[**To appear in:** ECSQARU/FAPR 97]

A logically sound method for uncertain reasoning with quantified conditionals

Gabriele Kern-Isberner

FernUniversität Hagen, Fachbereich Informatik
P.O. Box 940, D-58084 Hagen, Germany
e-mail: gabriele.kern-isberner@fernuni-hagen.de

Abstract. Conditionals play a central part in knowledge representation and reasoning. Describing certain relationships between antecedents and consequences by "if–then–sentences" their range of expressiveness includes commonsense knowledge as well as scientific statements. In this paper, we present the principles of maximum entropy resp. of minimum cross-entropy (*ME-principles*) as a logically sound and practicable method for representing and reasoning with quantified conditionals. First the meaning of these principles is made clear by sketching a *characterization* from a completely conditional-logical point of view. Then we apply the techniques presented to derive *ME–deduction schemes* and illustrate them by examples in the second part of this paper.

1 Introduction

Knowledge is often expressed by "if A then B"–statements (*conditionals*, written as $A \rightarrow B$), where the *antecedent* A describes a precondition which is known (or assumed) to imply the *consequence* B. Usually, a conditional supposes a special relationship between its antecedent and its consequence. The most characteristic property of a conditional is that its meaning is restricted to situations where its antecedent is true. Therefore to represent conditionals appropriately, the need for a third logical value u, interpreted as *undetermined* or *undefined*, arises, thus making it necessary to leave the area of classical two-valued logic (cf. [15]).

A lot of different approaches to a *logic of conditionals* have been made (for a survey, cf. [16]), also aiming at reflecting more general relationships between antecedent and consequence so as to capture the manifold meanings of commonsense conditionals. In general, conditionals are used to describe relationships that are assumed to hold *mostly* or *possibly*, or that are only *believed* to hold. Beside such qualitative approaches, the validity of conditionals may be quantified by *degrees of trueness* or *certainty* (cf. e.g. [2]). Cox [3] argued that a logically consistent handling of quantified conditionals is only possible within a probabilistic framework, where the degree of certainty associated with a conditional is interpreted as a conditional probability. In fact, probability theory provides a sound and convenient machinery to be used for knowledge representation and automated reasoning (cf. e.g. [5],[4], [14], [18], [23]). But its clear semantics and strict rules require a lot of knowledge to be available for an adequate modelling of problems.

In contrast to this, usually only relatively few relationships between relevant variables are known, due to incomplete information. Or maybe, an abstractional representation is intended, incorporating only fundamental relationships. In both cases, the knowledge explicitly stated is not sufficient to determine uniquely a probability distribution. One way to cope with this indetermination is to calculate upper and lower bounds for probabilities (cf. [23], [5]). This method, however, brings about two problems: Sometimes the inferred bounds are quite bad, and in any case, one has to handle intervals instead of single values.

An alternative way that provides best expectation values for the unknown probabilities and guarantees a logically sound reasoning is to use the *principle of maximum entropy* resp. the *principle of minimum cross entropy* to represent all available probabilistic knowledge by a unique distribution. If P is a probability distribution, its *entropy* is defined as $H(P) = -\sum_\omega p(\omega) \log p(\omega)$, and if Q is another distribution, $R(Q, P) = \sum_\omega q(\omega) \log \frac{q(\omega)}{p(\omega)}$ gives the *cross-* or *relative entropy of Q with respect to P*. Entropy is a numerical value for the uncertainty inherent to a distribution, and relative entropy measures the information-theoretic distance between P and Q (cf. e.g. [20], [13], [9], [8]). So if $\mathcal{R}$ is a set of conditionals, each associated with a probability, then the "best" distribution to represent $\mathcal{R}$ – and only $\mathcal{R}$ – is the one which fulfills all conditionals in $\mathcal{R}$ and has maximum entropy. By an analogous argument, if prior knowledge given by a distribution P has to be adjusted to new probabilistic knowledge $\mathcal{R}$, the one distribution should be chosen that satisfies $\mathcal{R}$ and has minimum relative entropy to P. In the sequel, the abbreviation *ME* will indicate representation of resp. adjustment to probabilistic knowledge at optimum (*M*inimum or *M*aximum) *E*ntropy.

There are three articles [21],[17],[11] that characterize optimum entropy distributions as sound bases for logically consistent inferences. In [11], the author chose a completely conditional-logical environment for this characterization, proving that optimum entropy representations are most appropriate for probabilistic conditionals. Together with the results of Cox [3], this establishes reasoning at optimum entropy as a most fundamental inference method in the area of quantified uncertain reasoning.

The characterization given in [12] rests on only four very basic and intelligible axioms for conditional reasoning, and it is developed step by step from a conditional-logical argumentation. In the first part of this paper, we will sketch this characterization for a deeper understanding of ME-reasoning. The representation of the ME-distribution central to the argumentation given then turns out to be not only of theoretical but also of practical use: Due to it, some deduction schemes for ME-inferring will be presented in the second part of this paper. For instance, we will show how knowledge is propagated transitively and give explicit ME-probability values for *cautious cut* and *cautious monotony*. These deduction schemes, however, are global, not local, i.e. all knowledge available has to be taken into account in their premises to give sound results. But they provide useful insights into the practice of ME-reasoning. Furthermore, a few examples will be given.

2 A probabilistic conditional language

Let $\mathcal{V} = \{V_1, V_2, V_3, \ldots\}$ be a finite set of binary propositional variables, and let $\mathcal{L}(\mathcal{V})$ be the propositional language over the alphabet $\mathcal{V}$ with logical connectives $\vee, \wedge$ resp. juxtaposition and $\neg$. For each V_i, $\dot{v}_i \in \{v_i, \bar{v}_i\}$ stands for one of the two possible outcomes of the variable (where negation is indicated by a bar). The set of all *atoms* (*complete conjunctions, elementary events*) over the alphabet $\mathcal{V}$ is denoted by Ω: $\Omega = \{\omega | \omega = \dot{v}_1 \dot{v}_2 \ldots, \ \dot{v}_i \in \{v_i, \bar{v}_i\}\}$. Each probability distribution P over $\mathcal{V}$ induces a probability function p on Ω and vice versa. Thus, given a distribution P, a probability can be assigned to each propositional formula $\mathrm{A} \in \mathcal{L}(\mathcal{V})$ via $p(\mathrm{A}) = \sum_{\omega : \mathrm{A}(\omega)=1} p(\omega)$, where $\mathrm{A}(\omega) = 1$ resp. $= 0$ means that A is true resp. false in the world described by ω.

