Probabilistic Abduction Without Priors

Didier Dubois [a] and Angelo Gilio [b]
and Gabriele Kern-Isberner [c]

[a] *Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, Toulouse, France*

[b] *Dip. Metodi e Modelli Matematici, Università di Roma ”La Sapienza”, Roma, Italy*

[c] *Dept. of Computer Science, University of Dortmund, Dortmund, Germany*

Abstract

This paper considers the simple problem of abduction in the framework of Bayes theorem, when the prior probability of the hypothesis is not available, either because there are no statistical data to rely on, or simply because a human expert is reluctant to provide a subjective assessment of this prior probability. This abduction problem remains an open issue since a simple sensitivity analysis on the value of the unknown prior yields empty results. This paper tries to propose some criteria a solution to this problem should satisfy. It then surveys and comments on various existing or new solutions to this problem: the use of likelihood functions (as in classical statistics), the use of information principles like maximum entropy, Shapley value, maximum likelihood. Finally, we present a novel maximum likelihood solution by making use of conditional event theory. The formal setting includes de Finetti's coherence approach, which does not exclude conditioning on contingent events with zero probability.

Key words: Bayes theorem, prior probability, maximum likelihood, Shapley value, entropy, imprecise probability, coherence

---

1 Introduction

Abductive reasoning tries to find plausible explanations for observed evidence. In the framework of probability theory, Bayes theorem may help solving the problem, provided that enough information is available, which is, however, rarely the case.

Email addresses: `dubois@irit.fr` (Didier Dubois),
`gilio@dmmm.uniroma1.it` (Angelo Gilio),
`gabriele.kern-isberner@cs.uni-dortmund.de` (Gabriele Kern-Isberner).

To put the basic problem of Bayesian abduction in more formal terms, assume H to be a Boolean proposition interpreted as a hypothesis, a disease, a fault, a cause, etc. pertaining to the state of a system. Let E be another proposition representing a hypothetically observed (that is, observable) fact, a symptom, an alarm, an effect, etc. Numerical assessments of positive conditional probability values $P(E|H) = a$ and $P(E|H^c) = b \leq a$ are supplied by an agent, who either uses available statistical data or proposes purely subjective assessments. The problem is to evaluate the relative plausibility of the hypothesis and its negation after observing event E. If a prior probability $P(H)$ is assigned and $b > 0$, the question is solved by Bayes theorem. But, due to sheer ignorance, suppose no prior probability $P(H)$ is assigned and observation E is made, or that probabilities a or b are set to zero. What can be said about the support given to hypotheses H vs. H^c upon observing E? For instance, a roadwork (H) might cause a traffic jam (E). Having assigned all necessary probabilities above and being stuck in a traffic jam, with which probability do we expect a roadwork ahead, i.e. how to estimate $P(H|E)$ in a reasonable way?

In this work, we discuss thoroughly past proposals for dealing with this problem and develop either new solutions or rigorous formalization of previously proposed solutions. To this aim, we review various approaches to probability theory, and to imprecise probabilities, such as maximum entropy, Shapley value, conditional events, de Finetti's coherence setting, possibility theory and the like, and include an in-depth comparison. As a main contribution of this paper, we present a novel maximum likelihood approach by making use of conditional event theory, viewing conditional probabilities as a kind of midvalues. Although we might want to deal with more complex abduction problems, investigating these methods in this simple context already helps clarifying substantial differences between them.

Here, by definition, we do not take for granted the Bayesian credo according to which whatever their state of knowledge, rational agents should produce a prior probability. Indeed the idea that point probability functions should be in one to one correspondence with belief states means that a probability degree *is equated* to a degree of belief. Then, in case of total ignorance about H, agents should assign equal probabilities to H and its complement, due to symmetry arguments. This claim can be challenged, and has been challenged by many scholars (e.g., [33,15,37,39]): Indeed agents must assign equal probabilities to H and its complement, when they know that the occurrence of H is driven by a genuine random process, and when they know nothing. The two epistemic states are different but result in the same probability assessment. Here, we take ignorance about H for granted, assuming $P(H)$ is unspecified (in other words the agent refuses to bet on a value of $P(H)$). We review what was done in the past, and what can legitimately be done to cope with ignorance, trying to formally justify various solutions to this problem.

Investigating the problem of abduction in a formal Bayesian framework here allows us to deal both with consistency based diagnosis (i.e. when evidence contradicts some hypotheses made about a system) and purely abductive reasoning (i.e. finding a minimal set of faults that explain the observations) in a common framework.

This paper is organized as follows: We put the problem into formal terms in the next section, and some criteria are laid bare for a solution to the problem to be acceptable. We continue by recalling three classical approaches for its solution. Afterwards, various information principles are applied to solve the problem, and compared with each other. Finally, we present our novel *relaxed maximum likelihood approach*. We conclude the paper with a summary and an outlook on further work.

2 Methodology

In this paper, the following notations are adopted: Ω is the sure event, AB is short for $A \wedge B$ (conjunction), and the complement of an event A is denoted A^c. Moreover, we use the same symbol to denote an event and its indicator.

2.1 Formalizing the problem

The basic variables in the problem are denoted

$$x = P(EH)\,; y = P(E^cH)\,; z = P(EH^c)\,; t = P(E^cH^c).$$

Let $\mathcal{P} = \{P, P(E|H) = a, P(E|H^c) = b\}$ be the set of probability functions described by the constraints expressing the available knowledge. The variables x, y, z, t are thus linked by the following constraints:

$$x + y + z + t = 1 \text{ (normalization)},$$
$$x = a(x + y) \text{ corresponding to } \; P(E|H) = a,$$
$$z = b(z + t) \text{ corresponding to } \; P(E|H^c) = b.$$

The problem of finding probability distributions which are solutions to a set of constraints has been addressed by many authors working on imprecise probabilities; see e.g. [8,9,14,17,28,31,24,38]. The set $\mathcal{P}$ is clearly a segment on a straight line in a 4-dimensional space (x, y, z, t), namely, the intersection of the hyperplanes with equations $x + y + z + t = 1$, $x = a(x + y)$ and $z = b(z + t)$.

In the most general case, assuming $0 < a, b < 1$, the constraints can be written

$$y = \frac{1-a}{a}\,x\,, \quad z = \frac{b}{1-b}\,t\,, \quad \frac{x}{a} + \frac{z}{b} = 1\,,$$

or equivalently

$$x = \frac{a}{1-b}(1-b-t)\,, \quad y = \frac{1-a}{1-b}(1-b-t)\,,$$
$$z = \frac{b}{1-b}\,t\,, \quad 0 \leq t \leq 1-b\,.$$

Then, the set $\mathcal{P}$ is the segment bounded by the probabilities $(a, 1-a, 0, 0)$ and $(0, 0, b, 1-b)$. It can be checked that this result still holds when $a = b = 1$: In this particular case, the constraints

$$x = a(x+y)\,, \quad z = b(z+t)\,, \quad x+y+z+t = 1\,,$$

become $x = x+y\,, \quad z = z+t\,, \quad x+z = 1$. Then, the set $\mathcal{P}$ is the segment bounded by the probabilities $(1, 0, 0, 0)$ and $(0, 0, 1, 0)$. We also note that, as it could be easily verified, the initial assessment $P(E|H) = a, P(E|H^c) = b$ is coherent in the sense of de Finetti [22] for every $a \in [0,1], b \in [0,1]$. The consistency of conditional probability assessments can be checked by a geometrical approach (see Appendix), or considering suitable sequences of probability functions (see, e.g., [13]). Hence, the constraints are always satisfiable.

Note that $P(E|H), P(E|H^c)$ often reflects generic knowledge (sometimes interpreted causally) expressing the probabilities of observing events of the form E when H occurs or when its contrary occurs, respectively. Then these probabilities refer to a population of situations where the occurrence of events of the form E was checked when H was present or absent. This population may be explicitly known (as in statistics) or not (for instance we know that birds fly but the concerned population of birds is ill-defined). On the contrary, the observation E is contingent, it pertains to the current situation, and nothing is then assumed on the probability of occurrence of events of the form E in the population. In this case, it is not legitimate to interpret the observation E as a (new) constraint $P(E) = 1$, which would mean that events of the form E are always the case, while we just want to represent the fact that event E has been observed now. [1]

Suppose the prior probability $P(H)$ is provided by an agent. Clearly it must be interpreted in a generic way (in general, events of the form H have this propensity to be present); otherwise, if $P(H)$ were only the contingent belief of the agent now, one may not be able to use it on the same grounds as the conditional probabilities so as to uniquely define a probability function in $\mathcal{P}$ (since we do not interpret the contingent but sure observation E as having probability 1). As a consequence, when the prior probability $P(H)$ is specified, our generic knowledge also includes the posterior probability $P(H|E)$, which we extract for the reference class E (as we

[1] Assuming $P(E) = 1$ is often done in probability kinematics [41], where conditioning is understood as revising a probability function

know the current situation is in the class of situations where E is true). In a second (inductive) step, the value $P(H|E)$ can be used by the agents for measuring their belief in the hypothesis H to be present now, given that E is observed.

A common, but fallacious objection to the above remark can be as follows: suppose that the agent interprets $P(E|H)$, $P(E|H^c)$ as contingent conditional belief degrees of observing E if H is present or not present in the current situation. In that case, since these values are interpreted as contingent uncertain beliefs, one may be tempted to interpret the observation of evidence in a strong way, as $P(E) = 1$, especially in the case where the prior probability of H is unknown. Unfortunately, the equality $P(E) = 1$ is inconsistent with $P(E|H) = a$ and $P(E|H^c) = b$ since they imply $a \geq P(E) \geq b$. So the formal framework cannot support the interpretation of $P(E|H), P(E|H^c)$ as contingent conditional belief degrees, unless we interpret conditioning as probability revision [41].

