



HAL
open science

Qualitative reasoning with imprecise probabilities

Didier Dubois, Lluís Godó, Ramon López de Màntaras, Henri Prade

► **To cite this version:**

Didier Dubois, Lluís Godó, Ramon López de Màntaras, Henri Prade. Qualitative reasoning with imprecise probabilities. *Journal of Intelligent Information Systems*, 1993, 2 (4), pp.319–363. 10.1007/BF00961659 . hal-04064867

HAL Id: hal-04064867

<https://hal.science/hal-04064867>

Submitted on 11 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Qualitative Reasoning with Imprecise Probabilities*

DIDIER DUBOIS

dubois@irit.fr, prade@irit.fr

Institut de Recherche en Informatique de Toulouse CNRS, Université Paul Sabatier, 31062 Toulouse Cedex, France

LLUIS GODO AND RAMON LÓPEZ DE MÀNTARAS

godo@ceab.es, mantaras@ceab.es

Institut d'Investigació en Intel·ligència Artificial, CEAB-CSIC, 17300 Blanes, Spain

HENRI PRADE

dubois@irit.fr, prade@irit.fr

Institut de Recherche en Informatique de Toulouse CNRS, Université Paul Sabatier, 31062 Toulouse Cedex, France

Abstract. This paper investigates the possibility of performing automated reasoning in probabilistic knowledge bases when probabilities are expressed by means of linguistic quantifiers. Data are expressed in terms of ill-known conditional probabilities represented by linguistic terms. Each linguistic term is expressed as a prescribed interval of proportions. Then instead of propagating numbers, qualitative terms are propagated in accordance with the numerical interpretation of these terms. The quantified syllogism, modeling the chaining of probabilistic rules, is studied in this context. It is shown that a qualitative counterpart of this syllogism makes sense and is fairly independent of the thresholds defining the linguistically meaningful intervals, provided that these threshold values remain in accordance with the intuition. The inference power is less than a full-fledged probabilistic constraint propagation device but corresponds better to what could be thought of as commonsense probabilistic reasoning. Suggestions that may improve the inferencing power in the qualitative setting are proposed.

Keywords: conditional probabilities, interval-valued probabilities, qualitative probabilities, linguistic quantifiers, syllogistic reasoning, nonmonotonic reasoning

1. Introduction

Precise values of probabilities are not always available. Experts often assess probabilities under the form of intervals (e.g., “between 80% and 90% of *As* are *Bs*”) or even linguistically (e.g., “almost all *As* are *Bs*”), or are only able to rank-order probability values, stating that one probability is certainly greater than another. Thus it raises the question of the possibility of reasoning with probabilities in a qualitative way. The main appeal of a qualitative approach (when such an approach is feasible), is that it requires less precision than a pure numerical

*This paper is an extended and revised version of a paper entitled “A Symbolic Approach to Reasoning with Linguistic Quantifiers” in *Proc. 8th Conf. Uncertainty in Artificial Intelligence* Dubois, D., Wellman, M.P., D’Ambrosio, B., and Smets, Ph. (Eds.), Morgan Kaufmann, pp. 74–82, 1992.

representation insofar as it still leads to meaningful conclusions in the reasoning process. Also, the qualitative approach allows us to have a better interface with human users, in a way more compatible with their own reasoning processes.

A major part of this paper deals with syllogisms involving linguistic quantifiers such as “few” and “most.” The extension of logical syllogisms to quantifiers distinct from the existential one (there exists..., denoted by \exists) and the universal one (for all..., denoted by \forall) currently used in first-order logic looks natural to be considered. In the nineteenth century philosophers such as De Morgan (1966, p. 9 and pp. 242–246) and Boole (1854, pp. 284–285) studied such quantified syllogisms. Quantifiers like “most” and “few” have been more recently considered mainly from a linguistic point of view (e.g. Peterson (1979)). Nowadays, numerical quantifiers receive attention from philosophical logic (see Van Benthem (1984), and Westerstahl (1989)). However, these works are devoted to integer-valued quantifiers like “4 As are Bs”, and mainly consider their embedding into a logical system involving conditionals. Integer-valued quantifiers have been implemented in semantic networks by Shapiro (1979), for instance.

Another, maybe more usual, approach to numerical quantifiers is to handle relative proportions in the description of conditional statements. Namely “most As are Bs” means that the relative cardinality $|A \cap B|/|A|$ lies in an interval which is not far from 1. More generally, this relative cardinality can be changed into a conditional probability $P(B|A)$. We then get close to probabilistic logics, as surveyed in Bacchus (1990). There have been some attempts in the past to handle conditional probabilities in inference systems. However, most of them have some limitations. The Bayesian approach to conditional probability always assumes a unique underlying joint probability (Pearl, 1988). In contrast, inference systems such as INFERNO (Quinlan, 1983) rely on some knowledge of bounds on absolute and conditional probabilities and compute bounds on other absolute probabilities from the former ones. The type of inference method considered here is about deriving bounds on conditional probabilities from the knowledge of other conditional probabilities only. Systematic results on quantified syllogisms have been obtained and this paper investigates whether these results can be exploited in a qualitative format, where conditional probabilities are changed into linguistic labels.

It is obvious that progress in the handling of conditional probabilities may have some impact on information systems. Very early, the encoding of expert knowledge has taken the form of weighted “if... then...” rules, and the main question has been to elucidate the meaning of such weights. It turns out that there is not a single meaning for all situations. But in many cases the weight is naturally interpreted as a numerical quantifier often roughly assessed in a linguistic way that corresponds to an estimation of the amount of exceptions to the rule. A proper handling of such linguistic quantifiers might help information systems supply answers to queries in a more realistic way than purely logic-oriented systems would do. Basically it would enable the information system to address queries of the form “how many As are Bs”, “how often As are Bs?”, “what is the probability

of B for a standard example of A ?", etc. In this respect our approach differs from those adapted to purely statistical information systems, where the analysis of a lot of examples has enabled a probability distribution to be completely determined, and stored under the form of, for instance, a Bayesian net. Here we assume that only some conditional probabilities are known in a rough way, and the question-answering problem is to infer conditional probability bounds (possibly in a linguistic format) from other already known ones, the knowledge of which is not enough to determine a unique probability distribution.

The idea of reasoning qualitatively with probabilities has been investigated along different lines by various researchers in artificial intelligence, especially in the last five years. A first family of approaches works with inequalities between probabilities. Wellman (1988, 1990) deals with qualitative probabilistic networks where positive influences between nodes capture statements of the form "A makes B more likely" (than "not A"), and reflect inequalities between probabilities of the type $P(B|Ax) \geq P(B|\bar{A}x)$ where x denotes the context. Neufeld (1990) uses $P(B|A) > P(B)$ for expressing that "A favors B," as a qualitative probability counterpart of a default rule. A second type of approach consists in defining probability-like functions which take their values in an abstract finite totally ordered semigroup rather than in $[0, 1]$ (Aleliunas, 1988; Xiang, Beddoes and Poole, 1990); however, these authors do not try to interpret the symbolic probability values as numerical subintervals of $[0, 1]$ nor to check the consistency of their calculus with this interpretation. Another kind of qualitative probability approach is Adams' (1975) conditional logic (see also Pearl (1988), Geffner (1992)) which manipulates probabilistic-like default rules "if A then B generally" understood as $P(B|A) \geq 1 - \epsilon$, where ϵ is infinitesimal, a very demanding interpretation. At a purely symbolic level let us also mention work on so-called conditional objects trying to give a meaning to the entity " $B|A$ " independently of the notion of probability, but still in agreement with this notion in the sense that $P(B|A)$ can indeed be considered as the probability of the entity " $B|A$ " (Goodman et al., 1991; Dubois and Prade, 1991).

The approach developed in this paper maintains an interpretation of qualitative (linguistic) probability values in terms of numerical intervals. Here, linguistic quantifiers such as *most*, *few*, etc., are viewed as imprecisely or fuzzily known conditional probabilities, i.e., terms represented by crisp or in the most general case, fuzzy, subintervals of $[0, 1]$ (Zadeh, 1985; Dubois and Prade, 1988; Spies, 1989). In the approach presented here, first an ordered set of elementary labels of quantifiers is chosen in order to provide a linguistic scale for conditional probabilities (or proportions) used in default rules like " Q As are Bs" where Q is viewed as the answer to the question: "how many As are Bs?". A qualitative algebra (Q -algebra) (Travé-Massuyès and Piera, 1989) is defined on the set of possible labels, built from the elementary labels constituting the scale. Inference rules which are the qualitative counterparts of numerical formulas for computing bounds on probabilities in quantified syllogisms or similar propagation rules, can be proposed for reasoning in qualitative probability networks.

The next section gives the necessary background about local patterns of inference used to propagate constraints on probabilities known to belong to intervals. Two inference rules, namely, the quantified syllogism and the generalized Bayes theorem, are reviewed. Section 3 illustrates the interest of the quantified syllogism on two cases in statistics and nonmonotonic reasoning respectively. Section 4 discusses how to build a set of linguistic labels to be used in the qualitative probability computations; qualitative versions of these rules of inference are defined and the robustness of the approach is discussed, i.e., to what extent the qualitative calculus remains unchanged when the numerical interpretation of the linguistic labels is slightly modified. Section 5 discusses the problems encountered when trying to extend to a qualitative setting a constraint propagation rule based on Bayes theorem. Section 6 gives one example that shows how the constraint propagation-based strategy, described in Section 2, can answer queries about linguistic conditional probabilities once adapted to the qualitative setting. Section 7 also indicates that the proposed approach, based on absolute orders of magnitude, is very limited for Bayesian nets, which leads to a suggestion of using relative orders of magnitude as *supplementary information*, on top of absolute orders of magnitude. These would lead to more precise results whenever products or quotient of qualitative probabilities need be performed.

2. Local propagation of interval-valued probabilities

Consider the following inference problem: given three predicates A, B, C which model some properties of interest, and given the precise knowledge of conditional probabilities $P(A|B), P(B|A), P(C|B), P(B|C)$, what is the value of the conditional probabilities $P(C|A)$ and $P(A|C)$ obtained from these four conditional probabilities? It corresponds to a numerical model of inference from uncertain pieces of knowledge, called the quantified syllogism, an example of which is

most students are young ; few young people are students

most young people are single ; about half singles are young

what can be said about students being single and singles being student?

where the terms “most,” “few,” “about half” above are understood precisely as, say, $P(\text{young}|\text{student}) = 0.9$, $P(\text{student}|\text{young}) = 0.2$, $P(\text{single}|\text{young}) = 0.8$, $P(\text{young}|\text{single}) = 0.5$, respectively. Exact bounds for $P(A|C)$ and $P(C|A)$ have been derived by Dubois et al. (1990) and are as follows, given the four probabilities $P(A|B), P(B|A), P(C|B), P(B|C)$ strictly lie between 0 and 1:

$$P(C|A) \geq P(B|A) \cdot \max \left(0, 1 - \frac{1 - P(C|B)}{P(A|B)} \right), \quad (1)$$

$$P(C|A) \leq \min \left(1, 1 - P(B|A) + \frac{P(B|A)P(C|B)}{P(A|B)}, \frac{P(B|A)P(C|B)}{P(A|B)P(B|C)}, \frac{P(B|A)P(C|B)}{P(A|B)P(B|C)}(1 - P(B|C)) + P(B|A) \right). \quad (2)$$

The complete proof of these bounds and their optimality is given in the Appendix (Theorems 1 and 2). Degenerated cases when $P(A|B) = 0$ or $P(B|C) = 0$ are also studied. With the above numerical values, we get that $P(\text{student} | \text{single}) \leq 5/36 = 0.1388$ and that $P(\text{single} | \text{student})$ is *left totally unknown*, i.e., belongs to $[0, 1]$ without being able to restrict the value.

Denoting by $P_*(C|A)$ and $P^*(C|A)$ the best lower and upper bounds, respectively, obtained from (1) and (2), it can be proved that very often $P_*(C|A) < P^*(C|A)$, i.e., the obtained *probability estimate is imprecise* (that is, not completely determined) as in the above example. In fact the only cases when $P^*(C|A) = P_*(C|A)$ i.e., which yield precise results are the following obvious cases (see Theorem 3 in the Appendix):

- $P(B|A) = P(C|B) = 1$; hence $P(C|A) = 1$. This is a form of the usual logical chaining of rules.
- $P(B|A) = P(A|B) = 1$; then $P(C|A) = P(C|B)$, because A and B are indiscernible.
- $P(C|B) = P(B|C) = 1$; then $P(C|A) = P(B|A)$, because C and B are indiscernible.
- $P(A|B) = P(B|C) = 1$; then $P(C|A) = P(C|B) \cdot P(B|A)$. This is because in that case, $P(B|A) = P(B)/P(A)$, $P(C|B) = P(C)/P(B)$, and $P(C|A) = P(C)/P(A)$.

This inference pattern can be extended to cope with incompletely specified input probabilities as follows:

As are Bs with $P(B A)$	Bs are As with $P(A B)$
$\in [P_*(B A), P^*(B A)];$	$\in [P_*(A B), P^*(A B)]$
Bs are Cs with $P(C B)$	Cs are Bs with $P(B C)$
$\in [P_*(C B), P^*(C B)];$	$\in [P_*(B C), P^*(B C)]$
As are Cs with $P(C A) \in ?$	Cs are As with $P(A C) \in ?$

where P_* and P^* , respectively, denote lower and upper bounds. Again, we are interested in computing (the tightest) bounds which can be deduced for $P(C|A)$ and $P(A|C)$.

The following bounds have been established and have been shown to be the tightest ones when no division by zero occurs in the expressions:

Lower bound:

$$P_*(C|A) = P_*(B|A) \cdot \max \left(0, 1 - \frac{1 - P_*(C|B)}{P_*(A|B)} \right). \tag{3}$$

Upper bound:

$$P^*(C|A) = \min \left(\begin{array}{l} 1, 1 - P_*(B|A) + \frac{P_*(B|A) \cdot P^*(C|B)}{P_*(A|B)}, \\ \frac{P^*(B|A) \cdot P^*(C|B)}{P_*(A|B) \cdot P_*(B|C)}, \\ \frac{P^*(B|A) \cdot P^*(C|B)}{P_*(A|B) \cdot P_*(B|C)} [1 - P_*(B|C)] + P^*(B|A), \\ \frac{P^*(C|B)}{P^*(C|B) + P_*(B|C) \cdot (P_*(A|B) - P^*(C|B))} \end{array} \right). \tag{4}$$

Related local patterns of inference for interval-valued conditional probabilities have been independently developed by Guntzer, Kießing and Thöne (1991), Thöne et al. (1991) and by Heinsohn (1991) in the contexts of deductive databases and terminological languages, respectively. Thöne, Guntzer and Kießing (1992) have pointed out the necessity of the additional term in (4). This is basically due to the fact that the third and fourth terms in (2) are linearly increasing with respect to $P(B|A)$ while the second term is linearly decreasing in $P(B|A)$ if $P^*(C|B) < P_*(A|B)$. See the Appendix (Theorem 4) for a complete proof, including degenerated cases.

