Efficient L1-Based Probability Assessments Correction: Algorithms and Applications to Belief Merging and Revision

Marco Baioletti and **Andrea Capotorti**
Dip. Matematica e Informatica - Università degli Studi di Perugia
{marco.baioletti, andrea.capotorti}@unipg.it

Abstract

In this article we define a procedure which corrects an incoherent probability assessment on a finite domain by exploiting a geometric property of L1-distance (known also as Manhattan distance) and mixed integer programming. L1-distance minimization does not produce, in general, a unique solution but rather a corrected assessment that could result an imprecise probability model. We propose a correction method for the merging of two separate assessments whose direct juxtaposition could be incoherent, and for the revision of beliefs where the core of the assessment must remain unchanged. A prototypical example on antidoping analysis guides the reader through this article to explain the various procedures.

Keywords. coherence, mixed-integer optimization, probability merging and revision, imprecise probability.

1 Introduction

The problem of correcting probability evaluations, especially on finite settings, has a long history and has been largely debated. Considering the significant amount of research on this subject, we can just mention two main "streams": one is the "right way" of assessing probability values, whose roots can be found in [4, 17, 20] while the other is the so called "calibration question" that stems from the seminal paper [31] and subsequent developments [15, 16]. More recently these two streams have been joined and faced with a unifying view by de Finetti's notion of coherence ([18] and in particular [27, pag. 361]). Hence several approaches have been proposed to deal with "incoherent" probabilities, for both unconditional and conditional values and by adopting different notions of "distances" and "scoring rules" (among the many, refer, e.g., to [6, 7, 8, 9, 28, 30]).

The risk of dealing with incoherent probability assessments is significantly present when the numerical evaluation comes from different sources of information and/or structural constraints limit the possible states (see, e.g., [5, 11, 12, 24, 33]). In this paper we come back to the fore of this argument leaving aside the more probabilistic approaches based on scoring rules that have a forecasting perspectives, by adopting the more aseptic approach based on geometrical distance minimization. In particular we will deal with the simple and easily understood L_1-distance, known also as "Manhattan" or "taxi-cab" metric. The main reason for using such metric is because we are able to propose an effective procedure (presented in Sec. 3), which is based on integer linear programming and hence is much more efficient than the correction procedures needed for other distances, for instance the quadratic programming for L_2-distance. L_1-distance minimization has moreover a simple interpretation, since it implies a direct minimal modification of each single value, permitting to use it for different purposes like the merging between two separate assessments (described in Sec. 4) and the revision of beliefs (depicted in Sec. 5).

The peculiarity of using L_1 minimization is the non-uniqueness, in general, of the solution and this could represent an alternative way of legitimating the adoption of imprecise probability models, in addition to the historical ones as stemming from buying/selling prices or desirability of gambles [35], or from extensions of coherent precise initial assessments [13, Chap.15]. In this paper, we assume that the initial assessments are precise, but this assumption could be easily generalized to initial imprecise probability assessments. However, assuming initial assessments as being precise is reasonable as it is consistent with usual estimate techniques which tend to express precise values.

In order for this paper to be as self-contained as possible, the next Section 2 briefly introduces the notion of probability assessments and formalizes the problem of their coherence. As already stated, the subsequent

Section 3 contains a proposal of a new correction algorithm based on L_1-distance minimization via mixed integer programming and on properties of convex polytopes, while Sections 4 and 5 legitimate its usefulness. A short concluding Section 6 closes the contribution.

2 Probability Assessments

A probability assessment on a finite domain can be expressed through a quadruple $\pi = (V, U, p, \mathfrak{C})$, where $V = \{X_1, \ldots, X_n\}$ is a finite set of propositional variables, representing any potential event of interest, U is a subset of V that contains the effective events taken into consideration, $p : U \to [0, 1]$ is a function which assigns a probability value to each variable in U, and $\mathfrak{C}$ is a finite set of logical constraints which lie among all the variables in V.

Note that the explicit presence of the set of variables V, even if the numerical assessment is given on the subset U, permits to extend an initial assessment to a larger domain without redefine the whole model, allowing a dynamical analysis. In this paper we will use it only on the merging application of Sec.4, but it is a good practice to allow this distinction also in static descriptions.

Since the Boolean logical setting in which we embed the assessment, in the sequel we will adopt the usual logical notation, with $\neg$, $\wedge$ and $\vee$ denoting the negation, disjunction and conjunction connectives, respectively; $\Rightarrow$ the material implication; $=$ the logical equivalence; $\top$ and $\bot$ the universal tautology and contradiction (sure and impossible events), respectively.