2.2 Requirements for cogent abduction without prior

The following conditions could be considered as minimal prerequisites for an abduction method to qualify as being reasonable:

(1) *The formal model should be faithful to the available information* : it should not select a unique probabilistic model if there is no reason for it. The assignment of a unique probabilistic prior in the situation of ignorance can be seen as a useful suggestion to apply Bayes theorem. But it then should be justified in some reasonable way.
(2) *The solution should be non-trivial*: the approach should not result in total ignorance, when $P(E|H) \neq P(E|H^c)$, since likelihood functions do express information. The fundamental Polya abductive pattern in the logical setting, whereby if H implies E and E is true then H becomes plausible should be retrieved.
(3) *The chosen approach should always provide a solution*: it should not lead to a logical contradiction, since likelihood functions are consistent with any prior probability, and the case where the prior probability is assigned is one of maximal information.
(4) *The method should be principled*: there should be a formal framework that can support the inference results of the abduction process, no ad hoc solution is searched for.
(5) *The solution should be intuitive and plausible*: the method should not yield an unreasonable result that commonsense would obviously dismiss.

In the following we first check if past proposals to the problem satisfy these requirements.

3 Three standard approaches

In the literature, three approaches exist that try to cope with ignorance of the prior probability. The first approach is based on varying the prior probability on the expression of $P(H|E)$ derived from Bayes theorem. Unfortunately the posterior probability remains totally unknown in this case, even if some zero probabilities prevent the standard approach from being carried out, as shown in the second approach, using de Finetti's coherence framework [22]. Another classical approach in non-Bayesian statistics relies on the relative values of $P(E|H)$ and $P(E|H^c)$ being interpreted as the likelihood of H and its complement. In this approach, the idea of computing a posterior probability is given up. The only way of ascertaining a hypothesis under this approach is by rejecting its complement. It turns out that this approach is consistent with possibility theory [19].

In the following, we recall these approaches in some detail.

3.1 Imprecise Bayes

The most obvious thing to do in the absence of prior is to perform sensitivity analysis on Bayes theorem. Let $P(H) = p$ be an unknown parameter. Then

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} = \frac{a \cdot p}{a \cdot p + b \cdot (1 - p)} \,.$$

But the value p is anywhere between 0 and 1. Clearly the corresponding range of $P(H|E)$ is $[0, 1]$. So this approach brings no information on the plausibility of the hypothesis, making the observation of evidence and the presence of the generic knowledge useless, in contradiction with requirements 2 and 5. Indeed, one feels prone to consider that evidence E should confirm H if for instance a is high and b is low. The above analysis presupposes $a \cdot p + b \cdot (1 - p) \neq 0$.

Two cases result in $a \cdot p + b \cdot (1 - p) = 0$. First the case when $P(E \mid H^c) = b = 0$ and $P(H) = p = 0$ (the case when $P(E \mid H) = a = 0$ implies $b = 0$ by construction); finally the case when $a = b = 0$, while $p > 0$.

First consider the case

$$P(E|H) = a > 0\,,\ P(E|H^c) = 0\,,\ P(H) = 0\,,$$

(so that $P(E) = 0$); what can be said about $P(H|E)$? It can be proved that $P(H|E) \in [0, 1]$. We observe that $P(E|H^c) = 0$ implies $z = 0$; $P(H) = 0$ implies $x + y$ = 0. So, $t = 1$. Since the only constraint acting on $P(H \mid E)$ is

$x = P(H \mid E)(x+z)$, the latter just reduces to $0 = P(H \mid E) \cdot 0$, so $P(H \mid E)$ is unconstrained.

Now, what can be said if we assume $P(E|H) = P(E|H^c) = 0$, so that $P(E) = 0$, without assuming $P(H) = 0$? It implies $x = z = 0$ so that $t + y = 1$. Then since by definition $x = P(H|E)(x+z)$, it all reduces to $0 = P(H|E) \cdot 0$ again. So, the range of $P(H|E)$ still remains $[0, 1]$.

However we cannot fully rely on the above approach in the case when some probabilities are zero. Suppose for instance the assessment $P(H \mid E \vee H) = \frac{1}{2}$ were added to the two current ones. It can be shown that the assessment $P(H \mid E) = p_1\,,\ P(H \mid E \vee H) = p_2$ is coherent if and only if $p_1 \leq p_2$. If there are no zero probabilities, this is obvious since we know that $P(H|E) \leq P(H|E \vee H)$. The conclusion $0 \leq P(H \mid E) \leq \frac{1}{2}$ is then obvious. But if

$$P(E|H) = a > 0\,,\ P(E|H^c) = 0\,,\ P(H) = 0\,,\ P(H|E \vee H) = \frac{1}{2}$$

which implies $P(E) = P(H) = P(E \vee H) = 0$, it follows $x = y = z = 0, t = 1$; then the assessments above, and $P(H|E) = \gamma$, lead to the constraints

$$\begin{aligned} & x = a(x+y)\,,\ z = 0(z+t)\,,\ x+y = 0\,, \\ & x+y = \frac{1}{2}(x+y+z)\,,\ x = \gamma(x+z)\,, \end{aligned}$$

that is

$$0 = a \cdot 0\,,\ 0 = 0 \cdot 1\,,\ x = y = 0\,,\ 0 = \frac{1}{2} \cdot 0\,,\ 0 = \gamma \cdot 0\,.$$

Hence, there is no way to obtain $P(H \mid E) \leq \frac{1}{2}$ since this system only implies $0 \leq \gamma \leq 1$. Only by using a general methodology, such as the coherence-based approach of de Finetti, may we be sure of properly handling all (explicit or implicit) constraints.

3.2 Coherence approach

The coherence setting of de Finetti [22] allows a sound handling of zero probabilities of conditioning events. In fact, zero probabilities on relevant relationships might occur easily in practice. For instance, Bernard [3] referred to a statistical investigation on the religious behaviour of people, in which no individuals were present that pray often, while not going to church regularly nor giving their children any religious education. Nevertheless, one might be interested in elaborating relationships between the event E = *pray_often* $\wedge\neg$*regular_church* $\wedge$ $\neg$*religious_education* and the event H = *Believing_in_paradise*, also studied in that investigation. If the expert directly evaluates probabilities by observed frequencies, she obtains $P(EH) = P(EH^c) = 0$ and so are $P(E|H)$ and $P(E|H^c)$. Then the conditional probability $P(H|E)$ is the indeterminate form $\frac{0}{0}$; hence a difficulty. Now suppose that some expert assesses the probabilities $P(E|H)$ and $P(E|H^c)$, based on some general information. These assessments would not depend on the fact that in some data collection experiment, the frequencies of relevant events are zero. In general, not excluding zero probabilities is often needed for hypothetical or counterfactual reasoning. However, the problem of zero probabilities could arise again, for instance in the (extreme) case in which the expert asserts $P(H) = P(E|H^c) = 0$. In such cases, direct reasoning as in the previous subsection in general may fail, as shown above.

So, our abduction problem must be treated in the coherence framework of de Finetti to make sure results we found in the zero probability cases are correct. Theoretical details on this approach are given in the Appendix. Two cases must be considered : $P(H) = b = 0$, and $P(H) > 0, a = b = 0$. We analyze the first pathological case. The second pathological case can be analyzed by similar reasoning. So consider the problem:

$$P(E|H) = a > 0\,,\ P(E|H^c) = 0\,,\ P(H) = 0\,.$$

It corresponds to the assignment $\mathbf{p} = (a, 0, 0, \gamma)$ to the tuple

$$\mathcal{F} = (E|H\,,\ E|H^c\,,\ H|\Omega\,,\ H|E).$$

To check coherence of $\mathbf{p}$, as a first step we have to consider the "constituents" (interpretations) generated by $\mathcal{F}$ and contained in the disjunction of the conditioning events $H \vee H^c \vee \Omega \vee E = \Omega$ (here, the sure event), which are

$$C_1 = EH\,,\ C_2 = E^cH\,,\ C_3 = EH^c\,,\ C_4 = E^cH^c\,,$$

(in the case we are examining, the complement C_0 of the disjunction of the conditioning events is not a constituent, as it coincides with the impossible event $\emptyset$). Let

$$E_1|H_1 = E|H, E_2|H_2 = E|H^c, E_3|H_3 = H|\Omega, E_4|H_4 = H|E.$$

To each basic constituent C_h assign a vector $Q_h = (q_{h1}, q_{h2}, q_{h3}, q_{h4})$, where, for each conditional indexed by $j = 1, 2, 3, 4$, component q_{hj} is defined as in equation (A.3) in the Appendix. Namely

- $h = 1$: $C_1 = EH \subseteq E_1H_1, E_3H_3, E_4H_4$;

- $h = 2$: $C_2 = E^c H \subseteq E_3 H_3, H_4^c$;
- $h = 3$: $C_3 = E H^c \subseteq H_1^c, E_2 H_2$;
- $h = 4$: $C_4 = E^c H^c \subseteq H_1^c, H_4^c$.