Another local inference rule is proposed in Amarger et al. (1991b). It takes advantage of an extended form of Bayes rule expressed in terms of conditional probabilities only, namely

$$\forall A_1, \dots, A_k, P(A_1|A_k) = P(A_k|A_1) \cdot \prod_{i=1, k-1} \frac{P(A_i|A_{i+1})}{P(A_{i+1}|A_i)} \tag{5}$$

(with all involved quantities positive), from which useful inequalities are obtained in the case where only lower and upper bounds are available.

The usual Bayes theorem is recovered when we choose $k = 3$, and $A_2 = U$, the referential set such that $A_1 = A \subseteq U$ and $A_3 = B \subseteq U$. Then (5) becomes

$$P(A|B) = \frac{P(B|A) \cdot P(A|U) \cdot P(U|B)}{P(U|A) \cdot P(B|U)} = \frac{P(B|A) \cdot P(A)}{P(B)}, \tag{6}$$

noticing that unconditional probabilities such as $P(A)$ can be written as $P(A|U)$, and $P(U|A) = 1$. Another interesting case is when $A_1 = A, A_2 = B, A_3 = C, A_4 = \bar{B}$, the complement of B . Then (5) becomes

$$P(A|\bar{B}) = \frac{(1 - P(B|A)) \cdot P(A|B) \cdot P(B|C) \cdot P(C|\bar{B})}{P(B|A) \cdot P(C|B) \cdot (1 - P(B|C))}, \tag{7}$$

i.e., $P(A|\bar{B})$ is a function of $P(B|A)$, $P(A|B)$, $P(B|C)$, $P(C|B)$, and $P(C|\bar{B})$ for any sets A, B, C .

The constraint propagation method which is used for the processing of a set of conditional probability bounds is the following: recursively apply the quantified syllogism to generate upper and lower bounds of missing probabilities. This step is performed until the probability intervals can no longer be improved. Then recursively the extended Bayes rule is applied to improve the bounds thus generated, and the whole procedure is continued until no improvement takes place. It has been observed that this constraint propagation method can sometimes give bounds which are as tight as the best ones computed by a global optimization method based on linear programming (see Amarger et al. (1991b)). However, not all of the laws of probability are captured by (3-4-5). It is possible to find cases when the combined use of the inference rules (3), (4), (5) is inefficient. Such cases can be found in Heinsohn (1991), for instance. Despite the incompleteness of this set of two propagation rules (the quantified syllogism and the generalized Bayes theorem), they have been proved to be generally efficient both in computation time and accuracy of probability bounds when only positive information relating classes to one another (without involving negation) is available, and simple queries of the form $P(A|B)$, with A and B being elementary categories, are solved (see Amarger et al. (1991b)). However more work is needed to clarify under what condition the two propagation rules are sufficient to obtain the tightest bounds.

These rules of inference are not capable of dealing with negation in a totally satisfactory way. Especially, if no knowledge about absolute probabilities is available, nothing can be known about probabilities of the form $P(A|\bar{B})$, given only $P(B|A), P(A|B)$. Namely, using Bayes theorem:

$$P(A|\bar{B}) = (1 - P(B|A)) \frac{P(A)}{1 - P(B)}$$

Moving $P(A)$ in $[0, 1]$ and $P(B)$ in $[0, 1)$, $P(A|\bar{B})$ ranges over the whole unit interval as well. Namely, fixing $P(B|A)$ and $P(A|B)$, only the quotient $P(A)/P(B) = P(A|B)/P(B|A) = \lambda$ is fixed. λ is a positive number as long as $P(A|B) \neq 0$ (or equivalently $P(B|A) \neq 0$). Then we have

$$\begin{aligned} P(A|\bar{B}) &= \frac{(1 - P(B|A)) \lambda P(B)}{1 - P(B)} \\ &= 0 \quad \text{if } P(B) = 0 \\ &= 1 \quad \text{if } P(B) = \frac{P(B|A)}{P(A|B) + P(B|A) - P(A|B) \cdot P(B|A)} \end{aligned}$$

Particularly, $P(A|\bar{B})$ and $P(A|B)$ are unrestricted and unrelated quantities when $P(A) \cdot P(B) \notin \{0, 1\}$.

Note that the knowledge of other conditional probabilities such as $P(C|B)$ and $P(B|C)$ may create some constraints on $P(A|\bar{B})$ since from (7), $P(A|\bar{B})$ is

proportional to $P(C|\bar{B})$. When the latter moves in $[0, 1]$, we get

$$0 \leq P(A|\bar{B}) \leq \frac{(1 - P(B|A)) \cdot P(A|B) \cdot P(B|C)}{P(B|A) \cdot P(C|B) \cdot (1 - P(B|C))},$$

i.e., we may get a nontrivial bound on the “negative” conditional probability $P(A|\bar{B})$ if the upper bound is smaller than 1.

3. Some applications of the quantified syllogism

In this section we show two examples of situations where the quantified syllogism rule looks useful: a classical question in statistics and the study of the approximation made when assuming that small probabilities behave like infinitesimals.

3.1. Can we apply statistics-based probabilities to individual cases?

The case $P(B|A) = 1$ (“all A s are B s”) and $P(B|C) = 1$ (“all C s are B s”) represents a typical case of statistical inference, when knowing from statistical data the probability $P(C|B)$, and considering some individual in class B , one tries to say something about its probability of being a C . Namely, B represents a population, C a subclass of this population for which the proportion or probability $P(C|B)$ is known. For instance, B represents the inhabitants of some city and C the proportion of individuals in that population that are older than 60. Now take an individual x_0 in B . What is the probability that he is more than 60? There are several ways of considering x_0 according to its peculiarities. Let A be the maximal subset of B containing individuals “just like x_0 .” It means that A is the subset of individuals in B that have enough common features with x_0 . Note that A can range from $\{x_0\}$ (if x_0 is so particular that nobody is like him) to B itself (if x_0 is viewed as having nothing special that makes him differ from other individuals in B). The problem is then: knowing $P(C|B)$, what is the probability that x_0 belongs to C ? This problem can be solved by computing $P(C|A)$ where A is a maximal subclass of B , such that x_0 is a typical element of A .

This problem corresponds to the quantified syllogism where $P(B|A) = 1$ and $P(B|C) = 1$, by definition. The probability to be computed, namely $P(C|x_0)$, is supposed to be equal to $P(C|A)$ and depends on $P(A|B)$; other probabilities in the syllogism are known.

This phenomenon can be precisely studied in an analytical way, letting $P(C|B) = \alpha$, and $P(A|B) = t \neq 0$ as $A \subset B$, and $P(A) \neq 0$ by assumption. The parameter t can be called a typicality index of the set A with respect to B . It expresses the probability that selecting at random an individual in B , it lies in A i.e., it is “like x_0 .” The commonsense saying that statistics should be cautiously used when making decisions about individual situations can be given a precise form thanks to the quantified syllogism. When $P(A|B) = t$, $P(B|A) = 1$, $P(C|B) = \alpha$, $P(B|C) = 1$, we get the following results

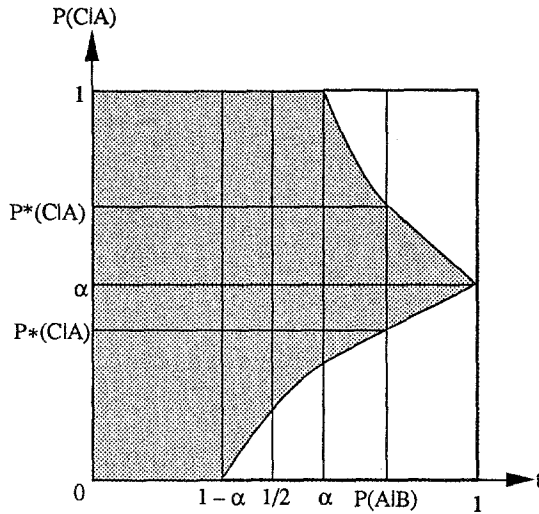


Figure 1. Statistical inference and typicality.

for $P(C|A)$ using (1-2):

$$P_*(C|A) = \max\left(0, 1 - \frac{1-\alpha}{t}\right); \quad P^*(C|A) = \min\left(1, \frac{\alpha}{t}\right) \quad (8)$$

We obtain the portions of hyperbola in Figure 1. As shown in this figure the only case when $P(C|A)$ must be equal to $P(C|B)$ is when $t = 1$, i.e., when the reference class of x_0 is B itself. In other words, all individuals in B should be like x_0 , which implies that B is homogeneous enough. This result nicely fits our intuition that one should not trust statistical knowledge based on a heterogeneous population. Let us consider the situation where $P(C|B) > 1/2$ as in Figure 1. If the degree of typicality $t \leq P(C|B) = \alpha$ then the probability $P(C|A)$ is no longer upper-bounded by a bound strictly less than 1. When the typicality t is low enough, that is $t \leq \min(P(C|B), 1 - P(C|B))$ nothing can be inferred for $P(C|A)$, i.e., $P_*(C|A) = 0$ and $P^*(C|A) = 1$. It corresponds to the case when A and C could be disjoint subsets of B . This phenomenon also suggests the existence of cases where despite the high values of some of the probabilities the result of the chaining is very imprecise, i.e., a big gap exists between $P_*(C|A)$ and $P^*(C|A)$.

3.2. A qualitative analysis of Adams' inference rules

Adams (1975) has proposed a probabilistic inference system based on the three rules:

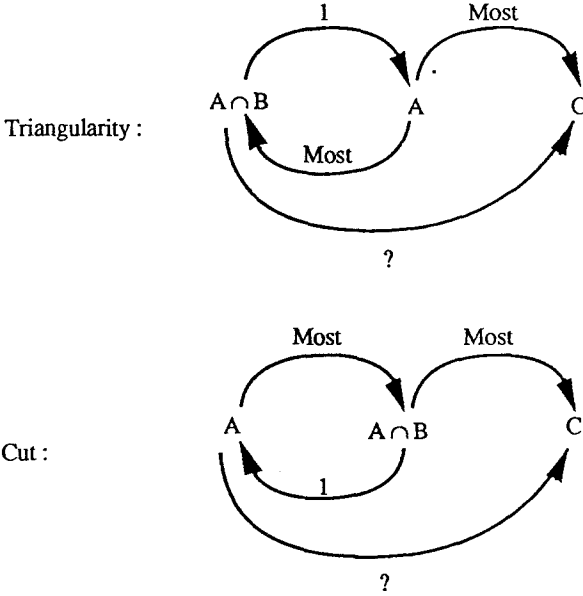


Figure 2. Triangularity and cut rules.

Triangularity: $A \rightarrow B, A \rightarrow C \Rightarrow (A \cap B) \rightarrow C$.

Cut: $A \rightarrow B, (A \cap B) \rightarrow C \Rightarrow A \rightarrow C$.

Disjunction: $A \rightarrow C, B \rightarrow C \Rightarrow (A \cup B) \rightarrow C$.

which are sound when $A \rightarrow B$ means that $P(B|A)$ is high (i.e., close to 1), and is understood as $P(B|A) \leq 1 - \epsilon$ where ϵ is arbitrarily small. These rules are used in Pearl (1988) to build a probabilistic-inference-like default logic. It is interesting to consider finistic semantics for these rules in relationship with imprecise probabilities. In this respect $A \rightarrow B$ will be interpreted as “most As are Bs.” First, it is easy to verify that triangularity and cut rules can be expressed in terms of the quantified syllogism, of which they are special cases, noticing that $P(B|A) = P(A \cap B|A)$. See Figure 2.

Taking “most” = $[1 - \alpha, 1)$ where $\alpha < 1/2$, we easily get the lower bound on $P(C|A \cap B)$ and $P(C|A)$ in each case by using the quantified syllogism

$$P_*(C|A \cap B) = \max \left(0, 1 - \frac{1 - P_*(C|A)}{P_*(A \cap B|A)} \right) = \frac{1 - 2\alpha}{1 - \alpha},$$

$$P_*(C|A) = P_*(B|A) \cdot P_*(C|A \cap B) = (1 - \alpha)^2.$$

There is a degradation of the lower bounds. However these lower bounds are greater than α when $\alpha \leq (3 - \sqrt{5})/2 \simeq 0.38$. Indeed this threshold is found when we compute α such that $(1 - 2\alpha)/(1 - \alpha) \geq \alpha$ or equivalently $(1 - \alpha)^2 \geq \alpha$.

The third axiom pertains to another kind of inference that does not directly relate to the quantified syllogism. In Amarger et al. (1991a) the following identity was obtained:

$$P(C|A \cup B) = \frac{P(C|A)/P(B|A) + P(C|B)/P(A|B) - P(C|A \cap B)}{1/P(B|A) + 1/P(A|B) - 1}. \tag{9}$$

Hence, a lower bound to $P(C|A \cup B)$ is obtained by putting $P(C|A \cap B) = 1$. When $P(C|A) \geq 1 - \alpha$, $P(C|B) \geq 1 - \alpha'$ where α and α' are independent values less than 1/2 (both express “most”), we get if $\alpha = \alpha'$

$$P(C|A \cup B) \geq \frac{K(1 - \alpha) - 1}{K - 1}$$

where $K = 1/P(B|A) + 1/P(A|B) \geq 2$. The right-hand term of the inequality is increasing with K . Hence the lower bound for $P(C|A \cup B) \geq 2(1 - \alpha) - 1 = 1 - 2\alpha$. More generally, $P(C|A \cup B) \geq 1 - \alpha - \alpha'$ when $P(C|A) \geq 1 - \alpha$, $P(C|B) \geq 1 - \alpha'$, as it can be checked from (9). On the whole, we have found finistic counterparts of Adams’ axioms that make it possible to quantify how inaccurate we are when we apply these axioms for commonsense reasoning with high probabilities (here $> 1/2$).

The three axioms can be summarized as

$$\begin{aligned} A \xrightarrow{\alpha} B, A \xrightarrow{\alpha} C &\implies (A \cap B) \xrightarrow{\frac{\alpha}{1-\alpha}} C, \\ A \xrightarrow{\alpha} B, (A \cap B) \xrightarrow{\alpha} C &\implies A \xrightarrow{2\alpha-\alpha^2} C, \\ A \xrightarrow{\alpha} C, B \xrightarrow{\alpha} C &\implies A \cup B \xrightarrow{2\alpha} C, \end{aligned}$$

where $A \xrightarrow{\alpha} B$ reads $P(B|A) \geq 1 - \alpha$. These rules enable probabilistic reasoning to be performed as a qualitative nonmonotonic logic, but where the validity of conclusions can be numerically assessed.

4. The qualitative quantified syllogism

4.1. Lattices of linguistic labels

Let us consider an ordered set of elementary labels of linguistic quantifiers that may account for any probability value. Each label corresponds to a subinterval of the unit interval, and the set of subintervals completely covers it. So a linguistic scale will be made of the labels of a collection of subintervals covering $[0, 1]$ of the form $\{0, (0, a_1], [a_1, a_2], \dots, [a_{n-1}, a_n], [a_n, 1), 1\}$. For convenience we shall still call a partition such a collection, although the intervals overlap at their edges, except in 0 and 1, which are dealt with separately due to their particular meanings corresponding to “none” and “all.”

