

A *state of belief* (or *epistemic state*, as Gärdenfors calls it in [6]) is characterized as a state of equilibrium which contains all explicitly stated as well as all implicitly derived knowledge. So in a probabilistic setting, probability distributions are usually considered to represent state of beliefs adequately (cf. also [13]).

As reasonable as it seems, this view brings about several problems: Firstly, a probability distribution can hardly represent lack of information. It is complete in the sense that to any (conditional) proposition, a probability is assigned. Speaking of *revision* as knowledge adding processes does not really make sense. Secondly, no distinction is made between explicit and implicit knowledge. All knowledge is considered on the same level, though e.g. adding evidential knowledge should be contrasted clearly to revising background knowledge (cf. [5]).

Voorbraak [16] uses *partial probabilistic models*, i.e. classes of probability functions, to study *conditioning* and *constraining* appropriately. If, however, one is not willing to give up the convenience of having a single distribution as "best" knowledge base, a way out of both those dilemmas described above may be offered by taking (properly defined) *probabilistic belief bases* as primitive representations of probabilistic knowledge from which *probabilistic belief states*, i.e. probability distributions, may be calculated. In [5] it is argued that belief sets arise from applying generic knowledge to evidence as sets of plausible consequences. There, as usual, evidential knowledge is restricted to certain facts, representing knowledge about a present *case*. This approach may be generalized quite easily by considering evidence as reflecting knowledge about the *context* under consideration, thus being again of a probabilistic nature. So the harsh distinction between background and evidential knowledge is mitigated, and fluctuation of knowledge is modelled more naturally: Our prior knowledge serves as a base for obtaining an adequate probabilistic description of the present context which may be used again as background knowledge.

Definition 1 *A* (probabilistic) belief state *is a distribution P (over the set of variables $\mathcal{V}$). A* (probabilistic) belief base *is a pair $(P, \mathcal{R})$, where P is a belief state (*background knowledge*), and $\mathcal{R} \subset \mathcal{L}^*$ is a P-consistent set of probabilistic rules (*context knowledge*).*
Let $\mathcal{BB}$ resp. $\mathcal{BS}$ be the set of all belief bases resp. of all belief states.

Besides the distinction between background and context knowledge two points have to be emphasized with this definition: Firstly, a probabilistic belief state, i.e. a distribution, represents an epistemic state, not only a *belief set* which is to contain merely certain beliefs. Darwiche and Pearl [3] also pointed out the importance of taking epistemic states as starting points for belief change operations. Secondly, certain beliefs, i.e. propositions or conditionals of probability 1, are not of major interest here. On principle, all information is assumed to be probabilistic, with probabilities $x \in [0,1]$, so that certain knowledge is dealt with as a borderline case.

4 From belief bases to belief states by probabilistic inference

The transition from belief bases to belief states has to be achieved by an adequate inference operation.

A *(probabilistic) inference operation* is an operation $C : 2^{\mathcal{L}^*} \to 2^{\mathcal{L}^*}$ on sets of probabilistic conditionals. C is called *reflexive* iff $\mathcal{R} \subset C(\mathcal{R})$, and *cumulative* iff $\mathcal{R} \subset \mathcal{R}' \subset C(\mathcal{R})$ implies $C(\mathcal{R}) = C(\mathcal{R}')$ for $\mathcal{R}, \mathcal{R}' \subset \mathcal{L}^*$. It satisfies *left logical equivalence* iff $C(\mathcal{R}) = C(\mathcal{S})$ whenever the sets $\mathcal{R}, \mathcal{S} \subset \mathcal{L}^*$ are probabilistically equivalent (cf. [11]). We call C *complete* iff it specifies for each set $\mathcal{R} \subset \mathcal{L}^*$ with $C(\mathcal{R}) \neq \emptyset, \mathcal{L}^*$ a unique distribution, i.e. iff there is a (unique) distribution $Q_{\mathcal{R}}$ such that $C(\mathcal{R}) = Th(Q_{\mathcal{R}})$.

