1st International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium, 29 June - 2 July 1999

Coherent Upper and Lower Bayesian Updating

Giulianella Coletti
University of Perugia, Italy
coletti@dipmat.unipg.it

Romano Scozzafava
University of Roma "La Sapienza", Italy
romscozz@dmmm.uniroma1.it

Abstract

The main concern of this paper is to show by means of suitable examples that a "naif" use of Bayesian updating can lead to wrong conclusions. Given some possible diseases (that could explain an initial piece of information) and a relevant tentative probability assessment, a doctor has usually at his disposal also a data base consisting of conditional probabilities $P(E|K)$, where each K is a disease and each evidence E comes from a suitable test. Once the coherence (à la de Finetti) of the whole assessment is checked, we want to suitably update the prior probabilities: since we do not assume that the diseases constitute a partition of the certain event Ω, the usual Bayes theorem cannot be applied. Then we proceed by referring to the relevant atoms (whose coherent probability assessment is, in a sense, "imprecise", since in general it is *not unique*). By checking again the coherence of the whole updated assessment, it turns out that we get upper and lower conditional probabilities. These steps are iterated until a degree of belief sufficient to make a diagnosis is reached: the coherence condition acts as a control tool on every stage.

Keywords. Coherence, Bayesian updating, upper and lower probabilities.

1 Introduction

Even if probability theory has not played, for a number of reasons, its proper role for uncertain knowledge representation and processing, the theoretical framework proposed by de Finetti [10] seems particularly flexible: it differs radically from the usual one (based on a measure-theoretic approach), which assumes that a unique probability measure is defined on the set of "elementary events", constituting the so-called sample space. De Finetti's approach allows instead to assess your (coherent) probability for as many or as few events as you feel able and interested, and this has many important theoretical and applied consequences; in particular, it makes simpler and more effective the "operational" aspects.

This – and, mainly, the generalization to conditional probability – has been discussed in many papers: see, for instance, [3], [2], [6] and, for a qualitative approach through a Bayes-like theorem, see [4]. Concerning the problem of medical diagnosis, which is the main concern of this paper (see Sects. 6 and 7), some preliminary results dealing with the difficulties arising when trying to get rid from the simplifying assumption of mutually exclusive and exhaustive diseases were already discussed in [5]. Conditional independence (see, e.g., [14]) is another usual assumption that our method aims at avoiding, when the motivation is just that of getting a simplification of the relevant setting. Our approach (the general theory is expounded in [6]) refers to learning from data as an operational procedure based on partial probability assessments and updating: the latter are not necessarily unique, and the procedure is ruled by coherence through an algorithm involving *linear* systems and *linear programming*. The computational difficulties can be strongly reduced by resorting to a technique (see [7]) based on the assignment *of zero probability* to some *conditioning* events H_i in such a way that the corresponding *direct* assessments $P(E_i|H_i)$ are coherent. What does *direct* assessment of a conditional probability $P(E|H)$ mean? Given an arbitrary family $\mathcal{C}$ of conditional events, a function $P(\cdot|\cdot)$, bound to *satisfy only the requirement of coherence*, can be defined on $\mathcal{C}$: so the knowledge (or the assessment) of the "joint" and the "marginal" probabilities $P(E \wedge H)$ and $P(H)$ – as in the "Kolmogorovian" framework – is not required. It is well known that such a function $P(\cdot|\cdot)$ satisfies on $\mathcal{C}$ the *axioms* (see Sect.3) of a conditional probability.

The so–called "imprecise probabilities" come naturally to the fore, since we are driven only by coherence and not by the "myth of uniqueness" of a probability assessment.

Concerning "imprecise probabilities", some remarks are now in order. In particular, we agree with de Finetti ([10], p. 368 of the English translation), whose answer to the question *"Do imprecise probabilities exist?"* is essentially (as we see it) YES and NO. To clarify this issue, let us take some excerpts from the quoted reference: "The question as it stands is rather ill-defined, and we must first of all make precise what we mean. In actual fact, there is no doubt that quantities can neither be measured, nor thought of as really defined with the absolute precision demanded by mathematical abstraction ... A subjective evaluation, like that involved in expressing a probability, attracts this criticism to an even greater degree ... It should be sufficient to say that all probabilities, like all quantities, are in practice imprecise, and that in every problem involving probability one should provide, just as one does for other measurements, evaluations whose precision is adequate in relation to the importance of the consequences that may follow ... The question posed originally, however, really concerns a different issue, one which has been raised by several authors: it concerns the possibility of cases in which one is not able to speak of a single value p for a given probability, but rather of two values, p' and p'', which bound an area of indeterminacy, $p' \leq p \leq p''$, possessing some essential significance ... An example of this occurs when one wishes to distinguish various hypotheses, and attributes different probabilities $P(E|H_i)$ to an event E, depending on the various hypotheses H_i ; if one then ignores the hypotheses, one can only conclude that the probability lies between the maximum and the minimum ... If we are dealing with hypotheses H_i about which we expect soon to have some information, it would be naive to assert that $P(E)$ will take on a value lying somewhere between the $P(E|H_i)$'s, since there is an infinite number of partitions into hypotheses, and the information which comes along might be anything at all ... The idea of translating the imprecision into bounds, $p' \leq p \leq p''$, even in the weaker sense proposed by Good (who regards p' and p'' not as absurd, rigid bounds, capable of *making the imprecision precise*, but merely as indications of maxima), is inadequate if one wishes to give an idea of the imprecision with which every quantity is known or can be considered. One should think of the imprecision in the choice of the function P ... for individual events ... not as isolated features, but with the connections deriving from logical or probabilistic relations".

Similar remarks concerning the function P are shared - through some subtle considerations concerning indeterminacy - by Williams (see [16], p.231), who claims also, at the beginning of the quoted paper: "It has been objected against the subjective interpretation of probability that it assumes that a subject's degree of belief $P(E)$ in any event or proposition E is an exact numerical magnitude which might be evaluated to any desired number of decimal places ... The same argument, however, would appear to show that no empirical magnitude can satisfy laws expressed in the classical logico-mathematical framework, so long as it is granted that indeterminacy, to a greater or lesser extent, is present in all empirical concepts".