Then, in geometrical terms, we get the points

$$Q_1 = (1, 0, 1, 1), Q_2 = (0, 0, 1, \gamma), Q_3 = (a, 1, 0, 0), Q_4 = (a, 0, 0, \gamma),$$

and, denoting by $\mathcal{I}$ the convex hull of $Q_1, \dots, Q_4$, we check the (necessary) coherence condition $\mathbf{p} \in \mathcal{I}$, that is the existence of a non-negative vector (x, y, z, t) such that

$$\mathbf{p} = xQ_1 + yQ_2 + zQ_3 + tQ_4\,, \quad x + y + z + t = 1.$$

This comes down to writing exactly the system of equations defining the set of probabilities $\mathcal{P}$ in section 2.1. Of course, the consistency of this system is in general necessary but not sufficient for the coherence of $\mathbf{p}$; it becomes sufficient when, like in the previous subsection, we exclude conditioning events of zero probability. In our pathological case the system becomes:

$$\begin{aligned} &x = a(x + y), \quad z = 0, \quad x + y = 0, \quad x = \gamma(x + z), \\ &x + y + z + t = 1, \; x \geq 0, \; y \geq 0, \; z \geq 0, \; t \geq 0. \end{aligned} \tag{1}$$

As $\mathbf{p} = Q_4$, the condition $\mathbf{p} \in \mathcal{I}$ is satisfied (i.e., the system is solvable), with $x = y = z = 0\,,\; t = 1$, for every $\gamma \in [0, 1]$. Notice that the solution of the above system, $(x, y, z, t) = (0, 0, 0, 1)$ is a probability function on the set of constituents $\{C_1, C_2, C_3, C_4\}$. With this probability function are associated, for the conditioning events H, H^c, Ω, E, the following probabilities

$$\begin{aligned} &P(H) = x + y = 0\,, \quad P(H^c) = z + t = 1\,, \\ &P(\Omega) = x + \cdots + t = 1\,, \quad P(E) = x + z = 0\,. \end{aligned}$$

Then, we must continue to check coherence on the sub-family of conditional events whose conditioning events have zero probability; that is, we have to check as a second step the coherence of the assessment $\mathbf{p}_0 = (a, \gamma)$ to $\mathcal{F}_0 = (E|H, H|E)$. Constituents in $H \vee E$ are $C_1 = EH\,, C_2 = E^c H\,, C_3 = E H^c$, with associated points: $Q_1 = (1, 1)\,, Q_2 = (0, \gamma)\,, Q_3 = (a, 0)$ (in the case we are examining, the disjunction of the conditioning events is $H \vee E$, so that $m = 3$ and $C_0 = (H \vee E)^c = H^c E^c$). As we can verify, the condition $\mathbf{p}_0 \in \mathcal{I}_0$ holds, that is

$$\begin{aligned} &\mathbf{p}_0 = xQ_1 + yQ_2 + zQ_3\,, \quad x + y + z = 1\,, \\ &x \geq 0\,,\; y \geq 0\,,\; z \geq 0\,, \end{aligned}$$

(whose geometrical meaning is that $\mathbf{p}_0$ belongs to the triangle $Q_1Q_2Q_3$). It amounts to solving the linear system

$$\begin{aligned}&x = a(x+y), \;\; x = \gamma(x+z),\\&x+y+z = 1, \; x \geq 0, \; y \geq 0, \; z \geq 0,\end{aligned} \tag{2}$$

solvable with

$$x = \frac{a\gamma}{a+\gamma(1-a)}\,, \quad y = \frac{(1-a)\gamma}{a+\gamma(1-a)}\,, \quad z = \frac{a(1-\gamma)}{a+\gamma(1-a)}\,,$$

for every $\gamma \in [0,1]$.

Notice that the vector $(x,y,z,t) = \left(\frac{a\gamma}{a+\gamma(1-a)},\, \frac{(1-a)\gamma}{a+\gamma(1-a)},\, \frac{a(1-\gamma)}{a+\gamma(1-a)},\, 0\right)$ is a probability function on the set of constituents $\{C_1, C_2, C_3, C_0\}$, where $C_0 = E^cH^c$. With this probability function, the following probabilities are associated with the conditioning events H, E:

$$\begin{aligned}P(H) &= x+y = \frac{\gamma}{a+\gamma(1-a)} \geq 0\,,\\P(E) &= x+z = \frac{a}{a+\gamma(1-a)} > 0\,.\end{aligned}$$

As the set of conditioning events with zero probability is empty or equal to $\{H\}$, the assessment $\mathbf{p}_0 = (a,\gamma)$ is coherent for every $\gamma \in [0,1]$; therefore, the initial assessment $\mathbf{p} = (a,0,0,\gamma)$ is coherent for every $\gamma \in [0,1]$. In other words, the range of $P(H|E)$ remains $[0,1]$.

We observe that in the above checking of coherence we used the following sequence of two probability functions:

$$P_0 = (0,0,0,1), \; P_1 = \left(\frac{a\gamma}{a+\gamma(1-a)},\, \frac{(1-a)\gamma}{a+\gamma(1-a)}, \frac{a(1-\gamma)}{a+\gamma(1-a)},\, 0\right).$$

We remark that: (i) at the second step the study is restricted to the sub-family $\mathcal{F}_0 = \{E|H, H|E\}$; (ii) the probabilities of the constituents not contained in (the disjunction of conditioning events) $E \vee H$ are equal to zero; in fact, the variable t, associated with $(E \vee H)^c = E^cH^c$, is zero. Using the notion of zero-layer (see [13]), with the above probability functions the following holds: If $\gamma > 0$, H^c (and of course Ω) belongs to the most normal zero-layer with level 0, while at the second zero-layer with level 1 are E and H; if $\gamma = 0$, H^c (and Ω) is in the zero-layer of level 0, E is at level 1, and the zero-layer of H is at level 2.

As shown above, differently from what happens in the usual approach to probabilistic reasoning, in the more general setting of coherence (where conditioning events are allowed to have zero probability) the admissibility of a given probability assessment on a family of conditional events amounts to the existence of (in general infinite) sequences of probability distributions defined on the relevant context (set of constituents). In fact, the machinery of coherence has the "minimal" aim of determining the set of all admissible probability assessments, without specifying a particular status for any of them. The choice of a particular assessment in such set mainly depends on how the expert weighs his information. In other words, coherence is a syntactic (not semantic) tool and, like with imprecise probabilities, it does not suggest a particular way of solving the abduction problem for $P(H|E)$.

3.3 Likelihood approach

Non-Bayesian statisticians (e.g. [20], [1]) consider $P(E|H)$ to be the likelihood of H, denoted by $L(H)$. When $P(E|H) = 1$, H is only fully plausible. When it is 0 (the probability $P(E|H^c)$ being positive) it rules out hypothesis H upon observing E. But there is no formal justification given to the notion of likelihood, usually, thus violating requirement 4. We are in a dilemma as the sensitivity approach is probabilistically founded but provides no information while the likelihood approach is informative but looks ad hoc in a probabilistic setting.

Note that the likelihood approach is also in agreement with a default Bayesian approach: in the absence of a prior probability, assume it is uniformly distributed. Then the posterior probability is $P(H|E) = \frac{a}{a+b}$, so that it is equivalent to renormalize the likelihood functions in the probabilistic style. This fact has been recurrently used to claim that the likelihood approach is like the Bayesian approach with a uniform prior. Even if the likelihood approach looks consistent with the uniform prior (Bayes) method, the former has no pretence to compute precise posterior probabilities: results it provides are informative only if one of a or b is small (and not the other). Saying that the likelihood approach is a special case of the Bayesian approach is like saying that an unknown probability distribution and a uniform probability distribution mean the same thing.

Dubois, Moral, and Prade [18] suggested that $L(H)$ can be viewed as an upper probability bound and also a degree of possibility: generally the quantity $P(A|B)$ is upper bounded by $\max_{x\in B} P(A|x)$, and as pointed out by [13], if set-function L is assumed to be inclusion-monotonic (as expected if we take it for granted that L means likelihood), then $P(A|B) = \max_{x\in B} P(A|x)$ is the only possible choice if only $P(A|x)$ is known for all x.

In this sense the likelihood approach, common in non-Bayesian statistics comes down to interpreting conditional probabilities in terms of possibility theory [19].

The quantity $P(E|H)$ can be used to eliminate assumption H if it is small enough in front of $P(E|H^c)$, but evaluating that $P(E|H) = 1$ is not sufficient to ascertain H.

4 Relying on information principles

One way out of the dilemma of abduction without priors is to introduce additional information by means of default assumptions that are part of the implicit background knowledge. The idea is that in the absence of prior probability, one finds a (default) probability measure in $\mathcal{P}$ in some way, relying on principles of information faithfulness, maximal independence assumptions, or symmetry assumptions, respectively [32]. Then the posterior beliefs of agents is dictated by the default probability thus selected. Unfortunately, as seen below the results obtained by means of the various principles are not fully consistent with each other.

4.1 The maximum likelihood principle

The maximum likelihood principle says that if an event occurred then this is because it was at the moment the most likely event. So the best probabilistic model in a given situation is the one which maximizes the probability of occurrence of the observed evidence. This principle is often used to pick a probability distribution in agreement with some data. For instance, assume we observe k heads and $n - k$ tails from tossing a coin n times. The probability function underlying the process is completely determined by the probability of heads, say x. To find the best value of x, one maximizes the likelihood $L(x) = P(E|x) = x^k \cdot (1 - x)^{n-k}$, where $E =$ "k heads and $n - k$ tails" and we find $x = \frac{k}{n}$. Interestingly, since x completely defines the probability measure P on $\{$tail, head$\}$, $P(E|x) = L(P)$, i.e. the likelihood of model P.

So, the maximum likelihood approach selects a plausible probabilistic model, with a view to solve the abduction problem in a second step. In our problem, we interpret $P(E)$ as the likelihood of the probability function P after E occurred. In our case, E occurred, so it is legitimate to establish the agent's posterior (contingent) belief about H assuming $P(E)$ is as large as possible under the constraints $P(E|H) = a < 1$ and $P(E|H^c) = b \leq a$. Again, in that case we interpret $P(E)$ as the likelihood of the probability function P to be selected among those such that $P(E|H) = a, P(E|H^c) = b$, while the non-Bayesian statistics approach directly chooses between H and H^c on the basis of their likelihoods. Here we first try to select a plausible probabilistic model, and then, solve the abduction problem.

Note that $P(E) = a \cdot p + b \cdot (1 - p)$ whose maximum is $P(E) = a$, which unfortunately enforces $p = 1$. It comes down to assuming $P(H) = 1$, so that $P(H|E) = 1$, too. This is clearly too strong to be credible, even under a weak interpretation of the posterior probability (H is present in the situation where E was observed). However note that in this approach the constraint $P(E) = a$ is not added to mean that the probability of E is indeed a in the population. It just assumes that the population of realizations relevant for the current situation is the one where E is as likely as possible, so that in the current situation, $\mathcal{P}$ can be restricted to $\{P \in \mathcal{P}, P(E)$ is maximal$\}$.