So to generate belief states from belief bases $(P, \mathcal{R})$, we need inference operations that depend in a reasonable way upon the background knowledge P. The following definition generalizes the well-known notion of an inference operation by taking prior knowledge explicitly into account:

Definition 2 *A* universal probabilistic inference operation $\mathbf{C}$ *assigns a complete probabilistic inference operation* C_P *to each distribution* P*:* $\mathbf{C} : P \mapsto C_P$*. It is said to be* reflexive *resp.* cumulative *iff all its involved inference operations have the corresponding property.* $\mathbf{C}$ preserves consistency *iff for each distribution* P *and for each* P*-consistent set* $\mathcal{R} \subset \mathcal{L}^*$*,* $C_P(\mathcal{R}) \neq \emptyset, \mathcal{L}^*$*. A universal probabilistic inference operation* $\mathbf{C}$ *is* founded *iff for each distribution* P *and for any* $\mathcal{R} \subset \mathcal{L}^*$*,* $P \models \mathcal{R}$ *implies* $C_P(\mathcal{R}) = Th(P)$*.*

The property of foundedness establishes a close and intuitive relation between a distribution P and its associated inference operation C_P, distinguishing P as its stable starting point. Each C_P is presupposed to be complete, so each element of $Im\, C_P := \{C_P(\mathcal{R}) | \mathcal{R} \subset \mathcal{L}^*, C_P(\mathcal{R}) \neq \emptyset, \mathcal{L}^*\}$, corresponds uniquely to a distribution: $C_P(\mathcal{R}) = Th(Q_{P,\mathcal{R}})$, which will be abbreviated by $Q_{P,\mathcal{R}} \equiv C_P(\mathcal{R})$.

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

The notion of *strong cumulativity* introduced below describes a cumulative behavior with respect to the distributions underlying the inference operations.

Definition 4 $\mathbf{C}$ *is* strongly cumulative *iff for each distribution P and for any distributions $P_1, P_2 \in Im\, C_P$, $P \leq_P P_1 \leq_P P_2$ implies: whenever $\mathcal{R}_1 \subseteq \mathcal{R}_2 \subseteq \mathcal{L}^*$ 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 states a relationship between inference operations based on different distributions, resp. probabilistic theories, thus linking up the inference operations of $\mathbf{C}$. In fact, it establishes an important coherence property for the universal inference operation: The intermediate knowledge bases $C_P(\mathcal{R}_1)$ may be used coherently as priors for further inferences. The name "strong cumulativity" is justified by the observation that together with foundedness, strong cumulativity implies cumulativity (cf. [10]).

Thus any universal probabilistic inference relation $P \mapsto C_P$ that preserves consistency gives rise to a map $* : \mathcal{BB} \to \mathcal{BS}$, $(P, \mathcal{R}) \mapsto P * \mathcal{R}$, where $P * \mathcal{R}$ is the unique distribution such that $C_P(\mathcal{R}) = Th(P * \mathcal{R})$ (we will use a prefix notation as well as an infix notation).

In [13], several probabilistic inference processes are presented and investigated in detail. Among them, inferences based on the principles of maximum entropy resp. of minimum cross-entropy (*ME-inference*) takes an outstanding position as particularly well-behaved inference operations (cf. also [14], [9]).

Let $*_e$ denote the *ME-inference 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

$$\min_{Q \models \mathcal{R}} R(Q, P) = \sum_{\omega} Q(\omega) \log \frac{Q(\omega)}{P(\omega)}.$$

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

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

where the α_i's are suitably chosen so as to make P_e fulfill $\mathcal{R}$ (cf. [9]).

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

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

Thus C_P^e is a complete inference operation preserving consistency, and $\mathbf{C}^e : P \mapsto C_P^e$ defines a universal inference operation which proves to be reflexive, founded and strongly cumulative (cf. [10]).

5 Changing probabilistic beliefs

Let us first recapitulate how the terms *revision* and *update* have been used in the literature. In [6], *revision* is the most general case of belief change, including giving up prior beliefs in establishing new information. Thus it is closely related to the process of *contracting beliefs*, this relationship being expressed by the identities of *Harper* and *Levi* (for more details, cf. e.g. [6]). *Expansion* is revision in the special case that there is no conflict between prior and new information, so contracting of beliefs is not necessary and expansion means mere affiliation of new knowledge. The so-called AGM-postulates for belief revision (cf. [1]) constitute an important benchmark for the formal description of revision operations. Katsuno and Mendelzon (cf. [8]), however, claimed that these postulates are only suitable if revision means incorporating new information about a *static world*, but not for modelling changes in belief that are induced by an *evolving world*. They modified the AGM postulates to obtain a new characterization of this latter kind of belief change which they called *updating*.