We extend $\mathcal{L}(\mathcal{V})$ to a *probabilistic conditional language* $\mathcal{L}^*$ by adding a *conditional operator* $\rightsquigarrow$: The wff in $\mathcal{L}^*$ are defined to have syntax $\mathrm{A} \rightsquigarrow \mathrm{B}[x]$ with $\mathrm{A}, \mathrm{B} \in \mathcal{L}(\mathcal{V}), x \in [0,1]$, and they are called *probabilistic conditionals* or *probabilistic rules*. Their semantics is based on P via conditional probabilities: We write $P \models \mathrm{A} \rightsquigarrow \mathrm{B}[x]$ iff $p(\mathrm{A}) > 0$ and $p(\mathrm{B}|\mathrm{A}) = \dfrac{p(\mathrm{AB})}{p(\mathrm{A})} = x$. By a simple probabilistic calculation, we see $p(\mathrm{B}|\mathrm{A}) = x$ iff $\dfrac{p(\mathrm{AB})}{p(\mathrm{A}\overline{\mathrm{B}})} = \dfrac{x}{1-x}$. *Probabilistic facts* are probabilistic rules $\top \rightsquigarrow \mathrm{B}$ with tautological antecedent. Thus the classical language $\mathcal{L}(\mathcal{V})$ may be embedded into $\mathcal{L}^*$ by identifying the classical propositional formula B with the probabilistic conditional $\top \rightsquigarrow \mathrm{B}[1]$.

3 Probabilistic representation and adjustment at optimum entropy

The *problem of adapting new probabilistic knowledge* may now be adequately formalized as follows, using the notation of the previous section:

(*) Given a (prior) distribution P and a set $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ of probabilistic conditionals, which (posterior) distribution P^* should be chosen such that it satisfies $P^* \models \mathcal{R}$ and that it is — in some sense — "nearest" to P?

To avoid inconsistencies between prior knowledge and new information, we assume P to be positive. The *representation problem*

(**) Given a set $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ of probabilistic conditionals, which distribution P^* that satisfies $P^* \models \mathcal{R}$ should be chosen to represent $\mathcal{R}$ best?

is a special case of (*) with a uniform distribution as prior distribution, starting from complete ignorance.

Let P_e denote the ME-solution to (*), that is P_e has minimum cross-entropy to P among all distributions Q with $Q \models \mathcal{R}$. Then P_e may be represented as

$$p_e(\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \le i \le n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^{1-x_i} \prod_{\substack{1 \le i \le n \\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} \alpha_i^{-x_i}, \tag{1}$$

for $\omega \in \Omega$ (cf. e.g [9], [11]). In particular, P_e satisfies $\mathcal{R}$, so $p_e(\mathrm{B}_i|\mathrm{A}_i) = x_i$, which is equivalent to postulating $\dfrac{p_e(\mathrm{A}_i\mathrm{B}_i)}{p_e(\mathrm{A}_i\overline{\mathrm{B}_i})} = \dfrac{x_i}{1-x_i}$ for all i, $1 \le i \le n$. Regarding (1), this last equation yields

$$\alpha_i = \frac{x_i}{1-x_i} \frac{\displaystyle\sum_{\omega:\mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1} p(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^{1-x_j} \prod_{\substack{j \neq i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^{-x_j}}{\displaystyle\sum_{\omega:\mathrm{A}_i\mathrm{B}_i(\omega)=1} p(\omega) \prod_{\substack{j \neq i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^{1-x_j} \prod_{\substack{j \neq i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^{-x_j}}, \quad 1 \le i \le n, \tag{2}$$

as a necessary (and sufficient) condition for any distribution of type (1) to satisfy $\mathcal{R}$. α_0 may be chosen adequately to ensure that p_e is a probability function, i.e. that $\sum_{\omega \in \Omega} p_e(\omega) = 1$. The next section will show that equations (1) and (2) are crucial for the understanding of ME-adjustment (cf. Theorem 8). All proofs may be found in [11].

4 Characterizing the principle of minimum cross entropy within a conditional-logical framework

4.1 The principle of conditional preservation

Following Calabrese [2], a conditional $\mathrm{A} \rightsquigarrow \mathrm{B}$ can be represented as an *indicator function* $(\mathrm{B}|\mathrm{A})$ on elementary events, setting

$$(\mathrm{B}|\mathrm{A})(\omega) = \begin{cases} 1 & : \quad \omega \in \mathrm{AB} \\ 0 & : \quad \omega \in \mathrm{A}\overline{\mathrm{B}} \\ u & : \quad \omega \notin \mathrm{A} \end{cases}$$

where u stands for *undefined.* This definition captures excellently the non-classical character of conditionals within a probabilistic framework. According to it, a conditional is a function that polarizes AB and $\mathrm{A}\overline{\mathrm{B}}$, leaving $\overline{\mathrm{A}}$ untouched.

As we saw earlier in section 2, conditional probabilities are ratios, the most fundamental of which being *elementary ratios*, i.e. ratios of probabilities of elementary events. They can be regarded as the atoms of the conditional-logical structure of a distribution, and products of elementary ratios provide a suitable means to investigate conditional structures. In statistics, logarithms of such expressions are used to measure the interactions between the variables involved (cf. [7], [24]).

Informally, the *principle of conditional preservation* is to state that all products of elementary ratios shall remain unchanged if there is no indication for change found in the set $\mathcal{R}$, representing new conditional information. As a consequence, all (statistical) interactions between variables are preserved as far as possible.

To make the idea of conditional preservation concrete, we have to devise a method to compare the elementary events involved in such products on the base of the information in $\mathcal{R}$. To this end, we will formalize the notion of *conditional structures* of elementary events resp. of multi-sets of elementary events by means of a group-theoretical representation.