Nevertheless it could be interesting to study coherence of a probability assessment possibly involving both "precise" and "imprecise" evaluations: the most genuine situation in an updating process is that in which we get – as (coherent) extension of an initial coherent assessment – an upper and a lower conditional probability; now, if we want to go on in the updating by taking into account new "information" (for example, some further conditional probability values), we need checking the "global" coherence – as lower and upper probability – of the new values and the previous upper and lower probability. The relevant theory is dealt with in [8] and an actual case is discussed at the end of Sect.6.

In Sections 2–5 we recall the main points of our approach: for a more detailed exposition, see [6], [9].

2 Coherent Probability

An *event* can be singled-out by a (nonambiguous) *proposition* E, that is a statement that can be either *true* or *false* (corresponding to the two "values" 1 or 0). Since in general it is not known whether E is true or not, we are *uncertain* on E. In our framework, probability is looked upon as an "ersatz" for the lack of information on the actual "value" of the event E, and it is interpreted as a measure of the *degree of belief* in E held by the *subject* that is making the assessment. So a careful distinction between the *meaning* of probability and all its possible *methods of evaluation* is essential.

The role of *coherence* is in fact that of ruling probability evaluations concerning "many" events, independently of any logical structure of the given family of events. Even if its intuitive *semantic* interpretation can be expressed in terms of a betting scheme, nevertheless this circumstance must not hide the fact that its role is essentially *syntactic*.

To illustrate the concept of coherence in the simpler case of unconditional events, consider an assessment $p_i = P(E_i)$, $i = 1, 2, ..., n$, on an *arbitrary* finite family

$$\mathcal{F} = \{E_1, ..., E_n\} ,$$

and denote by $A_1, ..., A_m$ the atoms generated by these events (*i.e.* made up with all possible intersections $E_1^* \wedge E_2^* \ldots \wedge E_n^*$, different from the impossible event $\emptyset$, obtained by putting in place of each E_i^*, for $i = 1, 2, \ldots, n$, the event E_i or its contrary E_i^c). This assessment is called *coherent* if the function P can be extended from $\mathcal{F}$ to the set of atoms, in such a way that P *is a probability* on the algebra generated by them. This clearly amounts to the existence of at least one solution of the following system, where $x_r = P(A_r)$,

$$(1) \qquad \begin{cases} \displaystyle\sum_{A_r \subseteq E_i} x_r = p_i\,, \qquad i = 1, 2, ..., n \\ \\ \displaystyle\sum_{r=1}^{m} x_r = 1\,, \quad x_r \geq 0\,, \quad r = 1, 2, ..., m\,. \end{cases}$$

We recall that a natural interpretation of $p_i = P(E_i)$ is to regard it as the amount paid to bet on the event E_i, with the proviso of getting back an amount 1 if E_i is true (the bet is won) or 0 if E_i is false (the bet is lost), so that, by paying p_i, the amount got back is just the *indicator* E_i (we use *the same symbol for both an event and its indicator*); moreover, it is possible (and useful) to consider, in a bet, also a "scale factor" (stake) λ_i, that is to refer to a payment $p_i\lambda_i$ to receive – when the bet is won – an amount λ_i (we were previously referring to the case $\lambda_i = 1$).

In general, given ν events $E_1, ..., E_\nu$ and ν real numbers $y_1, ..., y_\nu$, it is immediately seen that also a *discrete random variable*

$$(2) \qquad Y = \sum_{k=1}^{\nu} y_k E_k\,,$$

represents the amount got back in a combination of bets – on the ν events $E_1, ..., E_\nu$ – made by paying amounts $p_1 y_1, ..., p_\nu y_\nu$ (*i.e.* with stakes $y_1, ..., y_\nu$). Notice that the *total amount* paid, *i.e.*

$$(3) \qquad \mathbb{P}(Y) = \sum_{k=1}^{\nu} p_k y_k$$

is the so-called *prevision* (or *expectation*) of Y when the set $\{p_1, ..., p_\nu\}$ is a coherent probability assessment on the family $E_1, ..., E_\nu$. So, in the particular case that Y is just an event E (its indicator), we have $\mathbb{P}(Y) = P(E)$, *i.e.* prevision reduces to probability.

3 Conditional Events

Notice that it is not enough directing attention just toward an event E in order to assess "convincingly" its probability, but it is also essential taking into account *other* events which may possibly contribute in determining the "information" on E. Then the fundamental tool must be *conditional probability*, since the true problem is not that of assessing $P(E)$, but rather that of assessing $P(E|H)$, taking into account all the relevant "information" carried by some other event H. Dealing with conditional probability requires the introduction of *conditional* events $E|H$, with $H \neq \emptyset$ (where $\emptyset$ is the *impossible* event). An interpretation of $E|H$ in terms of a betting scheme (extending that given above for *unconditional* events) may help in clarifying its meaning.

If an amount p – which should suitably depend on $E|H$ – is paid to bet on $E|H$, we get, *when H is true*, an amount 1 if also E is true (the bet is won) and an amount 0 if E is false (the bet is lost), and *we get back the amount p if H turns out to be false* (the bet is called off). In short, introducing the *truth-value* $T(E|H)$ of a conditional event – recall that, for an (unconditional) event E, this is just its *indicator* $E = T(E|\Omega)$ – we may write, by elementary properties of indicators and introducing the symbol $t(E|H)$ in place of p for the "third" value of $T(E|H)$,

$$(4) \qquad T(E|H) = EH + t(E|H)H^c\,.$$

So a conditional event $E|H$ (or, better, its truth–value) can be seen as a particular case of a discrete random variable like (2), with $\nu = 2$, $E_1 = EH$, $E_2 = H^c$, and $y_1 = 1$, $y_2 = t(E|H)$: then its *prevision* is the "natural candidate" to be the *conditional probability* of $E|H$.

Moreover, to simplify the notation, we adopt (as we did in the unconditional case) *the same symbol for a conditional event and its truth–value* (the latter being the analogous of the indicator). With this proviso, eq. (4) can be written as

$$(5) \qquad E|H = EH + t(E|H)H^c\,,$$

where the function $t(E|H)$ depends in fact on the partition EH, E^cH, H^c of the certain event Ω.