Most of the work in describing and classifying belief change operations has been done for belief sets based on classical logics (cf. e.g. [7]). Gärdenfors [6] and Dubois and Prade [4] developed some axioms for probabilistic belief change, being mostly concerned with revising in the sense of establishing facts for certain. While Gärdenfors, however, claimed that conditioning corresponds to expansion (and thus to a certain case of revision) in a probabilistic framework, Dubois and Prade emphasize that conditioning is not revising but *focusing*, i.e. applying generic or background knowledge to the reference class describing properly the case under consideration. Paris [13] and Voorbraak [16] also consider probabilistic belief revision in the case of uncertain evidences.

For a belief change to be performed in the light of new information, the question whether this new information is contradictory to or consistent with prior knowledge is of great importance, as contradictoriness makes contraction of beliefs necessary (cf. [7]). Mostly, in the case of contradictory new information, revision is accomplished by contracting the negation of this information and then simply adding the new knowledge.

In the framework of probabilistic conditionals, however, things are more complicated because negation does not play such a clear role as in propositional two-valued logics. Though many authors agree that the negation of a conditional $\mathrm{A} \rightsquigarrow \mathrm{B}$ should be defined via $\neg(\mathrm{A} \rightsquigarrow \mathrm{B}) := \mathrm{A} \rightsquigarrow \overline{\mathrm{B}}$ (cf. e.g. [2]), which might be generalized to $\neg(\mathrm{A} \rightsquigarrow \mathrm{B}[x]) := \mathrm{A} \rightsquigarrow \overline{\mathrm{B}}[x]$, each probabilistic conditional $\mathrm{A} \rightsquigarrow \mathrm{B}[y]$ with $x \neq y$ is definitely different from $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$.

So to get a clear first view on probabilistic belief change, we will postpone the problem of contradictory information and assume, that all new information is consistent with what is already known. That means that for updating, new information should be consistent with respect to the prior distribution, and for revision, it should be consistent with background as well as with context knowledge. That corresponds to the idea of modeling *evolutions*, not *revolutions* of the world under consideration. Contractions of probabilistic beliefs will be dealt with in a further paper.

We adopt the distinction between revision and updating that was featured by Katsuno and Mendelzon and others: Revision means the procedure of learning new information about a static world, while updating applies to model the changes in an evolving world. Here *world* is used as a synonym for *context* which may be described by a set of probabilistic conditionals and is completely represented by a distribution. Thus revision requires the separation between background knowledge (i.e. a prior epistemic state) and (explicit, incomplete) context knowledge, whereas updating should best start from probabilistic epistemic states, i.e. from distributions. To realize these notions properly, we take revision (symbol: $\circ$) as an operation on belief bases and update (symbol: $\diamond$) as operating on belief states:

Notation 5 *Let $P \in \mathcal{BS}$, $(P, \mathcal{R}) \in \mathcal{BB}$, $\mathcal{R}_1 \subset \mathcal{L}^*$ such that $\mathcal{R}_1$ is P-consistent, $\mathcal{R}_2 \subset \mathcal{L}^*$ such that $\mathcal{R} \cup \mathcal{R}_2$ is P-consistent.*

Then $P \diamond \mathcal{R}_1$ denotes the update of P by $\mathcal{R}_1$, *and $(P, \mathcal{R}) \circ \mathcal{R}_2$ denotes the* revision of $(P, \mathcal{R})$ by $\mathcal{R}_2$.

To satisfy the *principle of categorical matching*, belief change should maintain the level of knowledge (cf. e.g. [7]), i.e. $P \diamond \mathcal{R}_1 \in \mathcal{BS}$ and $(P, \mathcal{R}) \circ \mathcal{R}_2 \in \mathcal{BB}$. Note that revision also induces a change of the corresponding belief state $P^* = P * \mathcal{R}$ to $(P^*)' = *((P, \mathcal{R}) \circ \mathcal{R}_2) \in \mathcal{BS}$.