Let Ω denote the set of all elementary events, and let $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ be a set of probabilistic conditionals. To each conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ in $\mathcal{R}$ we associate two symbols a_i, b_i. Let $F_\mathcal{R} = \langle a_1, b_1, \ldots, a_n, b_n \rangle$ be the free abelian group with generators $a_1, b_1, \ldots, a_n, b_n$, i.e. $F_\mathcal{R}$ consists of all elements of the form $a_1^{\mu_1} b_1^{\nu_1} \ldots a_n^{\mu_n} b_n^{\nu_n}$ with integers $\mu_i, \nu_i \in \mathcal{Z}$ (the ring of integers), and each element can be identified by its exponents, so that $F_\mathcal{R}$ is isomorphic to $\mathcal{Z}^{2n}$. The commutativity of $F_\mathcal{R}$ corresponds to the fact that the conditionals in $\mathcal{R}$ shall be effective all at a time, without assuming any order of application.

For each $i, 1 \leq i \leq n$, we define a function $\sigma_i : \Omega \to F_\mathcal{R}$ by setting

$$\sigma_i(\omega) = \begin{cases} a_i & \text{if} \quad (\mathrm{B}_i|\mathrm{A}_i)(\omega) = 1 \\ b_i & \text{if} \quad (\mathrm{B}_i|\mathrm{A}_i)(\omega) = 0 \\ 1 & \text{if} \quad (\mathrm{B}_i|\mathrm{A}_i)(\omega) = u \end{cases} .$$

$\sigma_i(\omega)$ represents the manner in which the conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ affects the elementary event ω. The neutral element 1 of $F_\mathcal{R}$ corresponds to the non-applicability of $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ in case that the antecedent A_i is not satisfied.

The function $\sigma = \sigma_\mathcal{R} : \Omega \to F_\mathcal{R}$, $\sigma(\omega) = \prod_{1 \leq i \leq n} \sigma_i(\omega)$ describes the all-over effect of $\mathcal{R}$ on ω. $\sigma(\omega)$ contains at most one of each a_i or b_i, but never both of them because each conditional affects ω in a well-defined way.

Definition 1. (i) Let $\omega \in \Omega$ be an elementary event. $\sigma(\omega) = \prod_{1 \leq i \leq n} \sigma_i(\omega) \in F_\mathcal{R}$ is called the *conditional structure* of ω with respect to $\mathcal{R}$.
(ii) Let $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ denote a multi-set of elementary events, with each r_j indicating the multiplicity of the corresponding ω_j. The element $\Omega_1^\sigma := \prod_{1 \leq j \leq m_1} \sigma(\omega_j)^{r_j} \in F_\mathcal{R}$ is called the *conditional structure of* Ω_1 .

Thus the conditional structure Ω_1^σ of a multi-set $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ is represented by a group element which is a product of the generators a_i, b_i of $F_\mathcal{R}$, with each a_i occurring with exponent $\sum_{k : \sigma_i(\omega_k) = a_i} r_k = \sum_{k : (\mathrm{B}_i|\mathrm{A}_i)(\omega_k) = 1} r_k$, and each b_i occurring with exponent $\sum_{k : \sigma_i(\omega_k) = b_i} r_k = \sum_{k : (\mathrm{B}_i|\mathrm{A}_i)(\omega_k) = 0} r_k$ (note that both sums may be zero in case that the corresponding conditional is not touched at all). So the exponent of a_i in Ω_1^σ indicates the number of elementary events in Ω_1 which *confirm* the conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i$, each event being counted with its multiplicity, and in the same way the exponent of b_i indicates the number of elementary events that *refute* $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i$. Ω_1^σ encodes this information in an elegant manner. Making use of the group structure allows to form products and thus provides a convenient representation and handling of conditional structures.

Definition 2. Two multi-sets $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ and $\Omega_2 = \{\nu_1 : s_1, \ldots, \nu_{m_2} : s_{m_2}\}$ of elementary events with equal cardinalities $\sum_{1 \le k \le m_1} r_k = \sum_{1 \le l \le m_2} s_l$ are *$\mathcal{R}$-equivalent* iff $\Omega_1^\sigma = \Omega_2^\sigma$, i.e. iff their conditional structures with respect to $\mathcal{R}$ are identical.

We are now in the position to formalize exactly what is to be understood by conditional preservation. Adaptations following this principle will be called *c-adaptations.*

Definition 3. Let P^* be a distribution which fulfills $\mathcal{R}$.
P^* is called a *c-adaptation of P to $\mathcal{R}$* iff it satisfies the following two conditions:

(i) For any $\omega \in \Omega$, $p^*(\omega) = 0$ if and only if there is a conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ in $\mathcal{R}$ such that either $x_i = 1$ and $\mathrm{A}_i\overline{\mathrm{B}_i}(\omega) = 1$, or $x_i = 0$ and $\mathrm{A}_i\mathrm{B}_i(\omega) = 1$;

(ii) the $\mathcal{R}$-equivalence of any two multi-sets $\Omega_1 = \{\omega_1 : r_1, \ldots, \omega_{m_1} : r_{m_1}\}$ and $\Omega_2 = \{\nu_1 : s_1, \ldots, \nu_{m_2} : s_{m_2}\}$ of elementary events implies

$$\frac{p^*(\omega_1)^{r_1} \ldots p^*(\omega_{m_1})^{r_{m_1}}}{p^*(\nu_1)^{s_1} \ldots p^*(\nu_{m_2})^{s_{m_2}}} = \frac{p(\omega_1)^{r_1} \ldots p(\omega_{m_1})^{r_{m_1}}}{p(\nu_1)^{s_1} \ldots p(\nu_{m_2})^{s_{m_2}}}.$$

Thus a c-adaptation is completely based both on the conditional structure of P and of $\mathcal{R}$, using the structure of P as a reference point and that of $\mathcal{R}$ as a guideline for changes. It realizes perfectly a conditional-logical approach to the adaptation problem (*):

Postulate (P1): conditional preservation
The solution P^* to (*) is a c-adaptation.

