Direct Algorithms for Checking Coherence and Making Inferences from Conditional Probability Assessments

Peter Walley, Renato Pelessoni [a] and Paolo Vicig [a]

[a] *Dipartimento di Matematica Applicata 'B. de Finetti', Università di Trieste, Trieste, Italy*

Abstract

We solve two fundamental problems of probabilistic reasoning: given a finite set of conditional probability assessments, how to determine whether the assessments are mutually consistent, and how to determine what they imply about the conditional probabilities of new events? These problems were posed in 1854 by George Boole, who gave a partial solution using algebraic methods. The two problems are fundamental in applications of the Bayesian theory; Bruno de Finetti solved the second problem for the special case of unconditional probability assessments in what he called 'the fundamental theorem of probability'. Using ideas from the theory of imprecise probabilities, we show that the general problems have simple, direct solutions which can be implemented using linear programming algorithms. Unlike earlier proposals, our methods are defined directly in terms of the assessments, without introducing unknown probabilities. Our methods work when any of the conditioning events may have probability zero, and they work when the assessments include imprecise (upper and lower) probabilities or previsions.

Key words: Bayesian inference; coherence; conditional probability; imprecise probability; lower probability; natural extension; probabilistic logic

1 Introduction

1.1 The fundamental problems of probabilistic reasoning

This paper is concerned with solving two fundamental problems of probabilistic reasoning:

The consistency problem: given an arbitrary finite collection of conditional probability assessments, how can we determine whether the assessments are mutually consistent?

The inference problem: how can we make inferences from the assessments concerning the conditional probabilities of other events?

In the simplest version of the problem, we suppose that conditional probabilities $P(A_i|B_i)$ are specified for $i = 1, 2, \ldots, k$, where A_i and B_i are arbitrary events with B_i non-null. The two problems are to check whether these conditional probability assessments are mutually consistent, and to calculate what they imply about a further conditional probability $P(A|B)$.

The consistency and inference problems are fundamental problems in probabilistic logic [24], in the Bayesian theory of inference [9], and in the theory of imprecise probability [31]. The problems are therefore of great importance in expert systems and artificial intelligence, where these theories are widely used [32]. In this paper we describe general algorithms that can be used to solve the two problems. The computations involve only linear programming. In most applications, each problem can be solved through a single linear program.

The algorithms that we propose for solving the consistency and inference problems are *direct*: they work by investigating certain kinds of linear combinations of the assessments. The algorithms that have been previously proposed for solving the two problems are *indirect*: they work by investigating certain kinds of precise probability measures that extend the assessments to other events or that dominate the assessments. The direct and indirect methods can be regarded as dual formulations of the same problem, but, not surprisingly, the direct methods turn out to be both theoretically and computationally simpler than the indirect ones. Our solutions are based on the concepts of coherence and natural extension from the theory of imprecise probabilities [31], which are both defined directly in terms of the assessments. In contrast, the previous methods are based on indirect definitions of consistency and inference: consistency is defined as compatibility with a precise conditional probability measure on a larger domain, and inferences are made by computing bounds on the possible extensions of the assessments to conditional probability measures on larger domains.

Two further properties that distinguish our algorithms from most of the previous proposals (exceptions are noted in subsection 1.4) are that they work when any of the conditioning events may have probability zero, and they work when any of the conditional probability assessments are imprecise. The standard techniques for checking consistency or making inferences [15–17,20,21,24,26], which involve a single linear program, are implicitly based on an assumption that the probabilities of all the conditioning events are bounded away from zero. When this assumption fails, the standard techniques may give incorrect answers. The assumption is frequently violated in practical applications, because the given assessments are usually uninformative about at least some of the probabilities of conditioning events, i.e., they imply the trivial lower bound of zero. For example, an assessment of a conditional probability $P(A|B)$ is, by itself, completely uninformative about the probability of B, and it is consistent with $P(B) = 0$. Further examples of zero probabilities are given in the following subsection.

It is also common in applications that not all probabilities can be assessed precisely, and methods are needed to deal with imprecise probabilities. This issue is discussed in subsection 1.3.

1.2 Numerical examples

In this subsection we describe two numerical examples which will be used throughout the paper to illustrate our algorithms. The first example, which is based on Example 6 of [17], shows that it is not difficult for zero probabilities to arise even in quite small examples.

Example 1 *Suppose that* $A_1, \ldots, A_5$ *are logically independent events, and the following precise, unconditional probability assessments are made:* $P(A_1) = 0.6$, $P(A_1^c \cup A_2) = 0.4$, $P(A_2 \cup A_3) = 0.8$, $P(A_3 \cap A_4) = 0.3$, $P(A_4^c \cup A_5) = 0.5$, *and* $P(A_2 \cup A_5) = 0.6$*. In this case the consistency problem is to determine whether the six assessments are coherent, and the inference problem is to determine what they imply about the probabilities of other events, for instance* $P(A_3)$, $P(A_4|A_3)$, *or* $P(A_1 \cap A_2|B)$ *where* $B = (A_1 \cap A_2) \cup (A_1^c \cap A_2^c)$.

It is immediately apparent that the six assessments are consistent with many events having zero probabilities. For example, it is clear from the first two assessments that $P(A_1 \cap A_2)$ *must be zero. If we knew that* $P(B)$ *was bounded away from zero then we could infer from the assessments that* $P(A_1 \cap A_2|B) = 0$*. However, it turns out that the assessments are also consistent with* $P(B) = 0$*, and it is not immediately clear what we can say about* $P(A_1 \cap A_2|B)$*. Because the conditioning event* B *may have probability zero, the standard algorithms cannot be relied upon to compute* $P(A_1 \cap A_2|B)$*. This is not an isolated example: there are many other events, such as* $A_3 \cap A_5$, $A_4 \cap A_5$, $(A_1 \cup A_4)^c$ *and* $A_2 \cap (A_1 \cup A_3 \cup A_5)$*, which may have probability zero here. If* C *is such an event and we want to make inferences about a conditional probability* $P(A|C)$*, we cannot in general use the standard linear programming algorithms. There may be similar difficulties in checking consistency, if any further assessment is made that is conditional on* C.

Example 2 *Suppose that three football teams, X, Y and Z, play a tourna-*

Table 1
Precise probability assessments $P(C_i)$ for the 27 events C_i in the football example.

(L,L,L)	(L,L,W)	(L,D,L)	(L,D,D)	(L,D,W)	(L,W,D)	(L,W,W)	(D,L,L)	(D,L,D)
0.015	0.04	0.02	0.045	0.04	0.055	0.06	0.015	0.02
(D,L,W)	(D,D,L)	(D,D,W)	(W,L,L)	(W,L,D)	(L,L,D)	(W,D,D)	(W,D,W)	(W,W,L)
0.04	0.02	0.05	0.01	0.02	0.02	0.04	0.055	0.04
(W,W,D)	(W,W,W)	(D,W,L)	(D,W,D)	(D,W,W)	(W,D,L)	(L,W,L)	(D,D,D)	(W,L,W)
0.04	0.06	0.04	0.05	0.07	0.02	0.04	0.035	0.04

ment involving three matches, X vs. Y, X vs. Z, and Y vs. Z. Two points are assigned for a win, and one for a draw. The winner is the team which gains the most points. If this criterion does not determine a unique winner, further rules are applied, based firstly on the number of goals scored, and finally using randomization if necessary.

A subject makes precise probability assessments for the 27 *events of the kind* $C_i = (R(X,Y), R(X,Z), R(Y,Z))$, *where* $R(\cdot,\cdot)$ *represents W (win), D (draw) or L (loss) for the first team in alphabetical order. For instance,* (W,W,L) *means that X wins against Y and against Z, and Y loses against Z. Suppose that the probability assessments are those given in Table 1, based on the subject's belief that team Z is weaker than teams X and Y, which are about equally strong. In this case, because the* 27 *events* C_i *form a partition, it easy to verify that the probability assessments are mutually consistent, by checking only that the numbers are non-negative and sum to one.*

The subject wishes to evaluate the probability of the event A, that team X wins the tournament. This is an example of an inference problem. In this case it is relatively easy to find the best possible lower and upper bounds for $P(A)$, *by summing the probabilities of all the events* C_i *that imply A to get the lower bound* 0.325, *and summing the probabilities of all the events* C_i *that are consistent with A to get the upper bound* 0.53.

All the events C_i *have been assigned positive probability in this example, so it may appear that no difficulty could arise from conditioning on an event of probability zero. But these difficulties can arise when we introduce other events. For example, let B denote the event that team X scores more goals in the tournament than each of the other teams. Then both B and its complement are consistent with all* 27 *events* C_i, *so it is consistent with the assessments in Table 1 that B has probability zero. Consequently difficulties may arise in calculating* $P(A|B)$.

1.3 More general versions of the problems

In general, a set of conditional probability assessments cannot be expected to determine a unique value for a further conditional probability $P(A|B)$, but only to determine upper and lower bounds for $P(A|B)$. That was just seen in the inferences about $P(A)$ in the football example, where we obtained only upper and lower bounds for $P(A)$. The upper and lower bounds are called *upper and lower probabilities.*

Because inferences usually need to be expressed in terms of upper and lower probabilities, it is natural to generalize the problem by allowing the initial assessments to be also in the form of upper and lower probabilities. In fact this generalization greatly extends the scope of the problem. There are many applications in which it is difficult or unreasonable to make precise probability assessments, because either there is little available information on which to base the assessments or the information is difficult to evaluate. In these problems we may be able only to assess upper and lower probabilities. For example, the precise probability assessments given in Table 1 are unrealistic since even a subject with very extensive information about the teams would find it difficult to assess precise probabilities for the 27 outcomes.

Because an assessment of an upper probability $\overline{P}(A|B)$ is equivalent to the assessment $\underline{P}(A^c|B) = 1 - \overline{P}(A|B)$ of a lower probability, we can assume that all the quantities assessed are lower probabilities. When the assessments of upper and lower probabilities coincide, the common value $\overline{P}(A|B) = \underline{P}(A|B)$ is called a *precise probability* and it is written as $P(A|B)$.

In the general formulation of the problem, we suppose that finitely many conditional lower probabilities $\underline{P}(A_i|B_i) = c_i$ are specified, for $i = 1, 2, \ldots, k$. We do not assume any structure for the collection of conditional events $\{A_1|B_1, A_2|B_2, \ldots, A_k|B_k\}$. A user is free to make whatever conditional probability assessments are most natural or convenient. All the relevant events, B_i and $A_i \cap B_i$ $(i = 1, 2, \ldots, k)$, can be identified with subsets of a possibility space Ω. If these events are not defined as subsets of a given possibility space, but instead the logical relationships amongst the events are specified, then we can define an appropriate possibility space whose atoms are all the events of the form $\bigcap_{i=1}^{k} D_i$ that are logically possible, where each D_i is chosen to be $A_i \cap B_i$, $A_i^c \cap B_i$ or B_i^c. For instance, the six unconditional assessments in Example 1 generate a partition which contains 20 atomic events. We assume that all the conditioning sets B_i are non-empty.

It is important to emphasize that this formulation allows assessments of precise probabilities, upper probabilities and unconditional probabilities. As previously noted, an assessment of an upper probability $\overline{P}(A|B) = c$ can be replaced by an equivalent assessment of a lower probability, $\underline{P}(A^c|B) = 1 - c$. A precise probability assessment $P(A|B) = c$, which is equivalent to specifying equal upper and lower probabilities $\overline{P}(A|B) = \underline{P}(A|B) = c$, can therefore be replaced by the two lower probability assessments $\underline{P}(A|B) = c$ and $\underline{P}(A^c|B) = 1 - c$. An unconditional lower probability assessment $\underline{P}(A) = c$ is equivalent to $\underline{P}(A|\Omega) = c$, i.e., equivalent to conditioning on the certain event

Table 2
Assessments of upper and lower conditional probabilities for five events $A|C_i$ in the football example.

	$A\|(L,W,L)$	$A\|(D,D,D)$	$A\|(D,W,W)$	$A\|(W,L,W)$	$A\|(W,D,L)$
$\overline{P}$	1	1	0.75	0.65	0.8
$\underline{P}$	0.6	0	0.25	0.4	0.65

Ω, and similarly for precise assessments of unconditional probability.

The formulation of the problem that we study here could be generalized further, to allow assessments of conditional upper and lower previsions (expectations) of random variables. See the Conclusions for details.

Example 3 *In the football example, suppose the subject recalls that* $P(A) = \sum_{i=1}^{27} P(A|C_i)P(C_i)$ *by the conglomerative property, and hence he could determine* $P(A)$ *precisely if he were able to assess the* 27 *values* $P(A|C_i)$. *The assessment of* $P(A|C_i)$ *is trivial for those* $A|C_i$ *which are impossible, like* $A|(L,L,D)$, *and those* $A|C_i$ *which are sure, like* $A|(W,W,D)$, *but there remain* 5 *events* $A|C_i$ *which are neither impossible nor sure. The subject might find it hard to give precise probabilities for these events because extra information, such as guessing the number of goals scored, is required. In particular, there is little information concerning the probability of* $A|(D,D,D)$. *So he adds* 10 *imprecise probability assessments, given in Table 2, to the* 27 *precise assessments in Table 1. Because each of the 27 precise assessments in Table 1 is equivalent to two assessments of lower probabilities, there are a total of* $2 \times 27 + 10 = 64$ *lower probability assessments.*

Now it is necessary to check the consistency of the system of 64 *lower probabilities, and to determine what the system implies about the probability of event* A. *These problems will be solved later in the paper. It turns out that, although the new assessments in Table 2 are quite imprecise, they substantially reduce the imprecision in the probability of* A.