6 Probabilistic belief revision

To set up axioms for probabilistic belief revision, we follow the ideas of [6] while observing that in the case of non-contradictory information, revision coincides with expansion. Due to distinguishing background knowledge from context information, we are able to compare the knowledge stored in different belief bases:

Definition 6 *An order $\sqsubseteq$ on $\mathcal{BB}$ is defined by: $(P_1, \mathcal{R}_1) \sqsubseteq (P_2, \mathcal{R}_2)$ iff $P_1 = P_2 = P$ and $\mathcal{R}_1 \subseteq \mathcal{R}_2$.*

The following *postulates* do not make use of the probabilistic inference operation but are to characterize pure base revision:

Let P be a distribution, $\mathcal{R}, \mathcal{R}', \mathcal{S} \subset \mathcal{L}^*$, and suppose that all necessary consistency preconditions are satisfied.

(PR1) $(P, \mathcal{R}) \circ \mathcal{S}$ is a belief *base*.

(PR2) If $(P, \mathcal{R}) \circ \mathcal{S} = (P, \mathcal{R}')$ then $\mathcal{S} \subset \mathcal{R}'$.

(PR3) $(P, \mathcal{R}) \sqsubseteq (P, \mathcal{R}) \circ \mathcal{S}$.

(PR4) $(P, \mathcal{R}) \circ \mathcal{S}$ is the smallest belief base (with respect to $\sqsubseteq$) satisfying (PR1)-(PR3).

(PR1) is the most fundamental axiom and coincides with the demand for categorical matching (see above). (PR2) is generally called *success*: the new context information is now represented. (PR3) states that revision should preserve prior knowledge. Thus it is crucial for revision in contrast to update. Finally, (PR4) is in the sense of *informational economy* (cf. [6]): No unnecessary changes should occur.

The following representation theorem may be proved easily:

Theorem 1 *The revision operator $\circ$ satisfies the axioms (PR1)-(PR4) iff*

$$(P, \mathcal{R}) \circ \mathcal{S} = (P, \mathcal{R} \cup \mathcal{S}). \tag{1}$$

So from (PR1)-(PR4), other properties of the revision operator also follow which are usually found among characterizing postulates:

Proposition 2 *Suppose $\circ$ to be defined via (1). Then it satisfies the following properties:*

(i) If $\mathcal{S} \subset \mathcal{R}$, then $(P, \mathcal{R}) \circ \mathcal{S} = (P, \mathcal{R})$;

(ii) If $(P_1, \mathcal{R}_1) \sqsubseteq (P_2, \mathcal{R}_2)$ then $(P_1, \mathcal{R}_1) \circ \mathcal{S} \sqsubseteq (P_2, \mathcal{R}_2) \circ \mathcal{S}$;

(iii) $(P, \mathcal{R}) \circ (\mathcal{S}_1 \cup \mathcal{S}_2) = ((P, \mathcal{R}) \circ \mathcal{S}_1) \circ \mathcal{S}_2$,

where $(P, \mathcal{R}), (P_1, \mathcal{R}_1), (P_2, \mathcal{R}_2)$ are belief bases and $\mathcal{S}, \mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{L}^*$.

(i) shows a minimality of change, while (ii) is stated in [6] as a monotonicity postulate. (iii) deals with the handling of iterated revisions.

In this paper, we investigate revision merely under the assumption that the new information is consistent with what was already known. Belief revision based on classical logics is nothing but expansion in this case, and theorem 1 indeed shows that revision should reasonably mean expanding context knowledge. But if we consider the belief states generated by two belief bases $*(P, \mathcal{R}) = P * \mathcal{R}$ and $*(P, \mathcal{R} \cup \mathcal{S}) = P * (\mathcal{R} \cup \mathcal{S})$, we see that the probabilities that $P * \mathcal{R}$ assigns to conditionals occurring in $\mathcal{S}$ will normally differ from those in P as well as from those actually stated by the new information. So the probabilistic belief in the conditionals in $\mathcal{S}$ is actually *revised*. Thus we prefer using the more general term here.