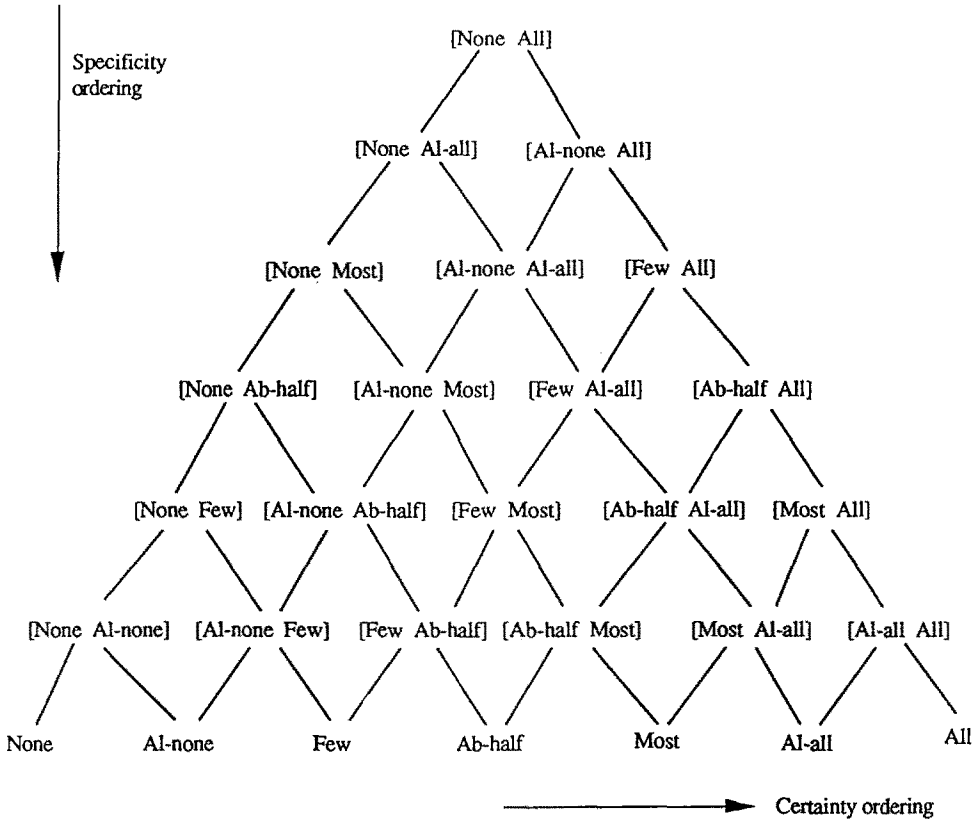


Figure 3. The biordered structure of qualitative probabilities.

Let \mathcal{P} be such a “partition” of $[0, 1]$ in subintervals representing quantifiers from a linguistic scale. By convention, both the linguistic scale and the corresponding partition will be denoted by \mathcal{P} . It seems reasonable that this linguistic scale should be symmetric with respect to 0.5 since the antonym of each linguistic quantifier in the scale should also be in the scale. Linguistic antonymy, for instance $ANT(Almost\ none) = Almost\ all$ or $ANT(Few) = Most$, is expressed at the numerical level by relations like $ANT([a, b]) = [1 - b, 1 - a]$, since intervals are used to represent the meaning of linguistic quantifiers. As a consequence, if $P(A)$ is the probability of an event A , linguistically qualified by $X \in \mathcal{P}$, then $P(\bar{A})$, the probability of the complementary event \bar{A} , should be qualified by $ANT(X) \in \mathcal{P}$ (see, e.g., Peterson (1979) for a systematic study of antonymy in linguistic quantifiers).

The universe of descriptions, U , induced by a partition \mathcal{P} is defined as the set of intervals that are the union of adjacent elements of \mathcal{P} . The set inclusion relationship (\subseteq) equips U with an ordered structure that has a lattice-like representation. Given $X, Y \in U$, X is said to be more specific than Y if and only if $X \subseteq Y$; it means that X and Y are in agreement but X gives tighter bounds than Y .

For instance, if we take parameters a and b to be smaller than 0.5 with $a < b$, then $[0, 1]$ can be (nonstrictly) symmetrically partitioned as

$$\mathcal{P} = \{0, (0, a], [a, b], [b, 1 - b], [1 - b, 1 - a], [1 - a, 1), 1\}$$

corresponding to the following linguistic quantifiers:

- $0 ::= \textit{None}$
- $(0, a] ::= \textit{Almost none (Al-none for short)}$
- $[a, b] ::= \textit{Few}$
- $[b, 1 - b] ::= \textit{About half (Ab-half for short)}$
- $[1 - b, 1 - a] ::= \textit{Most}$
- $[1 - a, 1) ::= \textit{Almost all (Al-all for short)}$
- $1 ::= \textit{All}$

The set \mathcal{P} constitutes the highest meaningful level of specificity with respect to the language. Between this level and the least specific one (i.e., the whole interval $[0, 1]$), the universe of description U contains several internal ordered levels of specificity. For example, with the seven terms defined above we have five levels in between: see the vertical ordering in Figure 3.

The set of elementary (most specific) linguistic quantifiers can also be ordered according to the usual certainty ordering in the unit interval:

$$\textit{None} \leq \textit{Almost none} \leq \textit{Few} \leq \textit{About half} \leq \textit{Most} \leq \textit{Almost all} \leq \textit{All}.$$

Higher level (less specific) elements of the universe U can be considered as intervals defined on the partition set, with respect to the certainty ordering. For instance,

$$[\textit{Few}, \textit{Most}] = \{X \in \mathcal{P} \mid \textit{Few} \leq X \leq \textit{Most}\}$$

and this is compatible with the above numerical interpretation in terms of probability intervals. The semantics of the higher level elements of the universe corresponds to the convex hull of the intervals attached to their edges. For instance, the (numerical) interpretation of $[\textit{Few}, \textit{Most}]$, i.e., “From few to most,” is the interval $[a, 1 - a]$. The certainty ordering can be partially extended to the whole universe U as well, by defining for every $[X_1, Y_1], [X_2, Y_2] \in U$,

$$[X_1, Y_1] \leq [X_2, Y_2] \text{ if and only if, } X_1 \leq X_2 \text{ and } Y_1 \leq Y_2,$$

giving rise to the horizontal ordering, shown also in Figure 3, which differs from the specificity ordering, and where the certainty ordering now increases from left

to right. It is worth noticing that, if we add the empty set \emptyset to U , and if we equip it with both the specificity (\subseteq) and the certainty (\leq) orderings, together with a suitable negation operator N , the structure of the qualitative algebra we are dealing with, $\{U, \leq, \text{none}, \text{all}, \subseteq, \emptyset, [\text{none}, \text{all}], N\}$, turns out to be a bilattice as discussed in Ginsberg (1988). Namely, $(U, \leq, \text{none}, \text{all})$ is the certainty lattice where “none” corresponds to “false” and “all” corresponds to “true,” and \leq is the “truth” ordering of Ginsberg. $(U, \subseteq, \emptyset, [\text{none}, \text{all}])$ is the specificity lattice where \emptyset means “contradictory” and $[\text{none}, \text{all}]$ means “unknown”, and \subseteq is the (perhaps misleadingly called) “uncertainty ordering” of Ginsberg. The negation performs a reflection around the $[\text{none}, \text{all}]-\emptyset$ axis (that goes through “ab-half”), and is such that for $X, Y \in \mathcal{P}$, $N([X, Y]) = [\text{ANT}(Y), \text{ANT}(X)]$, i.e., it is the extension of the antonym operation that relates $P(A)$ and $P(\bar{A})$ for two complementary events A and \bar{A} .

There would be another kind of qualitative term pertaining to degrees of probability, namely relative orders of magnitude. It makes sense to declare that the probability of some event A is negligible with respect to the probability of another event B . This type of information could be represented as linguistic odds on the scale $[1, +\infty)$. Consider three landmarks, r_1, r_2, r_3 where $1 < r_1 < r_2 < r_3 < +\infty$, and interpret them as follows:

$$\begin{aligned} \frac{1}{r_1} < \frac{P(A)}{P(B)} < r_1 &: P(A) \text{ is close to } P(B), \\ r_1 < \frac{P(A)}{P(B)} < r_2 &: P(B) \text{ is small with respect to } P(A), \\ r_2 < \frac{P(A)}{P(B)} < r_3 &: P(B) \text{ is very small with respect to } P(A), \\ r_3 < \frac{P(A)}{P(B)} &: P(B) \text{ is negligible with respect to } P(A). \end{aligned}$$

Although this type of information could be viewed as redundant with respect to absolute orders of magnitude, it is not. First relative orders of magnitude can be more expressive than absolute ones for describing extreme probabilities, while absolute orders of magnitude better represent linguistic terms referring to probabilities between say, 0.1 and 0.9. A second reason is that combining absolute probabilities by quotienting will usually deteriorate the precision of the resulting probabilities while relative probabilities are directly expressed as quotients. However in this paper we stick to absolute probabilities.

4.2. Computation of the qualitative syllogism table

In this section we will focus on the qualitative counterpart of the quantified syllogism inference pattern, introduced in the preceding section. We use the

following notations, where Q_i are linguistic labels.

$Q1$ A's are B's; $Q2$ B's are A's

$Q3$ C's are B's; $Q4$ B's are C's

$Q5$ A's are C's; $Q6$ C's are A's

This inference rule is interesting from the point of view of commonsense reasoning since it offers a precise model of chaining uncertain "if... then..." statements expressed by means of imprecise quantifiers or conditional probabilities. In the following we build the qualitative quantified syllogism function QS corresponding to that pattern for a given partition \mathcal{P} , i.e., we build a table providing qualitative values for $Q5$ and $Q6$ for any combination of possible qualitative values for $Q1, Q2, Q3$ and $Q4$ (of course we have the restriction that $Q1 = \text{none} \Leftrightarrow Q2 = \text{none}$ and $Q3 = \text{none} \Leftrightarrow Q4 = \text{none}$). Thus, the QS function must be defined as a mapping from $U \times U \times U \times U$ to U , U being the universe of description associated with the partition \mathcal{P} . However, as will be shown, it will be necessary to compute the table for any combination of elementary quantifiers only, i.e., to build a mapping from $\mathcal{P} \times \mathcal{P} \times \mathcal{P} \times \mathcal{P}$ to U .

The process of building the QS function is performed according to the following steps:

1. Consider a linguistic scale of linguistic quantifiers together with a suitable partition of the unit interval $[0, 1]$ that represents them. In what follows we will use the partition \mathcal{P} defined above with parameters $a = 0.2$, and $b = 0.4$, that is,

$0 ::= \text{None}$
 $(0, 0.2] ::= \text{Almost none}$
 $[0.2, 0.4] ::= \text{Few}$
 $[0.4, 0.6] ::= \text{About half}$
 $[0.6, 0.8] ::= \text{Most}$
 $[0.8, 1) ::= \text{Almost all}$
 $1 ::= \text{All}$

2. Consider all possible combinations of these linguistic values for $P(B|A)$, $P(A|B)$, $P(B|C)$, and $P(C|B)$.

3. For each of such combinations, compute the lower and upper bounds of $P(C/A)$ (and $P(A/C)$) using the numerical expression of the pattern given in Section 2. For example if we have

$Q1 = \text{Most}$

$Q2 = \text{Almost-all}$

$Q3 = \text{About half}$

$Q4 = \text{Almost-all}$

then we have

$$\begin{aligned}
 P^*(B|A) &= 0.8, & P_*(B|A) &= 0.6, \\
 P^*(A|B) &= 1, & P_*(A|B) &= 0.8, \\
 P^*(B|C) &= 0.6, & P_*(B|C) &= 0.4, \\
 P^*(C|B) &= 1, & P_*(C|B) &= 0.8,
 \end{aligned}$$

which gives for $Q5$ and $Q6$

$$\begin{aligned}
 P^*(C|A) &= 1, & P_*(C|A) &= 0.45, \\
 P^*(A|C) &= 1, & P_*(A|C) &= 0.3.
 \end{aligned}$$

4. These results are then approximated by the most specific element of the universe of description U (see Figure 3) which contains them. So, the interval $[0.45, 1]$ for $P(C|A)$ is approximated by the larger interval $[0.4, 1]$, that is, the resulting $Q5$ is set to $[About-half, All]$. In the same way, $Q6$ is approximated to $[Few, All]$.

In this way, we have partially defined the qualitative functions $Q5$ and $Q6$, i.e., defined as functions $Q5$ and $Q6$: $\mathcal{P} \times \mathcal{P} \times \mathcal{P} \times \mathcal{P} \rightarrow U$. Table 1 shows a small part of the table thus generated for $Q5$ and $Q6$.

5. Finally, the complete definition of the qualitative function $QS : U \times U \times U \times U \rightarrow U$ can be easily derived from the above partial one using the obvious property of interval analysis $f(I_1 \cup I_2) = f(I_1) \cup f(I_2)$ where I_1 and I_2 are subintervals of $[0, 1]$. Applying this property implies (i) any quantifier Qi can be considered as the union of the set Si of all elementary quantifiers below Qi in the universe of description (see Figure 3), and (ii) for any 4-tuple of quantifiers, the function QS is applied over all possible 4-tuples combining the elements of the sets Si . That is,

$$Q5 = QS(Q1, Q2, Q3, Q4) = \bigcup_{\substack{Si \in \mathcal{P} \\ Si \subseteq Qi}} QS(S1, S2, S3, S4).$$

However, this computation can be further simplified due to the following fact. The numerical expression of the quantified syllogism (see Section 2) shows that the lower bound for $Q5$ only depends on the lower bounds of $Q1, Q2$, and $Q4$, whereas the upper bound not only depends on the lower bounds of $Q2$ and $Q3$, the upper bound of $Q4$, but also on both the upper and lower bounds of $Q1$. Therefore, we can obtain the qualitative bounds for $Q5$ from the table defined on the set of elementary quantifiers in the following way:

$$\begin{aligned}
 Q5 &= QS(Q1, Q2, Q3, Q4) \\
 &= [QS(Q1_*, Q2_*, X, Q4_*), \bigcup_{\substack{S \in \mathcal{P} \\ S \subseteq Q_i}} QS(S, Q2_*, Q3_*, Q4^*)]
 \end{aligned}$$

Table 1. Instances of the quantified syllogism for the 7-element partition.