1.4 Previous work on the problems

George Boole [2,3] formulated a version of the inference problem, assuming that the given assessments are mutually consistent, as early as 1854. He recognized that, in general, the probability assessments will determine only upper and lower bounds for the conditional probability of a new event, and he suggested several algebraic methods for finding the upper and lower bounds. The most efficient of his methods involves using the assessments to determine a system of linear equality and inequality constraints on variables which represent the unknown probabilities of the possible atomic events, and then solving this system by successive elimination of variables (Fourier-Motzkin elimination). Because Boole formulated the problem in terms of unknown probabilities, his methods are what we call *indirect* solutions.

Whereas linear programming gives only numerical bounds for $P(A)$ in the inference problem, the methods of Boole and Hailperin [16] can be used to produce an analytical solution. For example, if A_1 and A_2 are logically independent, $A = A_1 \cup A_2$, and the assessments $P(A_1)$ and $P(A_2)$ are regarded as unspecified parameters, Boole's method gives the general bounds $\max\{P(A_1), P(A_2)\} \leq P(A) \leq P(A_1) + P(A_2)$. Boole's methods are described in detail in [16,18]. It appears to be possible to use the direct methods of this paper in a similar way to derive the rules of coherence for imprecise probabilities, by using Fourier-Motzkin elimination to successively remove the coefficients λ_i.

For the special case of unconditional probability assessments, the inference problem was solved by de Finetti [8], in what he later called 'the fundamental theorem of probability' [9]. The name he gave to this result shows how important the problem is in the Bayesian theory of probability. In effect, de Finetti's fundamental theorem shows that the inference problem can be solved (for precise, unconditional probabilities) by linear programming (LP). The consistency problem can be solved using de Finetti's concept of coherence, which again involves linear programming. LP approaches to the consistency and inference problems and to de Finetti's fundamental theorem were described by Hailperin [15] and by Bruno and Gilio [4]. Several generalizations of the fundamental theorem of probability were proposed in [21,22,31].

Linear programming solutions for the consistency and inference problems also have a fundamental role in the theory of *probabilistic logic* that was proposed by Nilsson [24,25]; see also [26]. Hansen, Jaumard et al. [17,20] have shown that these methods can be extended to cope with problems that involve a very large number of assessments and variables, by incorporating *column generation* methods which avoid explicit formulation of an underlying partition of atomic events. They call the consistency and inference problems *probabilistic satisfiability* problems. The survey paper [17] also includes a thorough survey of earlier work on the consistency and inference problems, and of computational algorithms for implementing the earlier solutions.

Hailperin [16] extended Boole's methods to deal with *conditional* probabilities, by regarding an assessment $P(A|B) = c$ as a linear constraint, of the form $P(A \cap B) = cP(B)$, on the unknown probabilities $P(A \cap B)$ and $P(B)$. In general, this constraint is equivalent to the constraint $P(A|B) = c$ only if $P(B) > 0$, so Hailperin's method effectively assumes that all conditioning events have positive lower probability. Most of the other methods for solving the consistency and inference problems have the same limitation [17,20–22].

As seen in subsections 1.1 and 1.2, it is quite common in practice for some conditioning events of interest to have zero lower probability: this does not mean that the probability of such an event is known to be zero, but merely that having probability zero is consistent with the assessments.

Methods for handling zero probabilities have been studied only quite recently. A general method for checking consistency of precise conditional probability assessments, which works when conditioning events may have zero probability, was developed in a series of papers by Coletti, Gilio and Scozzafava, including [5,6,13,14]. Computationally, this method requires solving a sequence of LP problems. Another solution for the inference problem in the case of precise conditional probabilities, which again involves a sequence of LP problems, was given by Vicig [30].

Many of the earlier studies, for example [5,13,15,17,20–22], have considered the possibility that probability assessments are *imprecise*, although most of these works contain only a brief discussion of imprecision. Walley [31] gave a detailed theory of imprecise conditional probabilities, including very general formulae for checking consistency and making inferences, on which the approach in this paper is based. These formulae apply also when conditioning on events of probability zero and when infinitely many conditioning events are involved. (See also [33,37].) The consistency problem for imprecise assessments was also studied in [5,13], using a definition of consistency that is similar to what we call 'avoiding uniform loss', and in [29], using a stronger definition which we call 'coherence'. Pelessoni and Vicig [27] suggested an algorithm for computing the least-committal coherent correction of imprecise assessments which avoid uniform loss, which can also be used to give a solution to the inference problem. Again these algorithms require solving a sequence of LP problems. The Pelessoni-Vicig algorithm is related, through duality, to the methods proposed in this paper; see subsection 3.7 for a comparison.

All the computational methods proposed in the work we have outlined are *indirect*, in the sense that they involve programming problems in which the variables are taken to be unknown probabilities, and their solutions rely on properties of precise conditional probabilities. In this paper we propose *direct* methods for solving the consistency and inference problems, using the theory developed in [31], which do not involve unknown probabilities.

1.5 Outline of the paper

The consistency problem, to check whether the assessments of lower probabilities or precise probabilities are mutually consistent, is defined and solved in Section 2. Consistency is characterized mathematically through a condition of 'avoiding uniform loss'. In the case where all assessments are precise, this condition is equivalent to de Finetti's definition of coherence. Two algorithms for verifying this condition are proposed in subsection 2.3. When some of the assessments are imprecise, avoiding uniform loss still characterizes a basic type of consistency, but there is a stronger notion of consistency, called *coherence*, which is considered in subsections 3.2 and 3.8.

Section 3 describes a solution to the inference problem. Inferences are made by calculating upper and lower probabilities for a new conditional event $A|B$, using a concept of natural extension. In the case where all assessments are precise and coherent, the natural extensions are the upper and lower bounds for the interval of coherent values of $P(A|B)$. Two algorithms for computing natural extensions are proposed in subsection 3.4. Section 3 also contains a study of the conditions under which natural extension can be computed exactly by solving a single linear program, an investigation of the dual problem, and methods for checking whether a system of imprecise probability assessments is coherent.

Brief suggestions for generalizations and further research are given in the concluding section 4.

1.6 Notation

We use the same symbol, A or B, to denote both an event and its indicator function (de Finetti's convention). Using this convention, write $G_i = B_i[A_i - c_i]$ for $i = 1, 2, \ldots, k$, where $c_i = \underline{P}(A_i|B_i)$. Here G_i is a random variable which plays an important role in the ensuing theory. It represents the net reward from a bet on the event A_i conditional on B_i at the rate $\underline{P}(A_i|B_i)$. (The bet is called off, and the reward is zero, unless B_i occurs.) The lower probability $\underline{P}(A_i|B_i)$ is interpreted as a *marginally acceptable* rate for betting on A_i conditional on B_i, meaning that the bet whose reward is $B_i[A_i - a]$ is acceptable whenever $a < \underline{P}(A_i|B_i)$. Hence the random reward G_i is at least marginally acceptable, but it is not necessarily strictly acceptable. For any positive ε, the random variable $G_i + \varepsilon B_i$, which is the net reward from a conditional bet at the lower rate $\underline{P}(A_i|B_i) - \varepsilon$, is strictly acceptable.

When $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$ is a k-vector, we write $\boldsymbol{\lambda} \geq \mathbf{0}$ to mean that $\lambda_i \geq 0$ for $i = 1, 2, \ldots, k$, and we write $\boldsymbol{\lambda} \succ \mathbf{0}$ to mean that $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\boldsymbol{\lambda} \neq \mathbf{0}$. We shall consider linear combinations of the random variables G_i, of the form $\sum_{i=1}^{k} \lambda_i G_i$. When $\boldsymbol{\lambda} \geq \mathbf{0}$, we define $S(\boldsymbol{\lambda})$, the support of $\boldsymbol{\lambda}$, to be the union of those conditioning events B_i for which $\lambda_i > 0$. If we interpret λ_i as the stake of a bet on A_i which is called off unless B_i occurs, $S(\boldsymbol{\lambda})$ can be interpreted as the event that at least one non-zero bet takes place. Let $I(\boldsymbol{\lambda}) = \{i : \lambda_i >$

$0, i = 1, \ldots, k\}$, so $S(\boldsymbol{\lambda}) = \bigcup_{i \in I(\boldsymbol{\lambda})} B_i$. If X is a bounded random variable and B is a non-empty subset of Ω, $\sup\,[X|B] = \sup\,\{X(\omega) : \omega \in B\}$ denotes the supremum possible value of X if B occurs.

2 The consistency problem

The first problem is to determine whether the given assessments $\underline{P}(A_i|B_i) = c_i$ $(i = 1, 2, \ldots, k)$ are mutually consistent. This problem will be solved in this section by formulating a concept of 'avoiding uniform loss', which is equivalent, in the special case where all the assessed probabilities are precise, to de Finetti's concept of coherence.

2.1 Avoiding uniform loss (AUL)

It was explained in subsection 1.6 that, for any positive ε, the random reward $G_i + \varepsilon B_i$ is strictly acceptable. Following Walley [31,33], we say that the assessments *incur uniform loss* if there is a positive linear combination of these acceptable gambles whose net reward cannot possibly be positive. If the assessments do not incur uniform loss, we say that they avoid uniform loss. Formally, the assessments *avoid uniform loss* (AUL) if and only if the parametric system of linear inequalities

$$\sum_{i=1}^{k} \lambda_i (G_i + \varepsilon B_i) \leq 0 \qquad \text{and} \qquad \boldsymbol{\lambda} \succ \mathbf{0} \tag{1}$$

has no solution $(\boldsymbol{\lambda}, \varepsilon)$ with $\varepsilon > 0$. The sum on the left-hand side of (1) is a random variable X, and we write $X \leq 0$ to mean that the value of the variable X is certainly less than or equal to 0, i.e., $X(\omega) \leq 0$ for all $\omega \in \Omega$.

Condition (1) is defined directly in terms of the assessments c_i, since the left-hand side of (1) is a positive linear combination of the gambles $G_i + \varepsilon B_i = B_i[A_i - (c_i - \varepsilon)]$. As already mentioned, the positive adjustment ε is needed to ensure that these gambles are acceptable. A slightly different way of interpreting ε is to regard it as an arbitrarily small reduction to the assessed lower probabilities c_i, which makes the assessments slightly more cautious. In other words, ε acts as a *perturbation* of the assessments.

If the assessments incur uniform loss, it is useful to identify a subset of mutually inconsistent assessments $I(\boldsymbol{\lambda}) = \{i : \lambda_i > 0, i = 1, \ldots, k\}$, where $\boldsymbol{\lambda} \succ \mathbf{0}$ is such that $\sum_{i=1}^{k} \lambda_i (G_i + \varepsilon B_i) \leq 0$ for some $\varepsilon > 0$. To achieve consistency, some of the assessments $\underline{P}(A_i|B_i)$ $(i \in I)$ must be reduced. Several ways of doing this are discussed later in subsection 2.4.

AUL is a simple consistency requirement. It is violated if and only if there is a positive linear combination of strictly acceptable conditional bets which cannot possibly result in a net gain. The discussion in [31, Ch. 7] indicates that AUL is the proper characterization of consistency in this problem. In the following discussion, we compare it with a weaker condition that is implicit in some of the literature. First we give an alternative characterization of AUL.

Lemma 1 *The assessments AUL if and only if*

$$\sup\,[\sum_{i=1}^{k} \lambda_i G_i | S(\boldsymbol{\lambda})] \geq 0 \quad \text{whenever} \quad \boldsymbol{\lambda} \succ \mathbf{0}. \tag{2}$$

PROOF. Using $S(\boldsymbol{\lambda}) = \bigcup\{B_i : \lambda_i > 0\}$ and writing $\tau_1 = \min\,\{\lambda_i : \lambda_i > 0, i = 1, \ldots, k\}$ and $\tau_2 = \sum_{i=1}^{k} \lambda_i$, we see that

$$\tau_1 S(\boldsymbol{\lambda}) \leq \sum_{i=1}^{k} \lambda_i B_i \leq \tau_2 S(\boldsymbol{\lambda}). \tag{3}$$

If the assessments incur uniform loss then there are $\boldsymbol{\lambda} \succ \mathbf{0}$ and $\varepsilon > 0$ such that $0 \geq \sum_{i=1}^{k} \lambda_i(G_i + \varepsilon B_i) \geq \sum_{i=1}^{k} \lambda_i G_i + \varepsilon\tau_1 S(\boldsymbol{\lambda})$, using (3), with $\tau_1 > 0$ since $\boldsymbol{\lambda} \succ \mathbf{0}$. Hence $\sup\,[\sum_{i=1}^{k} \lambda_i G_i | S(\boldsymbol{\lambda})] \leq -\varepsilon\tau_1 < 0$, so that (2) fails.

Conversely, if there is $\boldsymbol{\lambda} \succ \mathbf{0}$ such that $\sup\,[\sum_{i=1}^{k} \lambda_i G_i | S(\boldsymbol{\lambda})] = -\delta < 0$, let $\varepsilon = \delta/\tau_2$. (Here $\tau_2 > 0$ since $\boldsymbol{\lambda} \succ \mathbf{0}$.) Then $\sum_{i=1}^{k} \lambda_i(G_i + \varepsilon B_i) = \sum_{i=1}^{k} \lambda_i G_i + \varepsilon \sum_{i=1}^{k} \lambda_i B_i \leq -\delta S(\boldsymbol{\lambda}) + \varepsilon\tau_2 S(\boldsymbol{\lambda}) = 0$, using (3). Thus the failure of (2) implies that the assessments incur uniform loss. $\diamondsuit$

Another characterization is that the assessments AUL if and only if there is no $\boldsymbol{\lambda} \succ \mathbf{0}$ such that $\sum_{i=1}^{k} \lambda_i G_i \leq -S(\boldsymbol{\lambda})$. Another characterization, which can be obtained from the properties in section 3.2 or from results in [37], is that the assessments AUL if and only if there are precise conditional probabilities $\{P(A_i|B_i) : i = 1, \ldots, k\}$ which satisfy AUL (or, equivalently, de Finetti's definition of coherence) and $P(A_i|B_i) \geq c_i$ for $i = 1, \ldots, k$. This is an indirect characterization, in the sense that it refers to precise conditional probabilities with certain properties. This characterization is the basis for the algorithms previously introduced for checking AUL or similar conditions. In particular [13] takes the dominance condition in this characterization as a definition of coherence for imprecise probabilities, and [5] considers a condition similar to AUL but without assuming the conjugacy relation $\overline{P}(A|B) = 1 - \underline{P}(A^c|B)$. A different proof of essentially the same algorithm for checking the AUL condition is given in [27, sec. 3]. All these algorithms are indirect and rely on properties of precise conditional probabilities.