The next theorem provides a catchy and easy characterization of c-adaptations:

Theorem 4. *Let P^* be a distribution that satisfies $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$; let P be a positive prior distribution.*

P^* is a c-adaptation of P to $\mathcal{R}$ iff for each conditional $\mathrm{A}_i \rightsquigarrow \mathrm{B}_i[x_i]$ in $\mathcal{R}$, there is a pair of non-negative factors (α_i^+, α_i^-) associated with it such that

$$p^*(\omega) = \alpha_0 p(\omega) \prod_{\substack{1 \le i \le n \\ \mathrm{A}_i\mathrm{B}_i(\omega)=1}} \alpha_i^+ \prod_{\substack{1 \le i \le n \\ \mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1}} \alpha_i^- \tag{3}$$

for all elementary events ω, where $\alpha_0 > 0$ is a normalization factor.

From now on, we will use an operational notation to distinguish posterior distributions that satisfy Postulate (P1), i.e. the results of c-adaptations of P to $\mathcal{R}$: $P * \mathcal{R}$ shall denote any distribution of the form (3), where the factors α_i^+, α_i^-, $1 \le i \le n$, are solutions to

$$\frac{\alpha_i^+}{\alpha_i^-} = \frac{x_i}{1 - x_i} \frac{\sum\limits_{\omega:\mathrm{A}_i\overline{\mathrm{B}_i}(\omega)=1} p(\omega) \prod\limits_{\substack{j \ne i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^+ \prod\limits_{\substack{j \ne i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^-}{\sum\limits_{\omega:\mathrm{A}_i\mathrm{B}_i(\omega)=1} p(\omega) \prod\limits_{\substack{j \ne i \\ \mathrm{A}_j\mathrm{B}_j(\omega)=1}} \alpha_j^+ \prod\limits_{\substack{j \ne i \\ \mathrm{A}_j\overline{\mathrm{B}_j}(\omega)=1}} \alpha_j^-} \tag{4}$$

with $\alpha_i^- = 0$ for $x_i = 1$. (4) ensures that the resulting $P * \mathcal{R}$ fulfills all of the conditionals in $\mathcal{R}$.

4.2 The functional concept

In general, there will be many different posterior c-adaptations, corresponding to different solutions of (4). Of course, this situation is not very satisfactory because we intuitively feel that there are "good" solutions and "bad" solutions. For instance, if the prior distribution P already satisfies the information contained in $\mathcal{R}$ it would be reasonable to expect that the scheme does not alter any probability, yielding the prior as the posterior distribution. In general, the question for a "best" solution and the need to determine it uniquely arises. There should best be a function which calculates an appropriate posterior distribution from prior knowledge P and new conditional information $\mathcal{R}$.

The factors α_i^+ and α_i^- crucial for a c-adaptation are solutions to the equations (4) which reflect the complex logical interactions between the rules in $\mathcal{R}$ and the dependency from the P. Their common quotient $\alpha_i := \dfrac{\alpha_i^+}{\alpha_i^-}$ symbolizes the core of the impact the conditionals in $\mathcal{R}$ are to have on the new distribution, as well as it takes into account the influence of the prior P. Therefore it plays a key role, and we assume the factors α_i^+ and α_i^- to be functionally dependent on it: $\alpha_i^+ = F_i^+(\alpha_i) = F^+(x_i, \alpha_i)$ resp. $\alpha_i^- = F_i^-(\alpha_i) = F^-(x_i, \alpha_i)$. The functions F^+ and F^- are supposed to be sufficiently regular and are to follow a pattern independent of the specific form of a rule $\mathrm{A}_i \leadsto \mathrm{B}_i[x_i]$, thus realizing a fundamental inference pattern. We state these assumptions as

> *Postulate (P2): functional concept*
> The factors α_i^+ and α_i^- in (3) are determined by two regular real functions $F^+(x, \alpha)$ and $F^-(x, \alpha)$, defined for $x \in [0, 1]$ and non-negative real α, such that
> $$\alpha_i^+ = F^+(x_i, \alpha_i) \quad \text{and} \quad \alpha_i^- = F^-(x_i, \alpha_i),$$
> where $\alpha_i = \dfrac{\alpha_i^+}{\alpha_i^-}$ denotes their common quotient.

To symbolize the presence of the functional concept, we will use $*_F$ instead of $*$.

4.3 Logical consistency and representation invariance

Surely, the adaptation scheme (3) will be considered sound only if the resulting posterior distribution can be used as a prior distribution for further adaptations. This is a very fundamental meaning of *logical consistency.*

> *Postulate (P3): logical consistency*
> For any positive distribution P and any sets $\mathcal{R}_1, \mathcal{R}_2 \subset \mathcal{L}^*$, the (final) posterior distribution which arises from a two-step process of adjusting P first to $\mathcal{R}_1$ and then adjusting this intermediate posterior to $\mathcal{R}_1 \cup \mathcal{R}_2$ is identical to the distribution resulting from simply adapting P to $\mathcal{R}_1 \cup \mathcal{R}_2$ (provided that all adaptation problems are solvable).

More formally, the operator $*$ satisfies (P3) iff the following equation holds:

$$P * (\mathcal{R}_1 \cup \mathcal{R}_2) = (P * \mathcal{R}_1) * (\mathcal{R}_1 \cup \mathcal{R}_2)$$

Theorem 5. *If the adjustment operator $*_F$ satisfies the postulate (P3) of logical consistency then there is a regular function $c(x)$ such that*

$$F^-(x, \alpha) = \alpha^{c(x)} \quad \text{and} \quad F^+(x, \alpha) = \alpha^{c(x)+1}$$

for any positive real α and any $x \in (0, 1)$.

We surely expect the result of our adjustment process to be independent of the syntactic representation of probabilistic knowledge in $\mathcal{R}$:

> *Postulate (P4): representation invariance*
> If two sets of probabilistic conditionals $\mathcal{R}$ and $\mathcal{R}'$ are probabilistically equivalent the posterior distributions $P * \mathcal{R}$ and $P * \mathcal{R}'$ resulting from adapting a prior P to $\mathcal{R}$ resp. to $\mathcal{R}'$ are identical.