Q1 As are Bs	Q2 Bs are As	Q3 Cs are Bs	Q4 Bs are Cs	Q5 As are Cs	Q6 Cs are As
Most	Al-all	All	Ab-half	[None, Ab-half]	[Ab-half, All]
Most	Al-all	All	Most	[Few, Most]	[Most, All]
Most	Al-all	All	Al-all	[Ab-half, Most]	[Most, All]
Most	Al-all	All	All	[Most, Most]	[Al-all, Al-all]
Most	All	None	All	[Most, All]	[None, None]
Most	All	Al-none	None	[None, All]	[None, All]
Al-all	All	Al-all	Most	[Ab-half, All]	[Al-all, All]
Al-all	All	Al-all	Al-all	[Most, All]	[Al-all, All]
Al-all	All	Al-all	All	[Al-all, All]	[Al-all, All]
Al-all	All	All	None	[None, None]	[None, All]
Al-all	All	All	Al-none	[None, Al-none]	[None, All]
Al-all	All	All	Few	[None, Few]	[All, All]
Al-all	All	All	Ab-half	[Few, Ab-half]	[All, All]
Al-all	All	All	Most	[Ab-half, Most]	[All, All]
All	All	Ab-half	Most	[Most, Most]	[Ab-half, Ab-half]
All	All	Ab-half	Al-all	[Al-all, All]	[Ab-half, Ab-half]
All	All	Ab-half	All	[All, All]	[Ab-half, Ab-half]
All	All	Most	None	[None, None]	[None, All]

where $Q_i = [Q_{i*}, Q_{i*}^*]$, Q_{i*} and Q_{i*}^* are elementary quantifiers of U , i.e., elements of \mathcal{P} , and X can be any element of \mathcal{P} . Let us see an example of how this expression is applied.

$$\begin{aligned}
 &QS([Most, All], All, [None, All], [Ab-half, Almost-all]) \\
 &= [QS(Most, All, None, Ab-half), QS(Most, All, None, Almost-all) \\
 &\quad \cup QS(Almost-all, All, None, Almost-all) \cup QS(All, All, None, Almost-all)] \\
 &= [[Few, Most], [Ab - Half, Almost-all]] \cup \\
 &\quad [Most, Almost-all] \cup [Almost-all, Almost-all]] \\
 &= [Few, Almost-all]
 \end{aligned}$$

Remark. Note that in the above procedure, the qualitative calculation table for the quantified syllogism is computed by using the approximation step only at the end of the computation. Another approach one may think of would be to

have precomputed tables for the product and quotient operations and to use them in the calculation of the bounds. However, this latter approach would not be satisfactory because it yields results that are too imprecise. For instance, if one needs to compute the product $Q1 * Q2 * Q3$, where $Q1, Q2, Q3 \in \mathcal{P}$ and denoting $\mathcal{A}([a, b]) = \bigcap_{Q \in \mathcal{U}} \{Q \mid [a, b] \subseteq Q\}$, the function that maps interval $[a, b]$ to its “name” in the qualitative scale (approximation function), then generally

$$\mathcal{A}(Q1 * Q2 * Q3) \subset \mathcal{A}(Q1 * \mathcal{A}(Q2 * Q3))$$

because $\mathcal{A}(Q2 * Q3)$ strictly contains $Q2 * Q3$ generally. Even associativity may be lacking, i.e., $\mathcal{A}(Q1 * \mathcal{A}(Q2 * Q3)) \neq \mathcal{A}(\mathcal{A}(Q1 * Q2) * Q3)$.

Let us analyze the results obtained on the most elementary type of qualitative scale of linguistic quantifiers, i.e., {none, few, about half, most, all} where *few* is of the form $[\epsilon, \alpha]$ for some positive, infinitesimal value ϵ , α is some number in $(0, 1/2)$, *about half* is interpreted as $[\alpha, 1 - \alpha]$ and *most* is $[1 - \alpha, 1 - \epsilon]$. Note that the name “about half” is indeed short for “neither few nor most, but in between,” since the interval $[\alpha, 1 - \alpha]$ may be quite large. Table 2 gives the complete results when $\alpha = 0.3$ on 4-tuples which give nontrivial results for $P(C|A)$; the table is sorted by putting together the 4-tuples $(Q1 \ Q2 \ Q3 \ Q4)$ that lead to the same value of $Q5$. A first remark is that in many situations when none of the quantifiers mean “all,” no information is obtained on $P(C|A)$. This is especially true when both $P(A|B)$ and $P(B|A)$ take small qualitative values. Some results may look surprising. For instance, we read that nothing can be inferred from the four statements

- “all As are Bs,”
- “most Bs are As,”
- “all Cs are Bs,”
- “about half of the Bs are Cs”.

Especially, the lower bound $P_*(C|A) = 0$ is attained if pessimistic interpretations of “most” and “about half” are chosen, say 70% and 30% respectively. This case should not be surprising since it is the one studied in Section 3.1.

4.3. The robustness of qualitative probabilistic reasoning

Table 2 is obtained for a specific value of the threshold α between “few” and “half.” A legitimate question is whether such results are still valid for other values of the threshold. Let us start with qualitative tables for product and quotient, with “few” = $(0, \alpha]$, “most” = $[1 - \alpha, 1)$ “half” = $[\alpha, 1 - \alpha]$. The product table is given in Table 3, and the identities None * Q = None, All * Q = Q. The “?” indicates some ambiguity due to the choice of the value of α . Namely half * half = $[\alpha^2, (1 - \alpha)^2] \subseteq (0, \alpha]$ only if $(1 - \alpha)^2 \leq \alpha$ which requires $\alpha \geq d$,

Table 2. Compacted table of the quantified syllogism for the five element partition (“half” means “about half”).

$P(B A)$	$P(A B)$	$P(B C)$	$P(C B)$	$P(C A)$
none	none	[none, most]	[none, all]	[none, all]
few	few	[few, most]	[few, most]	"
few	half	[few, half]	[few, most]	"
few	most	[few, half]	half	"
[half, all]	most	[few, most]	half	"
[half, all]	[few, half]	[few, most]	[few, most]	"
all	[few, half]	all	[few, most]	"
all	most	all	half	"
few	[few, all]	none	none	[none, most]
few	half	most	most	"
[few, half]	most	few	few	"
most	[few, half]	all	[few, most]	"
most	most	all	half	"
few	half	most	[few, half]	[none, half]
few	most	half	few	"
few	most	most	half	"
half	[few, half]	all	[few, most]	"
half	[few, all]	none	none	"
[most, all]	most	[few, all]	few	"
few	[few, half]	all	[few, most]	[none, few]
few	most	most	few	"
few	most	all	[few, half]	"
half	most	all	few	"
most	[few, all]	none	none	"
none	none	all	[few, all]	[none, none]
all	[few, all]	none	none	"
few	few	most	all	[few, all]
few	[few, all]	[few, half]	all	"
[few, half]	most	[few, half]	most	"
half	most	most	most	"
few	half	most	all	[few, most]
few	all	few	[few, most]	"
few	all	half	most	"
half	all	[few, half]	[half, most]	"

Continued.

$P(B A)$	$P(A B)$	$P(B C)$	$P(C B)$	$P(C A)$
half	all	few	few	"
half	all	most	most	"
most	all	[few, most]	half	"
few	[most, all]	most	[most, all]	[few, half]
few	all	half	half	"
half	most	all	most	"
half	all	half	few	"
half	all	all	[half, most]	"
most	all	[few, most]	few	"
most	all	all	half	"
few	[few, all]	all	all	[few, few]
few	most	all	most	"
few	all	half	few	"
few	all	most	[few, half]	"
few	all	all	[half, most]	"
half	all	[most, all]	few	"
most	all	all	few	"
all	all	[few, all]	few	"
half	[few, all]	[few, most]	all	[half, all]
[most, all]	most	[few, most]	most	"
all	most	all	most	"
most	most	all	most	[half, most]
most	all	[few, all]	most	"
half	[few, all]	all	all	[half, half]
all	all	[few, all]	half	"
most	[few, all]	[few, most]	all	[most, all]
most	[few, all]	all	all	[most, most]
all	all	[few, all]	most	"
all	[few, all]	[few, all]	all	[all, all]

Table 3. Product table.

Product	few	half	most
few	few	few	few
half	few	?	[few, half]
most	few	[few, half]	?

Table 4. Quotient table.

Quotient	none	few	half	most	all
none	[none, all]	none	none	none	none
few	all	[few, all]	[few, all]	[few, half]	few
half	all	all	[half, all]	[half, all]	half
most	all	all	all	[most, all]	most
all	all	all	all	all	all

where $d = (\alpha - \sqrt{5})/2 \approx 0.382$. In that case $\text{half} * \text{half} = \text{few}$ and $\text{most} * \text{most} = [(1 - \alpha)^2, 1] \not\subset [\alpha, 1]$ when $\alpha > d$ so that $\text{most} * \text{most} = [\text{few}, \text{most}]$. The latter equality does not sound natural. On the contrary if $\alpha < d$, then

$$\text{half} * \text{half} = [\text{few}, \text{half}]; \quad \text{most} * \text{most} = [\text{half}, \text{most}].$$

From a commonsense point of view, it is not counterintuitive to require that “few” may mean a proportion less than 0.3 or so. Again “half” is here short for “neither few nor most but in between.” Hence it is clear that the product of qualitative probabilities is almost independent of the choice of the threshold α in $(0, 1/2)$. It fits the intuition and is completely threshold independent for “ α ” small enough.

The same problem can be solved for the quotient, and it leads to the almost-robust table (Table 4). The term half/half and few/most are given for $\alpha < d$. Only these terms change if α is larger. Note that the subdiagonal part of the table has been truncated to “all.”

In order to study the robustness of the quantified syllogism table, several runs of the program that generate this table have been done, with α varying between 0.025 and 0.38. Only a few lines of the qualitative table change (nine over $625 = 5^4$ distinct 4-tuples of quantifiers for $0.025 \leq \alpha \leq 0.35$). In order to get a better insight, it is interesting to consider a significant subpart of the table, where quantifiers are either “few,” “most”, or “half,” i.e., when $P(A|B)$, $P(B|A)$, $P(C|B)$, $P(B|C)$ are close to 0, close to 1, or neither close to 0 nor close to 1 but in between. In order to let the parameter α appear, we

Table 5. Parametrized instances of the quantified syllogism.

$P(B A)$	$P(A B)$	$P(B C)$	$P(C B)$	$P(C A)$
$V_0(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_1(\alpha)$	$V_0(\alpha)$	$V_0(\alpha/(1 - \alpha))$
$V_0(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_1(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_0(2\alpha)$
$V_0(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_0((2\alpha - \alpha^2)/(1 - \alpha))$
$V_0(\alpha)$	$V_1(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_0(\alpha)$	$V_0(\alpha/(1 - \alpha))$
$V_0(\alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_0(\alpha)$	$V_0(\alpha^2/(1 - \alpha)^2)$
$V_0(\alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_0(\alpha/(1 - \alpha))$
$V_0(\alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_0((\alpha/(1 - \alpha))^2 + \alpha)$
$V_{1/2}(1/2 - \alpha)$	$V_1(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_0(\alpha)$	$V_0(1/(2(1 - \alpha)))$
$V_{1/2}(1/2 - \alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_0(\alpha)$	$V_0(\alpha/(1 - \alpha))$
$V_1(\alpha)$	$V_1(\alpha)$	$V_0(\alpha)$	$V_0(\alpha)$	$V_0(2\alpha)$
$V_1(\alpha)$	$V_1(\alpha)$	$V_0(\alpha)$	$V_1(\alpha)$	$V_1(2\alpha)$
$V_1(\alpha)$	$V_1(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_0(\alpha)$	$V_0(2\alpha)$
$V_1(\alpha)$	$V_1(\alpha)$	$V_{1/2}(1/2 - \alpha)$	$V_1(\alpha)$	$V_1(2\alpha)$
$V_1(\alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_0(\alpha)$	$V_0(\alpha/((1 - \alpha)^2 + \alpha^2))$
$V_1(\alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_1(\alpha)$	$V_1(2\alpha)$

shall use the following notation,

$P(A|B)V_0(\alpha)$ means $P(A|B) \leq \alpha$,

$P(A|B)V_{1/2}(\alpha)$ means $1/2 - \alpha \leq P(A|B) \leq 1/2 + \alpha$,

$P(A|B)V_1(\alpha)$ means $P(A|B) \geq 1 - \alpha$,

where $V_x(\alpha)$ means “in the vicinity of x ” for $x = 0, 1/2, 1$, and $\alpha < 0.5$ quantifies the error (i.e., the actual imprecision). Using this notation, the above quantifiers “few,” “half,” and “most” can be represented by “ $P(A|B)V_0(\alpha)$,” “ $P(A|B)V_{1/2}(1/2 - \alpha)$,” and “ $P(A|B)V_1(\alpha)$ ” respectively. Notice that representing “few” by “ $P(A|B)V_0(\alpha)$ ” and “most” by “ $P(A|B)V_1(\alpha)$ ” includes $P(A|B) = 0$ and $P(A|B) = 1$ as possible interpretations of “few” and “most” respectively. This representation is adopted to simplify the computations, although it differs slightly from the previous ones where $P(A|B) = 0$ and $P(A|B) = 1$ were only interpreted as “none” and “all” respectively. Then by applying the optimal bounds on $P(C|A)$ as described in Section 2 on the $64 = 4^3$ 4-tuples of quantifiers, potential instability of the results was obtained only for the 15 cases in Table 5. It is easy to verify that for $\alpha \leq 1/3$,

$$\frac{\alpha}{1 - \alpha} \leq 1 - \alpha, \quad \frac{\alpha^2}{1 - \alpha^2} \leq \alpha,$$

Table 6. Robust linguistic inferences.

$P(B/A)$	$P(A/B)$	$P(B/C)$	$P(C/B)$	$P(C/A)$
few	half	most	few	[few, half]
few	half	most	half	[few, half]
few	half	most	most	[few, half]
few	most	half	few	[few, half]
few	most	most	few	few
few	most	most	half	[few, half]
few	most	most	most	[few, half]
half	most	half	few	[few, half]
half	most	most	few	[few, half]
most	most	few	few	[few, half]
most	most	few	most	[half, most]
most	most	half	few	[few, half]
most	most	half	most	[half, most]
most	most	most	few	[few, half]
most	most	most	most	[half, most]

$$\alpha + \frac{\alpha^2}{1 - \alpha^2} \leq 1 - \alpha, \quad 2\alpha \leq 1 - \alpha,$$

$$\frac{\alpha}{(1 - \alpha)^2 + \alpha^2} \leq 1 - \alpha \quad \left(\text{since } \frac{\alpha}{(1 - \alpha)^2 + \alpha^2} \leq 2\alpha \right),$$

and for $\alpha \leq \gamma = 1 - \sqrt{2}/2$, i.e., approximately 0.292, there also holds

$$\frac{2\alpha - \alpha^2}{1 - \alpha} \leq 1 - \alpha, \quad \frac{1}{2 - 2\alpha} \leq 1 - \alpha.$$

These inequalities guarantee that whatever the value of $\alpha \leq \gamma$, the value of $P(C/A)$, as shown in Table 5, remains within a given range (e.g., $[0, \alpha]$, $[0, 1 - \alpha]$, $[\alpha, 1]$) corresponding to a symbolic label even if there is a degradation of the result which, except in the fifth line of Table 5, is less specific than $V_0(\alpha)$ or $V_1(\alpha)$. Hence, we get the robust computation in Table 6 for the quantified syllogism (we only give here the 4-tuples that lead to an informative output). The table obtained for $\alpha = 0.3$ remains correct if smaller values of α are used but may not be optimally precise, i.e., obtained intervals from the table are possibly wider than they should be.