In any case, this approach violates requirement 5, as being counterintuitive. A way out of this difficulty will be proposed later on in this paper, as the *relaxed maximum likelihood approach*.

4.2 Maximizing entropy

A fairly popular informational principle is the maximization of entropy (e.g. [32]). Entropy quantifies the indeterminateness inherent to a probability distribution P by $H(P) = -\sum_\omega P(\omega)\log P(\omega)$. Given a set $\mathcal{R} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$ of probabilistic conditionals, the *principle of maximum entropy*

$$\max H(Q) = -\sum_\omega Q(\omega)\log Q(\omega),$$
$$\text{s.t. } Q \text{ is a distribution satisfying } \mathcal{R}$$

solves (uniquely) the problem of representing $\mathcal{R}$ by a probability distribution without adding information unnecessarily. The resulting distribution is denoted by $ME(\mathcal{R})$. The maximum entropy solution is often interpreted as a least committed probability, i.e. the one involving maximal indeterminateness in a subsequent decision process. In fact, maximum entropy processes conditional dependencies especially faithfully, and independence between events is implemented only if no information to the contrary can be derived. We will recall very briefly some facts on the principle of maximum entropy that are needed to solve the problem considered here; for further details, maximum entropy distribution, see e.g. [29].

Using well-known Lagrange techniques, we may represent $ME(\mathcal{R})$ in the form

$$ME(\mathcal{R})(\omega) = \lambda_0 \prod_{i:\omega\models A_iB_i} \alpha_i^{1-x_i} \prod_{i:\omega\models A_iB_i^c} \alpha_i^{-x_i} \tag{3}$$

with the α_i's being exponentials of the Lagrange multipliers, one for each conditional in $\mathcal{R}$, and λ_0 simply arises as a normalizing factor.

The maximum entropy solution to our problem can be computed as follows. Let

P_{me} be the maxent distribution in $\mathcal{P}$ and we use the notation $\alpha = \frac{a}{1-a}$, $\beta = \frac{b}{1-b}$. Here, the probabilistic information given is represented by $\mathcal{R} = \{(E|H)[a], (E|H^c)[b]\}$, so $P_{me} = ME(\mathcal{R})$. Let $\lambda_a^+ = \alpha^{1-a}$, $\lambda_a^- = \alpha^{-a}$, $\lambda_b^+ = \beta^{1-b}$, $\lambda_b^- = \beta^{-b}$ with a normalizing constant $\lambda_0 = (\alpha^{-a}(1-a)^{-1} + \beta^{-b}(1-b)^{-1})^{-1}$. Using equation (3) we get the following probabilities:

ω	$P_{me}(\omega)$	ω	$P_{me}(\omega)$
EH	$\lambda_0\lambda_a^+$	E^cH	$\lambda_0\lambda_a^-$
EH^c	$\lambda_0\lambda_b^+$	E^cH^c	$\lambda_0\lambda_b^-$

Now, it immediately follows that

$$P_{me}(H|E) = \frac{\alpha^{1-a}}{\alpha^{1-a} + \beta^{1-b}} = \frac{a \cdot a^{-a}(1-a)^{a-1}}{a \cdot a^{-a}(1-a)^{a-1} + b \cdot b^{-b}(1-b)^{b-1}},$$

and

$$P_{me}(H) = \lambda_0(\lambda_a^+ + \lambda_a^-) = \frac{a^{-a}(1-a)^{a-1}}{a^{-a}(1-a)^{a-1} + b^{-b}(1-b)^{b-1}}.$$

Furthermore, also the maxent probability of E can be calculated, and it turns out that this probability is obtained by ME-fusing the given probabilities a and b (in the sense of [30]):

$$P_{me}(E) = \lambda_0(\lambda_a^+ + \lambda_b^+) = \frac{a^{1-a}(1-a)^{a-1} + b^{1-b}(1-b)^{b-1}}{a^{-a}(1-a)^{a-1} + b^{-b}(1-b)^{b-1}}. \tag{4}$$

Remark A more elementary approach, only good for the particular problem at hand, is as follows. Every probability in $\mathcal{P}$ has the form

$$(ka, k(1-a), (1-k)b, (1-k)(1-b))$$

where $k = P(H)$, and its entropy amounts to

$$\begin{aligned} H = H(k) &= -ka \log ka - k(1-a) \log k(1-a) - (1-k)b \log(1-k)b \\ &\quad -(1-k)(1-b) \log(1-k)(1-b) \\ &= -k \log k - (1-k) \log(1-k) - ka \log a - k(1-a) \log(1-a) \\ &\quad -(1-k)b \log b - (1-k)(1-b) \log(1-b). \end{aligned}$$

The principle of maximum entropy selects the unique probability distribution P_{me} with maximum entropy in $\mathcal{P}$: so let us compute the value of k where the derivative

$$\frac{dH}{dk} = -(\log k - \log(1-k) + a\log a + (1-a)\log(1-a)$$
$$-b\log b - (1-b)\log(1-b))$$

vanishes. Solving $\frac{dH}{dk} = 0$ for k yields

$$\frac{k}{1-k} = \frac{b^b(1-b)^{1-b}}{a^a(1-a)^{1-a}},$$

which is equivalent to

$$k = \frac{b^b(1-b)^{1-b}}{a^a(1-a)^{1-a} + b^b(1-b)^{1-b}} = \frac{a^{-a}(1-a)^{a-1}}{a^{-a}(1-a)^{a-1} + b^{-b}(1-b)^{b-1}}.$$

With that k, we obtain, as expected,

$$P_{me}(H|E) = \frac{ka}{ka + (1-k)b} = \frac{a^{1-a}(1-a)^{a-1}}{a^{1-a}(1-a)^{a-1} + b^{1-b}(1-b)^{b-1}}.$$

Example 1 *We will study the example from the introduction in this framework, considering a roadwork as a possible explanation for a traffic jam. Here* $E =$ *traffic jam*, $H =$ *roadwork, and we assume* $P(E|H) = 0.9, P(E|H^c) = 0.2$. *Using the formal machinery from above, the maximum entropy probability* $P_{me}(H|E)$ *turns out to be* 0.791. *Therefore, roadwork appears to be a suitable explanation for the traffic jam.*

4.3 Shapley value as pignistic probability

The Shapley value was first proposed in cooperative game theory [34], to extract from a set of weighted coalitions of agents (a non-additive set-function), an assessment of the individual power of each agent (a probability distribution).

The Shapley value is defined as follows. Consider the lower probability function induced by the set $\mathcal{P}$, i.e $\forall A \subseteq \Omega (= \{E, E^c\} \times \{H, H^c\})$ $P_*(A) = \inf\{P(A), P \in \mathcal{P}\}$. For each permutation σ of elements of Ω, a probability distribution p_σ can be generated from P_*, letting, for $i = 2, \ldots n$,

$$p_\sigma(\omega_{\sigma(i)}) = P_*(\{\omega_{\sigma(1)}, \ldots, \omega_{\sigma(i)}\}) - P_*(\{\omega_{\sigma(1)}, \ldots, \omega_{\sigma(i-1)}\}).$$

The Shapley value is the average of these $n!$ (possibly identical) probability distributions, and it can be written, if $\omega \in \Omega$,

$$p_{Sh}(\omega) = \frac{s!(n-s-1)!}{n!} \sum_{S \subseteq \Omega, \omega \notin S} (P_*(S \cup \{\omega\}) - P_*(S)),$$

where s and n are cardinalities of S and Ω, respectively.

For convex capacities, it is the center of mass of the set $\mathcal{P}$ (which then coincides with set of probability functions$\{P \geq P_*\}$[35]). In the theory of belief functions, it is known as the "pignistic transformation" [37].

Selecting the Shapley value comes down to assuming that all probabilities in $\mathcal{P}$ are equally probable so that by symmetry the center of mass of this polyhedron can be chosen by default as the best representative probability function in this set. This is similar as replacing a solid by its center of mass for studying its kinematics. As shown above, $\mathcal{P}$ is a segment on a straight line, bounded by the probabilities $(a, 1-a, 0, 0)$ and $(0, 0, b, 1-b)$. It can be routinely proved using the above equation, that the Shapley value is the midpoint of this segment, i.e. $(\frac{a}{2}, \frac{1-a}{2}, \frac{b}{2}, \frac{1-b}{2})$. Under this default probability,

$$P_{Sh}(H|E) = \frac{a}{a+b}$$

that is, the Shapley value supplies the same response as the Bayesian approach where a uniform prior is assumed! This is not too surprising as the Shapley value can be seen as assuming a uniform metaprobability over the probability set induced by the constraints, and considering the average probability resulting from this meta-assessment. The above result suggests that assigning a uniform prior to assumptions and assuming a uniform metaprobability over the probability polygon come down to the same result.

Example 2 *We consider Example 1 in the Shapley framework. Here, we find easily* $P_{Sh}(H|E) = 0.818$. *The result is similar to that calculated by the maximum entropy approach.*

4.4 Comparative discussion

Contrary to the simple form, in some sense natural, of the Shapley value, the maximum entropy solution looks hard to interpret in the problem at hand, at first glance. But there is a similarity of form between them, except that the maxent solution distorts the influences of the probabilities a and b by the function

$$f(x) = \left(\frac{x}{1-x}\right)^{1-x}$$

so that the maxent solution for $P(H|E)$ takes the same form as the Shapley value, after distortion, namely, $\frac{f(a)}{f(a)+f(b)}$. Alternatively, one may see the maxent solution as defining a default prior, assuming $P(H) = \frac{w(a)}{w(a)+w(b)}$ depending on coefficients a and b, where $w(x) = x^{-x}(1-x)^{x-1}$, so that $P(H|E)$ takes the form $\frac{a \cdot w(a)}{a \cdot w(a)+b \cdot w(b)}$.