This implies (as we proved elsewhere [9]) that, given an *arbitrary* family $\mathcal{C}$ of conditional events, when the (ordinary) sum and product of two random variables representing the given conditional events is still of the kind (5), then $t(E|H)$ satisfies the axioms of a *conditional probability* (according to its most general view, that goes back to de Finetti [11], Rényi [15], Krauss [13], Dubins [12]).

So, in particular, if the set $\mathcal{C} = \mathcal{G} \times \mathcal{B}$ of conditional events $E|H$ is such that $\mathcal{G}$ is a Boolean algebra and $\mathcal{B} \subseteq \mathcal{G}$ is closed with respect to (finite) logical sums, then, putting $\mathcal{B}^o = \mathcal{B} \setminus \{\emptyset\}$, these axioms can be expressed as follows:

(i) $t(E|H) = t\big((E \wedge H)|H\big)$, for every $E \in \mathcal{G}$ and $H \in \mathcal{B}^o$,

(ii) $t(\cdot|H)$ is a (finitely additive) probability on $\mathcal{G}$ for any given $H \in \mathcal{B}^o$,

(iii) $t\big((E \wedge A)|H\big) = t(E|H) \cdot t\big(A|(E \wedge H)\big)$, for every $A \in \mathcal{G}$ and $E,\, H \in \mathcal{B}^o$, $E \wedge H \neq \emptyset$.

Condition *(i)* is equivalent to the following

(i') $t(H|H) = 1$, for every $H \in \mathcal{B}^o$.

So we can identify $t(E|H)$ with the conditional probability $P(E|H)$ and there is no need to introduce it as the ratio of the (unconditional) probabilities $P(E \wedge H)$ and $P(H)$, assuming positivity of the latter. This allows to deal with conditioning events of zero probability, a situation which in many respects represents a very crucial feature (even in the case of a finite family of events): dropping any positivity condition, the class of admissible conditional probability assessments and that of possible extensions are larger, the ensuing algorithms are more flexible, the management of *stochastic* independence (conditional or not) avoids many of the usual inconsistencies related to *logical* dependence.

4 Coherent Conditional Probability

What about an assessment P on an *arbitrary* set $\mathcal{C}$ of conditional events? Similarly to the case of unconditional probabilities, we will say that the assessment $P(\cdot|\cdot)$ is *coherent* if, given $\mathcal{C}' \supset \mathcal{C}$, with $\mathcal{C}' = \mathcal{G} \times \mathcal{B}^o$ ($\mathcal{G}$ a Boolean algebra, $\mathcal{B}$ an additive set), it can be extended from $\mathcal{C}$ to $\mathcal{C}'$ as a *conditional probability*. In [2], [6] and [9] it has been proved and extensively discussed the following

Theorem - Let $\mathcal{C}$ be an arbitrary finite family of conditional events and $\mathcal{A}_o$ denote the relevant set of atoms. For a real function P on $\mathcal{C}$ the following two statements are equivalent:

(i) P is a coherent conditional probability on $\mathcal{C}$;

(ii) there exists (at least) a class of probabilities $\{P_o, P_1, \ldots\}$, each probability P_α being defined on a suitable subset $\mathcal{A}_\alpha \subseteq \mathcal{A}_0$, such that for any $E_i|H_i \in \mathcal{C}$ there is a unique P_α with

$$\sum_{A_r \subseteq H_i}^{r} P_\alpha(A_r) > 0$$

$$(6) \qquad P(E_i|H_i) = \frac{\sum_{A_r \subseteq E_i \wedge H_i}^{r} P_\alpha(A_r)}{\sum_{A_r \subseteq H_i}^{r} P_\alpha(A_r)} \;;$$

moreover $\mathcal{A}_{\alpha'} \subset \mathcal{A}_{\alpha"}$ for $\alpha' > \alpha"$ and $P_{\alpha"}(A_r) = 0$ if $A_r \in \mathcal{A}_{\alpha'}$.

Notice that the classes of probabilities $\{P_\alpha\}$ are infinite in general; in particular we have the uniqueness only in the case that $\mathcal{C}$ is the Cartesian product of two Boolean algebras.

In the quoted papers it is proved in a constructive way the equivalence between conditions (i) and (ii). This proof sketches an algorithm to test coherence of an assessment P, based on the equivalence between condition (ii) and the compatibility of a sequence of systems (S_α) with unknowns $P_\alpha(A_r) \geq 0$, $A_r \in \mathcal{A}_\alpha$,

$$(S_\alpha) \quad \begin{cases} \sum_{A_r \subseteq E_i \wedge H_i}^{r} P_\alpha(A_r) = P(E_i|H_i) \sum_{A_r \subseteq H_i}^{r} P_\alpha(A_r) \\ \quad \big[\text{if } P_{\alpha-1}(H_i) = 0\big] \,, \\ \sum_{A_r \subseteq H_0^\alpha}^{r} P_\alpha(A_r) = 1 \end{cases}$$

where $P_{-1}(H_i) = 0$ for all H_i's, and H_o^α denotes, for $\alpha \geq 0$, the union of the H_i's such that $P_{\alpha-1}(H_i) = 0$; so, in particular,

$$(7) \qquad H_o^o = H_o = H_1 \cup \ldots \cup H_n \,.$$

On the basis of the previous results, the "*algorithm*" needed for an actual checking of the coherence of a probability assessment can be implemented along the following steps:

(**i**) given a set $\mathcal{C}$ of n conditional events $E_1|H_1$, ..., $E_n|H_n$, supply all the known logical relations among the events $E_1, \ldots, E_n$, $H_1, \ldots, H_n$ (so that the cardinality m of the relevant set $\mathcal{A}_o$ of atoms will be usually much less than 2^{2n});

(**ii**) given the assessment

$$(8) \qquad p_i = P(E_1|H_1), \ldots, p_n = P(E_n|H_n),$$

introduce the system (S_α);

(**iii**) Put $\alpha = 0$ in (S_α);

(**iv**) if (S_α) has no solutions, the assessment (8) is not coherent and must be revised;

(**v**) if (S_α) has a solution $P_\alpha(A_r)$ such that

$$(9) \qquad P_\alpha(H_i) = \sum_{A_r \subseteq H_i}^{r} P_\alpha(A_r) > 0$$

for every H_i specified in the first line of (S_α), then the assessment (8) is coherent and each $P(E_i|H_i)$ can be represented by (6), while if it has *only* solutions such that

$$(10) \qquad P_\alpha(H_i) = 0 \quad \text{for some} \quad H_i$$

proceed as follows:

(**vi**) represent by (6) those $P(E_i|H_i)$ such that H_i satisfies (9), then put $\alpha + 1$ in place of α, and go to steps (iv) and (v) until the exhaustion of the H_i's.