Usually some possible forms of logical constraints are: $\phi = \psi$, $\phi \implies \psi$ and $\phi = \bot$, where ψ and ϕ are boolean expressions involving the variables of V. But without loss of generality, we suppose that $\mathfrak{C}$ is expressed in conjunctive normal form, i.e., each element of $\mathfrak{C}$ is a disjunction of literals formed with variables in V, i.e., each element can be written as disjunctive clause

$$\left(\bigvee_{h \in H} X_h\right) \vee \left(\bigvee_{l \in L} \neg X_l\right)$$

for some $H, L \subseteq \{1, \ldots, n\}$, so that $\mathfrak{C}$ results as their conjunction.

For example, the constraint $X_i \implies X_j$ is expressed in $\mathfrak{C}$ by the clause $\neg X_i \vee X_j$.

A truth-value assignment α is a function from V to $\{0, 1\}$. Given a proposition ϕ, we write $\alpha \models \phi$ when α satisfies ϕ, otherwise we write $\alpha \not\models \phi$.

There are different, but equivalent, ways to define the coherence, i.e., the "rationality", of an assessment π: from semantical, syntactical or operational point of views (see, e.g., [12, 13, 18, 27]). Here we adopt the pragmatic way already used in [1], where a probability assessment $\pi = (V, U, p, \mathfrak{C})$ is coherent if there exists a probability distribution $\mu : 2^V \to [0, 1]$ on the set of all truth-value assignments 2^V which satisfies the following properties

1. for each $\alpha \in 2^V$, if there exists a constraint $c \in \mathfrak{C}$ such that $\alpha \not\models c$, then $\mu(\alpha) = 0$;
2. $\displaystyle\sum_{\alpha \in 2^V} \mu(\alpha) = 1$;
3. for each $X \in U$, $\displaystyle\sum_{\alpha \in 2^V, \alpha \models X} \mu(\alpha) = p(X)$.

The coherence of a probability assessment, called shortly CPA, has been already studied in [1, 2, 3, 32], albeit in a slightly different form, showing that checking if π is coherent is a NP-complete problem, even when the constraints in $\mathfrak{C}$ are binary (i.e., each of them involves only two variables).

The computational problem CPA is strictly related to the Probabilistic Satisfiability problem (PSAT [23]), where the probability assessment is defined on some finite set of propositions, instead that on the propositional variables. It can be proved that every instance of CPA can be easily translated as a PSAT instance, and that every PSAT can be formulated in a normal form, which is essentially a CPA instance [14].

There exist several algorithms to solve CPA and PSAT problems:

- A column-generation [23, 25] approach, where the problem is solved using linear programming techniques which exploit the sparsity of the solutions;
- CPA algorithm [1, 2], which is based on a symbolic manipulation which, in some cases, needs a further linear programming procedure;
- SAT-based approach [19], in which the problem is translated in a pure propositional satisfiability form (SAT);
- MIP-based approach [14], in which the problem is formulated as a mixed integer programming problem (MIP).

3 Correcting Probability Assessments

When a probability assessment $\pi = (V, U, p, \mathfrak{C})$ is not coherent, then it is possible to "correct" it in different ways, in order to obtain a coherent probability assessment π' which is as close as possible to π, according to a distance or a pseudo-distance function between probability assessments.

One possibility is to revise only the probability values, i.e., $\pi' = (V, U, p', \mathfrak{C})$, and to use a distance between probability assessments which is defined only in terms of p and p'.

Another possibility, which will not be taken into account in this paper, could be to revise (also) the logical constraints.

Since p and p' correspond to vectors of $\mathbb{R}^n$, where $n = |U|$, it is possible to use a distance d in $\mathbb{R}^n$. Then, chosen a distance d, a d-correction of a probability assessment $\pi = (V, U, p, \mathfrak{C})$ is a vector p' such that the probability assessment $\pi' = (V, U, p', \mathfrak{C})$ is coherent and $d(p, p')$ is minimized. We denote $\mathcal{C}_d(\pi)$ the sets of all the d-correction of π.

Clearly if π is coherent, then $\mathcal{C}_d(\pi) = \{p\}$, for any distance d of $\mathbb{R}^n$.

In general, given a probability assessment π, $\mathcal{C}_d(\pi)$ could have more than one element and in this case the operation of correcting a probability assessment leads to an imprecise probability model, the so called "credal set".