Note that $\log w(x)$ is the entropy of the probability distribution $(x, 1-x)$. So $w(x)$ is all the higher as the distance between x and 0.5 is smaller. So the prior probability selected by the maxent approach basically reflects the relative proximity from $P(E|H)$ to 0.5, and $P(E|H^c)$ to 0.5, regardless of their being greater or less than 0.5. For instance the cases where $a = b = 0.9$, $a = b = 0.1$ and where $a = 0.9, b = 0.1$ yield the same default prior probabilities. The value of weighting function w is not altered by exchanging a and $1-a$, (and b and $1-b$); $w(x)$ takes on values in $[1, 2]$ so that $P(H)$ lies in the interval $[\frac{1}{3}, \frac{2}{3}]$ with $P_{me}(H) = 0.5$ if and only if $a = b$ or $a = 1-b$. In other words, this weighting function shrinks the $[0, 1]$ range of prior probabilities symmetrically around 0.5. This makes maximum entropy more cautious, i.e. returning in general probabilities which are closer to 0.5, according to the maxent philosophy of not introducing determinateness unnecessarily. In the Shapley approach, $P_{Sh}(H) = 0.5$ is an invariant, independent of a and b.

As a and b approach the extreme probabilities 1 resp. 0, the maxent solution approaches the Shapley value. In fact, we have $P_{Sh}(H|E) = P_{me}(H|E)$ if and only if $a = b$, or $a = 1-b$. In the first case, H and E are statistically independent, in the second case, the influence of H on E is symmetrical – its presence makes E probable to the same extent as its absence makes it improbable, which can be understood as a generalization of logical equivalence to the probabilistic case. This reflects a strong symmetric dependence between E and H. What makes Shapley value bolder in the scope of maxent is that both approaches coincide only when E and H are either independent, or very strongly related. In fact, a (the degree of the presence of H) has a positive effect throughout on the probability $P_{Sh}(H|E)$ whereas b (the degree of the absence of H) has a negative effect. This means that increasing a or decreasing b always results in an increase of $P_{Sh}(H|E)$ which can be explained, e.g., by assuming H to be an essential cause of E.

As opposed to this, the maximum entropy probability processes information in a more unbiased way, i.e. without assuming either strong dependence or independence in general. But note, that when such a relationship seems plausible (in the cases $a = b$ or $a = 1-b$), then it coincides with the Shapley value.

A general comparison between the inference process based on center of mass propagation (resulting in the Shapley value) and that by applying the maxent principle was made in [32]. Paris showed that center of mass inference violates some properties that reasonable probabilistic inference processes should obey. More precisely, in general, center of mass inference can not deal appropriately with irrelevant information and with (conditional) independencies. For the problem that we focus on in this paper, however, the Shapley value seems to be as good a candidate for reasonable inference as the maximum entropy value, regarding invariance with respect to irrelevant evidence.

Overall, it seems that the maximum entropy approach is syntactically similar to both the Shapley approach (since there exists similar implicit default priors in both approaches) and the maximum likelihood approach (posterior probabilities are proportional to likelihoods or some function thereof) for solving the abduction problem.

However, the particular form of the maximum entropy solution is hard to interpret in the problem at hand. So, requirement 5 is met better by the Shapley value than by the maximum entropy solution. As to the other requirements, these two approaches are quite similar: Both always provide a solution, rely on formal frameworks and are non-trivial (requirements 2,3, and 4). Since they pick unique probability distributions for solving the problem, both methods violate requirement 1 in a strict sense, although one might argue that they do so for good reasons.

5 A relaxed maximum likelihood approach

The reason why the maximum likelihood fails is that maximizing $P(E)$ on $\mathcal{P}$ enforces $P(H) = 1$. It may mean that the available knowledge is too rigidly modelled as precise conditional probability values.

As pointed out in [21], the symbol $E|H$ stands as a three-valued entity, not a Boolean one as it distinguishes between examples EH, counterexamples E^cH and irrelevant situations H^c. Authors like [27] and [16] have claimed that $E|H$ can be identified with the pair $(EH, EH \vee H^c)$ of events (an interval in the Boolean algebra), or with the triple (EH, E^cH, H^c) that forms a partition of the universal set. And indeed (provided that $P(H) > 0$) $P(E|H)$ is a function of $P(EH)$ and $P(E \vee H^c)$; namely

$$P(E|H) = \frac{P(EH)}{P(EH) + 1 - P(E \vee H^c)}.$$

If $P(H) = 0$, i.e., $P(EH) = 0, P(E \vee H^c) = 1$, it can be verified that the assessment $(0, 1, z)$ on $\{EH, E \vee H^c, E|H\}$ is coherent for every $z \in [0, 1]$. Moreover, under minimal positivity conditions [27], $P(A|B) \leq P(C|D), \forall P$ if and only if $AB \subseteq CD$ and $A \vee B^c \subseteq C \vee D^c$ (or, equivalently, $AB \subseteq CD$ and $C^cD \subseteq A^cB$).

Now, it is important to realize that $E|H$ is a kind of mid-term between EH and $E \vee H^c$ since $P(E \vee H^c) \geq P(E|H) \geq P(EH)$. So it makes sense to interpret the conditional knowledge as $P(E \vee H^c) \geq a \geq P(EH)$ and $P(E \vee H) \geq b \geq P(EH^c)$, respectively. This is consistent with the original data due to the above remarks, which also show that the new formulation is a relaxation of the previous one.

According to the maximum likelihood principle, the default probability function should now be chosen such that $P(E) = x + z$ is maximal, under constraints:

$$P(E \vee H^c) \geq a \geq P(EH);\ P(E \vee H) \geq b \geq P(EH^c)$$

and we assume here a positive likelihood function $a \geq b > 0$. The problem then reads:

$$\text{maximize } x + z \text{ such that } 1 - y \geq a \geq x;\ \ 1 \geq x + y + z \geq b \geq z.$$

Since $a \geq x$, $b \geq z$, $x + z \leq a + b$, the maximal value of $P(E)$ is $P^*(E) = \min(1, a + b)$.

Now there may be more than one probability measure maximizing $P(E)$. In order to compute the posterior probability, $P(H|E)$, we are led to the problem of maximizing and minimizing $P(EH) = x$ subject to

$$1 - y \geq a \geq x,\ \ 1 \geq x + y + z \geq b \geq z,\ \ x + z = \min(1, a + b).$$

Proposition 1 *Under the conditional event approach, assuming a positive likelihood function* $P(E|H^c) = b \leq a = P(E|H)$, *for the maximum likelihood posterior probability,* $P(H|E)$, *we have*

(1) if $a + b \geq 1$ *then* $P(H|E) \in [1 - b, a]$.
(2) $P(H|E) = \frac{a}{a+b}$, *otherwise.*

Proof. When $a+b \geq 1$ then $x+z = 1$, then $y = 0$ is enforced. Hence the constraints of the problem reduce to:

$$a \geq x,\ b \geq 1 - x.$$

Then $x = P(EH) = P(H \mid E) \in [1 - b, a]$.
If $a + b < 1$, then $P(E) = x + z = a + b$. From this and $a \geq x, b \geq z$, it follows directly, that $x = a, z = b$ must hold, which yields $P(H|E) = \frac{x}{x+z} = \frac{a}{a+b}$. □

Example 3 *Studying our running example 1 in this framework is easily done. For the hypothesis roadwork, we have* $a + b \geq 1$, *so it is straightforward to see that here,* $P($*roadwork*$|$*traffic jam*$) \in [0.8, 0.9]$.

Framing the problem within the setting of the de Finetti coherence approach encompasses the case of zero probabilities. Given two quantities a and b in the interval $[0, 1]$, we assign the unspecified quantities $\mathbf{p} = (x,\ z,\ \alpha,\ \beta,\ \gamma,\ p)$ to the vector of conditional events $(EH,\ EH^c,\ E \vee H^c,\ E \vee H,\ E,\ H|E)$, where $EH = EH|\Omega$, and so o n. We want to obtain all the coherent values of $\mathbf{p}$ under the constraints $P(E \vee H^c) \geq a \geq P(EH)\,,\quad P(E \vee H) \geq b \geq P(EH^c)$ and the condition that

γ is maximum. Then, we obtain the following proposition, that takes into account all cases.

Proposition 2 *In the framework of coherence, the maximum likelihood posterior probability, $P(H|E)$, is only known to lie in the interval $[p', p'']$ such that :*

(1) if $a = b = 0$, then $p' = 0, p'' = 1$;
(2) if $a > 0,\ b = 0$, then $p' = p'' = 1$;
(3) if $a > 0,\ b > 0,\ a + b \geq 1$, then $p' = 1 - b, p'' = a$;
(4) $a > 0,\ b > 0,\ a + b < 1$, then $p' = p'' = \frac{a}{a+b}$.

Proof. To check the coherence of $\mathbf{p}$ we use again the constituents $C_1 = EH\ ,\ C_2 = E^cH\ ,\ C_3 = EH^c\ ,\ C_4 = E^cH^c$, and the associated 6-dimensional points $Q_1, \ldots,$ Q_4. For example, $Q_1 = (1, 0, 1, 1, 1, 1)$, and so on. Then, at the first step we check the condition $\mathbf{p} \in \mathcal{I}$, which amounts to the solvability of the following system (in the unknowns x, y, z, t, with non negative parameters α, β, γ, p)

$$\alpha = x + z + t\,, \beta = x + y + z\,,\ \ \gamma = x + z\,, x = p(x + z)\,,$$
$$x + y + z + t = 1\,,\ \ x \geq 0,\ y \geq 0,\ z \geq 0,\ t \geq 0\,,$$

subject to maximize $(x + z)$ when $x \leq a \leq x + z + t\,,\ \ z \leq b \leq x + y + z\,.$ Now, let us consider the different cases:

Case 1: $a = b = 0$; in this case $x = z = \max(x+z) = 0$; then, the system becomes

$$\alpha = t\,,\ \ \beta = y\,,\ \ \gamma = 0\,,\ \ 0 = p \cdot 0\,,$$
$$y + t = 1\,,\ \ x = 0,\ y \geq 0,\ z = 0,\ t \geq 0\,,$$

and, of course, is solvable for every $p \in [0, 1]$; hence the range of $P(H|E)$ is $[p', p''] = [0, 1]$.