The above procedure consists in solving a sequence of linear systems: the first one has $n + 1$ equations (where n is the cardinality of the set $\mathcal{C}$ of conditional events) and m unknowns (the number of relevant atoms). From a theoretical point of view, it seems that the "nicest" situation should correspond to the first statement under (v), *i.e.* to solve *only one system*, finding at the first step a solution satisfying (9). Nevertheless, notice that *to make easier the computational procedure* it is more suitable not looking for such a solution, but trying instead to solve "smaller" systems, in the sense of having a smaller number of equations (possibly only two), but – more important – a smaller number of unknowns, since the main computational problem is to build the atoms.

In fact, we should choose, at each step, *solutions in which there are many suitably chosen unknowns* $P_\alpha(A_r)$ *equal to zero* by taking those A_r's contained in as many as possible conditioning events H_i's (call H_j the remaining ones): the best situation would be when the A_r's are contained in all H_i's except one. Then each system would reduce to a system having only a few equations (possibly two) – that is only those which refer to the H_j's and which express $P(E_j|H_j)$ by means of the relevant probability P_α – plus the last one requiring that the sum of the probabilities of the atoms must be equal to 1.

In conclusion, since solving a system by giving the value zero to some unknowns (in such a way that some equations are trivially satisfied) is the same as solving a system with only the remaining equations, the ensuing nontrivial solutions may be clearly related only to the "bigger" atoms generated by *some* of the events: so we can adopt a strategy able to reduce (drastically, in the most common cases) the number of atoms needed to check coherence. This possibility – of checking coherence "locally" to get "global" coherence – is also strictly connected with the existence of logical relations among the given events, and it is then useful to find suitable subfamilies that may help to "decompose" the procedure: in other words, we need to build only the atoms generated by these subfamilies. An outline of this procedure is in [1].

5 Coherent Extensions

Another fundamental result is the following (also discussed in [6], [9]): if $\mathcal{C}$ is a given family of conditional events and P a corresponding assessment, then there exists a (possibly not unique) coherent extension of P to an arbitrary family $\mathcal{G}$ of conditional events, with $\mathcal{G} \supseteq \mathcal{C}$, *if and only if* P is coherent on $\mathcal{C}$. In particular, if $\mathcal{K} = \mathcal{C} \cup \{E|H\}$ and $P(E|H) = p$, coherent assessments of p are all values of a suitable closed interval $[p_*, p^*] \subseteq [0, 1]$, with $p_* \leq p^*$).

Then, given a finite family

$$\mathcal{C} = \{E_1|H_1, \ldots, E_n|H_n\}$$

of conditional events and a new one $E|H$, we proved in [6] (giving also the relevant algorithm, which has been improved in [9] exploiting zero probabilities) that these two bounds can be characterized as infimum and supremum, respectively, of probabilities $P(E_*|H_*)$ and $P(E^*|H^*)$ of suitable conditional events that are *logically dependent* on $\mathcal{C}$, *i.e.* such that the events E_*, H_*, E^*, H^* are union of atoms generated by $E_1, \ldots, E_n, H_1, \ldots, H_n$. In particular, $E_*|H_*$ and $E^*|H^*$ are, respectively, the "maximum" and the "minimum" conditional event logically dependent on $\mathcal{C}$ satisfying

$$E_*|H_* \subseteq\circ\ E|H \subseteq\circ\ E^*|H^*\,;$$

the inclusion $\subseteq\circ$ between conditional events is defined as

$$A|H \subseteq\circ\ B|K \iff AH \subseteq BK \quad \text{and} \quad B^cK \subseteq A^cH\,.$$

This inclusion has a *numerical* counterpart (proved in [6]), that is $A|H \leq B|K$: this is an inequality between the random variables representing the two conditional events by means of (5).

6 Some Crucial Examples

We start by analizing some examples of medical diagnosis to show how a hasty Bayesian updating of probability assessments can lead to wrong conclusions, so that it is better, in specific circumstance, to avoid its use.

Example 1. A patient feels serious generalized abdominal pains, fever and retchs. The doctor puts forth the following hypotheses concerning the possible relevant disease:

H_1 = *ileus*
H_2 = *peritonitis*
H_3 = *acute appendicitis, with an ensuing local peritonitis*

Moreover the doctor assumes a natural logical condition such as $H_3 \subset H_1^c \wedge H_2$. Correspondingly we have then five atoms

$$A_1 = H_1 \wedge H_2 \wedge H_3^c\,, \quad A_2 = H_1 \wedge H_2^c \wedge H_3^c\,,$$
$$A_3 = H_1^c \wedge H_2 \wedge H_3^c\,, \quad A_4 = H_1^c \wedge H_2 \wedge H_3\,,$$
$$A_5 = H_1^c \wedge H_2^c \wedge H_3^c\,.$$

The doctor initially gives these probability assessments:

$$P(H_1) = \frac{1}{2}\,, \quad P(H_2) = \frac{1}{5}\,, \quad P(H_3) = \frac{1}{8}\,.$$

This (partial) assessment is *coherent*, since the function P can be extended from the three given events to the set of relevant atoms in such a way that *P is a probability* on the algebra generated by them (see Sect.2), *i.e.* there exists a solution of the following system with unknowns $x_i = P(A_i)$

$$(S) \qquad \begin{cases} x_1 + x_2 = \frac{1}{2} \\ x_1 + x_3 + x_4 = \frac{1}{5} \\ x_4 = \frac{1}{8} \\ \sum_{i=1}^{5} x_i = 1 \\ x_i \geq 0 \,. \end{cases}$$