We use the term 'uniform loss' rather than 'sure loss' because violations of AUL cannot necessarily be exploited to produce a sure loss. Incurring uniform loss means that there is a positive linear combination of strictly acceptable bets which, *if any of them takes place*, must produce a net loss. That is, there is a net loss if $S(\boldsymbol{\lambda})$ occurs, but not otherwise. To illustrate that, consider the two assessments $\underline{P}(A|B) = 0.6$ and $\underline{P}(A^c|B) = 0.5$, where B is not certain to occur. By setting $\lambda_1 = \lambda_2 = 1$, so that $S(\boldsymbol{\lambda}) = B$, it is easily verified that these assessments incur uniform loss. However, there is no way to exploit the two assessments to produce a sure loss, because any bets based on the assessments must be conditional on B. If B fails to occur then all bets will be called off and nothing is lost or gained. Nevertheless the two assessments are obviously inconsistent.

We say that the assessments *avoid sure loss* (ASL) if $\sup\,[\sum_{i=1}^{k} \lambda_i G_i|\Omega] \geq 0$ whenever $\boldsymbol{\lambda} \succ \mathbf{0}$, or equivalently, there is no $\boldsymbol{\lambda} \succ \mathbf{0}$ such that $\sum_{i=1}^{k} \lambda_i G_i \leq -1$. This condition can be checked by solving a linear program. Clearly AUL implies ASL. The example in the preceding paragraph shows that AUL is a stronger condition than ASL, and it also shows that ASL is too weak to characterize 'consistency' of the assessments. A sufficient condition for ASL is that $\bigcup_{i=1}^{k} B_i \subset \Omega$ (since $\sum_{i=1}^{k} \lambda_i G_i = 0$ outside $\bigcup_{i=1}^{k} B_i$), and yet the assessments may be inconsistent in such cases. Conditions will be given later under which ASL is equivalent to AUL.

Many of the methods that have been proposed for dealing with conditional probability assessments implicitly use the ASL condition as a definition of consistency. Consequently they are unable to detect the type of inconsistency illustrated above, where the assessments satisfy ASL but not AUL. For further discussion of the difference between ASL and AUL, and examples which show that ASL is too weak, see [31].

In general, we cannot simplify the AUL condition (1) by setting $\varepsilon = 0$. That is illustrated by the following example.

Example 4 *Suppose that there is a single assessment* $\underline{P}(A_1|B_1) = 1$*, which is equivalent to the precise probability assessment* $P(A_1|B_1) = 1$*. This assessment avoids uniform loss, but taking* $\lambda_1 = 1$*, we see that* $\lambda_1 G_1 = B_1(A_1 - 1) \leq 0$*. Thus condition (1) is satisfied with* $\varepsilon = 0$*.*

In fact, whenever the assessments imply that some event B has probability zero, there is $\boldsymbol{\lambda} \succ \mathbf{0}$ such that $\sum_{i=1}^{k} \lambda_i G_i \leq -B \leq 0$. More generally, if the assessments imply that any non-trivial conditional probability $P(A|B)$ is precisely determined, i.e., $\underline{P}(A|B) = \overline{P}(A|B)$, then condition (1) is satisfied for $\varepsilon = 0$. This shows that the condition obtained by setting $\varepsilon = 0$ in (1) is not necessary for consistency.

2.2 Checking coherence of precise probability assessments

Suppose that all the probability assessments are precise, $P(A_i|B_i) = c_i$ for $i = 1, 2, \dots, m$. We say that the assessments are *de Finetti-coherent*, or *dF-coherent*, when

$$\sup\,[\sum_{i=1}^{m} \lambda_i G_i | S(\boldsymbol{\lambda})] \geq 0 \quad \text{for all real numbers} \quad \lambda_1, \dots, \lambda_m. \tag{4}$$

(See [19] for an equivalent definition.) This condition is identical to the necessary and sufficient condition (2) for AUL of conditional lower probabilities, except that now the coefficients λ_i are not required to be non-negative but are allowed to take any real values.

When all the probability assessments are precise, dF-coherence is equivalent to AUL. To see that, recall that any assessment of a precise conditional probability $P(A|B) = c$ is equivalent to two assessments of conditional lower probabilities, $\underline{P}(A|B) = c$ and $\underline{P}(A^c|B) = 1 - \overline{P}(A|B) = 1 - c$.

Lemma 2 *The precise probability assessments $P(A_i|B_i) = c_i$ $(i = 1, \dots, m)$ are dF-coherent if and only if the corresponding assessments of conditional lower probabilities, $\underline{P}(A_i|B_i) = c_i$ and $\underline{P}(A_i^c|B_i) = 1 - c_i$ $(i = 1, \dots, m)$, avoid uniform loss.*

PROOF. The assessments $\underline{P}(A_i|B_i) = c_i$ and $\underline{P}(A_i^c|B_i) = 1 - c_i$ have marginal gambles $G_i = B_i[A_i - c_i]$ and $G_i' = B_i[A_i^c - (1 - c_i)] = B_i[1 - A_i - 1 + c_i] = -B_i[A_i - c_i] = -G_i$. Including both G_i and $-G_i$ in (2), where the coefficients λ_j are required to be non-negative, is equivalent to allowing λ_j to take any real value. It follows from Lemma 1 that AUL is equivalent to dF-coherence. $\diamondsuit$

dF-coherence of precise probability assessments can therefore be verified by checking AUL.

2.3 Algorithms for checking consistency

To check the AUL condition, we need to determine whether the system of inequalities (1) has a solution $(\boldsymbol{\lambda}, \varepsilon)$ with $\varepsilon > 0$. Because (1) becomes weaker as ε decreases, it has a solution if and only if it has a solution for sufficiently small values of ε. That suggests the following algorithm for checking AUL.

Algorithm 1. Fix a very small positive value of ε, and check whether the system of linear inequalities (1) has a solution $\boldsymbol{\lambda}$. This involves a single linear program.

In practice, Algorithm 1 will almost always give the correct answer to the consistency problem, provided that ε is chosen to be sufficiently small. Generally ε should be chosen to be larger than the rounding error in computations, but also much smaller than the rounding in the assessments. The latter condition is generally easy to satisfy because probabilities are rarely specified to more than four decimal places. For example, ε can be taken to be 5 or 10 times the rounding error of computations. We have successfully used values of ε that range from 10^{-7} to 10^{-10} in standard optimization programs such as Lingo and the simplex package of Maple V, in computations with at least 10 floating-point digits. As shown below, however small ε is chosen to be, examples can be constructed in which Algorithm 1 fails to detect an inconsistency in the assessments. But in such examples, if ε is chosen to be not much larger than the rounding error of computations, the degree of inconsistency is so small as to be almost indistinguishable from rounding error and it will be difficult for any other algorithm to detect the inconsistency.

Using Lemma 2, Algorithm 1 can be used to check dF-coherence of *precise* conditional probability assessments. In the case of precise assessments, the small perturbation ε that is involved in Algorithm 1 is useful also from the point of view of numerical analysis. Precise assessments that are dF-coherent are *almost incoherent*, in the sense that an arbitrarily small perturbation can make them incoherent. Such perturbations can be introduced in computations through rounding errors, and this can cause serious problems for many algorithms. If the value of ε used in Algorithm 1 is larger than the rounding error of computations, it protects against this kind of numerical instability.

Whatever positive value of ε is used in Algorithm 1, examples can be constructed in which the system (1) has no solution $\boldsymbol{\lambda}$ but the assessments nevertheless incur uniform loss. In other words, when AUL fails, the supremum value of ε for which (1) has solutions may be arbitrarily close to zero. That can be seen in the following example.

Example 5 *Given any small positive value of ε, suppose that two assessments are made: $\underline{P}(A|B) = c$ and $\underline{P}(A^c|B) = 1 - c + \varepsilon$, where $0 < \varepsilon \le c \le 1$. Because $\underline{P}(A|B) + \underline{P}(A^c|B) > 1$, these assessments incur uniform loss, as can be easily proved by applying the definition. However, it can be seen that, for this value of ε, system (1) has no solution $\boldsymbol{\lambda}$.*

The next example shows that the supremum value of ε for which (1) has solutions may be very close to zero even when the assessments are specified only 'roughly', to just a few decimal places.

Example 6 *Suppose that six assessments of conditional upper and lower probabilities are made:* $\underline{P}(A) = 1/16$, $\underline{P}(C^c \cap D) = 0.01$, $\overline{P}(F) = 0.51$, $\overline{P}(A|B) = 0.49$, $\overline{P}(B|C) = 0.51$, *and* $\overline{P}(D|F) = 0.51$, *where the events are related by* $\emptyset \subset A \subset B \subset C \subset D \subset F \subset \Omega$ *and* $\subset$ *denotes strict inclusion. If we apply Algorithm 1 with* ε *greater than or equal to* 10^{-7}*, we find that the system (1) has no solution* $\boldsymbol{\lambda}$*. This might suggest that the assessments AUL. When* ε *is less than or equal to* 10^{-8}*, however, (1) does have a solution* $\boldsymbol{\lambda}$*, which shows that the assessments incur uniform loss. (In fact, they incur sure loss.) This behaviour could be detected if, after using Algorithm 1 with* ε *greater than or equal to* 10^{-7}*, we also checked for solutions of (1) with* $\varepsilon = 0$*: the discrepancy between the two answers warns us that further investigation is needed, using either smaller positive values of* ε *or Algorithm 2.*

We emphasize that the two preceding examples are artificial and were constructed to produce pathological behaviour of Algorithm 1. In the more realistic examples we have studied, Algorithm 1 has always produced the correct answer, even with only moderately small values of ε.

In cases where we need an absolute guarantee of consistency, Algorithm 1 can be extended in two ways. Note first that if Algorithm 1 has a solution $\boldsymbol{\lambda}$ then the assessments are definitely not consistent (they incur uniform loss). If Algorithm 1 has no solution $\boldsymbol{\lambda}$, the next step is to check whether the system (1) with $\varepsilon = 0$ has a solution $\boldsymbol{\lambda}$. This involves a second linear programming problem. If this second system has no solution then the assessments are definitely consistent (they satisfy AUL). In the remaining case, where the first system (with $\varepsilon > 0$) has no solution but the second system (with $\varepsilon = 0$) has a solution, we can either try smaller positive values of ε or use the following iterative algorithm, which is guaranteed to work in all cases.

Algorithm 2

(a) Set $I = \{1, 2, \ldots, k\}$.

(b)

$$\begin{aligned} &\text{Maximise} && \sum_{i \in I} \tau_i \\ &\text{subject to} && \boldsymbol{\lambda} \geq \mathbf{0}, \quad 0 \leq \tau_i \leq 1 \ (i \in I) \\ &\text{and} && \sum_{i \in I} \lambda_i G_i + \sum_{i \in I} \tau_i B_i \leq 0. \end{aligned} \tag{5}$$

(c) If $\tau_i = 1$ for all $i \in I$ then the assessments incur uniform loss. Otherwise, replace I by the subset $\{i \in I : \tau_i = 1\}$. If I is empty then the assessments AUL. Otherwise, return to (b).

The consistency of algorithm 2 relies upon the following three basic properties:

(i) The assessments incur uniform loss if and only if there exist $\boldsymbol{\lambda} \succ \mathbf{0}$ and $\tau_i \in \{0,1\}$ $(i = 1,\ldots,k)$ such that $\sum_{i=1}^{k} \lambda_i G_i + \sum_{i=1}^{k} \tau_i B_i \leq 0$, and $\tau_i = 1$ whenever $\lambda_i > 0$ (alternatively, iff there exist $\boldsymbol{\lambda} \succ \mathbf{0}$ and a non-empty $J \subset \{1,\ldots,k\}$ such that $\sum_{i\in J} \lambda_i G_i + \sum_{i\in J} B_i \leq 0$).
In fact, if the assessments incur uniform loss, define $m = min\,\{\lambda_i : \lambda_i > 0\}$ and divide all terms in (1) by εm, getting $\sum_{i=1}^{k}(\lambda_i/\varepsilon m)G_i+\sum_{i=1}^{k}(\lambda_i/m)B_i \leq 0$; hence, putting $\boldsymbol{\lambda'} = \boldsymbol{\lambda}/(\varepsilon m)$, $\tau_i = 1$ if $\lambda_i > 0$, $\tau_i = 0$ otherwise, and since $(\lambda_i/m)B_i \geq \tau_i B_i$, obtain $\sum_{i=1}^{k} \lambda'_i G_i + \sum_{i=1}^{k} \tau_i B_i \leq 0$. Conversely, if $\sum_{i=1}^{k} \lambda_i G_i + \sum_{i=1}^{k} \tau_i B_i \leq 0$ for $\boldsymbol{\lambda} \succ \mathbf{0}$ and $\tau_i \in \{0,1\}$ $(i = 1,\ldots,k)$ are such that $\tau_i = 1$ whenever $\lambda_i > 0$, then it is $\sum_{i:\lambda_i>0} \lambda_i\,(G_i + (1/\lambda_i)B_i) \leq 0$, and (1) is easily obtained from this putting $\varepsilon = min\,\{1/\lambda_i : \lambda_i > 0\}$.