Case 2: $a > 0,\ b = 0$; in this case $z = 0,\ \max(x + z) = \max\ \ x = a$ and the system can be written as

$$\alpha = a + t\,,\ \ \beta = a + y\,,\ \ \gamma = a\,,\ \ a = p \cdot a\,,$$
$$x + y + t = 1\,,\ \ x = a,\ y \geq 0,\ z = 0,\ t \geq 0\,.$$

Of course, the system is solvable if and only if $p = 1$, so $P(H|E) = p' = p'' = 1$.

Case 3: $a > 0,\ b > 0,\ a + b \geq 1$; in this case $\max(x + z) = 1,\ y = t = 0$ and the system becomes

$$\begin{aligned}
&\alpha = 1\,,\;\; \beta = 1\,,\;\; \gamma = 1\,,\;\; x = p\,,\\
&x + z = 1\,,\;\; 0 \le x \le a,\, y = 0,\, 0 \le z = 1 - x \le b,\, t = 0\,.
\end{aligned}$$

Then, the system is solvable for every $1 - b \le p = x \le a$, so that the range of $P(H|E)$ is $[p', p''] = [1 - b, a]$.

Case 4: $a > 0$, $b > 0$, $a + b < 1$; in this case $\max(x + z) = a + b$, $x = a$, $z = b$ and the system becomes

$$\begin{aligned}
&\alpha = a + b + t\,,\;\; \beta = a + b + y\,,\\
&\gamma = a + b\,,\;\; a = p(a + b)\,,\\
&a + b + y + t = 1\,,\;\; x = a,\, y \ge 0,\, z = b,\, t \ge 0\,.
\end{aligned}$$

Then, the system is solvable if and only if $p = \frac{a}{a+b}$, so that $P(H|E) = p' = p'' = \frac{a}{a+b}$.□

While confirming the results of the previous proposition, the coherence approach solves three additional cases with zero probabilities. When $a = 0$ and $b \neq 0$ or when $a \neq 0$ and $b = 0$, one of the assumptions H or its contrary are eliminated. When $a = b$, we get either $P(H|E) = P(H^c|E) = \frac{1}{2}$ if $a \in (0, \frac{1}{2})$; equal upper probabilities a on H and its contrary if $a > 1/2$; we also get the same result (upper probabilities 1, that is, total ignorance) for both $a = b = 0$ and for $a = b = 1$.

These results are not so surprising, even if new to our knowledge. This approach, in opposition to the ones in the previous section does not necessarily enforce a default prior. When $P(E|H)$ and $P(E|H^c)$ are large, we only find upper probabilities $P^*(H|E) = a$ and $P^*(H^c|E) = b$ (since the lower probability $P_*(H|E) = 1 - b$), which is in agreement with the interpretation of the likelihoods $L(H) = P(E|H)$ and $L(H^c) = P(E|H^c)$ as degrees of possibility (or upper probabilities). The larger they are the less information is available on the problem. In particular when $a = b = 1$, the likelihood function is a uniform possibility distribution on $\{H, H^c\}$ that provides no information (indeed $P(E|H) = P(E|H^c) = 1$ means that both H and H^c are possible). It is natural that the observation E should not inform at all about H in this case, that is, it is intuitively satisfying that $P(H|E) \in [0, 1]$ (total ignorance) even assuming $P(E) = 1$. If $a = b$ both increase from 0.5 to 1, our knowledge on the posterior evolves from equal probabilities on the hypothesis and its contrary to higher order uncertainty about them, ending up with total ignorance.

On the contrary, when $P(E|H)$ and $P(E|H^c)$ are small, the maximum likelihood solution in this case is a unique probability $P(H|E) = \frac{a}{a+b}$. This is the result obtained by the Bayesian approach under uniform priors and by the Shapley value of the probability sets induced by the likelihood functions. In this case the available knowledge, under maximum likelihood assumption, is rich enough to provide much information upon observing evidence, under the maximum likelihood principle.

When $P(E|H^c)$ is much smaller than $P(E|H)$, the maximum likelihood principle enables hypothesis H^c to be eliminated. It supplies a unique probability measure proportional to (a, b) if both values are small enough.

This new approach to handling abduction without priors has some advantages. It reconciliates the maximum likelihood principle (that failed due to an overconstrained problem) and the ad hoc likelihood-based inference of non-Bayesian statistics. But it also recovers the Shapley value and the Bayesian approach with a uniform prior in some situations. It confirms the possibilistic behavior of likelihood functions, being all the more uninformative as the likelihood of the hypothesis and of its complement are both close to 1. When they are both low but positive, the Bayesian approach with a uniform prior is recovered. When one of a and b is zero, then the hypothesis with zero likelihood is unsurprisingly disqualified by observing E. However, in the case when both likelihoods are zero or one, it results in total ignorance about the posterior probability of the hypothesis. So even if it is in partial agreement with some of the other techniques, this new approach is, in its spirit, also at odds with the maximum entropy method, with Shapley value and the uniform Bayes approach as well, all of which treat the cases $a = b < 0.5$ and $a = b > 0.5$ likewise.

The relaxed maximum likelihood approach is similarly well-behaved as Shapley value and maximum entropy, but it avoids sticking to the idea of selecting a unique probabilistic model. So, it satisfies all of our requirements, though it is not axiomatized, as Shapley value or the maximum entropy solution are. But it follows the maximum likelihood principle, as opposed to the simple ad hoc use of likelihood functions.

6 Conclusion

One of the traditional disputes in probability theory opposes the Bayesian approach whereby any state of knowledge can be characterized by a single probability function on the suitable space, and classical statistics where likelihood functions are often empirically estimated but subjective prior probabilities are not considered to be relevant information. The Bayesian approach has the merit of offering a complete and harmonious solution, but the price paid is, as already stressed in the past, that either a full data collection is needed or a debatable representation of ignorance in the form of prior probabilities must be adopted. The classical statistics approach

Method	Faithful	Non-trivial	Principled	Plausible
Uniform Bayes	NO	YES	YES	YES
Imprecise Bayes	YES	NO	YES	NO
Likelihood Reasoning	YES	YES	usually NO	YES
Maximum Likelihood	NO	NO	YES	NO
Maximum Entropy	NO	YES	YES	Debatable
Laplace-Shapley	NO	YES	YES	YES
Relaxed Maximum Likelihood	YES	YES	YES	YES

Fig. 1. Comparison of different approaches

may look as lacking formal foundations despite the existing rationales for this pragmatic approach. This paper has tried to put together many tools proposed in additive and non-additive probability theories so as to sort out the issue of unknown priors.

Several approaches to the problem of probabilistic abduction have been reviewed, and some novel solutions have been proposed, based on maximum entropy, Shapley value and maximum likelihood reasoning. This study suggests that the key issue is a suitable representation of the available probabilistic knowledge, and a suitable choice of a reasoning principle. Table 1 summarizes the performance of the considered approaches with respect to the criteria 1, 2, 4, 5 laid bare in the methodology section. As a result of this paper, some light is shed on the classical statistics approach and the maximum likelihood principle, by casting them in the framework of possibility and imprecise probability theories. The paper also shows the noticeable agreement between the use of Shapley value and the classical Bayesian assumption of uniform priors under ignorance. The maximum entropy approach is shown to significantly differ from the Bayesian tradition of uniform priors and the non-Bayesian approach based on likelihoods. Indeed, the selected $P(H)$ depends on the relative distance between the likelihoods of H and H^c and 0.5. The farther $P(E|H)$ to 0.5 compared to $P(E|H^c)$ the more informative H turns out to be. Only the maximum likelihood bluntly applied when likelihoods are known leads to a contradiction (criterion 3). Our new maximum likelihood approach under a relaxed interpretation of the causal knowledge provides an original solution to the probabilistic abduction problem that bridges the gap between the straightforward use of likelihood functions and the assumption of a uniform prior, being more informative than the pure sensitivity analysis approach but less precise than the Bayesian, Shapley and maxent solutions when the likelihoods are too high to enable any hypothesis rejection.

It could be interesting to develop the work made in this paper by applying the relaxed maximum likelihood approach to more general knowledge bases, and also for notions of coherence other than coherent inference. In particular, the case of multiple-valued universes for hypotheses and pieces of evidence is worth investigating.

More work is also needed to fully interpret the obtained results. In particular, a systematic comparative study of first principles underlying the Shapley value and the maximum entropy approach is certainly in order. We should also compare our results with what the imprecise probability school [39] and the Transferable Belief Model [36] have to say about this problem in a more careful way. The problem discussed in this paper is indeed closely connected with the issue of statistical inference with binomial data, when little knowledge about prior uncertainty is available. In the standard statistical literature, the so-called 'objective Bayesian aproach' (starting with works by Jeffreys) is devoted to the search of alleged uninformative priors. More recently, the imprecise Dirichlet model of Walley (see a survey paper by Bernard [2] and Utkin and Augustin [38]) deals with how to infer information about the parameter x ruling a binomial experiment, given some observations (modelled by likelihood functions $L(x)$ as in section 4.1) and prior information given under the form of a set of priors each having the form of a beta distribution. It is clearly related with our concerns here, since it includes the case when all possible beta priors are allowed. Here we do not assume any prior at all. Moreover Bernard's paper also recalls a set of principles that statistical inference without prior information should obey (symmetry, representation invariance, dependence on likelihood function, and coherence in the sense of Walley), that complement and refine the more general criteria discussed here.

Moreover, another point to study is the influence of irrelevant information on the results of the various approaches. Finally, in order to evaluate the cognitive plausibility (see Requirement 5) of the different approaches more thoroughly, psychological testing could be carried out with experts.

Acknowledgement. We thank the referees for their useful suggestions and comments.