For example, given λ, with $0 \leq \lambda \leq \frac{3}{40}$, a solution is

$$x_1 = \lambda\,, \quad x_2 = \frac{1}{2} - \lambda\,, \quad x_3 = \frac{3}{40} - \lambda\,,$$
$$x_4 = \frac{1}{8} \quad x_5 = \frac{3}{10} + \lambda.$$

The doctor considers now the event

E = *pressing in particular points of the abdomen does not increase pain*

and he gives the following relevant logical and probabilistic information

$$E \wedge H_3 = \emptyset$$
$$P(E|H_2) = \frac{2}{5}\,, \quad P(E|H_2^c) = \frac{1}{8}\,.$$

Obviously, the latter assignment is coherent, since it refers to a (trivial) partition (with respect to the *conditioning* events). If we update the (prior) probability of H_2 by means of the above likelihood (through Bayes theorem), we get

$$P(H_2|E) = \frac{4}{9}\,.$$

This new probability of H_2 is coherent with the previous probabilities of H_1 and H_3. To prove that, consider the atoms obtained when we take into account also the new event E :

$$B_1 = A_1 \wedge E\,, \quad B_2 = A_2 \wedge E\,, \quad B_3 = A_3 \wedge E\,,$$
$$B_4 = A_1 \wedge E^c\,, \quad B_5 = A_2 \wedge E^c\,, \quad B_6 = A_3 \wedge E^c\,,$$
$$B_7 = A_4 \wedge E^c\,, \quad B_8 = A_5 \wedge E\,, \quad B_9 = A_5 \wedge E^c\,.$$

To check coherence we consider the following system (T) (see the system (S_α) following the Theorem of Sect.4), with unknowns $y_i = P(B_i)$

$$(T) \qquad \begin{cases} y_1 + y_2 + y_4 + y_5 = \frac{1}{2} \\ y_7 = \frac{1}{8} \\ y_1 + y_3 = \frac{4}{9}(y_1 + y_2 + y_3 + y_8) \\ \sum_{i=1}^{9} y_i = 1 \\ y_i \geq 0 \,. \end{cases}$$

It is easily seen that the system (T) has (infinite) solutions and, since there are also solutions such that

$$y_1 + y_2 + y_3 + y_8 > 0\,,$$

this is sufficient to ensure that the assessment is coherent. This is true even if we take into account the updating of the probability of H_3, that is $P(H_3|E) = 0$: in fact this corresponds to ignoring the second equation of system (T).

But to consider this assessment as an updating of the previous one can be a too hasty (and wrong) conclusion, since the value of $P(H_2|E)$ has been obtained by considering in fact as "prior" the assessment

$$P(H_2) = \frac{1}{5}\,, \quad P(H_2^c) = \frac{4}{5}\,,$$

and not that actually given by the doctor, which involves also the evaluation of $P(H_1)$ and $P(H_3)$. The updating of that assessment *obviously* requires that the "whole" prior and the likelihood must be *jointly* coherent. Instead in this case coherence does not hold: considering the following system

$$(T') \qquad \begin{cases} y_1 + y_2 + y_4 + y_5 = \frac{1}{2} \\ y_1 + y_3 + y_4 + y_6 + y_7 = \frac{1}{5} \\ y_7 = \frac{1}{8} \\ y_1 + y_3 = \frac{2}{5}(y_1 + y_3 + y_4 + y_6 + y_7) \\ y_2 = \frac{1}{8}(y_2 + y_5) \\ \sum_{i=1}^{9} y_i = 1 \\ y_i \geq 0\,, \end{cases}$$

simple computations (solving for $y_1 + y_3$ the fourth and the second eq. and inserting this and the third eq. into the second one) show that it does not admit solutions, so that the assessment is not coherent.

The following example shows that even the "local" coherence of prior and "pseudoposterior" obtained in the previous example was just accidental.

Example 2. A patient feels a severe back-ache together with lack of sensitiveness and pain in the left leg; he had two years before a lung cancer that was removed by a surgical operation. The doctor considers the following exhaustive hypotheses concerning the patient situation:

H_1=*crushing of L5 and S1 vertebrae*
H_2=*rupture of the disc*
H_3=*inflammation of nerve-endings*
H_4=*bone tumor*

The doctor does not regard them as mutually exclusive; moreover, he assumes some logical relations:

$$H_4 \wedge (H_1 \vee H_2 \vee H_3) = \emptyset\,,$$

$$H_1 \wedge H_2 \wedge H_3 = \emptyset\,,$$
$$H_1 \wedge H_2^c \wedge H_3^c = \emptyset\,,$$
$$H_1^c \wedge H_2 \wedge H_3^c = \emptyset\,,$$
$$H_1^c \wedge H_2^c \wedge H_3 = \emptyset\,,$$

Correspondingly, we have only the four atoms

$A_1 = H_1 \wedge H_2 \wedge H_3^c \wedge H_4^c\,;\quad A_2 = H_1 \wedge H_2^c \wedge H_3 \wedge H_4^c\,;$
$A_3 = H_1^c \wedge H_2 \wedge H_3 \wedge H_4^c\,;\quad A_4 = H_1^c \wedge H_2^c \wedge H_3^c \wedge H_4\,.$

The doctor makes the following probabilistic assessments

$$P(H_1) = \frac{1}{3}\,,\ P(H_2) = \frac{1}{4}\,,\ P(H_3) = \frac{5}{12}\,,\ P(H_4) = \frac{1}{2}\,.$$

To check its coherence, we refer to the system (S') with unknowns $x_r = P(A_r)$,

$$(S') \qquad \begin{cases} x_1 + x_2 = \frac{1}{3} \\ x_1 + x_3 = \frac{1}{4} \\ x_2 + x_3 = \frac{5}{12} \\ x_4 = \frac{1}{2} \\ \displaystyle\sum_{i=1}^{4} x_i = 1 \\ x_i \geq 0\,. \end{cases}$$

This system has a unique solution

$$x_1 = \frac{1}{12}\,,\quad x_2 = \frac{1}{4}\,,\quad x_3 = \frac{1}{6}\,,\quad x_4 = \frac{1}{2}\,.$$

Let now E be the event

E=an X-ray test is sufficient for a reliable and decisive diagnosis

so that

$$H_4 \subset E\,,\ H_1 \wedge H_2 \subset E\,,\ H_1 \wedge H_3 \subset E^c\,,\ H_2 \wedge H_3 \subset E^c\,.$$