Here, as in general, two sets of rules $\mathcal{R}$ and $\mathcal{R}'$ are called *probabilistically equivalent* iff each rule in one set is derivable from rules in the other by elementary probabilistic calculation. Using the operational notation, we are able to express (P4) more formally: The adjustment operator $*$ satisfies (P4) iff

$$P * \mathcal{R} = P * \mathcal{R}'$$

for any two probabilistically equivalent sets $\mathcal{R}$ and $\mathcal{R}'$.

Theorem 6. *If the operator $*_F$ is to meet the fundamental demands for logical consistency (P3) and for representation invariance (P4), then F^- and F^+ necessarily must be*

$$F^+(x, \alpha) = \alpha^{1-x} \quad \text{and} \quad F^-(x, \alpha) = \alpha^{-x}.$$

4.4 Uniqueness

So far we have proved that the demands for logical consistency and for representation invariance uniquely determine the functions which we assumed to underly the adjustment process in the way we described in Subsection 4.2. We recognize that the posterior distribution necessarily is of the same *type* (1) as the ME-distribution if it is to yield sound and consistent inferences. But - are there possibly several different solutions of type (1), only one of which being the ME-distribution? The uniqueness which makes the characterization complete is now affirmed by the next theorem. This unique posterior distribution of type (1) must be the ME-distribution, $*_F$ then corresponds to ME-inference, and ME-inference is known to fulfill (P3) and (P4) as well as many other reasonable properties, cf. [10], [17], [21], [22].

Theorem 7. *There is at most one solution $P * \mathcal{R}$ of the adaptation problem of type (1).*

The following theorem summarizes our results in characterizing ME-adjustment within a conditional-logical framework:

Theorem 8. *Let $*_e$ denote the ME-adjustment operator, i.e. $*_e$ assigns to a prior distribution P and some set $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ of probabilistic conditionals the one distribution $P_e = P *_e \mathcal{R}$ which has minimal cross-entropy with respect to P among all distributions that satisfy $\mathcal{R}$ (provided the adjustment problem is solvable at all). Then $*_e$ yields the only adaptation of P to $\mathcal{R}$ that obeys the principle of conditional preservation (P1), realizes a functional concept (P2) and satisfies the fundamental postulates for logical consistency (P3) and for representation invariance (P4). $*_e$ is completely described by (1) and (2).*

Thus a characterization of the ME-principle within a conditional-logical framework is achieved, and its implicit logical mechanisms have been revealed clearly.

5 Some ME-deduction rules

After having proved ME-adjustment and ME-representation to be logically sound methods for uncertain reasoning, we will now leave the abstract level of argumentation and turn to concrete inference patterns. It must be emphasized, however, that ME-inferring is a *global*, not a *local* method: Only if *all* knowledge available is taken into account, the results of ME-inference are reliable to yield best expectation values. Thus it is not possible to use only partial information for reasoning, and then continue the process of adjusting from the obtained intermediate distribution with the information still left. It is important that in the two-step adjustment process $(P * \mathcal{R}_1) * (\mathcal{R}_1 \cup \mathcal{R}_2)$ dealt with in the consistency postulate (P3) (cf. section 4.3) the second adaptation step uses full information $\mathcal{R}_1 \cup \mathcal{R}_2$. In fact, the distributions $(P *_e \mathcal{R}_1) *_e (\mathcal{R}_1 \cup \mathcal{R}_2)$ and $(P *_e \mathcal{R}_1) *_e \mathcal{R}_2$ differ in general.

For this reason, the deduction rules to be presented in the sequel do not provide a convenient (and complete) calculus for ME-reasoning. But they illustrate effectfully the reasonableness of that technique by calculating explicitly inferred probabilities of rules in terms of given prior probabilities. In contrast to this, the inference patterns for deriving lower and upper bounds for probabilities presented in [5] and [23] are local, but they are afflicted with all problems typical to methods for inferring intervals, not single values (cf. section 1).

It must be pointed out that in principle, ME-reasoning is feasible for all consistent probabilistic representation and adaptation problems by iterative propagation, realized e.g. by the expert system shell SPIRIT (cf. [19]) far beyond the scope of the few inference patterns given below (also cf. Example 3).

We will use the following notation:

$$\frac{\mathcal{R} \ : \ \mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]}{\mathrm{A}_1^* \rightsquigarrow \mathrm{B}_1^*[x_1^*], \ldots, \mathrm{A}_m^* \rightsquigarrow \mathrm{B}_m^*[x_m^*]}$$

iff $\mathcal{R} = \{\mathrm{A}_1 \rightsquigarrow \mathrm{B}_1[x_1], \ldots, \mathrm{A}_n \rightsquigarrow \mathrm{B}_n[x_n]\}$ and $P_0 *_e \mathcal{R} \models \{\mathrm{A}_1^* \rightsquigarrow \mathrm{B}_1^*[x_1^*], \ldots, \mathrm{A}_m^* \rightsquigarrow \mathrm{B}_m^*[x_m^*]\}$, where P_0 is a uniform distribution of suitable size.

5.1 Chaining rules

Proposition 9 (Transitive Chaining). *Suppose A, B, C to be propositional variables, $x_1, x_2 \in [0,1]$. Then*

$$\frac{\mathcal{R}\ :\ a \leadsto b[x_1],\ b \leadsto c[x_2]}{a \leadsto c[\frac{1}{2}(2x_1x_2 + 1 - x_1)]} \tag{5}$$

Proof. According to equations (1) and (2), the posterior distribution $P_0 *_e \{a \leadsto b[x_1],\ b \leadsto c[x_2]\}$ may be calculated as shown in the following table:

ω	$P_0 * \mathcal{R}$
abc	$\alpha_0\alpha_1^{1-x_1}\alpha_2^{1-x_2}$
$ab\bar{c}$	$\alpha_0\alpha_1^{1-x_1}\alpha_2^{-x_2}$
$a\bar{b}c$	$\alpha_0\alpha_1^{-x_1}$
$a\bar{b}\bar{c}$	$\alpha_0\alpha_1^{-x_1}$
$\bar{a}bc$	$\alpha_0\alpha_2^{1-x_2}$
$\bar{a}b\bar{c}$	$\alpha_0\alpha_2^{-x_2}$
$\bar{a}\bar{b}c$	α_0
$\bar{a}\bar{b}\bar{c}$	α_0

with $\alpha_1 = \dfrac{x_1}{1-x_1}\dfrac{2}{\alpha_2^{1-x_2} + \alpha_2^{-x_2}} = \dfrac{x_1}{1-x_1}\alpha_2^{x_2}\dfrac{2}{\alpha_2 + 1}$ and $\alpha_2 = \dfrac{x_2}{1-x_2}$.
Now the probability of $a \leadsto c$ may be calculated in a straightforward manner:
$p^*(c|a) = \dfrac{p^*(ac)}{p^*(a)} = \dfrac{p^*(abc) + p^*(a\bar{b}c)}{p^*(abc) + p^*(ab\bar{c}) + p^*(a\bar{b}c) + p^*(a\bar{b}\bar{c})} =$
$= \dfrac{\alpha_1\alpha_2^{1-x_2} + 1}{\alpha_1\alpha_2^{-x_2}(\alpha_2 + 1) + 2} = \dfrac{1}{2}(2x_1x_2 + 1 - x_1)$, as desired.

Example 1. Suppose the propositional variables A, B, C are given the meanings A=*Being young*, B=*Being single*, and C=*Having children*, respectively. We know (or assume) that young people are usually singles (with probability 0.9) and that mostly, singles do not have children (with probability 0.85), so that $\mathcal{R} = a \leadsto b[0.9], b \leadsto \bar{c}[0.85]$. Using (5) with $x_1 = 0.9$ and $x_2 = 0.85$, ME-reasoning yields $a \leadsto \bar{c}[0.815]$ (the negation of C makes no difference). Therefore from the knowledge stated by $\mathcal{R}$ we may conclude that the probability of an individual not to have children if (s)he is young is best estimated by 0.815.

In many cases, however, rules must not be simply connected transitively as in Proposition 9 because definite exceptions are present. Let us consider the famous "Tweety the penguin"-example.

Example 2. Birds are known to fly mostly, i.e. $bird \leadsto fly[x_1]$ with a probability x_1 between 0.5 and 1, penguins are definitely birds, $penguin \leadsto bird[1]$, but no one has ever seen a flying penguin, so $penguins \leadsto fly[x_2]$ with a probability x_2 very close to 0. What may be inferred about Tweety who is known to be a bird *and* a penguin?

The crucial point in this example is that two pieces of evidence apply to Tweety, one being more specific than the other. The next proposition shows that ME-reasoning is able to cope with *categorical specificity.*

Proposition 10 (Categorical Specificity). *Suppose A, B, C to be propositional variables, $x_1, x_2 \in [0,1]$. Then*

$$\frac{\mathcal{R}\ :\ a \leadsto b[x_1],\ c \leadsto b[x_2],\ c \leadsto a[1]}{ac \leadsto b[x_2]}$$

Proof. Let $\alpha_1, \alpha_2, \alpha_3$ be the factors associated with the probabilistic conditionals $a \leadsto b[x_1], c \leadsto b[x_2], c \leadsto a[1]$. Using equation (2) we obtain $\alpha_3 = \frac{1}{0}\frac{\alpha_2^{1-x_2} + \alpha_2^{-x_2}}{\alpha_1^{1-x_1}\alpha_2^{1-x_2} + \alpha_1^{-x_1}\alpha_2^{-x_2}}$, thus, by convention, $\alpha_3^0 = 1$ and $\alpha_3^{-1} = 0$. This implies $\alpha_1 = \frac{x_1}{1-x_1}\frac{\alpha_2^{-x_2}+1}{\alpha_2^{1-x_2}+1}$ and $\alpha_2 = \frac{x_2}{1-x_2}\frac{\alpha_1^{-x_1}+0}{\alpha_1^{1-x_1}+0} = \frac{x_2}{1-x_2}\alpha_1^{-1}$, thus $\alpha_1\alpha_2 = \frac{x_2}{1-x_2}$. According to (1), the posterior probability of the conditional $ac \leadsto b$ can now be calculated as follows:
$p^*(b|ac) = \frac{p^*(abc)}{p^*(abc) + p^*(a\bar{b}c)} = \frac{\alpha_1^{1-x_1}\alpha_2^{1-x_2}}{\alpha_1^{1-x_1}\alpha_2^{1-x_2} + \alpha_1^{-x_1}\alpha_2^{-x_2}} = \frac{\alpha_1\alpha_2}{\alpha_1\alpha_2+1} = x_2.$

Proposition 10 states that specific information dominates more general information, as it is expected. In its proof, however, we used essentially that the specificity relation $c \leadsto a[1]$ is categorical. If its probability lies somewhere in between 0 and 1, the equational systems determining the α_i's become more complicated. But it can be solved at least by iteration, e.g. by the aid of SPIRIT (cf. [19]), if the conditional probabilities involved are numerically specified. Within a qualitative probabilistic context, Adam's ϵ-semantics [1] presents a method to handle exceptions and to take account of subclass specificity. Goldszmidt, Morris and Pearl [6] showed how reasoning based on infinitesimal probabilities may be improved by using ME-principles.

Example 3. A knowledge base is to be built up representing "Typically, students are adults", "Usually, adults are employed" and "Mostly, students are not employed" with probabilistic degrees of uncertainty $0.99 (< 1)$, 0.8 and 0.9, respectively. Let A, S, E denote the propositional variables A = Being an *A*dult, S = Being a *S*tudent, and E = Being *E*mployed. The quantified conditional information may be written as $\mathcal{R} = \{s \leadsto a[0.99], a \leadsto e[0.8], s \leadsto \bar{e}[0.9]\}$. From this, SPIRIT calculates $p^*(\bar{e}|as) = 0.8991 \approx 0.9$. So the more specific information s dominates a clearly, but not completely.