Table 7. Product.

Product	none	al-n	few	half	most	al-all	all
none	none	none	none	none	none	none	none
al-n	none	al-n	al-n	al-n	al-n	al-n	al-n
few	none	al-n	al-n	[al-n few]	[al-n few]	[al-n few]	few
half	none	al-n	[al-n few]	[al-n-few]	[few half]	[few half]	half
most	none	al-n	[al-n few]	[few half]	[few most]	[half most]	most
al-all	none	al-n	[al-n few]	[few half]	[half most]	[most al-all]	al-all
all	none	al-n	few	half	most	al-all	all

5. The qualitative version of Bayes theorem

The computation of Bayes theorem, namely $P(A|B) = P(A \cap B)/P(B)$ using absolute qualitative values usually yields very imprecise results. Consider, for instance, the quotient table given in Section 4.3. Only when both $P(A \cap B)$ and $P(B)$ carry the value “most” does the imprecision remain small, for the five-quantifier case. The same phenomenon can be observed in Tables 7 and 8, which show the qualitative product and qualitative quotient corresponding to the partition \mathcal{P} used in Section 2. Applying several times the qualitative version of Bayes theorem is likely to produce uninformative results due to the already strong deterioration of the precision of the results at the first step.

On the other hand, the generalized Bayes’ theorem (GBT), described in Section 2,

$$\forall A_1, \dots, A_k, P(A_1|A_k) = P(A_k|A_1) \prod_{i=1, k-1} \frac{P(A_i|A_{i+1})}{P(A_{i+1}|A_i)}$$

can be used, given a cycle $A_1, A_2, \dots, A_n, A_{n+1} = A_1$ in a probabilistic network, to improve the upper and lower bounds of $P(A_1|A_n)$ from the upper and lower bounds of the remaining $2n - 1$ conditional probabilities. Now, given a linguistic scale of numerical quantifiers determined by a partition of $[0, 1]$, the problem is to obtain the qualitative counterpart of this local inference rule. In this case, we cannot use the same method as we did in Section 4.1 with the quantified syllogism rule because here the number of arguments, i.e., the length of the involved cycle, is variable. This prevents us from having the qualitative inference pattern defined by only one table. Therefore the following two approaches have been investigated.

Table 8. Quotient.

Quotient	none	al-n	few	half	most	al-all	all
none	[none all]	none	none	none	none	none	none
al-n	all	[al-n all]	[al-n all]	[al-n half]	[al-n few]	[al-n few]	al-n
few	all	all	[half all]	[few all]	[few most]	[few half]	few
half	all	all	all	[most all]	[half all]	[half most]	half
most	all	all	all	all	[most all]	[half most]	most
al-all	all	all	all	all	all	[al-all all]	al-all
all	all	all	all	all	all	all	all

5.1. Using generalized Bayes theorem with qualitative operations

The first and more straightforward approach consists of replacing, in the above expression, the product and quotient operations by the corresponding qualitative ones, defined on a given universe of descriptions U , such as the ones defined in the above Tables 7 and 8. Therefore, given a cycle (A_1, \dots, A_k, A_1) with $A_1 = A, A_k = B$, the qualitative probability $QP(A|B)$, known to lie in the interval $[QP_*(A|B)_{old}, QP^*(A|B)_{old}]$, should be improved by letting

$$QP_*(A|B)_{new} = QP_*(A|B)_{old} \vee \left[QP_*(B|A) \cdot \frac{\prod_{i=1}^{k-2} QP_*(A_i|A_{i+1})}{\prod_{i=1}^{k-2} QP^*(A_{i+1}|A_i)} \right],$$

$$QP^*(A|B)_{new} = QP^*(A|B)_{old} \wedge \left[QP^*(B|A) \cdot \frac{\prod_{i=1}^{k-2} QP^*(A_i|A_{i+1})}{\prod_{i=1}^{k-2} QP_*(A_{i+1}|A_i)} \right],$$

where \wedge and \vee denote the min and max operations in the sense of the certainty ordering. But the computation of these quantities raises several problems.

(i) For a given cycle, find a proper ordering for the computation.

Especially, it is not always the case that

$$\frac{X_1 \cdot X_2}{X_3 \cdot X_4} \quad (\text{computing products first})$$

is equal to

$$\left(\frac{X_1}{X_3} \right) \cdot \left(\frac{X_2}{X_4} \right) \quad (\text{computing quotients first}).$$

Furthermore, because of the truncation effect of the quotient table, this second mode of computation may lead to incorrect results. Therefore, products should be computed first and the computation of quotients should be delayed as much as possible.

(ii) Since this operation must be done for all cycles one might look for the counterpart of a longest path algorithm, here with qualitative values. But this is tricky if we want to compute quotients only at the end of the shortest path procedure and keep the products of terms along cycles separate. In that case, the maximum operation

$$\frac{Q_1}{Q_2} \vee \frac{S_1}{S_2}$$

should be directly expressed as an operation \vee' between pairs (Q_1, Q_2) and (S_1, S_2) , furnishing a new pair of qualitative values. Moreover, longest-path algorithms make an extensive use of the distributivity of the product over the maximum. Here we would require a property such as

$$\frac{X_1}{X_2} \cdot \left(\frac{X_3}{X_5} \vee \frac{X_4}{X_6} \right) = \frac{X_1 \cdot X_3}{X_2 \cdot X_5} \vee \frac{X_1 \cdot X_4}{X_2 \cdot X_6}.$$

Clearly, this property does not always hold in our qualitative algebra.

But the basic question is whether this constraint propagation rule, which proved useful in the quantitative case leads to really improved qualitative probability bounds. This can be precisely studied on the five-quantifier case {none, few, half, most, all}.

The smallest expression to be computed with nonextreme probabilities is of the form

$$\frac{Q_1 \cdot Q_2 \cdot Q_3}{Q_4 \cdot Q_5},$$

with $Q_i \in \{\text{few, half, most}\}$. It is easy to check from the product and quotient tables (see Section 4.3) that

- (i) $Q_1 \cdot Q_2 \in \{\text{few, [few, most], [few, half], [half, most]}\}$,
- (ii) $Q_1 \cdot Q_2 \cdot Q_3$ can only belong to the same set as above,
- (iii) the only cases where the quotients can be significantly informative are

$$\frac{\text{few}}{\text{most}} = [\text{few, half}], \quad \frac{\text{half}}{\text{half}} = [\text{half, all}], \quad \frac{\text{half}}{\text{most}} = [\text{half, all}].$$

As a consequence $(Q_1 \cdot Q_2 \cdot Q_3)/(Q_4 \cdot Q_5)$ can give [few, all] at the very best. This is when $Q_1 \cdot Q_2 \cdot Q_3 = [\text{few, half}]$ and $Q_4 \cdot Q_5 = [\text{half, most}]$. This is not likely to be very useful for improving probability bounds. For example, in the 7-quantifier case, the best informative result can be shown to be [half, all] corresponding to the case when

$$\frac{Q_1 \cdot Q_2 \cdot Q_3}{Q_4 \cdot Q_5} = \frac{[\text{half, most}]}{[\text{half, most}]}, \quad \text{or} \quad \frac{[\text{half, most}]}{[\text{most, al-all}]}, \quad \text{or} \quad \frac{[\text{most, al-all}]}{[\text{most, al-all}]},$$

as shown in the product and quotient tables given for this case.

5.2. Using generalized Bayes theorem on fixed-length cycles

Due to the difficulties encountered with the first approach, we have studied a second approach to obtain a qualitative counterpart of GBT. As mentioned, it is not possible to associate a qualitative function to the GBT because cycles of variable length are involved. However, we can build, for different fixed lengths, the qualitative function corresponding to the generalized Bayes theorem for cycles with these fixed lengths. Therefore, we can think of approximating the “exact” qualitative GBT by using one of such partial computations with cycles of fixed length. The greater the number of considered cycles, the better the improvement (in the sense of precision) we may expect. We have then a classical trade-off between accuracy and computational cost. In doing that we found out that most of the time there was no substantial improvement of the qualitative bounds due to the increase of the length. So it is not worth working with long cycles.

For example, for cycles of length n ($n > 2$), the expression to be computed is of the form

$$\frac{Q_1 \cdot Q_2 \cdots Q_n}{Q_{n+1} \cdots Q_{2n-1}},$$

which involves $2n - 1$ operands. Again, it is not necessary to define such functions in the whole universe of descriptions U , but only on the set of elementary quantifiers \mathcal{P} . In particular, for $n = 3$, we would have

$$\frac{Q_1 \cdot Q_2 \cdot Q_3}{Q_4 \cdot Q_5} = \left[\frac{Q_{1*} \cdot Q_{2*} \cdot Q_{3*}}{Q_{4*} \cdot Q_{5*}}, \frac{Q_1^* \cdot Q_2^* \cdot Q_3^*}{Q_{4*} \cdot Q_{5*}} \right]$$

where $Q_i = [Q_{i*}, Q_i^*]$ and $Q_{i*}, Q_i^* \in \mathcal{P}$. Using this approach, much better results can be obtained with respect to the previous section. In particular, as we have seen in the five-quantifier case, the above expression gives [few, all] at the very best with nonextreme probabilities using qualitative products and quotients, whereas now more precise results, like [few, few], [few, half], [few, most], or even [half, all], can be obtained. For instance (for $\alpha \leq 0.3$), Table 9 lists some of these.

6. Symbolic constraint propagation

The quantified syllogism rule followed by the generalized Bayes rule with fixed-length cycles, as recomputed for a given linguistic partition of the unit interval, can be recursively applied to any set of linguistically quantified statements of the form $Q A_i$ s are B_i s which form a probabilistic network. It is then possible to generate new statements of that kind and to improve precision for the ones that were originally stated. Let us consider the following qualitative counterpart of a 5-predicate example of Amarger et al. (1991):

Table 9. Symbolic calculations of the generalized Bayes theorem.

Q_1	Q_2	Q_3	Q_4	Q_5	$Q_1 \cdot Q_2 \cdot Q_3 / Q_4 \cdot Q_5$
few	few	few	half	half	[few, few]
half	few	most	most	most	[few, half]
half	most	few	most	half	[few, most]
most	most	few	half	half	[few, all]
most	most	most	most	most	[half, all]

- Most to almost all students are sportsmen ($Q = [\text{most}, \text{al-all}]$).
- Almost all students are young ($Q = \text{al-all}$).
- Half of the sportsmen are students ($Q = \text{half}$).
- Almost all sportsmen are single ($Q = \text{al-all}$).
- At least almost all sportsmen are young ($Q = [\text{al-all}, \text{all}]$).
- Most to almost all singles are sportsmen ($Q = [\text{most}, \text{al-all}]$).
- Most singles are young ($Q = \text{most}$).
- Almost no singles have children ($Q = \text{al-none}$).
- Few young people are students ($Q = \text{few}$).
- Almost all young people are sportsmen ($Q = \text{al-all}$).
- At least almost all young people are single ($Q = [\text{al-all}, \text{all}]$).
- At most almost no young people have children ($Q = [\text{none}, \text{al-none}]$).
- At most almost no people who have children are single ($Q = [\text{none}, \text{al-none}]$).
- At most almost no people who have children are young ($Q = [\text{none}, \text{al-none}]$).

These statements are just examples and must not be examined as reflecting sociological reality. We have been experimenting with this seven-quantifier examples as well as with its adaptation to accommodate nine-quantifiers. We have also experimented with different ways of partitioning, i.e., different intervals to represent these quantifiers, and with cycles of different length (3 and 4) when applying the generalized Bayes rule.

The process of symbolic constraint propagation is the following one. The quantified syllogism rule is run until no improvement of the quantifiers (that is narrowing of the intervals $[Q_i, Q_j]$) nor new statements can be generated. Then the generalized Bayes rule is applied over all cycles of a particular length (3 or 4). Only in the case that some improvement of the quantifiers is obtained using this rule, the process starts again.

What follows is a representative selection of results, showing some interesting aspects. Let us consider a seven-element partition as follows:

Table 10. Input data.

Input data	student	sport	single	young	children
student	[1.000, 1.000]	[0.700, 0.900]	[0.000, 1.000]	[0.850, 0.950]	[0.000, 1.000]
sport	[0.400, 0.600]	[1.000, 1.000]	[0.800, 0.850]	[0.900, 1.000]	[0.000, 1.000]
single	[0.000, 1.000]	[0.700, 0.900]	[1.000, 1.000]	[0.600, 0.800]	[0.050, 0.100]
young	[0.250, 0.350]	[0.800, 0.900]	[0.900, 1.000]	[1.000, 1.000]	[0.000, 0.050]
children	[0.000, 1.000]	[0.000, 1.000]	[0.000, 0.050]	[0.000, 0.050]	[1.000, 1.000]

partition : (0, 0.2, 0.4, 0.6, 0.8, 1),

$P = \{ \text{none, al-none, few, half, most, al-all, all} \}.$

The following results were obtained:

- At least few students are single ($Q = [\text{few, all}]$).
- Not more than few sportsmen have children ($Q = [\text{none, few}]$).
- From almost-none to half singles are students ($Q = [\text{al-none, half}]$).

It is worth noticing that the generalized Bayes rule gave no improvement on cycles with length up to 4. Let us consider now a nine-element partition as follows:

partition : (0, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 1),

$P = \{ \text{none, al-none, v-few, few, half, most, v-many, al-all, all} \},$

where v-few stands for “very few” ([0,1, 0,2]) and v-many stands for “very many” ([0.8, 0.9]). Using the same input data, more results are obtained:

- At least half of the students are single ($Q = [\text{half, all}]$).
- Not more than half of the students have children ($Q = [\text{none, half}]$).
- Not more than very few sportsmen have children ($Q = [\text{none, few}]$).
- From very few to half singles are students ($Q = [\text{v-few, half}]$).
- Most to very many singles are sportsmen ($Q = [\text{most, v-many}]$).

The last statement is a conclusion obtained by the application of the generalized Bayes rule. These results are consistent but stronger than those obtained with the seven-element partition. It is interesting to compare these results with those of the numerical procedure that directly handles interval probabilities given in Table 10 and 11 under the form of incidence matrices.

The main difference between the numerical and symbolic results appears in the last row. The symbolic inference device was not able to deduce that

Table 11. Probabilities after constraint propagation.