7 Probabilistic updating

Updating is to be understood as the process of altering a belief state appropriately when the world described by it *evolves* in the sense that some probabilities may change, but the "structure" of the world is supposed to be maintained as far as possible. Typical situations for updating occur when knowledge about a prior world is to be adapted to more recent information (e.g. a demographic model gained from statistical data of past periods should be brushed up by new dates), or when a special context is to be modeled by the use of suitable general knowledge (e.g. when designing a medical diagnostic system by combining theoretical knowledge and actual dates obtained in a certain hospital). Thus updating amounts to what is usually called "adjustment of (prior or background) knowledge to (new or context) information". This, however, coincides with the intuitive task of the inference operation $*$ itself. For reasons of coherence and unambiguity, it appears adequate to realize updating via probabilistic inference:

$$P \diamond \mathcal{R} := P * \mathcal{R} \tag{2}$$

for a distribution P and a P-consistent set $\mathcal{R} \subset \mathcal{L}^*$.

The close connection between belief change on one hand and nonmonotonic logics, i.e. the study of nonmonotonic inference operations, on the other hand, has been known already for a couple of years (cf. e.g. [12], [7]) within the framework of classical logics. In the sequel, we will elaborate similar relations between these areas in a probabilistic setting by establishing postulates for belief change and compare them to properties of (universal) nonmonotonic inference operations.

Katsuno and Mendelzon list in [8] eight axioms (U1)-(U8) that an updating operation should reasonably obey. As the most striking difference to revision, updating is not expected to preserve prior knowledge. This is due to the idea of an evolving world that updating is to describe.

In the following, we will translate those *postulates* of [8] which do not make use of classical connectives into a probabilistic framework .

(PU1) $P \diamond \mathcal{R} \models \mathcal{R}$.

(PU2) If $P \models \mathcal{R}$ then $P \diamond \mathcal{R} = P$.

(PU3) If $\mathcal{R}_1$ and $\mathcal{R}_2$ are probabilistically equivalent, then $P \diamond \mathcal{R}_1 = P \diamond \mathcal{R}_2$.

(PU4) If $P \diamond \mathcal{R}_1 \models \mathcal{R}_2$ and $P \diamond \mathcal{R}_2 \models \mathcal{R}_1$ then $P \diamond \mathcal{R}_1 = P \diamond \mathcal{R}_2$.

(PU5) $P \diamond (\mathcal{R}_1 \cup \mathcal{R}_2) = (P \diamond \mathcal{R}_1) \diamond (\mathcal{R}_1 \cup \mathcal{R}_2)$.

Postulates (PU1), (PU2) and (PU3) correspond to (U1), (U2) and (U4), respectively, constituting basic properties of probabilistic belief change. (PU4) states that two updating procedures with respect to sets $\mathcal{R}_1$ and $\mathcal{R}_2$ should result in the same knowledge base if each update represents the new information of the other. This property is called *reciprocity* in the framework of nonmonotonic logics (cf. [11]).

(PU5) is new insofar as it deals with *iterative updating*. This is generally problematic because updates from different starting points are hardly comparable. (PU5) demands that at least, updating any intermediate world $P \diamond \mathcal{R}_1$ by the full information $\mathcal{R}_1 \cup \mathcal{R}_2$ should lead to the same result as updating P by $\mathcal{R}_1 \cup \mathcal{R}_2$ in one step. The rationale behind this axiom is that if the information about the new world drops in in parts, updating any provisional state of belief by the full information should result unambigously in a final belief state.

Note that in general, the updates $(P \diamond \mathcal{R}_1) \diamond \mathcal{R}_2$ and $(P \diamond \mathcal{R}_1) \diamond (\mathcal{R}_1 \cup \mathcal{R}_2)$ will differ because $(P \diamond \mathcal{R}_1) \diamond \mathcal{R}_2 \not\models \mathcal{R}_1$ – updates do not maintain prior context information, in contrast to revision (cf. postulate (PR3)).

(PU5) has proved to be a crucial property for the characterization of ME-inference (cf. [9]) but actually goes back to [15]. It is worth noticing that this postulate originally inspired by studying ME-inference is equivalent to a postulate which quite recently was proposed in the context of iterated belief revisions: (PU5) is a probabilistic version of axiom (C1) in [3].