The doctor assigns the likelihood

$$P(E|H_1) = \frac{1}{3}\,,\quad P(E|H_1^c) = \frac{1}{6}\,.$$

If we update the (prior) probability $P(H_1)$ by the above likelihood through Bayes' theorem, we get $P(H_1|E) = \frac{1}{2}$. But now (contrary to the previous situation of *Example 1*) this updated probability of H_1 is not coherent with the given probabilities of H_2 and H_3. Notice in fact that the atoms obtained when we take into account the new event E are exactly those generated by the events H_i, so that to check coherence we need to study the solvability of the system $(S")$ with unknowns $x_r = P(A_r)$

$$(S") \qquad \begin{cases} x_1 = \frac{1}{2}(x_1 + x_4) \\ x_4 = \frac{1}{2} \\ x_1 + x_3 = \frac{1}{4} \\ x_2 + x_3 = \frac{5}{12} \\ \displaystyle\sum_{i=1}^{4} x_i = 1 \\ x_i \geq 0\,. \end{cases}$$

But the first two equations give $x_1 = \frac{1}{2}$, hence $x_3 < 0$, so this system is inconsistent.

Remark – The next example shows that it is possible to update (prior) probability by Bayes rule, also in unusual situations (such as when we assume that the diseases are not mutually exclusive), if coherence of the "global" (*i.e.* prior and likelihood together) assessment holds.

Example 3. A patient arrives at the hospital showing symptoms of choking. The doctor considers the following hypotheses concerning the patient situation:

H_1=cardiac insufficiency
H_2=asthma attack
H_3=cardiac lesion caused by asthma

The doctor does not regard them as mutually exclusive; moreover, he assumes the following natural logical relation:

$$H_3 \subset H_1 \wedge H_2\,.$$

Correspondingly, we have the atoms

$$A_1 = H_1 \wedge H_2 \wedge H_3\,;\quad A_2 = H_1 \wedge H_2 \wedge H_3^c\,;$$
$$A_3 = H_1^c \wedge H_2 \wedge H_3^c\,;\quad A_4 = H_1 \wedge H_2^c \wedge H_3^c\,;$$
$$A_5 = H_1^c \wedge H_2^c \wedge H_3^c\,.$$

The doctor makes the probability assessments

$$P(H_1) = \frac{1}{2}\,,\ P(H_2) = \frac{1}{3}\,,$$
$$P(H_3) = \frac{1}{5}\,,\ P(H_1 \vee H_2) = \frac{3}{5}\,.$$

To check its coherence, we refer to the following system with unknowns $x_r = P(A_r)$

$$\begin{cases} x_1 + x_2 + x_4 = \frac{1}{2} \\ x_1 + x_2 + x_3 = \frac{1}{3} \\ x_1 = \frac{1}{5} \\ x_1 + x_2 + x_3 + x_4 = \frac{3}{5} \\ \displaystyle\sum_{i=1}^{5} x_i = 1 \\ x_i \geq 0 \end{cases}$$

which has a unique solution

$$x_1 = \frac{1}{5}\,,\ x_2 = \frac{1}{30}\,,\ x_3 = \frac{1}{10}\,,\ x_4 = \frac{4}{15}\,,\ x_5 = \frac{2}{5}\,.$$

Let now E be the event

E=taking the medicine M against asthma does not reduce choking symptoms

so that

$$H_2 \wedge H_1^c \wedge E = \emptyset\,.$$

The doctor now draws out from his data base the likelihood

$$P(E|H_1) = \frac{3}{10}\,, \quad P(E|H_3^c) = \frac{1}{3}\,.$$

Then the process of learning from data starts by building the new atoms

$$B_1 = A_1 \wedge E^c\,, \quad B_2 = A_2 \wedge E^c\,, \quad B_3 = A_3 \wedge E^c = A_3\,,$$

$$B_4 = A_4 \wedge E^c\,, \quad B_5 = A_5 \wedge E^c\,, \quad B_6 = A_1 \wedge E\,,$$

$$B_7 = A_2 \wedge E\,, \quad B_8 = A_4 \wedge E\,, \quad B_9 = A_5 \wedge E\,,$$

and to check coherence we need to consider the following system (T") with unknown $y_i = P(B_i)$

$$\begin{cases} y_1 + y_2 + y_4 + y_6 + y_7 + y_8 = \frac{1}{2} \\ y_1 + y_2 + y_3 + y_6 + y_7 = \frac{1}{3} \\ y_1 + y_6 = \frac{1}{5} \\ y_1 + y_2 + y_3 + y_4 + y_6 + y_7 + y_8 = \frac{3}{5} \\ y_6 + y_7 + y_8 = \frac{3}{10}(y_1 + y_2 + y_4 + y_6 + y_7 + y_8) \\ y_7 + y_8 + y_9 = \frac{1}{3}(y_2 + y_3 + y_4 + y_5 + y_7 + y_8 + y_9) \\ \sum_{i=1}^{9} y_i = 1 \\ y_i \geq 0\,. \end{cases}$$

This system has solutions, *i.e.*, given λ and μ with

$$\frac{7}{60} \leq \mu \leq \frac{7}{30}\,, \qquad \frac{7}{30} - \mu \leq \lambda \leq \frac{4}{15} - \mu\,,$$

a solution is

$$y_1 = \frac{19}{60} - \mu\,,\ y_2 = \lambda + \mu - \frac{7}{30}\,,\ y_3 = \frac{1}{10}\,,\ y_4 = \frac{4}{15} - \lambda\,,$$

$$y_5 = \frac{2}{5} - \mu\,, \quad y_6 = \mu - \frac{7}{60}\,, \quad y_7 = \frac{4}{15} - \lambda - \mu\,,$$

$$y_8 = \lambda\,, \quad y_9 = \mu\,.$$

Then "global" coherence (*i.e.*, referring to prior and likelihood jointly) allows Bayesian updating $P(H_i|E)$, $i = 1, 2, 3$. But notice that in this case we lack the usual representation ("disintegration" formula)