(ii) If τ is a solution in (b), it must have $\tau_i = 0$ or $\tau_i = 1$ for all $i \in I$.
To prove (ii), suppose that $(\boldsymbol{\lambda'}, \boldsymbol{\tau'})$ satisfies the constraints (5), but that $0 < \tau'_i < 1$ for some $i \in I$. Then $(\boldsymbol{\lambda'}, \boldsymbol{\tau'})$ cannot be a solution in (b), because putting $s = min\,\{\tau'_i : \tau'_i > 0\}$, $\boldsymbol{\lambda} = \boldsymbol{\lambda'}/s$, $\tau_i = 0$ if $\tau'_i = 0$, $\tau_i = 1$ if $\tau'_i > 0$ $(i \in I)$ then $\sum_{i=1}^{k} \tau_i > \sum_{i=1}^{k} \tau'_i$, and $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ satisfies the constraints (5), since $\sum_{i\in I} \lambda_i G_i + \sum_{i\in I} \tau_i B_i = (1/s)\,(\sum_{i\in I} \lambda'_i G_i + \sum_{i\in I} s\tau_i B_i) \leq (1/s)\,(\sum_{i\in I} \lambda'_i G_i + \sum_{i\in I} \tau'_i B_i) \leq 0$.

(iii) Let $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ be an optimal solution in (b) and let $(\boldsymbol{\lambda'}, \boldsymbol{\tau'})$ satisfy (5). Then $\tau_i = 1$ whenever $\tau'_i > 0$.
To prove (iii), note that also $((\boldsymbol{\lambda} + \boldsymbol{\lambda'})/2, (\boldsymbol{\tau} + \boldsymbol{\tau'})/2)$ satisfies (5). It ensues from the proof of (ii) that there exists a solution $(\boldsymbol{\lambda''}, \boldsymbol{\tau''})$ in (b) such that $\tau''_i \in \{0,1\}$, $\tau''_i = 1$ iff $(\tau_i + \tau'_i)/2 > 0$ iff $\tau_i > 0$ or $\tau'_i > 0$ $(i \in I)$. If it were $\tau'_i > 0$ and $\tau_i = 0$ for some i, this would then imply $\sum_{i\in I} \tau''_i > \sum_{i\in I} \tau_i$, contradicting the optimality of $(\boldsymbol{\lambda}, \boldsymbol{\tau})$.

Algorithm 2 checks whether the assessments incur uniform loss by using (i) and by successively reducing the set I. Consider the set of indexes i corresponding to positive λ_i in at least one $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ making the assessments incur uniform loss by (i). Reducing I produces progressively more stringent bounds on this set. In fact, at each iteration of step (b), I is replaced by $I' = \{i \in I : \tau_i = 1\}$ and, by (ii) and (iii), step (b) produces a solution $\boldsymbol{\tau}$ with as many components as possible equal to 1. This implies that I' is the unique largest subset of I with the property that $\sum_{i\in I} \lambda_i G_i + \sum_{i\in I'} B_i \leq 0$. It implies also that a necessary condition for $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ to make the assessments incur uniform loss according to (i) is that $\lambda_i = 0$ for each $i \in I - I'$. For instance, if at the first iteration $I' \neq I = \{1,\ldots,k\}$, then τ_i (and hence λ_i) must be zero for $i \in I - I'$ in any such $(\boldsymbol{\lambda}, \boldsymbol{\tau})$. Consequently, $\sum_{i\in I'} \lambda_i G_i + \sum_{i\in I'} B_i \leq 0$ must hold for these $(\boldsymbol{\lambda}, \boldsymbol{\tau})$, and in fact, after replacing I by I', this is precisely inequality (5) in step (b) of the second iteration. The same argument applies to the next iterations. When the algorithm terminates, $\tau_i = 1$ for all $i \in I$ (where possibly I is empty), so that $I' = I$. Therefore $\sum_{i\in I} \lambda_i G_i + \sum_{i\in I} B_i \leq 0$. By (i), if I is non-empty then the assessments incur uniform loss, whereas if I is empty they AUL, because any $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ which could make the assessments incur uniform loss should be such that $\lambda_i = 0$ for $i \in \{1, \ldots, k\} - I' = \{1, \ldots, k\}$, that is $\boldsymbol{\lambda} = \mathbf{0}$. But this is not consistent with $\boldsymbol{\lambda} \succ \mathbf{0}$ in (i).

Reducing the set I in step (c) is equivalent to reducing the set of assessments. Since the number of assessments must be reduced at each stage, the algorithm is guaranteed to give an answer in at most k steps, i.e., after solving at most k linear programs of the form (5). If the assessments incur sure loss then Algorithm 2 requires only one step, and similarly if there is no set B_j such that $\sum_{i=1}^{k} \lambda_i G_i + B_j \leq 0$. If the assessments incur uniform loss, the algorithm reveals (at the last step) the largest subset I which does so.

As explained in the preceding subsection, dF-coherence of m precise probability assessments can be verified by checking AUL. Therefore, Algorithms 1 and 2 can be used also to verify dF-coherence. Algorithm 1 involves a single linear program with $2m$ variables, while Algorithm 2 requires solving at most $2m$ linear programs.

A sufficient condition for AUL to be equivalent to ASL is that all the conditioning events B_i have probabilities that are bounded away from zero (formally, the *natural extensions* $\underline{E}(B_i)$, defined in Section 3, are non-zero for $i = 1, \ldots, k$). This holds, for instance, if all the assessments are of unconditional probabilities, as in Example 1. In this case, AUL can be checked by solving a single linear program: the assessments AUL if and only if there is no $\boldsymbol{\lambda} \succ \mathbf{0}$ such that $\sum_{i=1}^{k} \lambda_i G_i \leq -1$. For the assessments in Example 1, this system has no solution $\boldsymbol{\lambda} \succ \mathbf{0}$, and therefore the assessments are consistent. This can also be verified by applying Algorithm 1 or Algorithm 2. More generally, if $\underline{E}(B_i | \bigcup_{i=1}^{k} B_i) > 0$ for all $i = 1, 2, \ldots, k$, the assessments AUL if and only if there is no $\boldsymbol{\lambda} \succ \mathbf{0}$ such that $\sum_{i=1}^{k} \lambda_i G_i + \bigcup_{i=1}^{k} B_i \leq 0$, which can be checked by solving a single linear program.

Example 7 *Consider the football example, with just the* 10 *upper and lower probability assessments given in Table 2, plus one further assessment that* $\overline{P}(A|B) = 0.5$, *where A denotes the event that team X wins the tournament, and B is the event that X finishes equal first on points and also Y loses against Z. Here B is the union of the two outcomes (L, W, L) and (W, D, L). Using this fact, it is easy to see that the three assessments* $\overline{P}(A|B) = 0.5$, $\underline{P}(A|(L, W, L)) = 0.6$ *and* $\underline{P}(A|(W, D, L)) = 0.65$ *are mutually inconsistent, and therefore the system of* 11 *upper and lower probabilities incurs uniform loss. This can be checked using Algorithm 1, with any value of ε smaller than 0.05. It is also easy to see that this system avoids sure loss, because the union of the* 11 *conditioning events is not certain to occur. Again, the ASL condition is too weak to detect the inconsistency in the assessments.*

Now suppose that, in addition to these 11 *imprecise probability assessments,* *we consider the 27 precise assessments in Table 1. In this case the assessments incur sure loss: combining the three assessments that incur uniform loss with* $P(L, W, L) = 0.04$ *produces a sure loss. The precise probability assessments tell us that the event B has positive probability, and this is enough to turn the uniform loss (on the set* $S(\boldsymbol{\lambda}) = B$*) into a sure loss. This illustrates that the difference between AUL and ASL is essentially concerned with whether or not the union of conditioning events,* $S(\boldsymbol{\lambda})$*, has lower probability zero.*

Computations for this and the later examples were done using both the optimization program Lingo and Maple V (release 5.1). We wrote a Maple program that constructs an appropriate space Ω, sets up the relevant LP problems with varying values of ε, and solves the LP problems.

2.4 Methods for modifying inconsistent assessments

Suppose that the k assessments $\underline{P}(A_i|B_i) = c_i$ $(i = 1, 2, \ldots, k)$ are inconsistent, in the sense that they incur uniform loss. Then it is useful to have an automatic method for modifying the assessments in some kind of minimal way, so that the modified assessments AUL. One method is suggested by the definition of AUL (1), in which the positive parameter ε can be regarded as a constant reduction to each of the assessments. The method is to reduce each assessment by a constant amount ε_1, where ε_1 is defined to be the minimum value of ε such that the reduced assessments $\underline{P}(A_i|B_i) = c_i - \varepsilon$ $(i = 1, 2, \ldots, k)$ AUL. (It can be seen from (1) that the infimum such ε does achieve AUL. A similar method for modifying unconditional probability assessments was suggested in [17].) It can be shown that ε_1 is the maximum value in the nonlinear program: maximize ε subject to the system of constraints (1).

If $c_i < \varepsilon_1$ then the modified assessment $c_i - \varepsilon_1$ is negative, so that $G_i + \varepsilon_1 B_i \geq 0$ and the modified assessment does not contribute to the uniform loss. Hence we can make a slightly different modification, to $\underline{P}(A_i|B_i) = \max\{c_i - \varepsilon_1, 0\}$, without changing AUL.

Another approach, which seems a little more natural, is to make a multiplicative modification instead of an additive one. In that case we replace $\underline{P}(A_i|B_i) = c_i$ by $\underline{P}(A_i|B_i) = (1 - \varepsilon_2)c_i$ $(i = 1, 2, \ldots, k)$, where ε_2 is minimal such that the modified assessments AUL. The new assessments correspond to forming an ε_2-*contamination neighbourhood* of the original assessments. Again, ε_2 can be computed as the maximum value of a nonlinear program: maximize ε subject to $\sum_{i=1}^{k} \lambda_i(G_i + \varepsilon c_i B_i) \leq 0$ and $\boldsymbol{\lambda} \succ \mathbf{0}$.

In both approaches, it suffices to modify only those assessments $\underline{P}(A_i|B_i)$ for $i \in I$, where I is the largest subset of $\{1, 2, \ldots, k\}$ which incurs uniform loss. This is the set that is computed in Algorithm 2.

The values ε_1 and ε_2 can be regarded as measures of the *degree of inconsistency* of the assessments. They could be defined in the same way in the case where the assessments AUL: in that case both values are non-positive, and $-\varepsilon_1$ and $-\varepsilon_2$ measure the *degree of consistency* of the assessments.

As a numerical example, consider the first part of Example 7, involving 11 assessments. The minimal reductions which achieve AUL are found to be $\varepsilon_1 = 0.05$ and $\varepsilon_2 = 1/11$. The modified assessments that are produced by the two methods are quite similar.

3 The inference problem

3.1 Natural extension

Given the assessments of conditional lower probabilities $\underline{P}(A_i|B_i) = c_i$ $(i = 1, 2, \ldots, k)$, we make inferences by calculating further conditional lower and upper probabilities, which will be denoted by $\underline{E}(A|B)$ and $\overline{E}(A|B)$. The symbol E stands for 'Extension'. Here $A|B$ need not be a new conditional event: it may agree with one of the conditional events $A_i|B_i$ for which assessments are made. We always assume that the conditioning event B is non-null. Again the upper probabilities are determined by lower probabilities through the conjugacy relation $\overline{E}(A|B) = 1 - \underline{E}(A^c|B)$, so we concentrate on the lower probability $\underline{E}(A|B)$.

The quantity $\underline{E}(A|B)$ represents what can be inferred from the assessments concerning the conditional lower probability $\underline{P}(A|B)$. Recall that $\underline{P}(A|B)$ is interpreted as a supremum (marginally acceptable) rate for betting on A conditional on B. We therefore define the *natural extension* $\underline{E}(A|B)$ to be the supremum rate for betting on A conditional on B that can be constructed from the assessments through positive linear combinations of strictly acceptable bets. See [31–33] for further discussion of this idea.

Formally, the *natural extension* $\underline{E}(A|B)$ is defined to be the supremum value of μ for which there are $\varepsilon > 0$ and $\boldsymbol{\lambda} \geq \mathbf{0}$ such that

$$B(A - \mu) \geq \sum_{i=1}^{k} \lambda_i (G_i + \varepsilon B_i). \tag{6}$$

The supremum may or may not be achieved by some μ. Here $B(A - \mu)$ represents the net reward from a bet on A conditional on B at betting rate μ, and $\sum_{i=1}^{k} \lambda_i (G_i + \varepsilon B_i)$ is the net reward from a positive linear combination of bets on A_i conditional on B_i at rates $\underline{P}(A_i|B_i) - \varepsilon$. The positive adjustment

ε is needed to ensure that these bets are strictly acceptable. Note that (6) directly characterizes the natural extension in terms of linear combinations of the assessments $\underline{P}(A_i|B_i)$.

There are always values of $(\mu, \boldsymbol{\lambda}, \varepsilon)$ which satisfy the constraints in the definition of natural extension. For example, $\mu = 0, \boldsymbol{\lambda} = \mathbf{0}$ and any positive value of ε satisfy the constraints, and this shows that always $\underline{E}(A|B) \geq 0$.

To define the natural extension, it is not necessary that the assessments AUL. However, assessments that incur uniform loss may produce bad inferences, in the sense that $\underline{E}(A|B)$ may be infinite: this happens if and only if $B \subseteq S$, where S is the largest set $S(\boldsymbol{\lambda})$ on which the assessments incur uniform loss.

An alternative characterization of the natural extension is given in the following lemma, which is analogous to Lemma 1.

Lemma 3 *The natural extension* $\underline{E}(A|B)$ *is the supremum value of* μ *for which there is* $\boldsymbol{\lambda} \geq \mathbf{0}$ *such that*

$$\sup \, [\sum_{i=1}^{k} \lambda_i G_i - B(A-\mu) | S(\boldsymbol{\lambda}) \cup B] < 0. \tag{7}$$

Here the supremum is not achieved: the set of μ*-values which satisfy these constraints is the open interval* $(-\infty, \underline{E}(A|B))$.