Saturated network	student	sport	single	young	children
student	[1.000, 1.000]	[0.900, 0.900]	[0.607, 1.000]	[0.850, 0.850]	[0.000, 0.271]
sport	[0.400, 0.400]	[1.000, 1.000]	[0.850, 0.850]	[0.900, 0.958]	[0.000, 0.154]
single	[0.222, 0.366]	[0.700, 0.700]	[1.000, 1.000]	[0.800, 0.800]	[0.050, 0.100]
young	[0.350, 0.350]	[0.834, 0.888]	[0.900, 0.900]	[1.000, 1.000]	[0.000, 0.050]
children	[0.000, 0.099]	[0.000, 0.127]	[0.000, 0.050]	[0.000, 0.044]	[1.000, 1.000]

almost no person having children is a student, and very few are sportsmen. However, better results with respect to this particular problem were obtained when choosing partitions where the upper bound for the quantifier “almost none” has been significantly reduced. For instance, if we consider the seven-element partition

$$\text{partition : } (0, 0.05, 0.4, 0.6, 0.95, 1),$$

$$P = \{\text{none, al-none, few, half, most, al-all, all}\},$$

the statement

- Not more than most people with children are sportmen ($Q = [\text{none, most}]$)

can be now inferred. On the contrary, nothing can be now inferred about the proportion of students who are single, which was inferred to be [few, all] with the first seven-element partition considered. Much better results are obtained when introducing two additional intermediate quantifiers in the last partition, i.e., when considering a nine-quantifier partition as

$$\text{partition : } (0, 0.05, 0.2, 0.4, 0.6, 0.8, 0.95, 1),$$

$$P = \{\text{none, al-none, v-few, few, half, most, v-many, al-all, all}\}.$$

With respect to the previous nine-quantifier partition, although we lose a little bit of information about the proportion of students who are single (from [half, all] to [few, all]) and that no improvement on the proportion of singles who are sportmen is obtained, a significant improvement is obtained with the inferred statements:

- Not more than few students have children ($Q = [\text{none, few}]$).
- Not more than few children are sportmen ($Q = [\text{none, few}]$).

Moreover, Bayes rule applied on cycles of length 4 allows us to further infer

- Not more than very few children are student ($Q = [\text{none, v-few}]$).

All these results show that, as one could expect, the finer the partition is, the better (more precise) the results are. But it is also true that, as the number of elements of the partition increase, results become more sensitive with respect to the values defining the intervals of the partition, i.e., the representation of the meaning of the quantifiers. As a matter of fact, recall that in Section 4.3 the quantified syllogism table was proved to be “almost” robust for the five-element partitions. In contrast, coming back to the above example, if, in the last partition, we move the upper bound for “almost none” from 0.05 to 0.01, the inferential power is drastically reduced with respect to the one provided by the first seven-element partition presented in this section due to a loss of sensitivity in the neighborhood of 0. Of course there would remain the problem of choosing the “best partition” with a given number of terms. But it is not clear that we have the freedom of choosing the partition (the meanings of the terms depend on the person who uses them).

7. Propagating qualitative probabilities in Bayesian networks: the smoke-alarm problem

The approach to reason qualitatively with imprecise probabilities developed so far in this paper can also be applied to propagate qualitative probability values in Bayesian networks. As pointed out by Xiang et al. (1990), the two inference rules needed in such networks are the Bayes rule and the reasoning-by-case rule. This last rule, expressed as

$$P(A|C) = P(A|B, C) \cdot P(B|C) + P(A|\neg B, C) \cdot P(\neg B|C).$$

allows the computation of a conditional probability by partitioning the condition C into the exclusive cases of B true and B false.

Using the same approach as in Section 5.2, these numerical rules easily admit a qualitative counterpart. The case of Bayes rule reduces to compute qualitatively a three-argument expression of the type

$$B(Q1, Q2, Q3) = Q1 \cdot Q2/Q3$$

that is, a special case of the qualitative version of GBT with only three operands. For the reasoning-by-case rule, it should be noticed that only three independent arguments are involved because $P(\neg B|C) = 1 - P(B|C)$ and thus it leads to the computation of the following qualitative expression:

$$G(Q1, Q2, Q3) = (Q1 - Q2) \cdot Q3 + Q2 = Q3 \cdot Q1 + (1 - Q3) \cdot Q2$$

Note again (see remark in Section 4.2) that in doing so, the qualitative computation of such expressions involves an approximation only at the end, and so

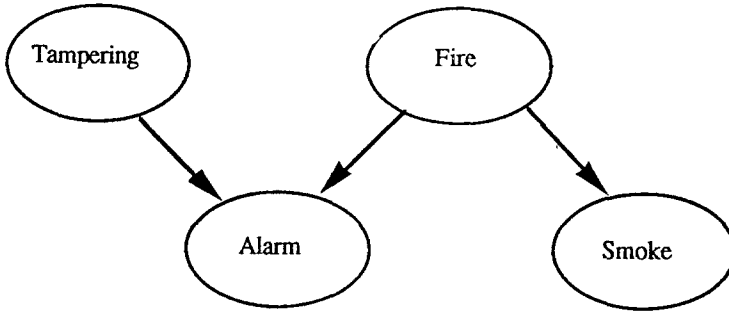


Figure 4. Smoke-alarm problem.

we avoid introducing imprecision as much as possible. It is also the case that, having defined such functions on the set of elements of a partition \mathcal{P} , we can safely extend them to the whole universe of description U by letting

$$\begin{aligned}
 B(Q_1, Q_2, Q_3) &= [B(Q_{1*}, Q_{2*}, Q_{3*}), B(Q_{1*}, Q_{2*}, Q_{3*})] \\
 G(Q_1, Q_2, Q_3) &= [G(Q_{1*}, Q_{2*}, Q_{3*}) \wedge G(Q_{1*}, Q_{2*}, Q_{3*}), \\
 &\quad G(Q_{1*}, Q_{2*}, Q_{3*}) \vee G(Q_{1*}, Q_{2*}, Q_{3*})]
 \end{aligned}$$

where $Q_i = [Q_{i*}, Q_i^*]$, with $Q_{i*}, Q_i^* \in \mathcal{P}$.

As an example, consider the well-known smoke-alarm problem (Figure 4) discussed also in Xiang et al. (1990), with the following numerical data (supposedly based on empirical experience):

$$\begin{aligned}
 P(\text{fire}) &= 0.01, & P(\text{alarm} \mid \neg\text{fire}, \text{tampering}) &= 0.85, \\
 P(\text{tampering}) &= 0.02, & P(\text{alarm} \mid \neg\text{fire}, \neg\text{tampering}) &= 0.0001, \\
 P(\text{alarm} \mid \text{fire}, \text{tampering}) &= 0.5, & P(\text{smoke} \mid \text{fire}) &= 0.9, \\
 P(\text{alarm} \mid \text{fire}, \neg\text{tampering}) &= 0.99, & P(\text{smoke} \mid \neg\text{fire}) &= 0.01.
 \end{aligned}$$

We take the following scale of qualitative probabilities:

- impossible = 0
- extremely unlikely = (0, 0.005]
- very unlikely = [0.005, 0.2]
- unlikely = [0.2, 0.4]
- maybe = [0.4, 0.6]
- likely = [0.6, 0.8]
- very likely = [0.8, 0.995]
- extremely likely = [0.995, 1)
- sure = 1

Given such a partition, tables corresponding to the qualitative Bayes and reasoning-by-case rules respectively are generated, and then the qualitative constraint propagation can be performed. For instance, $P(\text{alarm} | \text{fire})$ (whose computed value is 0.98) can be qualitatively evaluated by using the reasoning-by-case table, according to the following expression:

$$\begin{aligned}
 P(\text{alarm}|\text{fire}) &= \\
 G(P(\text{alarm}|\text{fire}, \text{tampering}), P(\text{alarm}|\text{fire}, \neg\text{tampering}), P(\text{tampering})) &= \\
 G(\text{maybe}, \text{very_likely}, \text{very_unlikely}) &= [\text{likely}, \text{very_likely}]
 \end{aligned}$$

which is an acceptable result, i.e., it is quite informative and compatible with the exact value 0.98. In general, in those cases where the only inference rule involved is reasoning-by-case, i.e., when we estimate probabilities of effects given causes, the results are acceptable. However, when we need to estimate probabilities of causes given effects, i.e., when Bayes rule is involved, results are much more imprecise, especially when dealing with small probability values. For instance, the qualitative value of $P(\text{fire}|\text{smoke})$, whose exact computed value is 0.47, is computed as follows:

$$\begin{aligned}
 P(\text{smoke}) &= G(P(\text{smoke} | \text{fire}), P(\text{smoke} | \neg\text{fire}), P(\text{fire})) \\
 &= G(\text{very_likely}, \text{very_unlikely}, \text{very_unlikely}) \\
 &= [\text{very_unlikely}, \text{unlikely}], \\
 P(\text{fire} | \text{smoke}) &= B(P(\text{smoke} | \text{fire}), P(\text{fire}), P(\text{smoke})) \\
 &= B(\text{very_likely}, \text{very_unlikely}, [\text{very_unlikely}, \text{unlikely}]) \\
 &= [\text{very_unlikely}, \text{sure}].
 \end{aligned}$$

Clearly, the result is almost uninformative. This is mainly due to the fact that Bayes rule is very sensitive to the relative orders of magnitude of the probabilities involved but not so sensitive to their absolute values. So we think that this type of situation should be approached, as mentioned in the last part of Section 4.1, by adding qualitative information concerning the relative orders of magnitude for the variables involved in the network. This issue will be the subject of future studies.

8. Concluding remarks

We have shown that a qualitative calculus for the probabilistic scale “none,” “few,” “from few to most,” “most,” “all” can be developed in agreement with a numerical interpretation of probabilities, provided that the intended numerical meaning of “few” is less than 30 percent in any case and the one of “most” is more than 70 percent. These thresholds are quite in agreement with commonsense reasoning, which seems to disagree that “most A ’s are B ’s” if less than 70 percent of A ’s are B ’s, or that “few A ’s and B ’s” when there are more than 30 percent

of A 's which are B 's. However, it does not mean that humans are currently able to provide the correct (in the sense of probability calculus) qualitative values given by the rules derived in this paper. It is well known (e.g., Kahneman et al. (1980)), that humans find it difficult not only to correctly assess probabilities, but also to make accurate inferences from them.

One might wonder whether fuzzy intervals are useful or not in the modeling of linguistic quantifiers. Clearly the use of precise thresholds to delimit the extensions of "few," "half," "most," is somewhat arbitrary. However, since the linguistic computation tables obtained here are partially independent of the choice of the threshold, it turns out that using fuzzy partitions instead of nonfuzzy ones would not make much difference here, especially if a fuzzy partition is viewed as an imprecise specification of the thresholds between the meanings of the basic terms. Nevertheless, fuzzy intervals remain useful in the scope of specifying numbers in probabilistic networks, with respect to understanding linguistic values, rather than reasoning *with* linguistic values. Indeed, when looking for the numerical interpretation of linguistic quantifiers, fuzzy intervals look like a more faithful model than crisp ones. But then the constraint propagation algorithms must be adapted to handle fuzzy upper and lower probabilities in the numerical setting. Applying fuzzy arithmetic to the quantified syllogism rule (as done by Dubois and Prade (1988)) appears to be in total contrast with defining linguistic counterparts of numerical constraint propagation rules, as done here. In the above-mentioned paper no qualitative computation is performed: starting from linguistic data, a numerical translation of these data is done by means of fuzzy numbers, and a quantitative computation on fuzzy numbers is performed. Here we tried to remain at a symbolic level for the computations themselves, in an attempt to avoid the translation step, but trying anyway to remain in accordance with the numerical counterparts of the terms if we had them.

The present work, despite its limitations, can be useful to deal with information systems containing partial statistical knowledge like ill-known contingency tables, or binary relations between attributes, weighted by conditional probabilities. The numerical version of the proposed inference rules may help addressing queries that relate two attributes whose relationships are known only via other attributes in the data base. The linguistic version of the inference rules gives a precise idea about what is lost when using symbolic frequency information (like "most" and "few") instead of the numeric frequency data. Specifically we have found that Bayes rule does not get along well with absolute orders of magnitude. Yet, it may be considered as rather surprising that, on examples such as in Section 6, linguistic labels can be propagated in accordance with the laws of probability, regardless of their precise numeric counterpart, and give results that, although suboptimal, remain somewhat informative.

Acknowledgments

This work has been partially supported by the ESPRIT-Basic Research Action N°. 3085 “DRUMS” (Defeasible Reasoning and Uncertainty Management Systems).

Appendix

Let A, B, C be three events corresponding to subsets of U . Suppose we start with the knowledge of $P(A|B), P(B|A), P(C|B), P(B|C)$, and we want to compute the resulting values of $P(C|A)$ and $P(A|C)$. In the following we shall assume that $P(A) \cdot P(B) \cdot P(C) > 0$. The case when one of $P(A), P(B), P(C)$ is zero is not considered because in that case some of the conditional probabilities are undefined. To see that this assumption is self-consistent note that the only constraint induced by $P(B|A)$ and $P(A|B)$ on $P(A)$ and $P(B)$ is of the form $\beta P(A) = \alpha P(B)$ with $\beta = P(B|A), \alpha = P(A|B)$. Hence if $P(A) \in (0, 1], P(B) \in (0, \beta/\alpha]$ provided that $0 < \beta < \alpha$. When $\beta = \alpha = 0$ then $P(A \cap B) = 0$ and $P(A), P(B)$ can be given any value in $[0, 1]$. Hence $P(A) > 0$, and $P(B) > 0$ can always be assumed. This assumption also implies $P(A \cap B) = 0 \Leftrightarrow P(A|B) = P(B|A) = 0$. Lastly, $P(C)$ is not constrained by $P(A|B)$ and $P(B|A)$. Note that assuming $P(A) > 0, P(B) > 0, P(A|B) > 0$, and $P(B|A) = 0$ leads to an inconsistency with respect to Bayes theorem.

THEOREM 1. Assume $P(A) \cdot P(B) \cdot P(C) > 0$, and $P(B|A), P(A|B), P(B|C), P(C|B)$ are fixed; then the following inequality holds:

$$\begin{aligned}
 P(C|A) &\geq P(B|A) \cdot \max\left(0, 1 - \frac{1 - P(C|B)}{P(A|B)}\right) && \text{if } P(B|A) > 0, P(A|B) > 0, \\
 &\geq 0 && \text{if } P(B|A) = P(A|B) = 0, \\
 P(C|A) &\leq \min\left(1, 1 - P(B|A) + \frac{P(B|A)P(C|B)}{P(A|B)}, \frac{P(B|A)P(C|B)}{P(A|B)P(B|C)}, \right. \\
 &\quad \left. \frac{P(B|A)P(C|B)}{P(A|B)P(B|C)}(1 - P(B|C)) + P(B|A)\right) && \text{if } P(B|A) > 0, P(A|B) > 0, \\
 & && P(C|B) > 0, P(B|C) > 0, \\
 &\leq 1 - P(B|A) && \text{if } P(B|C) = P(C|B) = 0, \\
 &\leq 1 && \text{otherwise.}
 \end{aligned}$$

Proof.