5.2 Cautious monotony and cautious cut

Obviously, ME-logic is nonmonotonic: conjoining the antecedent of a conditional with a further literal may alter the probability of the conditional dramatically (cf. Example 3). But a weak form of monotony is reasonable and can indeed be proved:

Proposition 11 (Cautious Monotony). *Suppose A, B, C to be propositional variables, $x_1, x_2 \in [0,1]$. Then*

$$\frac{\mathcal{R} \,:\, a \rightsquigarrow b[x_1],\, a \rightsquigarrow c[x_2]}{ab \rightsquigarrow c[x_2]} \tag{6}$$

Proof. Let α_1 be the ME-factor belonging to the first conditional, and α_2 that of the second one. Then immediately $\alpha_1 = \dfrac{x_1}{1-x_1}$ and $\alpha_2 = \dfrac{x_2}{1-x_2}$, by (2), so that $p^*(c|ab) = \dfrac{\alpha_1^{1-x_1}\alpha_2^{1-x_2}}{\alpha_1^{1-x_1}\alpha_2^{1-x_2} + \alpha_1^{1-x_1}\alpha_2^{-x_2}} = \dfrac{\alpha_2}{\alpha_2+1} = x_2$.

(6) illustrates how ME-propagation respects conditional independence (cf. [22]): $p^*(c|ab) = p^*(c|a) = x_2$.

The monotony inference rule deals with *adding* information to the antecedent. Another important case arises if literals in the antecedent have to be *deleted*. Of course we cannot expect the classical cut rule to hold. But, as in the case of monotony, a *cautious cut rule* may be proved:

Proposition 12 (Cautious Cut). *Suppose A, B, C to be propositional variables, $x_1, x_2 \in [0,1]$. Then*

$$\frac{\mathcal{R} \,:\, ab \rightsquigarrow c[x_1],\, a \rightsquigarrow b[x_2]}{a \rightsquigarrow c[\frac{1}{2}(2x_1x_2 + 1 - x_2)]}$$

Proof. Let α_1, α_2 be the ME-factors associated with the conditionals $ab \rightsquigarrow c[x_1]$, $a \rightsquigarrow b[x_2]$ in $\mathcal{R}$.
Again by using (2), we see $\alpha_1 = \dfrac{x_1}{1-x_1}$ and $\alpha_2 = \dfrac{x_2}{1-x_2}\alpha_1^{x_1}\dfrac{2}{\alpha_1+1}$. According to (1), the probability of the conditional in question may be calculated as follows:
$p^*(c|a) = \dfrac{\alpha_1^{1-x_1}\alpha_2^{1-x_2} + \alpha_2^{-x_2}}{\alpha_1^{1-x_1}\alpha_2^{1-x_2} + \alpha_1^{-x_1}\alpha_2^{1-x_2} + 2\alpha_2^{-x_2}} = \dfrac{\alpha_1^{1-x_1}\alpha_2 + 1}{\alpha_1^{-x_1}\alpha_2(\alpha_2+1) + 2} =$
$= \dfrac{\alpha_1 \dfrac{x_2}{1-x_2}\dfrac{2}{\alpha_1+1} + 1}{2\dfrac{x_2}{1-x_2} + 2} = \dfrac{1}{2}(2x_1x_2 + 1 - x_2).$

5.3 Conjoining literals in antecedent and consequence

The following deduction schemes deal with various cases of inferring probabilistic conditionals with literals in antecedents and consequences being conjoined. Three of them – *Conjunction Left, Conjunction Right, (ii) and (iii)* – are treated in [23] under similar names, thus allowing a direct comparison of ME-inference to probabilistic local bounds propagation. Cautious monotony (6) may be found in that paper, too, where it is denoted as *Weak Conjunction Left.* We will omit the straightforward proofs.

Proposition 13 (Conjunction Right). *Suppose A, B, C to be propositional variables, $x_1, x_2 \in [0,1]$. Then the following ME-inference rules hold:*

$$(i) \quad \frac{\mathcal{R} \,:\, a \rightsquigarrow b[x_1],\, a \rightsquigarrow c[x_2]}{a \rightsquigarrow bc[x_1 x_2]}$$

$$(ii) \quad \frac{\mathcal{R} \,:\, a \rightsquigarrow b[x_1],\, ab \rightsquigarrow c[x_2]}{a \rightsquigarrow bc[x_1 x_2]}$$

$$(iii) \quad \frac{\mathcal{R} \,:\, a \rightsquigarrow b[x_1],\, b \rightsquigarrow c[x_2]}{a \rightsquigarrow bc[x_1 x_2]}$$

Proposition 14 (Conjunction Left). *Suppose A, B, C to be propositional variables, $x_1, x_2 \in [0,1]$. Then*

$$\frac{\mathcal{R} \,:\, a \rightsquigarrow b[x_1],\, a \rightsquigarrow bc[x_2]}{ab \rightsquigarrow c[\frac{x_2}{x_1}]}$$

5.4 Reasoning by cases

The last inference scheme presented in this paper will show how probabilistic information obtained by considering exclusive cases is being processed at maximum entropy:

Proposition 15 (Reasoning by cases). *Suppose A, B, C to be propositional variables, $x_1, x_2 \in [0,1]$. Then*

$$\frac{\mathcal{R} \,:\, ab \rightsquigarrow c[x_1],\, a\bar{b} \rightsquigarrow c[x_2]}{\begin{array}{c} a \rightsquigarrow b[(1 + \dfrac{x_1^{x_1}(1-x_1)^{1-x_1}}{x_2^{x_2}(1-x_2)^{1-x_2}})^{-1}], \\ a \rightsquigarrow c[x_1(1 + \dfrac{x_1^{x_1}(1-x_1)^{1-x_1}}{x_2^{x_2}(1-x_2)^{1-x_2}})^{-1} + x_2(1 + \dfrac{x_2^{x_2}(1-x_2)^{1-x_2}}{x_1^{x_1}(1-x_1)^{1-x_1}})^{-1}] \end{array}}$$