As already stated in te Introduction, several distance choice are possible. Among the many, in this paper we focus on the L_1 distance defined as

$$d_1(p, p') = \sum_{i=1}^{n} |p(X_i) - p'(X_i)|$$

and we denote $\mathcal{C}_{d_1}(\pi)$ as $\mathcal{C}(\pi)$. Whether this could be the best distance and how it performs with respect to the others is not directly considered. Rather its use as a tool is considered as it is reasonable and easily interpretable by users so that technical aspects connected with its adoption are addressed. Our interest in L_1 distance is that with its adoption translating the optimization problem into a linear problem by using both integer and real variables is possible. This last represents a computational advantage compared to other distances that imply implementation of non linear (quadratic, logarithmic, etc.) optimizations tools.

The resulting mixed integer program $\mathcal{P}1$ is built similarly to the method described in [14]. Let us suppose that $U = \{X_1, \ldots, X_n\}$. Moreover let $m = |\mathfrak{C}|$.

The real variables of $\mathcal{P}1$ are

- b_{ij}, for $i = 1, \ldots, n$ and $j = 1, \ldots, n+1$.
- q_j, for $j = 1, \ldots, n+1$
- r_i, s_i, for $i = 1, \ldots, n$

all of them are non-negative (as usual in linear programming).

The program $\mathcal{P}1$ also has the integer variables

- a_{ij}, for $i = 1, \ldots, n$ and $j = 1, \ldots, n+1$

which are constrained to 0 or 1.

The constraints of $\mathcal{P}1$ are

1. for each $i = 1, \ldots, n$,
$$\sum_{j=1}^{n+1} b_{ij} = p(X_i) + (r_i - s_i)$$
2. for $i = 1, \ldots, n$ and $j = 1, \ldots, n+1$,
$$0 \leq b_{ij} \leq a_{ij}, \quad a_{ij} - 1 + q_j \leq b_{ij} \leq q_j$$
3. $$\sum_{i=1}^{n+1} q_j = 1$$
4. for $i = 1, \ldots, n$,
$$r_i \leq 1, \quad s_i \leq 1$$

Moreover, for each clause c_i (for $i = 1, \ldots, m$), where $c_i = \bigvee_{h \in H_i} X_h \vee \bigvee_{l \in L_i} \neg X_l$, and for each $j = 1, \ldots, n+1$ the following linear constraint is added

$$\sum_{h \in H_i} a_{h,j} + \sum_{l \in L_i} (1 - a_{l,j}) \geq 1$$

Finally, the objective function to be minimized is

$$\sum_{i=1}^{n} (r_i + s_i)$$

Solving the linear program $\mathcal{P}1$ is equivalent to correcting the probability p, because every correction p' of p is a coherent probability assessment, hence it can be written as a convex combination of at most $n+1$ atoms, i.e., truth assignments which satisfy the logical constraint $\mathfrak{C}$ (for more details refer, e.g., to [25]). The binary variables a_{ij} are a representation of these atoms, because of the constraint 4., while the real variables q_j are the coefficients of the convex combination. The role of constraint number 2. is to set $b_{ij} = a_{ij} \cdot q_j$ (for $i = 1, \ldots, n$ and $j = 1, \ldots, n+1$), without using the multiplication, otherwise $\mathcal{P}1$ would not be a linear problem. The variables r_i, s_i are slack variables, which represent, respectively, the positive and the negative difference between $p(X_i)$ and $p'(X_i)$, as implied by the constraint 1. Finally, the objective function corresponds to minimize the L_1-distance between p and p', i.e., $\sum_{i=1}^n |p(X_i) - p'(X_i)|$.

From a theoretical point of view, to find the correction is a computational hard problem. Indeed given a probability assessment $\pi = (V, U, p, \mathfrak{C})$ and a real non-negative number D, it is a NP-complete problem to check if there exists a coherent probability assessment $\pi' = (V, U, p', \mathfrak{C})$ such that $d_1(p, p') \leq D$. The proof of NP-containment is easy because any solution of $\mathcal{P}1$ provides a succinct certificate for the existence of π' i.e., the values of a_{ij}, r_i and s_i's. While the NP–hardness derives from the fact that the coherence of π (which is a NP-complete problem) can be tested by posing $D = 0$.

Anyway, the actual implementations of MIP solvers make possible to solve probability correction problems of reasonable size in a feasible amount of time.

The optimal value δ for the objective function corresponds to the minimum possible correction on p and any coherent probability assessment $\pi' = (V, U, p', \mathfrak{C})$ such that $d_1(p, p') = \delta$ is a possible solution i.e., p' is an element of $\mathcal{C}(\pi)$.

In many situations $\mathcal{C}(\pi)$ has more than one element and the MIP problem is able to find just one solution, which could not be a good representative of all the elements of $\mathcal{C}(\pi)$, as it happens when it is an extreme value. Hence the following procedure to generate all the elements of $\mathcal{C}(\pi)$ is proposed.