For representing updating operations satisfying the postulates stated above, we will make use of the relationship between probabilistic updating and probabilistic inference:

Proposition 3 *Suppose updating is being realized via universal probabilistic inference, that means (2) holds.*

(i) $\diamond$ *satisfies (PU1) iff* $*$ *is reflexive.*

(ii) $\diamond$ *satisfies (PU2) iff* $*$ *is founded.*

(iii) $\diamond$ *satisfies (PU3) iff* $*$ *satisfies left logical equivalence.*

(iv) Assuming reflexivity resp. the validity of (PU1), $\diamond$ *satisfies (PU4) iff* $*$ *is cumulative.*

(v) Assuming foundedness resp. the validity of (PU2), $\diamond$ *satisfies (PU5) iff* $*$ *is strongly cumulative.*

From this proposition, a representation result follows in a straightforward manner:

Theorem 4 *If $\diamond$ is defined by (2), it satisfies all of the postulates (PU1)-(PU5) iff $*$ is reflexive, founded, strongly cumulative and satisfies left logical equivalence.*

8 Revision vs. updating

We have already described in detail the different ideas underlying revision and updating processes. We will now look upon the formal parallels resp. differences. For an adequate comparison, we have to observe the changes of belief *states* that are induced by revision of belief *bases*. Again making use of the universal inference operation $*$, (R2) and (R3) translate into

(R2') $*((P, \mathcal{R}) \circ \mathcal{S}) \models \mathcal{S}$.

(R3') $*((P, \mathcal{R}) \circ \mathcal{S}) \models \mathcal{R}$.

While (R2') parallels (PU1), (R3') establishes the crucial difference between revision and updating: revision preserves prior knowledge, thus it is, in this sense, *monotonic* while updating does not, neither in a classical nor in a probabilistic framework.

The intended effects of revision and updating on a belief state $P * \mathcal{R}$ that is generated by a belief base $(P, \mathcal{R})$ are made obvious by informally writing

$$(P * \mathcal{R}) \circ \mathcal{S} = P * (\mathcal{R} \cup \mathcal{S}), \tag{3}$$

while

$$(P * \mathcal{R}) \diamond \mathcal{S} = (P * \mathcal{R}) * \mathcal{S}. \tag{4}$$

This reveals clearly the difference, but also the relationship between revision and updating: Revising $P*\mathcal{R}$ by $\mathcal{S}$ results in the same state of belief as updating P by (the full context information) $\mathcal{R} \cup \mathcal{S}$.

Unfortunately, the representation of a probabilistic belief state by a belief base is not unique, different belief bases may generate the same belief state. So we could not use the intelligible formula (3) to define probabilistic revision but had to go back to belief bases in order to separate background and context knowledge unambigously. It is interesting to observe, however, that strong cumulativity, together with foundedness, ensures at least a convenient independence of revision from background knowledge: If two belief bases $(P_1, \mathcal{R}), (P_2, \mathcal{R})$ with different prior distributions but the same context knowledge give rise to the same belief state $P_1 * \mathcal{R} = P_2 * \mathcal{R}$, then – assuming strong cumulativity and foundedness – $P_1 * (\mathcal{R} \cup \mathcal{S}) = (P_1 * \mathcal{R}) * (\mathcal{R} \cup \mathcal{S}) = (P_2 * \mathcal{R}) * (\mathcal{R} \cup \mathcal{S}) = P_2 * (\mathcal{R} \cup \mathcal{S})$.

So strong cumulativity and foundedness guarantee a particular well-behavedness with respect to inference, updating and revision.

9 Focusing

According to [5], focusing means applying generic knowledge to a reference class appropriate to describe the context of interest. So in a probabilistic setting, focusing is best be done by conditioning which, however, is normally used for revision, too. So revision and focusing are supposed to coincide in the framework of Bayesian probabilities though they differ conceptually: Revision is not only *applying knowledge* but means incorporating a new constraint so as to *refine knowledge.*

Making use of ME-inference which is known to generalize the Bayesian approach (cf. e.g. [13]), it is indeed possible to realize this conceptual difference appropriately. To make this clear, we have to consider belief changes induced by some certain information A[1], that is, we learn proposition A with certainty. The following proposition reveals the connections of both revision and updating to conditioning which is usually considered to be the "classical" belief change operation in probabilistics.