Proof. The ME-factors α_1 and α_2 associated with the conditionals in $\mathcal{R}$ (in order of appearance above) are computed to be $\alpha_1 = \dfrac{x_1}{1-x_1}$ and $\alpha_2 = \dfrac{x_2}{1-x_2}$. Following (1) we thus obtain

$$p^*(b|a) = \frac{\alpha_1^{1-x_1} + \alpha_1^{-x_1}}{\alpha_1^{1-x_1} + \alpha_1^{-x_1} + \alpha_2^{1-x_2} + \alpha_2^{-x_2}} = \frac{\alpha_1^{-x_1}(\alpha_1 + 1)}{\alpha_1^{-x_1}(\alpha_1+1) + \alpha_2^{-x_2}(\alpha_2+1)} =$$
$$= \frac{1}{1 + \dfrac{x_1^{x_1}(1-x_1)^{1-x_1}}{x_2^{x_2}(1-x_2)^{1-x_2}}},$$
as desired. The probability of the second conditional $a \rightsquigarrow c$ is proved by applying the fundamental probabilistic equality $p^*(c|a) = p^*(c|ab)p^*(b|a) + p^*(c|a\bar{b})p^*(\bar{b}|a)$, and using the information given by $\mathcal{R}$.

6 Concluding remarks

We showed that using the principles of optimum entropy provides a powerful and sound machinery for probabilistic reasoning. The presentation of an ME-distribution given by (1) and (2) is crucial for understanding the logical mechanisms that underlie ME-adjustment, as it is important for practical reasoning. The inference patterns presented in Section 5 only make use of the principle of maximum entropy to represent conditional knowledge, and perhaps, this will be a major field of application for ME-reasoning. But, as the axiom (P3) necessary for the characterization shows, the process of adjusting prior probabilistic knowledge to new information in a logically consistent way – and thus using the more general cross-entropy – is indispensable for the whole principle.

One of the most striking features of ME-reasoning is its thoroughness: In fact, only one method – that of minimizing cross-entropy – is used to realize representation, adaptation and instantiating of probabilistic knowledge. The last statement arises from the fact that instantiating a distribution P with respect to evidence A means taking the conditional distribution $P(\cdot|\mathrm{A})$, and this amounts obviously to calculating the ME-distribution $P *_e \{\mathrm{A}[1]\}$.

Considering the principle of maximum entropy as subordinate to the principle of minimum cross-entropy apparently means to accept that total ignorance is represented by uniform distributions. This has been disputed for decades if not even for centuries (cf. e.g. [9]). The calculations in this paper, however, based on the formulas (1) and (2), show clearly how ignorance is handled by ME-reasoning: *Non-knowledge* is realized as *non-occurrence*. In fact, the normalizing factor α_0 which includes the prior uniform probabilities is always cancelled in inferring posterior *conditional* probabilities. Only if *facts* are to be deduced, α_0 affects the posterior probabilities, but merely representing the probabilistic convention $P^*(\Omega) = 1$.

References

1. E.W. Adams. *The Logic of Conditionals.* D. Reidel, Dordrecht, 1975.
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. R.T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946.
4. D. Dubois and H. Prade. Conditional objects and non-monotonic reasoning. In *Proceedings 2nd Int. Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 175–185. Morgan Kaufmann, 1991.
5. D. Dubois, H. Prade, and J.-M. Toucas. Inference with imprecise numerical quantifieres. In Z.W. Ras and M. Zemankova, editors, *Intelligent Systems - state of the art and future directions*, pages 52–72. Ellis Horwood Ltd., Chichester, England, 1990.
6. M. Goldszmidt, P. Morris, and J. Pearl. A maximum entropy approach to non-monotonic reasoning. In *Proceedings AAAI-90*, pages 646–652, Boston, 1990.

7. I.J. Good. Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.*, 34:911–934, 1963.
8. A.J. Grove, J.Y. Halpern, and D. Koller. Random worlds and maximum entropy. *J. of Artificial Intelligence Research*, 2:33–88, 1994.
9. E.T. Jaynes. *Papers on Probability, Statistics and Statistical Physics.* D. Reidel Publishing Company, Dordrecht, Holland, 1983.
10. R.W. Johnson and J.E. Shore. Comments on and correction to "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy". *IEEE Transactions on Information Theory*, IT-29(6):942–943, 1983.
11. G. Kern-Isberner. Characterizing the principle of minimum cross-entropy within a conditional logical framework. Informatik Fachbericht 206, FernUniversität Hagen, 1996.
12. G. Kern-Isberner. Conditional logics and entropy. Informatik Fachbericht 203, FernUniversitaet Hagen, 1996.
13. S. Kullback. *Information Theory and Statistics.* Dover, New York, 1968.
14. S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities in graphical structures and their applications to expert systems. *Journal of the Royal Statistical Society B*, 50(2):415–448, 1988.
15. N.Rescher. *Many-Valued Logic.* McGraw-Hill, New York, 1969.
16. D. Nute. *Topics in Conditional Logic.* D. Reidel Publishing Company, Dordrecht, Holland, 1980.
17. J.B. Paris and A. Vencovská. A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, 14:183–223, 1990.
18. J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, Ca., 1988.
19. W. Rödder and C.-H. Meyer. Coherent knowledge processing at maximum entropy by spirit. In E. Horvitz and F. Jensen, editors, *Proceedings 12th Conference on Uncertainty in Artificial Intelligence*, pages 470–476, San Francisco, Ca., 1996. Morgan Kaufmann.
20. J.E. Shore. Relative entropy, probabilistic inference and AI. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 211–215. North-Holland, Amsterdam, 1986.
21. 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.
22. J.E. Shore and R.W. Johnson. Properties of cross-entropy minimization. *IEEE Transactions on Information Theory*, IT-27:472–482, 1981.
23. H. Thöne, U. Güntzer, and W. Kiessling. Towards precision of probabilistic bounds propagation. In D. Dubois, M.P. Wellmann, B. D'Ambrosio, and P. Smets, editors, *Proceedings 8th Conference on Uncertainty in Artificial Intelligence*, pages 315–322, San Mateo, Ca., 1992. Morgan Kaufmann.
24. J. Whittaker. *Graphical models in applied multivariate statistics.* John Wiley & Sons, New York, 1990.

This article was processed using the LaTeX macro package with LLNCS style