(i) Lower bound. Let us start with the following inequality which holds in probability theory:

$$\max(0, P(A \cap B) + P(B \cap C) - P(B)) \leq P(A \cap C). \tag{A1}$$

This inequality is obtained using the additivity property of probability as

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C). \quad (\text{A2})$$

Noticing that $P(A \cup B \cup C) + P(A \cap C) - P(A) - P(C) \geq 0$, we can claim that $P(B) - P(A \cap B) - P(B \cap C) + P(A \cap B \cap C) \geq 0$, and (A1) is obtained by turning $P(A \cap B \cap C)$ into $P(A \cap C)$ in the preceding inequality. Now assuming $P(A) > 0$, (A1) leads to

$$\max \left(0, P(B|A) + \frac{P(B \cap C)}{P(A)} - \frac{P(B)}{P(A)} \right) \leq P(C|A).$$

Now

$$\begin{aligned} \frac{P(B \cap C)}{P(A)} &= \frac{P(B \cap C)}{P(B)} \cdot \frac{P(B)}{P(A \cap B)} \cdot \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(C|B) \cdot P(B|A)}{P(A|B)} \end{aligned} \quad \text{provided that } P(B) > 0, P(A \cap B) > 0$$

and

$$\frac{P(B)}{P(A)} = \frac{P(B) \cdot P(A \cap B)}{P(A) \cdot P(A \cap B)} = \frac{P(B|A)}{P(A|B)} \quad \text{if } P(A \cap B) > 0.$$

Hence we get the lower bound of $P(C|A)$ in the general case. The only case when $P(A \cap B) = 0$ can occur when $P(B|A) = P(A|B) = 0$. In that case (A1) becomes the trivial inequality $P(A \cap C) \geq 0$, hence, $P(C|A) \geq 0$.

(ii) Upper bound. Consider again (A2) and note that $P(A \cup B \cup C) - P(B) - P(C) + P(B \cap C) \geq 0$. Hence $P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) \geq 0$, and we arrive at

$$P(A \cap C) \leq P(A) - P(A \cap B) + P(B \cap C). \quad (\text{A3})$$

Moreover, the following bound also exists

$$P(A \cap C) \leq P(A) \quad (\text{A4})$$

and it is not redundant with respect to (A3). By symmetry between A and C , we also have the two other nontrivial, independent bounds:

$$P(A \cap C) \leq P(C) - P(B \cap C) + P(A \cap B) \quad (\text{A5})$$

$$P(A \cap C) \leq P(C). \quad (\text{A6})$$

Again, since it is always possible to assume $P(A) > 0$, (A3)–(A6) lead to

$$P(C|A) \leq 1 - P(B|A) + \frac{P(B \cap C)}{P(A)},$$

$$\begin{aligned}
 P(C|A) &\leq 1, \quad \text{which is trivial,} \\
 P(C|A) &\leq \frac{P(C)}{P(A)} - \frac{P(B \cap C)}{P(A)} + P(B|A), \\
 P(C|A) &\leq \frac{P(C)}{P(A)}.
 \end{aligned}$$

Now we use the following identities:

$$\begin{aligned}
 \frac{P(B \cap C)}{P(A)} &= \frac{P(B \cap C)}{P(A)} \cdot \frac{P(A \cap B)}{P(A \cap B)} \cdot \frac{P(B)}{P(B)} \\
 &\qquad \qquad \qquad \text{provided that } P(A \cap B) > 0, \\
 &= \frac{P(C|B) \cdot P(B|A)}{P(A|B)} \\
 &\qquad \qquad \qquad \text{provided that } P(B|A) > 0 \text{ (or equivalently } P(A|B) > 0), \\
 \frac{P(C)}{P(A)} &= \frac{P(C) \cdot P(B) \cdot P(A \cap B) \cdot P(B \cap C)}{P(A) \cdot P(B) \cdot P(A \cap B) \cdot P(B \cap C)} \\
 &= \frac{P(B|A)P(C|B)}{P(A|B)P(B|C)} \\
 &\qquad \qquad \qquad \text{if } P(B \cap C) > 0, P(A \cap B) > 0, \\
 &\qquad \qquad \qquad \text{i.e. if } P(A|B) > 0, P(B|C) > 0, \\
 \frac{P(B)}{P(A)} &= \frac{P(B|A)}{P(A|B)} \quad \text{if } P(A \cap B) > 0, \text{ i.e., } P(A|B) > 0.
 \end{aligned}$$

Hence we get the upper bound in the general case.

If $P(B|A) = P(A|B) = 0$, then (A3) becomes trivial, (A5) becomes

$$P(A \cap C) \leq P(C) - P(B \cap C)$$

and subsumes (A6) and then

$$\frac{P(A \cap C)}{P(A)} \leq \frac{P(C)}{P(A)} - \frac{P(B \cap C)}{P(A)}.$$

But it is possible to choose $P(A) = P(C) - P(B \cap C)$ since $P(A)$ can be freely chosen. Hence $P(C|A)$ is only upper-bounded by 1.

If $P(B \cap C) = 0$, i.e., $P(B|C) = P(C|B) = 0$, then $P(C)$ can be freely chosen. Taking $P(C) = 1$ leads to trivial inequalities in (A5) and (A6). The only nontrivial inequality is (A3) that yields $P(C|A) \leq 1 - P(B|A)$.

THEOREM 2. The bounds given in Theorem 1 are the best ones (i.e., are optimal).

Proof.

(i) Lower bound. Let x_1, x_2, \dots, x_7 be nonnegative quantities proportional to the probabilities of the elements of the partition induced by A, B, C as in Figure A1.

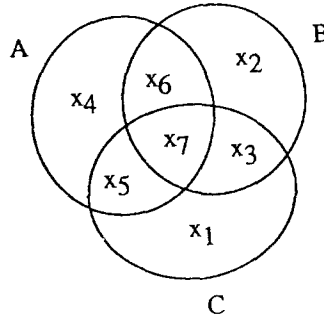


Figure A1.

The knowledge of $P(B|A)$ is equivalent to the linear constraint

$$P(A \cap \bar{B}) \cdot P(B|A) = P(A \cap B) \cdot (1 - P(B|A)),$$

which is written

$$(x_4 + x_5) \cdot P(B|A) = (x_6 + x_7) \cdot (1 - P(B|A)). \tag{A7}$$

And similarly we have

$$(x_2 + x_3)P(A|B) = (x_6 + x_7)(1 - P(A|B)), \tag{A8}$$

$$(x_1 + x_5)P(B|C) = (x_3 + x_7)(1 - P(B|C)), \tag{A9}$$

$$(x_2 + x_6)P(C|B) = (x_3 + x_7)(1 - P(C|B)). \tag{A10}$$

Note that changing all values x_i into $k \cdot x_i$ for $k > 0$ yields another solution to (A7)–(A10). Let us assume that the lower bound $P(B|A) \cdot (1 - (1 - P(C|B))/P(A|B))$ can be attained. It means that (A1) holds with equality, i.e.,

$$P(A \cap B) + P(B \cap C) - P(B) = P(A \cap C).$$

It is easy to verify that it forces $x_2 = x_5 = 0$. To show that this situation can occur, start with any original distribution x_1, \dots, x_7 satisfying (A7)–(A10) and make the following change: $X_1 = x_1 + x_5, X_2 = 0, X_3 = x_2 + x_3, X_4 = x_4 + x_5, X_5 = 0, X_6 = x_2 + x_6$. Replacing x_i by $X_i, i = 1, 6$ in (A7)–(A10) leads to

$$X_4 \cdot P(B|A) = (X_6 - x_2 + x_7)(1 - P(B|A)),$$

$$X_3 \cdot P(A|B) = (X_6 - x_2 + x_7)(1 - P(A|B)),$$

$$X_1 \cdot P(B|C) = (X_3 - x_2 + x_7)(1 - P(B|C)),$$

$$X_6 \cdot P(C|B) = (X_3 - x_2 + x_7)(1 - P(C|B)).$$

Hence the distribution $X_1, X_2, \dots, X_6, X_7 = x_7 - x_2$ is feasible, insofar as $x_7 \geq x_2$, and leaves the data unchanged while saturating the lower bound, since

$X_2 = X_5 = 0$. Note that $P(A \cap B) + P(B \cap C) - P(B) = x_7 - x_2 \leq 0$ when $x_7 \leq x_2$; hence, the bound (A1) becomes ineffective when $x_7 \leq x_2$. In that case let us prove that $P(C|A) = 0$ is always possible to get; that is, $x_5 + x_7 = 0$. Consider the new distribution $X'_1 = x_1 + x_5$, $X'_3 = x_3 + x_7$, $X'_4 = x_4 + x_5$, $X'_5 = 0$, $X'_6 = x_6 + x_7$, $X'_7 = 0$, and $X'_2 = x_2 - x_7 \geq 0$. Making these substitutions in (A7)–(A10) leads to a feasible assignment such that $P(C|A) = 0$. Hence, the set of solutions allows the reduction of (A1) to a zero lower bound.

Lastly, consider the case when $P(A \cap B) = 0$, i.e., $x_6 = x_7 = 0$. Then $x_2 \geq x_7$ always holds and the last substitution always makes sense.

(ii) Upper bound. Again let us saturate (A3), (A4), (A5), and (A6) in turn. Saturation of (A3) leads to $P(A \cap C) = P(A) - P(A \cap B) + P(B \cap C)$, i.e., $x_5 + x_7 = x_4 + x_5 + x_6 + x_7 - x_6 - x_7 + x_3 + x_7$. Hence, $x_4 + x_3 = 0$, i.e., $x_4 = x_3 = 0$. Consider the new set of weights $X_1 = x_1 - x_4$, $X_2 = x_2 + x_3$, $X_3 = 0$, $X_4 = 0$, $X_5 = x_4 + x_5$, $X_6 = x_6 - x_3$, $X_7 = x_7 + x_3$. These weights satisfy (A7)–(A10). It is feasible to make this change if $x_6 \geq x_3$ and $x_1 \geq x_4$. Then

$$P(C|A) = 1 - P(B|A) - \frac{P(B|A) \cdot P(C|B)}{P(A|B)}$$

is attained.

Saturation of (A4) leads to $P(A \cap \bar{C}) = 0$, i.e., $x_4 = x_6 = 0$. From a distribution of weights x_1, \dots, x_7 satisfying (A7)–(A10) let us compute the new distribution $X_1 = x_1 - x_4$, $X_2 = x_2 + x_6$, $X_3 = x_3 - x_6$, $X_4 = 0$, $X_6 = 0$, $X_7 = x_6 + x_7$. It is feasible to make this change if $x_1 \geq x_4$, $x_3 \geq x_6$.

Saturation (A5) reads

$$P(\bar{A} \cap C) - P(B \cap C) + P(A \cap B) = 0 = x_1 + x_3 - x_7 - x_3 + x_6 + x_7$$

i.e., $x_1 = x_6 = 0$. This is obtained by the following new set of weights that satisfies (A7)–(A10): $X_1 = 0$, $X_2 = x_2 + x_6$, $X_3 = x_3 - x_6$, $X_4 = x_4 - x_1$, $X_5 = x_1 + x_5$, $X_6 = 0$, $X_7 = x_6 + x_7$. This set of weights is feasible if and only if $x_3 \geq x_6$ and $x_4 \geq x_1$.

Saturation of (A6) reads $P(C \cap \bar{A}) = 0$, i.e., $x_1 = x_3 = 0$. This is obtained by the following new set of weights: $X_1 = 0$, $X_2 = x_2 + x_3$, $X_3 = 0$, $X_4 = x_4 - x_1$, $X_5 = x_1 + x_5$, $X_6 = x_6 - x_3$, $X_7 = x_7 + x_3$. This set of weights is feasible if and only if $x_4 \geq x_1$ and $x_6 \geq x_3$.

Clearly, the four situations correspond to a partition of the set of feasible weights for constraints (A7)–(A10), according to the signs of $x_4 - x_1$ and $x_6 - x_3$. Hence any set of weights verifying (A7)–(A10) can be modified in such a way that one of the upper bounds is reached without modifying the input conditional probabilities. Therefore, bounds are optimal in the general case.

The pathological case $P(A|B) = P(B|A) = 0$ leads to $x_6 + x_7 = 0$. Then $P(C|A) = 1$ requires $P(\bar{C} \cap A) = 0$, i.e., $x_4 = 0$. But x_4 is not restricted by (A9) and (A10) and can always be taken as 0.

The pathological case $P(B|C) = P(C|B) = 0$ leads to $x_3 + x_7 = 0$, i.e., $x_3 = x_7 = 0$. Hence (A7)–(A10) become

$$\begin{aligned} (x_4 + x_5)P(B|A) &= x_6(1 - P(B|A)), \\ x_2P(A|B) &= x_6(1 - P(B|A)). \end{aligned}$$

The first bound $P(C|A) = 1$ obtained if $P(A \cap \bar{C}) = 0$, i.e., $x_4 = 0 = x_6$; but this implies $P(B|A) = 0$. Otherwise the only attainable upper bound is when $P(C \cap A) = P(A) - P(B \cap A) = P(A \cap \bar{B})$, which reads $x_5 = x_4 + x_5$, i.e., $x_4 = 0$. Then given any assignment of weights (x_2, x_4, x_5, x_6) , the new set of weights $X_2 = x_2, X_4 = 0, X_5 = x_4 + x_5, X_6 = x_6$ leads to attaining the upper bound.

THEOREM 3. The only cases when the result of the quantified syllogism is precise, i.e., $P^*(C|A) = P_*(C|A)$, are as follows:

- (i) $P(B|A) = P(C|B) = 1$ and then $P(C|A) = 1$,
- (ii) $P(A|B) = P(B|A) = 1$ and then $P(C|A) = P(C|B)$,
- (iii) $P(A|B) = P(B|C) = 1$ and then $P(C|A) = P(C|B) \cdot P(B|A)$,
- (iv) $P(C|B) = P(B|C) = 1$ and then $P(C|A) = P(B|A)$.

Proof. Note that $P^*(C|A) = 0$ is impossible. Indeed when $P(B|A) = P(A|B) = 0$, $P(C|A)$ is totally unknown; and all terms in the upper bound of $P(C|A)$ are strictly positive if $P(B|A) > 0$.

Let us check when $P^*(C|A) = P_*(C|A) > 0$. Note that $P(B|A)$ and $P(A|B)$ are positive.

- (i) $P^*(C|A) = P_*(C|A) = 1$. This reads

$$P(B|A) \cdot \left(1 - \frac{1 - P(C|B)}{P(A|B)}\right) = 1,$$

which forces $P(B|A) = P(C|B) = 1$.

- (ii)

$$P^*(C|A) = P_*(C|A) = 1 - P(B|A) + \frac{P(B|A) \cdot P(C|B)}{P(A|B)}.$$

Using the value of $P_*(C|A)$ and simplifying, we get the condition

$$P(A|B) \cdot (1 - P(B|A)) + P(B|A) \cdot (1 - P(A|B)) = 0,$$

which forces $P(B|A) = P(A|B) = 1$. Then $P(C|A) = P(C|B)$.