PROOF. First suppose that $(\mu, \boldsymbol{\lambda}, \varepsilon)$ satisfy the system of inequalities (6), and $\eta < \mu$. Also let $S(\boldsymbol{\lambda}) = \bigcup\{B_i : \lambda_i > 0\}$, and $\tau_1 = \min \{\lambda_i : \lambda_i > 0\}$ or $\tau_1 = 1$ if $\boldsymbol{\lambda} = \mathbf{0}$. Then, using (3), $\sum_{i=1}^k \lambda_i G_i - B(A-\eta) \leq -\varepsilon \sum_{i=1}^k \lambda_i B_i - B(\mu-\eta) \leq -\varepsilon\tau_1 S(\boldsymbol{\lambda}) - B(\mu-\eta) \leq -\delta[S(\boldsymbol{\lambda}) \cup B]$, where $\delta = \min \{\varepsilon\tau_1, \mu-\eta\} > 0$. (This holds also if $\boldsymbol{\lambda} = \mathbf{0}$ since then $S(\boldsymbol{\lambda}) = \emptyset$.) Thus $\sup \, [\sum_{i=1}^k \lambda_i G_i - B(A-\eta)|S(\boldsymbol{\lambda}) \cup B] \leq -\delta < 0$, so that $(\eta, \boldsymbol{\lambda})$ satisfy (7), for every $\eta < \mu$. It follows that the quantity defined in the lemma is at least as large as $\underline{E}(A|B)$.

For the reverse inequality, suppose that μ and $\boldsymbol{\lambda} \geq \mathbf{0}$ satisfy (7). Then there is $\delta > 0$ such that $\sum_{i=1}^k \lambda_i G_i - B(A-\mu) \leq -\delta[S(\boldsymbol{\lambda}) \cup B]$. (Here all the terms are zero outside $S(\boldsymbol{\lambda}) \cup B$.) If $\boldsymbol{\lambda} \succ \mathbf{0}$ then $\tau_2 = \sum_{i=1}^k \lambda_i > 0$, so $\varepsilon = \delta/\tau_2 > 0$, and from (3) $\varepsilon \sum_{i=1}^k \lambda_i B_i \leq \varepsilon\tau_2 S(\boldsymbol{\lambda}) = \delta S(\boldsymbol{\lambda}) \leq \delta[S(\boldsymbol{\lambda}) \cup B]$. (This also holds if $\boldsymbol{\lambda} = \mathbf{0}$ since then $\sum_{i=1}^k \lambda_i B_i = 0$.) It follows that $\sum_{i=1}^k \lambda_i G_i - B(A-\mu) \leq -\varepsilon \sum_{i=1}^k \lambda_i B_i$, which shows that $(\mu, \boldsymbol{\lambda}, \varepsilon)$ satisfy the system (6) that defines $\underline{E}(A|B)$. This proves that $\underline{E}(A|B)$ is at least as large as the quantity defined in the lemma. $\diamondsuit$

3.2 Properties of natural extension

Here we outline the most important properties of natural extension; proofs of these results are in [34].

(a) For all events A and B, $\underline{E}(A|B) \geq 0$. (This was proved in subsection 3.1.)
(b) If the assessments AUL then $\underline{E}(A|B) \leq 1$ for all events A and B. If the assessments incur uniform loss then there is at least one assessment $\underline{P}(A_i|B_i)$ such that $\underline{E}(A_i|B_i) = \infty$. This shows that AUL can be characterized in terms of natural extension: the assessments AUL if and only if $\underline{E}(A_i|B_i) \leq 1$ for $i = 1, \ldots, k$.
(c) $\underline{E}(A_i|B_i) \geq \underline{P}(A_i|B_i)$ for $i = 1, \ldots, k$.
(d) Say that a finite collection of conditional lower probabilities is *coherent* if each conditional lower probability agrees with the corresponding natural extension of the collection. (This is equivalent to the definitions of Williams [37] and Walley [31,33].) For example, the assessments are coherent if and only if $\underline{E}(A_i|B_i) = \underline{P}(A_i|B_i)$ for $i = 1, \ldots, k$. For precise probabilities, coherence is equivalent to dF-coherence and to AUL, but for imprecise probabilities coherence is stronger than AUL. If the assessments AUL and their natural extensions $\underline{E}(A|B)$ are defined for any finite collection of conditional events, then these natural extensions are coherent.
(e) If the assessments are coherent then $\underline{E}(A|B)$ is the minimal value of $\underline{P}(A|B)$ that is coherent with the assessments, i.e., their *minimal coherent extension*. Thus any coherent collection of conditional lower probabilities can be coherently extended to any other conditional events, and $\underline{E}(A|B)$ is the minimal coherent extension.
(f) If the assessments AUL then the natural extension is the lower envelope of all collections of precise conditional probabilities that dominate the assessments and AUL. (Recall that, for precise probabilities, AUL is equivalent to dF-coherence.) Formally, let Γ index the non-empty set of all collections of precise conditional probabilities $(P_\gamma(A|B), P_\gamma(A_1|B_1), \ldots, P_\gamma(A_k|B_k))$ which satisfy AUL and $P_\gamma(A_i|B_i) \geq \underline{P}(A_i|B_i)$ for $i = 1, \ldots, k$. Then $\underline{E}(A|B) = \min\{P_\gamma(A|B) : \gamma \in \Gamma\}$. This property gives an indirect characterization of natural extension, in terms of a set of precise conditional probabilities.

3.3 Making inferences from precise probability assessments

As explained earlier, m assessments of precise conditional probabilities can be replaced by $2m$ equivalent assessments of conditional lower probabilities. Suppose that the assessments are dF-coherent, which is equivalent to AUL. To calculate what the assessments imply about a further conditional probability $P(A|B)$, we compute the natural extensions $\underline{E}(A|B)$ and $\underline{E}(A^c|B)$. The next result shows that these two natural extensions give a complete solution to the problem of making inferences about $P(A|B)$. Thus natural extension solves the Bayesian problem of inference.

Lemma 4 *Suppose that all the conditional probability assessments are precise and dF-coherent. Let* $\overline{E}(A|B) = 1 - \underline{E}(A^c|B)$. *Then the range of values* $P(A|B)$ *that are dF-coherent with the assessments is the closed interval* $[\underline{E}(A|B), \overline{E}(A|B)]$.

PROOF. Because the assessments are precise, any precise conditional probabilities that dominate the assessments must coincide with them. Hence the set $\{P_\gamma(A|B) : \gamma \in \Gamma\}$ in 3.2(f) is the set of all values $P_\gamma(A|B)$ that are dF-coherent with the assessments. By result 3.2(f), this set has minimum value $\underline{E}(A|B)$, and similarly its maximum is $\overline{E}(A|B)$. By applying the definition of dF-coherence, as in [29, Lemma 5.2.1], it follows that the range of dF-coherent values is the closed interval $[\underline{E}(A|B), \overline{E}(A|B)]$. $\diamondsuit$

3.4 Algorithms for computing natural extension

Suppose that the assessments AUL. (In practical applications, this should be verified first by using the algorithms in subsection 2.3.) To compute the natural extension $\underline{E}(A|B)$ from the definition (6), we must solve the following problem:

$$\begin{aligned} &\text{maximise} && \mu \\ &\text{subject to} && \varepsilon > 0, \quad \boldsymbol{\lambda} \geq \mathbf{0} \\ &\text{and} && \sum_{i=1}^{k} \lambda_i (G_i + \varepsilon B_i) + \mu B \leq AB. \end{aligned} \tag{8}$$

This is a parametric linear programming problem with scalar parameter ε. Problems of this type, in which the parameter appears in the matrix of linear constraints, are not easily solvable in general [11].

We now discuss some practical methods to solve this problem. The simplest method, which will work in almost all practical applications, is to fix a sufficiently small value of ε and to solve the resulting linear program (8).

Algorithm 3. Fix a very small positive value of ε and solve the linear program (8). If ε is sufficiently small, the maximized value of μ agrees with $\underline{E}(A|B)$ to a very close approximation.

The advice given in subsection 2.3 concerning the choice of ε applies also to Algorithm 3. For fixed ε, the linear program (8) involves $k+1$ variables $(\mu, \lambda_1, \ldots, \lambda_k)$ and at most $k + ||\Omega||$ linear constraints, where $||\Omega||$ denotes the cardinality of the possibility space Ω. Because $AB \geq 0$, applying the simplex method to solve (8) does not require us to run the two-phase method (or another method) for initialization, since a starting point is found at once by adding the slack variables.

Again we emphasize that Algorithm 3 will give the correct value of $\underline{E}(A|B)$, to a degree of approximation that is much better than that of the assessments, in almost all practical problems. The maximum value produced by Algorithm 3 is guaranteed to be a lower bound for $\underline{E}(A|B)$, but again we can construct examples in which it may not be a good approximation to the correct value. As in Section 2, we can extend the simple (one-step) algorithm to cope with the rare cases in which it may fail.

Let $\mu^*(\varepsilon)$ denote the maximum value of μ when (8) is solved for a fixed value of ε. Assuming that the assessments AUL, for any fixed positive ε the supremum $\mu^*(\varepsilon)$ is achieved by some $(\mu, \boldsymbol{\lambda})$, because the feasible region is closed and non-empty (the origin is always feasible) and by property 3.2(b) the objective function μ is bounded from above by 1. A crucial point is that, because the strength of the constraints in (8) becomes weaker when ε decreases, $\mu^*(\varepsilon)$ is a non-increasing function of ε. It follows that $\underline{E}(A|B)$ is the limit of the maxima $\mu^*(\varepsilon)$ as $\varepsilon \to 0$ from above. It also follows that this limit is no larger than $\mu^*(0)$. Also, by a well known result from parametric linear programming [7,10], μ^* can have only finitely many points of discontinuity. So $\mu^*(\varepsilon)$ can approximate $\underline{E}(A|B)$ arbitrarily closely by taking ε to be sufficiently small.

By the preceding results, $\mu^*(0) \geq \underline{E}(A|B) \geq \mu^*(\varepsilon)$ for every $\varepsilon > 0$. Thus we can find upper and lower bounds for $\underline{E}(A|B)$ by solving two linear programming problems of the form (8), with the values $\varepsilon = 0$ (to give the upper bound) and $\varepsilon = \varepsilon_1$ (to give the lower bound), where ε_1 is the small positive value used in Algorithm 3. If the difference between the upper and lower bounds is negligible then the upper bound can be adopted as the solution. It appears that, in many problems, the function μ^* is right-continuous at 0, in which case the upper bound is the exact solution. In particular, when the natural extension is unconditional, so that $B = \Omega$, the exact value is $\underline{E}(A) = \mu^*(0)$. More generally, whenever the probability of B is bounded away from zero, so that $\underline{E}(B) > 0$, we have $\underline{E}(A|B) = \mu^*(0)$. More general conditions under which $\underline{E}(A|B) = \mu^*(0)$ are given in section 3.6.

Some caution is needed when computing $\mu^*(0)$, because the corresponding linear programming problem is not always bounded and it is possible that $\mu^*(0) = +\infty$. For example, given the single assessment $\overline{P}(B) = 0$ and an event A that is logically independent of B, it can be verified from (8) that the natural extension is $\underline{E}(A|B) = 0$, but that $\mu^*(0) = +\infty$. A necessary and sufficient condition for $\mu^*(0) = +\infty$ is that $\overline{E}(B|\bigcup_{i=1}^k B_i \cup B) = 0$; this will be proved in subsection 3.7.

If the difference between the upper and lower bounds $\mu^*(0)$ and $\mu^*(\varepsilon)$ is non-negligible, so that μ^* appears to be discontinuous at 0, Algorithm 3 (with a sufficiently small positive ε) still gives an approximate solution. In rare cases, Algorithm 3 may yield a poor approximation to $\underline{E}(A|B)$ because ε is not sufficiently small; this can happen when μ^* has a discontinuity at some value that lies between 0 and ε. The next example, which is a modification of Example 6, shows that this can happen even when ε is much smaller than the 'roughness' of the assessments.

Example 8 *Suppose that six assessments of conditional upper and lower probabilities are made:* $\underline{P}(A\cap T^c|H\cap T^c) = 1/16$, $\underline{P}(C^c\cap D) = 0.01$, $\overline{P}(F) = 0.51$, $\overline{P}(A|B) = 0.49$, $\overline{P}(B|C) = 0.51$, *and* $\overline{P}(D|F) = 0.51$, *where the events are related by* $\emptyset \subset T \subset A \subset B \subset C \subset D \subset F \subset H \subset \Omega$ *and* $\subset$ *denotes strict inclusion. We can verify that these assessments AUL, by checking that the system (1) has no solution* $\boldsymbol{\lambda}$ *when* $\varepsilon = 0$.

Suppose that we wish to compute the natural extension $\underline{E}(H^c|H^c \cup T)$. *If we apply Algorithm 3 with* $\varepsilon = 10^{-8}$ *(or a larger value), we obtain the maximum value* $\mu^*(10^{-8}) = 0$, *which would suggest that* $\underline{E}(H^c|H^c\cup T)$ *is close to* 0. *But in fact* $\underline{E}(H^c|H^c\cup T) = \mu^*(0) = 15/16$, *and we obtain a very close approximation to this value by taking* ε *to be* 10^{-9} *or smaller. Here* μ^* *is right-continuous at* 0 *but it has a large discontinuity at a very small value of* ε, *between* 10^{-9} *and* 10^{-8}. *This behaviour can be detected if, after running Algorithm 3 with a very small positive value of* ε, *we run it again with* $\varepsilon = 0$. *If* ε *has not been chosen to be sufficiently small, the large discrepancy between* $\mu^*(\varepsilon)$ *and* $\mu^*(0)$ *warns us to try smaller values or to use Algorithm 4.*

The following theory leads to an iterative algorithm which always gives the exact value of $\underline{E}(A|B)$. As noted earlier, $\underline{E}(A|B) = \lim_{\varepsilon\downarrow 0} \mu^*(\varepsilon) \leq \mu^*(0)$, but if μ^* is not right-continuous at 0 then $\underline{E}(A|B) < \mu^*(0)$. The crucial step in an exact algorithm to compute $\underline{E}(A|B)$ is to find a subset of the assessments, indexed by $I \subseteq \{1, 2, \ldots, k\}$, which determines $\underline{E}(A|B)$ and for which the function μ_I^* (defined using only the subset I) is right-continuous at 0 so that $\underline{E}(A|B) = \mu_I^*(0)$. Then $\underline{E}(A|B)$ can be constructed from the subset of assessments $\{\underline{P}(A_i|B_i) : i \in I\}$, without using the other assessments, and we can set $\varepsilon = 0$ in the computations. The appropriate set I is identified in the next lemma.