References

[1] V. Barnett. *Comparative Statistical Inference* . J. Wiley, New York, 1973.

[2] J.-M. Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *Int. J. Approx. Reasoning*, 39:123–150, 2005.

[3] J.M. Bernard. Implicative analysis for multivariate binary data using an imprecise Dirichlet model. In G. de Cooman, F.G. Cozman, S. Moral, and P. Walley, editors, *Proc. of ISIPTA '99*, Ghent, Belgium, 29 June – 2 July 1999.

[4] V. Biazzo and A. Gilio. A generalization of the fundamental theorem of de Finetti for imprecise conditional probability assessments. *International Journal of Approximate Reasoning*, 24:251–272, 2000.

[5] V. Biazzo and A. Gilio. On the linear structure of betting criterion and the checking of coherence. *Annals of Mathematics and Artificial Intelligence*, 35:83–106, 2002.

[6] V. Biazzo, A. Gilio, T. Lukasiewicz, and G. Sanfilippo. Probabilistic Logic under Coherence, Model-Theoretic Probabilistic Logic, and Default Reasoning in System . *Journal of Applied Non-Classical Logics*, 12(2):189–213, 2002.

[7] V. Biazzo, A. Gilio, T. Lukasiewicz, and G. Sanfilippo. Probabilistic Logic under Coherence: Complexity and algorithms. *Annals of Mathematics and Artificial Intelligence*, 45:35–81, 2005.

[8] A. Cano and S. Moral. A review of propagation algorithms for imprecise probabilities. In G. de Cooman, F.G. Cozman, S. Moral, and P. Walley, editors, *Proc. of ISIPTA '99*, Ghent, Belgium, 29 June – 2 July 1999.

[9] A. Capotorti. Configurations of locally strong coherence in the presence of conditional exchangeability (the case of cardinality $k \leq 3$). In F.G. Cozman, R. Nau, and T. Seidenfeld, editors, *Proc. of ISIPTA '05*, Pittsburgh, Pennsylvania, USA, 20–23 July 2005.

[10] G. Coletti. Coherent numerical and ordinal probabilistic assessments. *IEEE Trans. on Systems, Man, and Cybernetics*, 24(12):1747–1754, 1994.

[11] G. Coletti and R. Scozzafava. Characterization of coherent conditional probabilities as a tool for their assessment and extension. *Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 4(2):103–127, 1996.

[12] G. Coletti and R. Scozzafava. Conditioning and inference in intelligent systems. *Soft Computing*, 3(3):118–130, 1999.

[13] G. Coletti and R. Scozzafava. *Probabilistic logic in a coherent setting*. Kluwer Academic Publishers, 2002.

[14] F.G. Cozman. Computing posterior upper expectations. In G. de Cooman, F.G. Cozman, S. Moral, and P. Walley, editors, *Proc. of ISIPTA '99*, Ghent, Belgium, 29 June – 2 July 1999.

[15] D.Dubois and H.Prade. Modeling uncertain and vague knowledge in possibility and evidence theories. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, pages 303–318. North-Holland, Amsterdam, 1990.

[16] D.Dubois and H.Prade. Conditional objects as nonmonotonic consequence relationships. *IEEE Trans. on Systems, Man and Cybernetics*, 24(12):1724–1740, 1994.

[17] C. de Campos and F. Cozman. Computing lower and upper expectations under epistemic independence. *Int. J. Approx. Reasoning*, 44(3):244–260, 2007.

[18] D. Dubois, S. Moral, and H. Prade. A semantics for possibility theory based on likelihoods. *J. of Mathematical Analysis and Applications*, 205:359–380, 1997.

[19] D. Dubois and H. Prade. *Possibility theory*. Plenum Press, New-York, 1988.

[20] W. F. Edwards. *Likelihood*. Cambridge University Press, Cambridge, U.K., 1972.

[21] B. De Finetti. La logique de la probabilité. In *Actes du Congrès Inter. de Philosophie Scientifique*, volume 4, Paris, 1936. Hermann et Cie Editions.

[22] B. De Finetti. *Theory of Probability*, volume 1. Wiley, London, 1974. English translation of *Teoria delle probabilità* (1970).

[23] D. Gale. *The Theory of Linear Economic Models*. McGraw-Hill, New York, 1960.

[24] C. Geyer, R. Lazar, and G. Meeden. Computing the joint range of a set of expectations. In F.G. Cozman, R. Nau, and T. Seidenfeld, editors, *Proc. of ISIPTA '05*, Pittsburgh, Pennsylvania, USA, 20–23 July 2005.

[25] A. Gilio. Criterio di penalizzazione e condizioni di coerenza nella valutazione soggettiva della probabilità. In *Bollettino Unione Matematica Italiana (7)*, volume 4-B, pages 645–660, 1990.

[26] A. Gilio. Algorithms for precise and imprecise conditional probability assessments. In G. Coletti, D.Dubois, and R. Scozzafava, editors, *Mathematical Models for Handling Partial Knowledge in Artificial Intelligence*, pages 231–254. Plenum Publ. Co., New York, 1995.

[27] I.R. Goodman, H.T. Nguyen, and E.A. Walker. *Conditional Inference and Logic for Intelligent Systems: A Theory of Measure-Free Conditioning*. North-Holland, Amsterdam, 1991.

[28] P. Hansen, B. Jaumard, M.P. de Aragao, F. Chauny, and S. Perron. Probabilistic satisfiability with imprecise probabilities. In G. de Cooman, F.G. Cozman, S. Moral, and P. Walley, editors, *Proc. of ISIPTA '99*, Ghent, Belgium, 29 June – 2 July 1999.

[29] G. Kern-Isberner. *Conditionals in nonmonotonic reasoning a nd belief revision*. Springer, Lecture Notes in Artificial Intelligence LNAI 2087, 2001.

[30] G. Kern-Isberner and W. Rödder. Belief revision and information fusion on optimum entropy. *International Journal of Intelligent Systems*, 2004.

[31] R. Lazar and G. Meeden. Exploring imprecise probability assessments based on linear costraints. In J.M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *Proc. of ISIPTA '03*, Lugano, Switzerland, 14–17 July 2003. Carleton Scientific.

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

[33] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeto, 1976.

[34] L.S. Shapley. A value for n-person games. In Kuhn and Tucker, editors, *Contributions to the Theory of Games, II*, pages 307–317. Princeton Univ. Press, 1953.

[35] L.S. Shapley. Core of convex games . *Int. J. Game Theory*, 1:11–26, 1971.

[36] P. Smets. Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. *Int. J. of Approximate Reasoning*, 9:1–35, 1993.

[37] P. Smets and R. Kennes. The transferable belief model . *Artificial Intelligenc*, 66:191–234, 1994.

[38] L.V. Utkin and T. Augustin. Decision making under incomplete data using the imprecise dirichlet model. *Int. J. Approx. Reasoning*, 44(3):322–338, 2007.

[39] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.

[40] P. Walley, R. Pelessoni, and P. Vicig. Direct Algorithms for Checking Coherence and Making Inferences from Conditional Probability Assessments. *Journal of Statistical Planning and Inference*, 126(1):119–151, 2004.

[41] P.M. Williams. Bayesian conditionalization and the principle of minimum information. *British J. for the Philosophy of Sciences*, 31:131–144, 1980.

A Appendix: Coherent conditional probability assessments

We recall basic results on the handling of linear equations in checking coherence of conditional probability assessments stemming from conditional probability assessments, as pioneered by de Finetti (see, e.g., [5], [10], [11–13], [22], [25,26], [40]). Given an arbitrary family of conditional events $\mathcal{K}$ and a real function $P: \mathcal{K} \to \mathcal{R}$, let us consider a sub-family $\mathcal{F} = \{E_1|H_1, \ldots, E_n|H_n\} \subseteq \mathcal{K}$, and the vector $\mathbf{p} = (p_1, \ldots, p_n)$, where $p_i = P(E_i|H_i)\,, \quad i = 1, \ldots, n\,.$ We denote by $\mathcal{H}_n$ the disjunction $H_1 \vee \cdots \vee H_n$. Notice that, as $E_iH_i \vee E_i^cH_i \vee H_i^c = \Omega\,, \;\; i = 1, \ldots, n$, by expanding the expression $\bigwedge_{i=1}^n (E_iH_i \vee E_i^cH_i \vee H_i^c)$, we can represent Ω as the disjunction of 3^n logical conjunctions, some of which may be impossible. The remaining ones are the constituents generated by the family $\mathcal{F}$. We denote by $C_1, \ldots, C_m$ the constituents contained in $\mathcal{H}_n$ and (if $\mathcal{H}_n \neq \Omega$) by C_0 the further constituent $\mathcal{H}_n^c = H_1^c \cdots H_n^c$, so that $\mathcal{H}_n = C_1 \vee \cdots \vee C_m\,, \quad \Omega = \mathcal{H}_n^c \vee \mathcal{H}_n = C_0 \vee C_1 \vee \cdots \vee C_m\,, \quad m + 1 \leq 3^n\,.$

Coherence with betting scheme: Using the same symbols for the events and their indicators, with the pair $(\mathcal{F}, \mathbf{p})$ we associate the random gain

$$\mathcal{G} = \sum_{i=1}^n s_i H_i (E_i - p_i)\;,$$

where $s_1, \ldots, s_n$ are n arbitrary real numbers. Let g_h be the value of $\mathcal{G}$ when C_h is true. Of course $g_0 = 0$ (notice that g_0 will play no role in the definition of coherence). We denote by $\mathcal{G}|\mathcal{H}_n$ the restriction of $\mathcal{G}$ to $\mathcal{H}_n$; hence $\mathcal{G}|\mathcal{H}_n \in \{g_1, \ldots, g_m\}$, $\min\, \mathcal{G}|\mathcal{H}_n = \min\{g_1, \ldots, g_m\}$, $\max\, \mathcal{G}|\mathcal{H}_n = \max\{g_1, \ldots, g_m\}$.