$$P(E) = \sum_{i=1}^{3} P(H_i) P(E|H_i)\,,$$

since the hypotheses do not constitute a partition. On the other hand $P(E)$ can be obtained by summing the probabilities of the atoms contained in E, that is

$$P(E) = P(B_6) + P(B_7) + P(B_8) + P(B_9) =$$

$$= y_6 + y_7 + y_8 + y_9 = \frac{3}{20} + \mu\,,$$

and so it ranges in the interval $[\frac{4}{15}, \frac{23}{60}]$. Therefore, by straightforward computations, we have

$$\frac{9}{23} \leq P(H_1|E) \leq \frac{9}{16}\,, \qquad 0 \leq P(H_3|E) \leq \frac{7}{23}\,,$$

while the evaluation of $P(H_2|E)$ requires resorting (since the likelihood $P(E|H_2)$ is not given) to

$$P(H_2 \wedge E) = P(B_6) + P(B_7) = \frac{3}{20} - \lambda\,,$$

to be divided by $P(H_2)$, so that, finally, we obtain

$$0 \leq P(H_2|E) \leq \frac{9}{20}\,.$$

Clearly,

$$p_* = \{\frac{9}{23}, 0, 0\}\ ,\ p^* = \{\frac{9}{16}, \frac{9}{20}, \frac{7}{23}\}$$

are, respectively, a (coherent) lower and a (coherent) upper conditional probability of $H_i|E$ ($i = 1, 2, 3$). Recall that a *coherent lower (upper) conditional probability* is a function P_* (P^*) such that there exists a non-empty family $\mathcal{P} = \{P_i\}$ of coherent conditional probabilities on $\mathcal{C}$ whose lower (upper) envelope is P_* (P^*).

Now we could go on by updating again the conditional probabilities $P(H_i|E)$, given, for example, the new event

F=taking the medicine M against asthma increases tachycardia

with

$$\emptyset \neq E \wedge F\ ,\ H_2 \wedge H_1^c \wedge F = \emptyset\,.$$

Obviously, the doctor has in the data base some relevant likelihood: nevertheless the checking of coherence cannot proceed as before, since we start now with an upper and lower probability (the "old" posteriors, that now should be taken as priors). So the Theorem of Sect.4 cannot be applied: for the sake of brevity, we do not report all the necessary computations, that are based on the following Theorem, given in [8]:

Theorem - Let $\mathcal{C}$ be an arbitrary finite family of conditional events. For a real function P_* on $\mathcal{C}$ the following two statements are equivalent:

(i) the function P_* is a coherent conditional lower probability on $\mathcal{C}$;

(ii) denoting by $\mathcal{A}_o$ the relevant set of atoms, there exists, for any $E_i|H_i \in \mathcal{C}$ (at least) a class of probabilities $\{P_o^i, P_1^i, \ldots\}$, each probability P_α^i being defined on a suitable subset $\mathcal{A}_\alpha^i \subseteq \mathcal{A}_0$, such that for any $E_j|H_j \in \mathcal{C}$ there is a unique P_α^i with

$$\sum_{A_r \subseteq H_j}^{r} P_\alpha^i(A_r) > 0$$

$$P_*(E_j|H_j) \geq \frac{\sum_{A_r \subseteq E_i \wedge H_i}^{r} P_\alpha^i(A_r)}{\sum_{A_r \subseteq H_j}^{r} P_\alpha^i(A_r)} \quad \text{if} \quad j \neq i$$

$$P_*(E_j|H_j) = \frac{\sum_{r \atop A_r \subseteq E_i \wedge H_i} P^i_\alpha(A_r)}{\sum_{r \atop A_r \subseteq H_j} P^i_\alpha(A_r)} \quad \text{if} \quad j = i$$

and moreover $\mathcal{A}^i_{\alpha'} \subset \mathcal{A}^i_{\alpha"}$ for $\alpha' > \alpha"$, while $P^i_{\alpha"}(A_r) = 0$ if $A_r \in \mathcal{A}^i_{\alpha'}$.

Actually, it is possible to build the classes $\{P^i_\alpha\}$ as solutions of sequences of systems (one for each conditional event $E_i|H_i \in \mathcal{C}$) similar to the following :

$$\begin{cases} \sum_{r \atop A_r \subseteq E_i \wedge H_i} P^i_\alpha(A_r) = P_*(E_i|H_i) \sum_{r \atop A_r \subseteq H_i} P^i_\alpha(A_r)\,, \\ \quad [\text{if } P^i_{\alpha-1}(H_i) = 0] \\ \sum_{r \atop A_r \subseteq E_j \wedge H_j} P^i_\alpha(A_r) \geq P_*(E_j|H_j) \sum_{r \atop A_r \subseteq H_i} P^i_\alpha(A_r)\,, \\ \quad [\text{if } P^i_{\alpha-1}(H_j) = 0] \\ \sum_{r \atop A_r \subseteq H^\alpha_{0,i}} P^i_\alpha(A_r) = 1 \end{cases} .$$

This procedure suggests, as in the case of probabilities, an actual algorithm to prove the consistency (coherence) of a lower probability assessment on a finite set of events.

Obviously, a partial lower probability assessment has more "chance" than a probability assessment to fulfil consistency (*i.e.*, coherence), since systems with inequalities have more solutions than those with only equalities. But the relevant check is computationally more burdensome (in fact we must repeat the same procedure n times, where n is the number of the given conditional events).

Remark - The partial assessment (lower probability) p_* contains two values *equal to zero*: now, since these must be taken – to go on in the updating process – as new "prior" assignments, we can realize how is crucial (also from a practical point of view) to have a theory that privileges the possibility of managing *conditioning* events of zero probability (which appear in the "new" likelihood).