Lemma 5 *Let* I *be the largest subset of* $\{1, 2, \ldots, k\}$ *with* $\underline{E}(B|\bigcup_{i\in I} B_i \cup B) > 0$. *(A unique largest subset exists because if two sets* I_1 *and* I_2 *have this property then so does* $I_1 \cup I_2$.*) Then* $\underline{E}(A|B)$ *is the maximum value in the following linear program:*

$$\begin{array}{ll} \text{maximise} & \mu \\ \text{subject to} & \lambda_i \geq 0 \quad (i \in I) \\ \text{and} & \sum_{i \in I} \lambda_i G_i + \mu B \leq AB. \end{array} \tag{9}$$

PROOF. First suppose that $(\mu, \boldsymbol{\lambda})$ satisfy condition (7) of Lemma 3. Then simple manipulation of (7) shows that, for some $\delta > 0$,

$$\sup \left[\sum_{i=1}^{k} \lambda_i G_i - (S(\boldsymbol{\lambda}) \cup B)(B - \delta) | S(\boldsymbol{\lambda}) \cup B\right] < 0.$$

(This holds if $\mu \geq 0$ in (7). If $\mu < 0$, the same inequality can be obtained by dividing $\boldsymbol{\lambda}$ by $1-\mu$.) This implies that $\underline{E}(B|S(\boldsymbol{\lambda}) \cup B) > \delta > 0$. It follows that, in (7), $S(\boldsymbol{\lambda}) \subseteq \bigcup_{i \in I} B_i$, so $I(\boldsymbol{\lambda}) \subseteq I$. This means that any $\boldsymbol{\lambda}$ which satisfies the conditions of Lemma 3 must have $\lambda_i = 0$ whenever $i \notin I$. Since Lemma 3 characterizes the natural extension $\underline{E}(A|B)$, this implies that $\underline{E}(A|B)$ can be computed by natural extension from the subset of assessments indexed by I.

Let $\mu_I^*(\varepsilon)$ denote the maximum value in the modified linear program that is obtained from (8) by adding the constraints that $\lambda_i = 0$ whenever $i \notin I$. Because this is equivalent to reducing the set of assessments to I, and $\underline{E}(B| \bigcup_{i \in I} B_i \cup B) > 0$ by definition of I, the sufficient condition for right-continuity of μ^* in the later Lemma 7 implies that μ_I^* is right-continuous at 0 and hence that $\underline{E}(A|B) = \mu_I^*(0)$. This gives the characterization (9). $\diamondsuit$

Lemma 5 shows that, given the set I, $\underline{E}(A|B)$ can be computed through the single linear program (9). To use this result in practice, we need to be able to determine I. It is not obvious that the definition of I given in Lemma 5 is useful, because it requires finding the natural extensions $\underline{E}(B| \bigcup_{i \in J} B_i \cup B)$ for various sets J. The next lemma gives some other characterizations of I which are more useful, especially (c) which is used in Algorithm 4.

Lemma 6 *The set I, defined in Lemma 5, is characterized by each of the following conditions.*

(a) $I = \{i : \underline{E}(B|B_i \cup B) > 0, i = 1, 2, \ldots, k\}$.
(b) I is the largest subset of $\{1, 2, \ldots, k\}$ for which there is $\boldsymbol{\lambda} \geq \mathbf{0}$ such that $\sup \left[\sum_{i \in I} \lambda_i G_i | \bigcup_{i \in I} B_i \cap B^c\right] < 0$.
(c) I is the largest subset of $\{1, 2, \ldots, k\}$ for which there is $\boldsymbol{\lambda} \geq \mathbf{0}$ such that $\sup \left[\sum_{i \in I} \lambda_i G_i + \sum_{i \in I} B_i | B^c\right] \leq 0$.

PROOF.

(a) $j \in I$ implies that $\underline{E}(B|B_j \cup B) > 0$, since otherwise $\underline{E}(B|\bigcup_{i\in I} B_i \cup B) \leq \underline{E}(B|B_j \cup B) \leq 0$ by coherence of the natural extensions. Conversely, suppose that $\underline{E}(B|B_j \cup B) > 0$ and let $J = I \cup \{j\}$. Using coherence of the natural extensions, $\underline{E}(B_j \cup B|\bigcup_{i\in J} B_i \cup B) \geq \underline{E}(B|\bigcup_{i\in I} B_i \cup B) > 0$. Also $B \subseteq B_j \cup B \subseteq \bigcup_{i\in J} B_i \cup B$, so it follows, again using coherence of the natural extensions, that $\underline{E}(B|\bigcup_{i\in J} B_i \cup B) \geq \underline{E}(B|B_j \cup B)\ \underline{E}(B_j \cup B|\bigcup_{i\in J} B_i \cup B)$ > 0. This shows that $J \subseteq I$, hence $j \in I$.

(b) By definition, I is maximal such that $\underline{E}(B|\bigcup_{i\in I} B_i \cup B) > 0$. Using Lemma 3, this condition is equivalent to the existence of $\boldsymbol{\lambda} \geq \mathbf{0}$, $\alpha > 0$ and $\varepsilon > 0$ such that

$$\sum_{i=1}^{k} \lambda_i G_i - (\cup_{i\in I} B_i \cup B)(B - \alpha) + \varepsilon(S(\boldsymbol{\lambda}) \cup [\cup_{i\in I} B_i] \cup B) \leq 0. \tag{10}$$

On the set B, (10) is equivalent to $\sum_{i=1}^{k} \lambda_i G_i + \alpha + \varepsilon \leq 1$, and on B^c it is equivalent to $\sum_{i=1}^{k} \lambda_i G_i + \alpha \bigcup_{i\in I} B_i + \varepsilon(S(\boldsymbol{\lambda}) \cup \bigcup_{i\in I} B_i) \leq 0$. If $(\boldsymbol{\lambda}, \alpha, \varepsilon)$ satisfy the condition on B^c then so do $(\delta\boldsymbol{\lambda}, \delta\alpha, \delta\varepsilon)$ whenever $\delta > 0$, and by taking δ to be sufficiently small the condition on B can also be satisfied. This shows that the condition on B is redundant. Hence, writing $I(\boldsymbol{\lambda}) = \{i : \lambda_i > 0\}$ and rewriting the condition on B^c, I is maximal such that there is $\boldsymbol{\lambda} \geq \mathbf{0}$ with $\sup\ [\sum_{i\in I(\boldsymbol{\lambda})} \lambda_i G_i | \bigcup_{i\in I \cup I(\boldsymbol{\lambda})} B_i \cap B^c] < 0$. But we can replace I by $I \cup I(\boldsymbol{\lambda})$ without changing this condition, so the maximal I must contain $I(\boldsymbol{\lambda})$. Hence we obtain the characterization in (b).

(c) By (b), I is maximal such that $\sup\ [\sum_{i\in I} \lambda_i G_i + \varepsilon \bigcup_{i\in I} B_i | B^c] \leq 0$ for some $\boldsymbol{\lambda} \geq \mathbf{0}$, $\varepsilon > 0$. This is equivalent to (c), as can be seen by multiplying this inequality by $k\varepsilon^{-1}$ and using $k \bigcup_{i\in I} B_i \geq \sum_{i\in I} B_i$. ♢

By using Lemma 6(c) together with the method of Algorithm 2, we can determine I by an iterative procedure in at most k steps. We then use I in the linear program (9) to determine $\underline{E}(A|B)$. This produces the following exact algorithm.

Algorithm 4

(a) Set $I = \{1, 2, \ldots, k\}$.

(b)

$$\begin{array}{ll} \text{Maximise} & \sum_{i\in I} \tau_i \\ \text{subject to} & \boldsymbol{\lambda} \geq \mathbf{0},\quad 0 \leq \tau_i \leq 1\ (i \in I) \\ \text{and} & \sup\ [\sum_{i\in I} \lambda_i G_i + \sum_{i\in I} \tau_i B_i | B^c] \leq 0. \end{array} \tag{11}$$

(c) If $\tau_i = 1$ for all $i \in I$ then go to (d). Otherwise, replace I by the subset $\{i \in I : \tau_i = 1\}$. If I is non-empty then return to (b).

(d) Solve the linear program (9). The maximized value of μ in (9) is the exact value of $\underline{E}(A|B)$.

Algorithm 4 works in a similar way to Algorithm 2, by successively reducing the set I. At each iteration of (b), I is replaced by its unique largest subset, I', with the property that $\sup\,[\sum_{i\in I}\lambda_i G_i+\sum_{i\in I'} B_i|B^c]\leq 0$. (As in Algorithm 2, the solution $\boldsymbol{\tau}$ in (b) must have $\tau_i=0$ or $\tau_i=1$ for all $i\in I$, so $\boldsymbol{\tau}$ has as many components as possible equal to 1.) The final set I, used in step (d), satisfies $I'=I$ because $\tau_i=1$ for all $i\in I$, so I is the set characterized in Lemma 6(c). It then follows from Lemma 5 that step (d) gives the correct value of $\underline{E}(A|B)$.

Algorithm 4 is analogous to the iterative Algorithm 2 for checking AUL. It differs in that we do not require $\sum_{i\in I}\lambda_i G_i+\sum_{i\in I}\tau_i B_i\leq 0$ everywhere on Ω, but only on B^c. Thus computing the natural extension $\underline{E}(A|B)$ is similar to checking AUL using the reduced possibility space B^c.

Algorithm 4 can be made more efficient in many problems by first running Algorithm 3 to determine an initial set $I_1=\{i:\lambda_i>0\}$, and then imposing the extra constraints $\tau_i=1$ for all $i\in I_1$, in each application of the linear program (11). (This works because the maximal set I must contain I_1.) Provided that I_1 is non-empty, this reduces the number of variables involved in (11).

The modified algorithm is as follows.

Algorithm 5

(a) Use Algorithm 3 to determine I_1. Set $J=\{1,2,\ldots,k\}\cap I_1^c$. If J is empty then go to (d).

(b)

$$\begin{aligned}&\text{Maximise}\quad \sum_{i\in J}\tau_i\\ &\text{subject to}\quad \boldsymbol{\lambda}\geq\mathbf{0},\quad 0\leq\tau_i\leq 1\ (i\in J),\\ &\text{and}\quad \sup\,[\sum_{i\in I_1\cup J}\lambda_i G_i+\sum_{i\in I_1}B_i+\sum_{i\in J}\tau_i B_i|B^c]\leq 0.\end{aligned}\tag{12}$$

If $\tau_i=1$ for all $i\in J$ then go to (d).

(c) Replace J by the subset $\{i\in J:\tau_i=1\}$. If J is non-empty then return to (b).

(d) Use $I_1\cup J$ instead of I in step (d) of Algorithm 4, giving $\underline{E}(A|B)$ as the maximized value of μ in (9).

In almost all practical problems, the final set J in step (d) will be empty.

Example 9 *In the football example, the subject wishes to evaluate the probability of event A, that team X wins the tournament. First consider just the precise probability assessments in Table 1. By applying Algorithm 3 or 4 to compute the natural extensions* $\underline{E}(A) = 0.325$ *and* $\underline{E}(A^c) = 0.47$, *we obtain the lower and upper bounds* 0.325 *and* 0.53 *for* $P(A)$. *These bounds can also be obtained quite easily by summing the probabilities of the events C_i that imply A, and of those C_i that are consistent with A.*

Now consider the effect of combining the 10 *imprecise probability assessments in Table 2 with those in Table 1. By applying Algorithm 3 or 4 again to compute the natural extensions, we now obtain* $\underline{E}(A) = 0.3955$ *and* $\overline{E}(A) = 0.4945$. *Although the* 10 *extra assessments are quite imprecise, they substantially reduce the interval* $[\underline{E}(A), \overline{E}(A)]$. *Again the LP algorithms were not really needed here: because of the simple structure of the events involved, the natural extensions could have been calculated through the simple formulae* $\underline{E}(A) = \sum_{i=1}^{27} \underline{P}(A|C_i)P(C_i)$ *and* $\overline{E}(A) = \sum_{i=1}^{27} \overline{P}(A|C_i)P(C_i)$.

Example 10 *Consider the six assessments of precise unconditional probabilities that were given in Example 1. Using Algorithm 3 or 4, we can compute the natural extensions to any conditional or unconditional events. For example, we obtain the natural extensions* $\underline{E}(A_3) = 0.4$ *and* $\overline{E}(A_3) = 0.8$ *concerning* A_3, *and* $\underline{E}(A_4|A_3) = 0.375$ *and* $\overline{E}(A_4|A_3) = 0.75$ *concerning* $A_4|A_3$. *By Lemma 7, all these values can be computed by setting* $\varepsilon = 0$ *in Algorithm 3, because in each case the conditioning event has positive lower probability. Further examples, including natural extensions which cannot be obtained by setting* $\varepsilon = 0$, *are given in Example 12.*

3.6 Continuity of μ^* at zero

When the function μ^* is right-continuous at 0, the natural extension is given by $\underline{E}(A|B) = \mu^*(0)$, which can be found by setting $\varepsilon = 0$ in (8) and solving the linear program. Continuity of μ^* at 0 therefore simplifies the computational problem. First we give an example to show that μ^* can be discontinuous at 0, so that the problem cannot always be simplified in this way.