Then, the function P defined on $\mathcal{K}$ is said *coherent* if and only if, for every integer n, for every finite sub-family $\mathcal{F} \subseteq \mathcal{K}$ and for every $s_1, \ldots, s_n$, one has

$$\min\ \mathcal{G}|\mathcal{H}_n\ \leq 0 \leq \max\ \mathcal{G}|\mathcal{H}_n\ , \tag{A.1}$$

or equivalently

$$\max\ \mathcal{G}|\mathcal{H}_n\ \geq 0 \quad (\min\ \mathcal{G}|\mathcal{H}_n\ \leq 0)\,. \tag{A.2}$$

Coherence with penalty criterion: de Finetti [22] has proposed another operational definition of probabilities, which can be extended to conditional events [25]. With the pair $(\mathcal{F}, \mathbf{p})$ we associate the loss $\mathcal{L} = \sum_{i=1}^{n} H_i(E_i - p_i)^2$; we denote by L_h the value of $\mathcal{L}$ if C_h is true. If You specify the assessment $\mathbf{p}$ on $\mathcal{F}$ as representing your belief's degrees, You are required to pay a penalty L_h when C_h is true. Then, the function P is said *coherent* if and only if do not exist an integer n, a finite sub-family $\mathcal{F} \subseteq \mathcal{K}$, and an assessment $\mathbf{p}^* = (p_1^*, \ldots, p_n^*)$ on $\mathcal{F}$ such that, for the loss $\mathcal{L}^* = \sum_{i=1}^{n} H_i(E_i - p_i^*)^2$, associated with $(\mathcal{F}, \mathbf{p}^*\ )$, it results $\mathcal{L}^* \leq \mathcal{L}$ and $\mathcal{L}^* \neq \mathcal{L}$; that is $L_h^* \leq L_h\ ,\ \ h = 1, \ldots, m\,$, with $L_h^* < L_h$ in at least one case.
Notice that the betting scheme and the penalty criterion are equivalent [25]; this means that a probability assessment $\mathbf{p}$ is coherent under the betting scheme if and only if it is coherent under the penalty criterion.
If P is coherent, then it is called a *conditional probability* on $\mathcal{K}$. Notice that, if P is coherent, then P satisfies all the well known properties of conditional probabilities (while the converse is not true; see [26], Example 8; or [13], Example 13).

We can develop a geometrical approach to coherence by associating, with each constituent C_h contained in $\mathcal{H}_n$, a point $Q_h = (q_{h1}, \ldots, q_{hn})$, where

$$q_{hj} = \begin{cases} 1, \ \text{if} \ \ C_h \subseteq E_j H_j, \\ 0, \ \text{if} \ \ C_h \subseteq E_j^c H_j, \\ p_j, \ \text{if} \ \ C_h \subseteq H_j^c. \end{cases} \tag{A.3}$$

Denoting by $\mathcal{I}$ the convex hull of the points $Q_1, \ldots, Q_m$, based on the penalty criterion, the following result can be proved ([25])

Theorem 1 The function P is coherent if and only if, for every finite sub-family $\mathcal{F} \subseteq \mathcal{K}$, one has $\mathbf{p} \in \mathcal{I}$.

Notice that, if $\mathcal{F} = \{E \mid H\}$, $\mathbf{p} = (P(E \mid H)) = (p)$, we have

$$\mathbf{p} \in \mathcal{I} \iff \begin{cases} p = 0, & EH = \emptyset, \\ p = 1, & EH = H, \\ p \in [0,1], & \emptyset \subset EH \subset H. \end{cases}$$

Then, by Theorem 1, it immediately follows

Corollary 3 A probability assignment $P(E \mid H) = p$ is coherent iff it holds that

$$\begin{cases} p = 0, & EH = \emptyset, \\ p = 1, & EH = H, \\ p \in [0,1], & \emptyset \subset EH \subset H. \end{cases}$$

The betting scheme and the penalty criterion are *equivalent*, due to the following results:
(i) the condition $\mathbf{p} \in \mathcal{I}$ amounts to solvability of the following system $(\mathcal{S})$ in the unknowns $\lambda_1, \dots, \lambda_m$

$$\begin{cases} \sum_{h=1}^{m} q_{hj}\lambda_h = p_j \ , \quad j = 1, \dots, n\,; \\ \sum_{h=1}^{m} \lambda_h = 1 \ , \quad \lambda_h \geq 0\,, \ h = 1, \dots, m. \end{cases}$$

(ii) let $\mathbf{x} = (x_1, \dots, x_m)$, $\mathbf{y} = (y_1, \dots, y_n)^t$ and $A = (a_{ij})$ be, respectively, a row $m-$vector, a column $n-$vector and a $m \times n-$matrix. The vector $\mathbf{x}$ is said *semipositive* if $x_i \geq 0, \forall\, i,$ and $x_1 + \cdots + x_m > 0$. Then, we have (cf. [23], Theorem 2.9)

Theorem 2 Exactly one of the following alternatives holds.
(i) the equation $\mathbf{x}A = 0$ has a *semipositive* solution;
(ii) the inequality $A\mathbf{y} > 0$ has a solution.

We observe that, choosing $a_{ij} = q_{ij} - p_j, \forall\, i, j$, the solvability of $\mathbf{x}A = 0$ means that $\mathbf{p} \in \mathcal{I}$, while the solvability of $A\mathbf{y} > 0$ means that, choosing $s_i = y_i, \forall\, i$, one has $\min\ \mathcal{G}|\mathcal{H}_n\ > 0$ (and hence $\mathbf{p}$ would be incoherent). Therefore, applying Theorem 2 with $A = (q_{ij} - p_j)$, we obtain $\max\ \mathcal{G}|\mathcal{H}_n\ \geq 0$ iff $(\mathcal{S})$ is solvable, that is, $\max\ \mathcal{G}|\mathcal{H}_n\ \geq 0$ iff $\mathbf{p} \in \mathcal{I}$.

Checking coherence. It could seem that, in order to verify coherence, we should check the condition $\mathbf{p} \in \mathcal{I}$ for every $\mathcal{F} \subseteq \mathcal{K}$ (which tends to become intractable).

We show that this is not the case, by restricting the attention to the checking of coherence of the assessment $\mathbf{p}$ on $\mathcal{F}$. Let S be the set of solutions $\Lambda = (\lambda_1, \ldots, \lambda_m)$ of the system $(\mathcal{S})$. Then, define $\Phi_j(\Lambda) = \Phi(\lambda_1, \ldots, \lambda_m) = \sum_{r:C_r \subseteq H_j} \lambda_r \;, \quad j = 1, \ldots, n\,;\ M_j = \max\ \Phi_j(\Lambda)\;, \quad j = 1, \ldots, n\,; \quad I_0 = \{j \in J\,:\, M_j = 0\}$. Notice that I_0 is a strict subset of $\{1, \ldots, n\}$. We denote by $(\mathcal{F}_0, \mathbf{p}_0)$ the pair associated with I_0.
Given the pair $(\mathcal{F}, \mathbf{p})$ and a subset $J \subset \{1, \ldots, n\}$, we define $\mathcal{H}_J = \bigvee_{j \in J} H_j$.
Moreover, we denote by $(\mathcal{F}_J, \mathbf{p}_J)$ the pair associated with J and by $(\mathcal{S}_J)$ the corresponding system.
We observe that $(\mathcal{S}_J)$ is solvable if and only if $\mathbf{p}_J \in \mathcal{I}_J$, where $\mathcal{I}_J$ is the convex hull associated with the pair $(\mathcal{F}_J, \mathbf{p}_J)$. Then, it can be proved:

Theorem 3 Given the assessment $\mathbf{p}$ on $\mathcal{F}$, assume that $(\mathcal{S})$ is solvable, i.e. $\mathbf{p} \in \mathcal{I}$, and let J be a subset of $\{1, \ldots, n\}$. If there exists a solution $(\lambda_1, \ldots, \lambda_m)$ of $(\mathcal{S})$ such that $\sum_{r:C_r \subseteq \mathcal{H}_J} \lambda_r > 0$, then $(\mathcal{S}_J)$ is solvable, i.e. $\mathbf{p}_J \in \mathcal{I}_J$.

Theorem 4 Given the assessment $\mathbf{p}$ on $\mathcal{F}$, assume that $(\mathcal{S})$ is solvable, i.e. $\mathbf{p} \in \mathcal{I}$. Then, for every $J \subset \{1, \ldots, n\}$ such that $J \setminus I_0 \neq \emptyset$, one has $\mathbf{p}_J \in \mathcal{I}_J$.

By the previous results, we obtain:

$$\mathbf{p} \text{ coherent } \Longleftrightarrow \begin{cases} \mathbf{p} \in \mathcal{I}\ ; \\ \text{if } I_0 \neq \emptyset, \text{ then } \mathbf{p}_0 \text{ is coherent.} \end{cases}$$

Then, we can check coherence by the following procedure:

Algorithm 1 Let the pair $(\mathcal{F}, \mathbf{p}\ )$ be given.

(1) Construct the system $(\mathcal{S})$ and check its solvability;
(2) If the system $(\mathcal{S})$ is not solvable then $\mathbf{p}$ is not coherent and the procedure stops, otherwise compute the set I_0;
(3) If $I_0 = \emptyset$ then $\mathbf{p}$ is coherent and the procedure stops; otherwise set $(\mathcal{F}, \mathbf{p}) = (\mathcal{F}_0, \mathbf{p}_0)$ and repeat steps 1-3.

Notice that similar results and methods can be used for checking generalized coherence and for propagation of imprecise conditional probability assessments ([26]). The coherence-based approach to probabilistic reasoning with imprecise probabilities has been studied in many papers ([4,5], [13], [39], [40]). In particular, modelling uncertainty by conditional probability bounds, the relationship between coherence-based probabilistic reasoning and model-theoretic probabilistic reasoning has been examined in [6]. In [7], among other things, a complete study of the complexity of coherence-based probabilistic reasoning has been made.