- (iii)

$$P_*(C|A) = P^*(C|A) = \frac{P(B|A) \cdot P(C|B)}{P(A|B) \cdot P(B|C)}.$$

Performing as in (ii) leads to the condition

$$P(B|C) \cdot (1 - P(A|B)) + P(C|B) \cdot (1 - P(B|C)) = 0,$$

which forces $P(A|B) = 1$ and $P(B|C) = 1$. Then

$$P(C|A) = \frac{P(C)}{P(A)} = \frac{P(C)}{P(B)} \cdot \frac{P(B)}{P(A)} = P(C|B) \cdot P(B|A).$$

(iv)

$$P_*(C|A) = P^*(C|A) = \frac{P(B|A) \cdot P(C|B)}{P(A|B) \cdot P(B|C)} \cdot (1 - P(B|C)) + P(B|A).$$

A similar treatment leads to the condition

$$P(C|B) \cdot (1 - P(B|C)) + P(B|C) \cdot (1 - P(C|B)) = 0,$$

which forces $P(C|B) = P(B|C) = 1$. Then $P(C|A) = P(B|A)$.

THEOREM 4. When $P(A|B)$, $P(B|C)$, $P(B|A)$, $P(C|B)$ are imprecisely known, say $P(A|B) \in [P_*(A|B), P^*(A|B)]$, etc., then the optimal bounds on $P(C|A)$ are given as follows:

$$P_*(C|A) = P_*(B|A) \cdot \max \left(0, 1 - \frac{1 - P_*(C|B)}{P_*(A|B)} \right) \quad \text{if } P_*(A|B) > 0 \tag{A11}$$

$$= P_*(B|A) \quad \text{if } P_*(A|B) = 0, \quad P_*(C|B) = 1 \tag{A12}$$

$$= 0 \quad \text{otherwise} \tag{A13}$$

$$P^*(C|A) = \min \left(1, P^*(B|A) \left(1 + \frac{P^*(C|B)}{P_*(A|B)} \left(\frac{1}{P_*(B|C)} - 1 \right) \right) \right) \quad \text{if } P_*(A|B) > 0, \quad P_*(B|C) > 0, \quad P^*(C|B) > P_*(A|B) \tag{A14}$$

$$= \min \left(1 + P_*(B|A) \left(\frac{P^*(C|B)}{P_*(A|B)} - 1 \right), \frac{P^*(C|B)}{P^*(C|B) + P_*(B|C)(P_*(A|B) - P^*(C|B))}, \frac{P^*(B|A)P^*(C|B)}{P_*(A|B)P_*(B|C)} \right) \quad \text{if } P_*(A|B) > 0, \quad P_*(B|C) > 0, \quad P^*(C|B) < P_*(A|B) \tag{A15}$$

$$= \min \left(1, \frac{P^*(B|A)}{P_*(B|C)} \right) \quad \text{if } P^*(C|B) = P_*(A|B) > 0, \quad P_*(B|C) > 0 \tag{A16}$$

$$= \min \left(1, 1 + P_*(B|A) \left(\frac{P^*(C|B)}{P_*(A|B)} - 1 \right) \right) \quad \text{if } P_*(B|C) = 0, \quad P_*(A|B) > 0, \quad P^*(B|A) > 0 \tag{A17}$$

$$= P^*(B|A) \quad \text{if } P_*(A|B) = 0, \quad P_*(B|C) = 1, \quad P^*(B|A) > 0, \quad P^*(C|B) > 0 \tag{A18}$$

$$= 1 - P_*(B|A) \quad \text{if } P_*(B|A) = 0, \quad P^*(C|B) = 0, \quad P^*(B|A) > 0, \tag{A19}$$

$$= 1 \quad \text{otherwise} \tag{A20}$$

Proof. Consider the regular form of the lower bound, with $x = P(A|B)$, $y = P(B|A)$, $z = P(C|B)$. It is of the form

$$f(x, y, z) = y \cdot \max \left(0, 1 - \frac{1 - z}{x} \right).$$

This function is increasing with x, y, z as long as $x > 0$, on $[0, 1]^3$. This proves (A11). If x can reach 0 but $P_*(C|B) < 1$ then the term $1 - (1 - z)/x$ can get arbitrarily negative and the bound (A13) is obtained. When $P_*(C|B) = 1$, then the original bound in Theorem 1 when $P(C|B)$ is precise and is 1, reads $P(C|A) \geq P(B|A)$ provided that $P(B|A) > 0, P(A|B) > 0$. Hence $f(x, y, 1) = y, \forall x > 0$. Hence $\lim_{x \rightarrow 0} f(x, y, 1) = y$ and the bound (A12) is obtained, since if $y > 0$, we cannot have $x = 0$ exactly ($P(B|A) > 0 \Rightarrow P(A|B) > 0$). In this case $P_*(A|B) = 0$ means that this probability is allowed to be infinitesimal without being zero.

For the upper bound let us consider $t = P(B|C)$, and the function

$$g(x, y, z, t) = \min \left(1, 1 - y + \frac{yz}{x}, \frac{yz}{xt}, \frac{yz}{xt} - \frac{yz}{x} + y \right).$$

It comes from the upper bound of $P(C|A)$ in the precise case. It is always decreasing with x and increasing with z . But the behavior is less simple with y and t . Let us write the second term $1 + y(z/x - 1)$. It is increasing with y when $z > x$ and decreasing otherwise. The last term can be written $y[1 + z/x(1/t - 1)]$ and is clearly increasing with y , and decreasing with t ; moreover this term is clearly subsumed by yz/xt when $z < x$.

In summary, when $P_*(A|B) > P_*(B|C) > 0$ we have

$$\sup_{x,t,z} g(x, y, z, t) = \min \left(1, 1 - y + \frac{yP^*(C|B)}{P_*(A|B)}, \frac{yP^*(C|B)}{P_*(A|B)P_*(B|C)}, y \left(1 + \frac{P^*(C|B)}{P_*(A|B)} \left(\frac{1}{P_*(B|C)} - 1 \right) \right) \right).$$

If $P^*(C|B) > P_*(A|B)$ then the third term above is not involved and we can substitute $y = P^*(B|A)$ elsewhere since the second and fourth terms are then increasing with y . This explains (A14) since the second term is above 1.

1. When $P^*(C|B) > P_*(A|B)$, the fourth term is greater than the third, so that we must compute

$$\sup_y \min \left(1 - y \left(1 - \frac{P^*(C|B)}{P_*(A|B)} \right), y \frac{P^*(C|B)}{P_*(A|B)P_*(B|C)} \right)$$

where the first term and the second one are respectively linearly decreasing and increasing. If y is unrestricted, the optimum is attained when the terms are equal:

$$1 - y \left(1 - \frac{P^*(C|B)}{P_*(A|B)} \right) = y \frac{P^*(C|B)}{P_*(A|B)P_*(B|C)}.$$

hence for

$$y^* = \frac{P_*(A|B)P_*(B|C)}{P^*(C|B) + P_*(B|C)(P_*(A|B) - P^*(C|B))}.$$

But this optimum makes sense only if $P_*(B|A) \leq y^* \leq P^*(B|A)$. In that case

$$P^*(C|A) = \frac{P^*(C|B)}{P^*(C|B) + P_*(B|C)(P_*(A|B) - P^*(C|B))}.$$

If $y^* \leq P_*(B|A)$, the maximum is obtained by letting $y = P_*(B|A)$ in $g(P_*(A|B), y, P_*(B|C), P^*(C|B))$, and is equal to $1 - P_*(B|A)(1 - P^*(C|B)/P_*(A|B))$ (second term). If $y^* \geq P^*(B|A)$, then we must let $y = P^*(B|A)$ in $g(P_*(A|B), y, P_*(B|C), P^*(C|B))$ and the optimum is equal to $P^*(B|A)P^*(C|B)/P_*(A|B)P_*(B|C)$. On the whole, we get (A15), and (A16) follows obviously.

Let us now assume that only $P_*(B|C) = 0$. Then the term z/t can become as large as needed in $g(x, y, z, t)$ (fixing $z > 0$ and choosing t small enough). Hence the third and fourth terms in g can grow arbitrarily large. Only the second term remains bounded. If $P^*(C|B) > P_*(A|B)$ then this term is bigger than 1 for any y . Otherwise its bound is obtained by letting $y = P_*(B|A)$. This proves expression (A17). If $P^*(C|B) = 0$ too, then, from Theorem 1, $P(C|A)$ is bounded from above only by $1 - P_*(B|A)$, and this again justifies (A17).

Assume now that $P_*(A|B) = 0$. Then, if $P^*(B|A) > 0$, the term y/x can grow arbitrarily. Now if $P^*(C|B) > 0$ and $P_*(B|C) \neq 1$, all terms in g can grow arbitrarily and $P^*(C|A) = 1$. This proves (A20). If $P^*(B|A) = 0$, then (A20) applies again because of Theorem 1.

Assume $P_*(A|B) = 0$, $P^*(B|A) > 0$, $P_*(B|C) = 1$, and $P_*(C|B) > 0$. Then for all $x > 0$, $\sup_{z,x} g(x, y, 1, z) = y$, and (A18) holds.

Lastly assume $P_*(A|B) = 0$, $P^*(B|A) > 0$, and $P^*(C|B) = 0$. Then from Theorem 1, $P^*(C|A) \leq 1 - P^*(B|A)$, and (A19) holds.

References

Adams, E.W. (1975). *The Logic of Conditionals*, Dordrecht Riedel.
 Aleliunas, R. (1988). A new normative theory of probabilistic log. In *Proceedings of 7th Biennial Conf. of Canadian Society for Computational Studies of Intelligence (AICSCSI'88)* (R. Goedel, ed.), Edmonton, Alberta, Canada, June 6-10, 67-74.
 Amarger, S., Dubois, D., and Prade, H. (1991a). Handling imprecisely-known conditional probabilities. In *Proc. of the UNICOM Seminar "AI and Computer Power—The Impact on Statistics*. Brunel Conf. Center, West London, March 13-14, 1991. 1-33, Final proceedings to be published by Chapman & Hall.
 Amarger, S., Dubois, D., and Prade, H. (1991b). Constraint propagation with imprecise conditional probabilities. In *Proc. of the 7th Conf. on Uncertainty in Artificial Intelligence*. Los Angeles, CA, July 13-15, 26-34.

- Bacchus, F. (1990). *Representing and Reasoning with Probabilistic Knowledge*, Cambridge, MA: MIT Press.
- Boole, G. (1854). *The Laws of Thought*, MacMillan, New York. (Reprinted by Dover, New York, 1958.)
- De Morgan, A. (1966). "On the Syllogism" and Other Logical Writings. London: Routledge & Kegan Paul Ltd. (First published between 1846 and 1868.)
- Dubois, D., and Prade, H. (1988). On fuzzy syllogisms. *Computational Intelligence (Canada)*, 4-2, 171-179.
- Dubois, D., and Prade, H. (1991). "Conditional objects and non-monotonic reasoning. In *Proc. of the 2nd Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'91)*. Cambridge, MA., April 22-25, 175-185.
- Dubois, D., Prade, H., and Toucas, J.M., (1990). Inference with imprecise numerical quantifiers. In Ras, Z.W., and Zemankova, M. (Eds.), *Intelligent Systems—State of the Art and Future Direction*. Ellis Horwood, 52-72.
- Geffner, H. (1992), *Default Reasoning: Causal and Conditional Theories*. Cambridge, MA: MIT Press.
- Ginsberg, M.L. (1988). "Multivalued logics," *Computational Intelligence*, 4, pp. 265-316.
- Goodman, I.R., Gupta, M.M., Nguyen, H.T., and Rogers, G.S. (Eds.) (1991). *Conditional Logic in Expert Systems*. Amsterdam: North-Holland.
- Güntzer, U., Kießing, W., and Thöne, H. (1991). New directions for uncertainty reasoning in deductive databases. In *Proc. ACM SIGMOD Inter. Conf. on Management of Data, Denver*, 178-187.
- Heinsohn, J. (1991). A hybrid approach for modeling uncertainty in terminological logics. In Kruse, R., Siegel, P. (Eds.), *Symbolic and Quantitative Approaches to Uncertainty* (Proc. of the Europ. Conf. ECSQAU, Marseille, October, 1991), Lecture Notes in Computer Science, 548, Berlin Springer Verlag, 198-205. (Complete version: Research Report n° RR-91-24, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Saarbrücken, Germany, 1991).
- Kahneman, D., Slovic, P., and Tversky, A. (1980). *Judgement under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Neufeld, E. (1990). A probabilistic commonsense reasoner. *Int. J. of Intelligent Systems*, 5, 565-594.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.
- Peterson, P. (1979). On the logic of few, many and most. *Notre Dame J. Formal Logic*, 20, 155-179.
- Quinlan, J.R. (1983). INFERNO, a cautious approach to uncertain inference. *The Computer Journal*, 26, 255-269.
- Shapiro, S.C. (1979). Numerical quantifiers and their use in reasoning with negative information. In *Proc. of the 6th Inter. Joint Conf. on Artificial Intelligence (IJCAI'79)*, Tokyo, Japan 791-796.
- Spies, M. (1989). *Syllogistic Inference under Uncertainty*. München, Germany: Psychologie Verlags Union.
- Thöne, H., Güntzer, U., and Kießing, W. (1991). Probabilistic reasoning with facts and rules in deductive databases. In Kruse, R., Siegel, P. (Eds.), *Symbolic and Quantitative Approaches to Uncertainty* (Proc. of the Europ. Conf. ECSQAU, Marseille, October, 1991), Lecture Notes in Computr Science, 548, Springer Verlag, Berling, 333-337.
- Thöne, H., Güntzer, U., and Kießing, W. (1992). Towards precision of probabilistic bounds propagation. In Dubois, D., Wellman, M.P., D'Ambrosio, B., and Smets, P. (Eds.), *Proc. of the 8th Conf. on Uncertainty in Artificial Intelligence*, Morgan and Kaufmann, 315-322.
- Travè-Massuyès, L., and Piera, N. (1989). The orders of magnitude models as qualitative algebras. In *Proceedings of 11th Inter. Joint Conf. on Artificial Intelligence (IJCAI-89)*, Detroit, Michigan, Aug. 20-25, 1261-1266.
- Van Benthem, J. (1986), *Essays on Logical Semantic*. Amsterdam: D. Reidel.
- Wellman, M.P. (1988). Qualitative probabilistic networks for planning under uncertainty. In Lemmer, J.F., and Kanal, L.N. (Eds.), *Uncertainty in Artificial Intelligence 2*. Amsterdam: North-Holland, 197-208, 1988.

- Wellman, M.P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44, 257–303.
- Westerstahl, D. (1989). Quantifiers in formal and natural languages. In Gabbay, D., and Guentner, F. (Eds.), *Handbook of Philosophical Logic, Vol. IV*. Amsterdam: D. Reidel 1–131.
- Xiang, X., Beddoes, M.P., and Poole, D. (1990). Can uncertainty management be realized in a finite totally ordered probability algebra. In Henrion, M., Shachter, R.D., Kanal, L.N., and Lemmer, J.F. (Eds.), *Uncertainty in Artificial Intelligence 5*. Amsterdam: North-Holland, 41–57.
- Zadeh, L.A. (1985). Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions. *IEEE Trans. on Systems, Man and Cybernetics*, 15–6 754–763.