Example 11 *Suppose that the only assessment is* $\underline{P}(B^c|A^c) = 1$, *where* $\emptyset \subset A \subset B \subset \Omega$, *and we wish to compute* $\underline{E}(A|B)$. *To do so we find* $\mu^*(\varepsilon)$, *which, by (8), is the maximum value of* μ *such that* $B(A - \mu) \geq \lambda(G + \varepsilon A^c)$ *for some* $\lambda \geq 0$, *where* $G = A^c(B^c - 1)$. *By considering the values on* $B^c, B \cap A^c$ *and*

A, we obtain the three inequalities:

$$0 \geq \lambda\varepsilon, \quad -\mu \geq \lambda(-1+\varepsilon), \quad 1-\mu \geq 0. \tag{13}$$

Because $\lambda \geq 0$ and $\varepsilon > 0$, the first inequality gives $\lambda = 0$ and then the second inequality gives $\mu \leq 0$. Hence we obtain $\mu^*(\varepsilon) = 0$ for all $\varepsilon > 0$, giving the natural extension $\underline{E}(A|B) = 0$. The assessment is completely uninformative about $P(A|B)$.

But setting $\varepsilon = 0$ in (13) gives $\mu \leq \lambda$ and $\mu \leq 1$, and hence $\mu^*(0) = 1$. Thus the function μ^* has a large discontinuity at 0.

Next we give sufficient conditions for μ^* to be right-continuous at 0. (Note that we cannot simply apply the sufficient conditions given in the operations research literature [1] for continuity of the maximum μ^*, because the optimality region and feasible region for our problem (in the variables μ and $\boldsymbol{\lambda}$) are not always bounded.) The following results show that right-continuity at 0 depends essentially on the new conditioning event B having positive probability conditional on $(\cup_{i=1}^k B_i \cup B)$. In many problems, $\cup_{i=1}^k B_i \cup B = \Omega$, and then the results require that B has positive unconditional (lower) probability.

Lemma 7 *Assume that the assessments AUL. Then a sufficient condition for $\underline{E}(A|B) = \mu^*(0)$, i.e., for right-continuity of μ^* at 0, is that $\underline{E}(B| \cup_{i=1}^k B_i \cup B) > 0$. Hence it is sufficient that $\underline{E}(B) > 0$.*

PROOF. Let $\gamma = \mu^*(0)$. First assume that γ is finite. In that case the supremum γ is achieved, and hence there is $\boldsymbol{\rho} \geq \mathbf{0}$ such that $B(A-\gamma) \geq \sum_{i\in I} \rho_i G_i$. Assuming that $\underline{E}(B| \cup_{i=1}^k B_i \cup B) > 0$, there is $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\tau > 0$ such that $B - \tau(\cup_{i=1}^k B_i \cup B) \geq \sum_{i\in I} \lambda_i G_i$. Given any $\delta > 0$, let $\varepsilon = k^{-1}\delta\tau / \max\,\{\rho_i + \delta\lambda_i : i = 1, \ldots, k\}$, so $\varepsilon > 0$. Then $B(A-\gamma+\delta) \geq \sum_{i\in I}(\rho_i + \delta\lambda_i)G_i + \delta\tau(\cup_{i=1}^k B_i) \geq \sum_{i\in I}(\rho_i + \delta\lambda_i)(G_i + \varepsilon B_i)$. (This holds also if all values of ρ_i and λ_i are zero, since then the last term is zero for all $\varepsilon > 0$.) By definition of the natural extension, $\underline{E}(A|B) \geq \gamma - \delta$. Since δ is arbitrarily small, $\underline{E}(A|B) \geq \gamma$, and it follows that $\underline{E}(A|B) = \gamma = \mu^*(0)$ since always $\underline{E}(A|B) \leq \mu^*(0)$. Thus μ^* is right-continuous at zero. The same argument shows that $\mu^*(0)$ must be finite, because otherwise γ can be chosen to be arbitrarily large and then $\underline{E}(A|B) \geq \gamma$ contradicts $\underline{E}(A|B) \leq 1$. The second statement in the lemma follows from the coherence property $\underline{E}(B| \cup_{i=1}^k B_i \cup B) \geq \underline{E}(B|\Omega) = \underline{E}(B)$. $\diamondsuit$

Alternatively, Lemma 7 can be derived from the necessary and sufficient condition for continuity of μ^* at 0 that is stated in subsection 3.7. Compare the sufficient condition in Lemma 7 with the weaker condition $\overline{E}(B| \cup_{i=1}^k B_i \cup B) > 0$, which is necessary for continuity of μ^* at 0 (assuming that the assessments

AUL). In fact, if this condition fails then $\mu^*(0) = \infty$. (That is easy to verify from the definition of natural extension.)

To check the condition in Lemma 7, we must compute a natural extension. The following stronger condition is much easier to verify.

Corollary 1 *Assume that the assessments AUL. A sufficient condition for right-continuity of* μ^* *at* 0 *is that* B *contains* $\cup_{i=1}^k B_i$*. Hence it is sufficient that* $B = \Omega$*, i.e., that we are computing an unconditional natural extension.*

It follows that, in the case $B = \Omega$, the problem of computing the natural extension to an unconditional lower probability $\underline{E}(A)$ requires just the single linear program (8) with $\varepsilon = 0$. In this case, the natural extension is given by the formulae $\underline{E}(A) = \sup\{\mu : \mu \leq A - \sum_{i=1}^k \lambda_i G_i, \boldsymbol{\lambda} \geq \mathbf{0}\} = \sup\{\inf[A - \sum_{i=1}^k \lambda_i G_i|\Omega] : \boldsymbol{\lambda} \geq \mathbf{0}\}$. Several examples of such computations have been given in subsection 3.5.

Assuming that the assessments AUL, it can be shown that a necessary and sufficient condition for μ^* to be right-continuous at zero is that, if we add the extra precise assessment $P(A|B) = \underline{E}(A|B)$ to the k given assessments, the natural extension of these $k+1$ assessments satisfies $\overline{E}(B| \cup_{i=1}^k B_i \cup B) > 0$. An equivalent condition is that there is $\varepsilon > 0$ such that adding the extra assessment $\underline{P}(B| \cup_{i=1}^k B_i \cup B) = \varepsilon$ to the given assessments does not increase the natural extension $\underline{E}(A|B)$. However, these conditions are relatively difficult to verify.

In the special case where the assessments AUL and a precise probability $P(B| \cup_{i=1}^k B_i \cup B)$ is assessed, or is precisely determined by the assessments, it is necessary and sufficient for right-continuity of μ^* at 0 that this probability is *positive*. In this case, again we can set $\varepsilon = 0$ in the linear program (8).

Example 12 *Consider again the six assessments in Example 1. Here we investigate the continuity of* μ^* *at zero, for three events of interest:* $A = (A_1 \cap A_2)^c, B = (A_1 \cap A_2) \cup (A_1^c \cap A_2^c)$*, and the conditional event* $A|B$*. To investigate the behaviour of* μ^**, we solved the LP problem (8) for several values of* ε*, including zero. The results are reported in Table 3.*

First consider the results for A*. The second column of the table shows that* $\underline{E}(A) = 1$ *and that* μ^* *is continuous at zero, which is as expected from the preceding results since this is an unconditional natural extension. The values of* $\mu^*(\varepsilon)$ *for small* ε *are very good approximations to* $\underline{E}(A)$*. The result* $\underline{E}(A) = 1$ *implies that* $\overline{E}(A_1 \cap A_2) = 0$*, hence every coherent extension of the assessments must give precise probability zero to* $A_1 \cap A_2$*.*

The third column of the table shows that $\underline{E}(B) = 0$ *and again, because we are computing an unconditional natural extension,* μ^* *is continuous at zero.*

Table 3
Values of $\mu^*(\varepsilon)$ for different values of ε in Example 1, with $A = (A_1 \cap A_2)^c$ and $B = (A_1 \cap A_2) \cup (A_1^c \cap A_2^c)$.

	A	B	$A\|B$
$\varepsilon = 10^{-1}$	0.8	0	0
10^{-2}	0.98	0	0
10^{-3}	0.998	0	0
10^{-4}	0.9998	0	0
10^{-5}	0.99998	0	0
10^{-6}	0.999998	0	0
10^{-7}	0.9999998	0	0
0	1	0	1

In both these cases it would have been sufficient to compute just $\mu^*(0)$.

The third example, concerning $A|B$, involves conditioning on an event of lower probability zero. From the fourth column of Table 3, the result is $\underline{E}(A|B) = 0$, which means that $\overline{E}(A_1 \cap A_2|B) = 1$. Here μ^* has a large discontinuity at 0.

3.7 The dual problem

As seen in subsection 3.6, it is useful to study the LP problem obtained from (8) by setting $\varepsilon = 0$. The dual of this LP problem is closely related to the algorithm of Pelessoni and Vicig [27]. To simplify the formulae in this subsection, we consider computing the natural extension of the assessments to $\underline{E}(A_0|B_0)$. Let $s = ||\Omega||$.

By applying a general form of the duality theorem of linear programming [28, p. 91,eq. 22], the dual problem is

$$\text{minimize} \quad \sum_{j=1}^{s} x_j A_0(\omega_j) B_0(\omega_j) \tag{14}$$

$$\text{subject to} \quad \sum_{j=1}^{s} B_i(\omega_j) \left[A_i(\omega_j) - \underline{P}(A_i|B_i)\right] x_j \geq 0 \quad (i = 1, \ldots, k) \tag{15}$$

$$\sum_{j=1}^{s} B_0(\omega_j) x_j = 1 \tag{16}$$

$$\text{and} \quad x_j \geq 0 \quad (j = 1, \ldots, s). \tag{17}$$

Every vector $(x_1, \ldots, x_s)$ in the feasible region of this LP problem is proportional to a conditional probability distribution P, and $x_j = P(\omega_j | \bigcup_{i=0}^k B_i)/P(B_0 | \bigcup_{i=0}^k B_i)$ $(j = 1, \ldots, s)$. Using this fact, the inequalities (15) impose the dominance conditions $P(A_i|B_i) \geq \underline{P}(A_i|B_i)$ $(i = 1, \ldots, k)$, while (16) implies that $P(B_0 | \bigcup_{i=0}^k B_i) > 0$. The feasible region of the problem identifies the set $\mathcal{M}$ of probability distributions dominating $\underline{P}$ and such that $P(B_0 | \bigcup_{i=0}^k B_i) > 0$. It also follows that the objective function in (14), which is minimized over all $P \in \mathcal{M}$, is equal to $P(A_0|B_0)$. (See [27] for details and also for the sequel. In [27] it is assumed that $A_0|B_0$ is included in the set of conditional events for which assessments are made, but this assumption is not restrictive because, if it is not satisfied, we can add an uninformative assessment $\underline{P}(A_0|B_0) = 0$ without changing the problem.)

By the well-known strong duality theorem, if the LP problem (14–17) is feasible then its optimal value and that of its primal problem are equal. Hence, the following characterization of continuity of μ^* at zero follows: μ^* is continuous at zero if and only if there exists a probability distribution $P \geq \underline{P}$ such that $P(B_0 | \bigcup_{i=0}^k B_i) > 0$ and $P(A_0|B_0) = \underline{E}(A_0|B_0)$.

When problem (14–17) is infeasible, it follows from the meaning of the feasible region and from property 3.2(f) that $\overline{P}(B_0 | \bigcup_{i=0}^k B_i) = 0$. Conversely, if $P(B_0 | \bigcup_{i=0}^k B_i) = 0$ for every $P \geq \underline{P}$, then $\mathcal{M}$ is empty. It follows from duality theory that the primal problem (which is always feasible) must be upper unbounded, which means that $\mu^*(0) = +\infty$. We therefore obtain the characterization of unboundedness of μ^* at zero that was given in section 3.4.

Finding a solution of problem (14–17) is the final step in the Pelessoni-Vicig algorithm [27] if and only if all probability distributions P that dominate $\underline{P}$ have $P(B_0 | \bigcup_{i=0}^k B_i) > 0$. (This condition is checked in the algorithm through a LP problem.) If not, a finite sequence of LP problems finds the exact value for $\underline{E}(A_0|B_0)$. The latter alternative happens also in instances when μ^* is continuous at zero. In all cases, the number of LP problems that need to be solved in Algorithm 4 is less than or equal to the number required in the Pelessoni-Vicig algorithm.

3.8 Checking coherence of imprecise probability assessments

By 3.2(d), the lower probability assessments $\underline{P}(A_i|B_i) = c_i$ $(i = 1, \ldots, k)$ are coherent if and only if $\underline{E}(A_i|B_i) = c_i$ for $i = 1, \ldots, k$. To check coherence of the assessments, it therefore suffices to compute the natural extensions $\underline{E}(A_i|B_i)$ $(i = 1, \ldots, k)$. This involves k problems of the form (8), one for each assessment. Using Algorithm 3, it requires solving k linear programs.

This method of checking coherence can be simplified slightly, using result

3.2(c) that $\underline{E}(A_i|B_i) \geq \underline{P}(A_i|B_i)$ for $i = 1, \ldots, k$. Hence it suffices to check that $\underline{E}(A_i|B_i) \leq c_i$ for $i = 1, \ldots, k$. Using the definition of natural extension, this is true if and only if the system of inequalities $\varepsilon > 0$, $\boldsymbol{\lambda} \geq \mathbf{0}$ and $G_j - \varepsilon B_j \geq \sum_{i=1}^{k} \lambda_i (G_i + \varepsilon B_i)$ has no solution $(\varepsilon, \boldsymbol{\lambda})$, for each $j = 1, \ldots, k$. Again this can be determined in practice by k applications of Algorithm 3.

If the assessments include both precise and imprecise (lower probability) assessments, to check coherence of the whole system it suffices to check AUL of the system and to check that $\underline{E}(A_i|B_i) \leq \underline{P}(A_i|B_i)$ for each of the lower probability assessments.