7 Generalized Bayes Inference

The procedure applied to the previous specific examples can be put forth in general, to handle uncertainty in the process of automatic medical diagnosis. The proposed interactive procedure initially refers to

(i) a family of hypotheses (that is, *events* represented by suitable propositions) supplied by the physician: they correspond to possible diseases H_i ($i = 1, 2, ..., n$) which could explain a given initial piece of information referring to the specific situation (anamnesis). No structure and no simplifying and unrealistic assumption (such as mutual exclusiveness and exhaustivity) is required for this family of events;

(ii) all logical relations between these hypotheses, either already included in the knowledge base, or given by the doctor on the basis of the specific situation;

(iii) a probability assessment on the given set of hypotheses. Clearly, this is not a complete assessment, since these events have been chosen as the most natural according to the doctor's experience: they do not constitute, in general, a partition of the certain event Ω, and so the extension to other events of these probability evaluations is not necessarily unique. Moreover, a doctor often assigns degrees of belief directly to *sets of hypotheses* (for example, one may suspect that one of the diseases the patient suffers from is an infectious one, but he is not able to commit any belief to particular infectious diseases);

(iv) a data base consisting of conditional events $E|K$ and their relevant probabilities $p = P(E|K)$, where each event K is a possible disease which is in some way related to the given hypotheses H_i, while each evidence (acquired or assumed) E (that can be looked on as an event) comes from a suitable evidential test. These probabilities p could have been obtained by means of relevant frequencies and should be recorded in files of the diagnostic center.

Then, once this preliminary preparation has been done, the first step of our procedure consists in building the family of atoms (generated by the hypotheses $H_1, H_2, ..., H_n$): they are a partition of the certain event, but they are not the "natural" events to which the doctor is willing to assign probabilities. Nevertheless these atoms are the main tool for checking the coherence of the relevant assessment: in fact coherence amounts to finding on the set of atoms (by solving a linear system) a probability distribution (not necessarily unique) compatible with the given assignment. If the assessment turns out not being coherent, the doctor can be driven to a different assignment based on the relevant mathematical relations contained in the corresponding linear system.

On the contrary, coherence of the probabilities $P(H_i)$ allows to go on by checking now the coherence of the whole assessment including also the probabilities $P(E|K)$. This requires the introduction of new atoms, possibly taking into account all logical relations involving the evidences E and the hypotheses H_i. In particular, some of the latter may coincide with some K. As the previous examples (Sect.6) have shown, the whole assignment (prior probabilities and likelihood) can be incoherent even if the two separate assessment were not.

On the basis of the results obtained by means of the evidential tests, the doctor can now update the probabilities of the hypotheses H_i, *i.e.* he assesses the conditional probabilities $P(H_i|E)$. Then he needs to check again coherence of the whole assessment including the latter and the former probability evaluations.

When prior probabilities and likelihood are jointly coherent, the doctor can get formulas representing each posterior probability (of a disease H_i given an evidence E) by Bayes' theorem

$$P(H_i|E) = \frac{P(H_i)P(E|H_i)}{P(E)},$$

but the denominator $P(E)$ cannot by computed by the usual "disintegration" formula, since the H_i's are not a partition. Nevertheless we can express $P(E)$ in terms of the atoms, but this representation is not unique, since the corresponding linear system may have more than just one solution: computing upper and lower bounds of $P(E)$ we get, respectively, lower and upper bounds for the posterior probabilities $P(H_i|E)$.

8 Conclusions

Thanks also to some crucial examples concerning medical diagnosis, we have been able to show that in our setting we can privilege a gradual assignment of the relevant probabilities: these values are not necessarily unique (they may be also the result of coherent extensions to new conditional events) and possibly belong to suitable closed intervals (for *unconditional* events, this result is known as *de Finetti's fundamental theorem of probabilities*: it essentially rules the set $\mathcal{P}$ of all probabilities extending (coherently) a given P). Usually, probability assessments are not uniquely singled-out by the initial information and data that have been taken into account in each relevant situation, so that we need dealing also with upper and lower probabilities.

References

[1] A. Capotorti and B. Vantaggi. An algorithm for coherent conditional probability assessments. *Proc. IV Congresso Nazionale SIMAI, Giardini Naxos (Messina)*, Vol.2:144–148, 1998.

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

[3] G. Coletti, A. Gilio and R. Scozzafava. Conditional events with vague information in expert systems. *Lecture Notes in Computer Sciences* (Eds. B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh). Springer-Verlag, Berlin, n.521:106–114, 1991.

[4] G. Coletti and R. Scozzafava. Qualitative Bayesian inference. *Proc. of the Section on "Bayesian Statistical Science"*, San Francisco, Amer. Statist. Assoc., Alexandra:184–191, 1993.

[5] G. Coletti and R. Scozzafava. Learning from data by coherent probabilistic reasoning. *Proc. 3rd Int. Symp. on "Uncertainty Modeling and Analysis"*, Maryland (Ed. B.M. Ayyub), IEEE Comp. Soc. Press:535–540, 1995.

[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*, 4:103–127, 1996.

[7] G. Coletti and R. Scozzafava. Exploiting zero probabilities. *Proc. EUFIT '97, Aachen, ELITE Foundation*:1499–1503, 1997.

[8] G. Coletti and R. Scozzafava. Conditional measures: old and new. *"New Trends in Fuzzy Systems"*, World Scientific:107–120, 1998.

[9] G. Coletti and R. Scozzafava. Conditioning and Inference in Intelligent Systems. *Soft Computing*, to appear, 1999.

[10] B. de Finetti. *Teoria della probabilità*, Einaudi, Torino, 1970. – *Engl. transl.*: *Theory of Probability*, Wiley, Chichester, 1974.

[11] B. de Finetti. Sull'impostazione assiomatica del calcolo delle probabilità. *Annali Univ. Trieste*, 19:3–55, 1949. – *Engl. transl.*: Ch.5 in *Probability, Induction, Statistics*, Wiley, London, 1972.

[12] L.E. Dubins. Finitely additive conditional probabilities, conglomerability and disintegration. *Ann. Probab.*, 3:89–99, 1975.

[13] P.H. Krauss. Representation of conditional probability measures on Boolean algebras. *Acta Math. Acad. Scient. Hungar.*, 19:229–241, 1968.

[14] S.L. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their applications to expert systems. *Journal of the Royal Statistical Society*, Ser.B, 50:157–189, 1988.

[15] A. Rényi. On conditional probability spaces generated by a dimensionally ordered set of measures. *Theor. Probab. Appl.*, 1:61–71, 1956.

[16] P.M. Williams. Indeterminate probabilities. *Formal Methods in the Methodolgy of Empirical Sciences* (Eds. M. Przelecki, K. Szaniawski, and R. Vojcicki). Reidel, Dordrecht:229–246, 1976.