Proposition 5 *Let P be a distribution, $\mathcal{R} \subset \mathcal{L}^*$ a P- consistent set of probabilistic conditionals, and suppose $A[1]$ to be a certain probabilistic fact.*

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

(ii) $*_e((P, \mathcal{R}) \circ \{\mathrm{A}[1]\}) = P *_e (\mathcal{R} \cup \{\mathrm{A}[1]\}) = P(\cdot|\mathrm{A}) *_e \mathcal{R}$.

Thus, in the context of ME-inference, updating with a certain fact results in conditioning, while revision shows a behavior similar to *imaging*: All evidential knowledge is shifted to the "nearest" A-world. So focusing, realized via conditioning, turns out to be a special case of updating: The certain fact $A[1]$ may be considered not only as a proper reference class, but also as representing a new world the prior knowledge has to be adapted to.

The proposition above shows that, in a (generalized) Bayesian framework, a proper distinction between focusing and revision is possible.

10 Concluding remarks

The approach to probabilistic belief change being made in this paper uses belief bases and belief states, thus allowing to distinguish between explicitly stated and derived beliefs. Belief bases represent background knowledge as different from context knowledge, and both are combined to yield a distribution by a nonmonotonic inference operation. We defined belief revision as a change in context knowledge, and belief update as adapting a prior distribution to a new context. Both operations may be realized by virtue of the underlying probabilistic inference operation. It is only the different ways in which it is applied that yield different results in belief change while allowing a convenient homogeneity of methods.

As a paricularly well-behaved probabilistic inference operation, we feature reasoning at optimum entropy (ME-inference). One of its characterizing properties, namely (PU5), is stated here as a postulate for handling iterated updates and turned out to be crucial in other frameworks, too (cf. [3]). Making use of the ME-methods, we show that revision and focusing may be modelled differently in Bayesian probabilistics, thus obtaining results similar to those of Dubois and Prade in the context of upper and lower probabilities.

Though the concern of this paper is with probabilistic belief change, probabilistic knowledge is actually used only in quite a formal way. So the approach to realize revision and updating developed here, in particular with respect to the concept of a universal inference operation, may be of interest for other knowledge representation formalisms which are rich enough for distinguishing between background and evidential knowledge, as well as for iterated belief change in general.

References

[1] C.E. Alchourrón, P. Gärdenfors, and P. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

[2] P.G. Calabrese. Deduction and inference using conditional logic and probability. In I.R. Goodman, M.M. Gupta, H.T. Nguyen, and G.S. Rogers, editors, *Conditional Logic in Expert Systems*, pages 71–100. Elsevier, North Holland, 1991.

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

[4] D. Dubois and H. Prade. A survey of belief revision and updating rules in various uncertainty models. *Intern. Journal of intelligent systems*, 9:61–100, 1994.

[5] D. Dubois and H. Prade. Focusing vs. belief revision: A fundamental distinction when dealing with generic knowledge. In *Proceedings First International Joint Conference on Qualitative and Quantitative Practical Reasoning, ECSQARU-FAPR'97*, pages 96–107, Berlin Heidelberg New York, 1997. Springer.

[6] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, Mass., 1988.

[7] P. Gärdenfors and H. Rott. Belief revision. In D.M. Gabbay, C.H. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 35–132. Oxford University Press, 1994.

[8] H. Katsuno and A.O. Mendelzon. On the difference between updating a knowledge base and revising it. In *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning, KR'91*, pages 387–394, San Mateo, Ca., 1991. Morgan Kaufmann.

[9] G. Kern-Isberner. Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence*, 1998. (to appear).

[10] G. Kern-Isberner. Nonmonotonic reasoning in probabilistics, 1998. submitted.

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

[12] 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*, pages 185–205, Berlin Heidelberg New York, 1991. Springer.

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

[14] 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.

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

[16] F. Voorbraak. Probabilistic belief expansion and conditioning. Technical Report LP-96-07, Institute for Logic, Language and Computation, University of Amsterdam, 1996.