Example 13 *Consider the assessments for the football example given in Tables 1 and 2, which are equivalent to a system of* 64 *lower probability assessments. To check the coherence of this system, (a) we used Algorithm 1 to verify that the assessments AUL, and (b) we used Algorithm 3 to compute the natural extensions* $\underline{E}(A|C_i)$ *and* $\underline{E}(A^c|C_i)$ *corresponding to each of the lower and upper probabilities in Table 2. For (b), in each application of Algorithm 3 we can set* $\varepsilon = 0$, *since we know from Table 1 that each conditioning event* C_i *has positive probability. We cannot set* $\varepsilon = 0$ *in step (a), because precise probability assessments are involved and therefore problem (1) has solutions for* $\varepsilon = 0$ *(see the comments following Example 4). We find that all the natural extensions agree with the corresponding upper or lower probabilities in Table 2, and therefore the* 64 *assessments are coherent.*

If the assessments AUL then their natural extensions $\{\underline{E}(A_i|B_i) : i = 1, \ldots, k\}$ are always coherent. If the assessments AUL but they are not coherent then their natural extensions can be regarded as coherent 'corrections' of the assessments, in the sense that at least one lower probability assessment must be increased (or an upper probability decreased) to achieve coherence. In fact the natural extensions are produced by making the minimal corrections of this type that achieve coherence.

Example 14 *Consider the* 10 *assessments of conditional upper and lower probabilities for the football example, given in Table 2, plus one further assessment that* $\overline{P}(A|B) = 0.625$, *where A and B are defined in Example 7. We can check that the* 11 *assessments AUL, by using Algorithm 1 or checking that (1) has no solution for* $\varepsilon = 0$. *But we find, using Algorithm 3 or 4, that the natural extension* $\overline{E}(A|(L, W, L))$ *is* 0.625, *which differs from the assessment* $\overline{P}(A|(L, W, L)) = 1$. *Thus the* 11 *assessments are incoherent. Reducing the assessment to* $\overline{P}(A|(L, W, L)) = 0.625$ *does achieve coherence.*

Another way of achieving coherence in this example is to increase the assessment of $\overline{P}(A|B)$ *slightly, from* 0.625 *to* 0.65. *It can be verified, using the same procedure, that these* 11 *assessments are coherent. In particular, we now obtain* $\overline{E}(A|(L, W, L)) = \overline{P}(A|(L, W, L)) = 1$. *Here* $\overline{E}(A|(L, W, L))$ *is computed* *via $\underline{E}(A^c|(L,W,L))$, which produces another example of a function μ^* that is discontinuous at zero.*

3.9 Theoretical approaches to the computational problem

The problem (8) is a LP parametric problem in the scalar parameter ε. Formally, it is a special case of the following parametric linear program

$$\text{maximise} \quad \mathbf{c}^T\mathbf{x} \tag{18}$$
$$\text{subject to} \quad H(\varepsilon)\mathbf{x} \leq \mathbf{b} \tag{19}$$

where each element of the matrix H is function of the scalar parameter ε (typically a linear or polynomial function). Let us call this problem $\mathcal{P}(\varepsilon)$, and let $f(\varepsilon)$ be the optimum value of $\mathcal{P}(\varepsilon)$ as a function of ε. Problem $\mathcal{P}(\varepsilon)$ is usually hard to solve (see [11] for a review, and also [12]).

Much of the work in the literature has focused on the problem of the continuity of $f(\varepsilon)$, and several sufficient conditions have been proposed [1,23,36]. A customary requirement for them is that the optimality region or the feasible region of the problem is bounded. These conditions do not always hold in our case, as shown in the following example.

Example 15 *Suppose that assessments are given for $A_i|B_i$ $(i = 1, \ldots, k)$ and that $\bigcup_{i=1}^k B_i \subset A^c$. Then, the inequalities in (8) become $\sum_{i=1}^k \lambda_i(G_i + \varepsilon B_i) \leq 0$ when $B = 0$, $\sum_{i=1}^k \lambda_i(G_i + \varepsilon B_i) \leq -\mu$ when $B = 1$ and $A = 0$, $\sum_{i=1}^k \lambda_i(G_i + \varepsilon B_i) \leq 1 - \mu$ when $AB = 1$. Observing that the third inequality reduces to $\mu \leq 1$ and that $\mu \geq 0$, it is easily seen that if $(\mu, \lambda_1, \ldots, \lambda_2)$ is an optimal solution of (8), then so is $(\mu, t\lambda_1, \ldots, t\lambda_2)$, for all $t > 1$.*

Some of the sufficient conditions for the continuity of the optimal value have some probabilistic meaning when applied to our problem and this helps in showing that they may be too weak in our framework. For instance, it can be checked that the conditions given in [36] imply but are more restrictive than the condition $\underline{P}(B|\bigcup_{i=1}^k B_i \cup B) > 0$, which in turn is sufficient to solve the natural extension problem in one step by means of the algorithm proposed in [27].

Another way of tackling a parametric linear programming problem is the sensitivity analysis approach: given a solution of the problem for a given ε, to which corresponds an optimal basis β in the simplex method, the critical region $R_\beta = \{\varepsilon : \beta \text{ is an optimal basis for } \mathcal{P}(\varepsilon)\}$ is defined. It has been shown in [7,10,11,35] that

Proposition 1 *For each optimal basis β, R_β is a finite union of intervals; further, if ε is an interior point of R_β, f is continuous at ε.*

Corollary 2 *The number of points of discontinuity of f is finite.*

A procedure for determining all R_β is given in [35], for the case, which includes our problem, where the matrix H is a polynomial function of ε. It requires employing symbolic computations and finding the real roots of polynomials of possibly high degrees, roots which form the elements of the set $E = \bigcup_\beta E_\beta$, where E_β is the set of the endpoints of the intervals in R_β and the union is over all the optimal bases β.

An alternative method is suggested in [10] to detect whether a given value of ε belongs to E. This could be applied to $\varepsilon = 0$, but the computational burden remains anyway high, and further, even if ε belongs to E this does not necessarily imply that the optimal value is discontinuous at ε. For instance, ε might belong to E_β for some optimal basis β, being an interior point of the critical region for some other optimal basis.

This survey suggests that the method for solving our problem outlined in 3.4 seems to be a good compromise amongst theoretical needs, practical viability and precision of the results obtained.

4 Conclusions

We have described several direct algorithms for checking consistency of conditional probability or lower probability assessments and for making inferences from them. The algorithms are quite general, but there is scope for further generalization in the following respects.

(a) **From probabilities to previsions**: Instead of assuming that conditional lower probabilities $\underline{P}(A_i|B_i)$ are assessed, we can allow any assessments of conditional lower previsions $\underline{P}(X_i|B_i)$, where X_i is a simple random variable (one which has only finitely many possible values). This formulation is more general, because the lower probability of an event A can be identified with the lower prevision of its indicator function. Similarly we may need to calculate the natural extension of the assessments to a lower prevision $\underline{E}(X|B)$. In fact, it follows from results in [31] that this more general problem can be solved by the methods described in this paper: the definitions of AUL and natural extension can be generalized from events to simple random variables by replacing A_i by X_i in (1) and (6), and replacing A by X in (6) [33].

(b) **Independence judgements**: The algorithms need to be generalized to allow judgements of independence or conditional independence [31,32], as well as conditional probability assessments. We are currently investigating algorithms of this kind.

(c) **Infinitely many assessments**: We may want to allow infinitely many assessments of conditional lower probabilities or previsions. This introduces new complications because the definitions of AUL and natural extension should be generalized to include conglomerative conditions [31].

(d) **Very large LP problems**: When the number of assessments is large, the possibility space Ω that they generate (as outlined in subsection 1.3) may be so large that the LP problems become intractable. In such cases, it may be useful to combine our algorithms with 'row generation' methods which are dual to the column generation methods in [20], to enable us to solve the LP problems without specifying Ω explicitly.

Acknowledgements

Peter Walley wishes to thank Dipartimento di Matematica Applicata 'B. de Finetti', Università di Trieste, for supporting a visit to Trieste in 1998, when most of this research was carried out, and also FAPESP and Escola Politécnica, Universidade de São Paulo, for support in the later stages of this project.

References

[1] B. Bereanu, The continuity of the optimum in parametric programming and application to stochastic programming, *Journal of Optimization Theory and Applications*, **18** (1976) 319–333.

[2] G. Boole, *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities* (Macmillan, London, 1854). Reprinted in 1958 by Dover, New York.

[3] G. Boole, On the conditions by which solutions of questions in the theory of probabilities are limited, *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* **4** (8) (1854) 91–98.

[4] G. Bruno and A. Gilio, Applicazione del metodo del simplesso al teorema fondamentale per le probabilità nella concezione soggettiva, *Statistica* **40** (1980) 337–344.

[5] G. Coletti, Coherent numerical and ordinal probabilistic assessments, *IEEE Transactions on Systems, Man and Cybernetics* **24** (1994) 1747–1754.

[6] G. Coletti and R. Scozzafava, Characterization of coherent conditional probabilities as a tool for their assessment and extension, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **4** (1996) 103–127.

[7] W. Dinkelbach, *Sensitivitätsanalysen und Parametrische Programmierung* (Springer-Verlag, Berlin, 1969).

[8] B. de Finetti, Problemi determinati e indeterminati nel calcolo delle probabilità, *Rendiconti Reale Accademia dei Lincei* **6** (1930) 367–373.

[9] B. de Finetti, *Theory of Probability, Volume 1* (Wiley, London, 1974).

[10] R. M. Freund, Postoptimal analysis of a linear program under simultaneous changes in matrix coefficients, *Mathematical Programming Study* **24** (1985) 1–13.

[11] T. Gal, Linear parametric programming - a brief survey, *Mathematical Programming Study* **21** (1984) 43–68.

[12] T. Gal, *Postoptimal Analyses, Parametric Programming and Related Topics* (de Gruyter, Berlin, 1995).

[13] A. Gilio, Algorithms for precise and imprecise conditional probability assessments, in: G. Coletti, D. Dubois and R. Scozzafava, eds., *Mathematical Models for Handling Partial Knowledge in Artificial Intelligence* (Plenum Press, New York, 1995) 231–254.

[14] A. Gilio, Algorithms for conditional probability assessments, in: D.A. Berry, K.M. Chaloner and J.K. Geweke, eds., *Bayesian Statistics and Econometrics* (Wiley, New York, 1996) 29–39.

[15] T. Hailperin, Best possible inequalities for the probability of a logical function of events, *American Mathematical Monthly* **72** (1965) 343–359.

[16] T. Hailperin, *Boole's Logic and Probability*, second enlarged edition. Studies in Logic and the Foundations of Mathematics **85** (North-Holland, Amsterdam, 1986).

[17] P. Hansen and B. Jaumard, Probabilistic satisfiability, Technical Report G-96-31, Les Cahiers du GERAD, Montréal (1996).

[18] P. Hansen, B. Jaumard and M. Poggi de Aragão, Boole's conditions of possible experience and reasoning under uncertainty, *Discrete Applied Mathematics* **60** (1995) 181–193.

[19] S. Holzer, On coherence and conditional prevision, *Boll. Un. Mat. Ital.* Serie **VI** (IV-C) (1985) 441–460.

[20] B. Jaumard, P. Hansen and M. Poggi de Aragão, Column generation methods for probabilistic logic, *ORSA Journal on Computing* **3** (1991) 135–148.

[21] F. Lad, *Operational Subjective Statistical Methods* (Wiley, New York, 1996).

[22] F. Lad, J. M. Dickey and M. A. Rahman, The fundamental theorem of prevision, *Statistica* **50** (1990) 19–38.

[23] D. H. Martin, On the continuity of the maximum in parametric linear programming, *Journal of Optimization Theory and Applications* **17** (1975) 205–210.

[24] N. J. Nilsson, Probabilistic logic, *Artificial Intelligence* **28** (1986) 71–87.

[25] N. J. Nilsson, Probabilistic logic revisited, *Artificial Intelligence* **59** (1993) 39–42.

[26] G. Paass, Probabilistic logic, in: P. Smets, A. Mamdani, D. Dubois and H. Prade, eds., *Non-Standard Logics for Automated Reasoning* (Academic Press, London, 1988) 213–251.

[27] R. Pelessoni and P. Vicig, A consistency problem for imprecise conditional probability assessments, in: *Proceedings of IPMU'98* (E.D.K., Paris, 1998) **2**, 1478–1485.

[28] A. Schrijver, *Theory of Linear and Integer Programming* (Wiley, New York, 1986).

[29] P. Vicig, An algorithm for imprecise conditional probability assessments in expert systems, in: *Proceedings of IPMU'96* (Proyecto Sur de Ediciones, Granada, 1996) **1**, 61–66.

[30] P. Vicig, Upper and lower bounds for coherent extensions of conditional probabilities given on finite sets, *Quad. n.7/97*, Dip. Mat. Appl. 'B. de Finetti', Univ. di Trieste (1997).

[31] P. Walley, *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London, 1991).

[32] P. Walley, Measures of uncertainty in expert systems, *Artificial Intelligence* **83** (1996) 1–58.

[33] P. Walley, Coherent upper and lower previsions, *The Imprecise Probabilities Project* (1997) http://ensmain.rug.ac.be/∼ipp/documentation/

[34] P. Walley, An introduction to the theory of natural extension, unpublished manuscript (1998).

[35] E. Weickenmeier, Zur Lösung parametrischer linearer Programme mit polinomischen Parameterfunktionen, *Zeitschrift für Operation Research* **22** (1978) 131–149.

[36] A. C. Williams, Marginal values in linear programming, *J. Soc. Indust. Appl. Math.* **11** (1963) 82–94.

[37] P. M. Williams, Notes on conditional previsions, Research Report, School of Math. and Phys. Science, University of Sussex (1975).