Let $\mathcal{Q}$ be the set of all vectors $q \in \mathbb{R}^n$ such that the probability assessments $(V, U, q, \mathfrak{C})$ are coherent. $\mathcal{Q}$ forms a convex polytope whose extremal points are exactly the atoms, i.e., all truth-value assignments α which satisfy the logical constraints $\mathfrak{C}$.

Let $\mathcal{B}_\pi(\delta)$ be the ball of all vectors $q \in \mathbb{R}^n$ such that $d(p, q) \leq \delta$, with p the numerical probability assessment present in π. Such ball $\mathcal{B}_\pi(\delta)$ is a convex set whose extremal points are the points $p \pm \delta e_i$, where e_i is the i-th vector of the canonical basis, for $i = 1, \ldots, n$.

Then $\mathcal{C}(\pi)$ is a convex set of $\mathbb{R}^n$, because it is the intersection (see [29]) between the convex sets $\mathcal{Q}$ and $\mathcal{B}_\pi(\delta)$.

It is possible to describe $\mathcal{C}(\pi)$ in terms of its extremal points $q_1, \ldots, q_s$, indeed any element of $q \in \mathcal{C}(\pi)$ can be expressed as

$$q = \sum_{i=1}^{s} \lambda_i q_i$$

for some coefficients $\lambda_1, \ldots, \lambda_s \in \mathbb{R}$, such that $0 \leq \lambda_i \leq 1$, for $i = 1, \ldots, s$, and $\sum_{i=1}^s \lambda_i = 1$.

As a starting point, let us find a particular element $\bar{p} \in \mathcal{C}(\pi)$, which has the property that

$$\max_{i=1,\ldots,n} |\bar{p}(X_i) - p(X_i)| \qquad (1)$$

is minimum, among all the coherent assessments such that

$$d_1(\bar{p}, p) = \delta. \qquad (2)$$

This optimization problem can be formulated as a MIP problem $\mathcal{P}2$. All the constraints and the variables of $\mathcal{P}1$ are reported in $\mathcal{P}2$. Moreover, $\mathcal{P}2$ contains a new real variable z, which is subject to the constraints $r_i + s_i \leq z$, for $i = 1, \ldots, n$ (hence $z \geq \max_{i=1,\ldots,n}(r_i + s_i)$), and the new additional constraint $\sum_{i=1}^n (r_i + s_i) = \delta$ (which represents the equality (2)). In this way, the $\mathcal{P}2$ objective function to be minimized is simply z, since it equates (1).

The corrected assessment $\bar{\pi} = (V, U, \bar{p}, \mathfrak{C})$ differs from π by δ and tries to spread this difference as much as possible among the variables of U. Moreover, $\bar{p}$ is, in some sense, the most "central" point of $\mathcal{C}(\pi)$.

Using $\bar{p}$, it is possible to find the face F_1 of the polytope $\mathcal{Q}$ where $\mathcal{C}(\pi)$ lies. The face F_1 is itself a convex set with at most $n + 1$ atoms as extremal points, which can be found as a part of the solutions of $\mathcal{P}2$ (i.e., the optimal values of a_{ij}).

By looking at the signs of $\bar{p}(X_i) - p(X_i)$, for $i = 1, \ldots, n$, it is also possible to determine the face F_2 of $\mathcal{B}_\pi(\delta)$ which contains $\mathcal{C}(\pi)$. Indeed, F_2 is a convex set with at most n extremal points of the form

$$p + sign(\bar{p}(X_j) - p(X_j)) \cdot \delta \cdot e_j. \qquad (3)$$

The extremal points $Q = \{q_1, \ldots, q_s\}$ of $\mathcal{C}(\pi)$ can be easily found by means of the following procedure.

- let E_1 be the extremal points of F_1 and E_2 be the extremal points of F_2
- compute H_1 as the H-representation of F_1
- compute H_2 as the H-representation of F_2
- let $H = H_1 \cup H_2$, the H-representation of $F_1 \cap F_2 = \mathcal{C}(\pi)$
- compute Q as the V-representation of H

where the V-representation of a convex set C is the set of its extremal points, while the H-representation of C is a set H of half-spaces such that $C = \bigcap_{h \in H} h$. It is possible to convert from the V-representation of C to its H-representation by means of a face enumeration algorithm, while the inverse conversion is performed by a vertex enumeration algorithm [21].

Both steps can be computed in polynomial time as shown, for instance, in [21].

Let us summarize the whole process with the following pseudo-code where FaceEnum and VertexEnum are suitable procedures to compute the H and V representations.

procedure Correct
Input: assessment $(V, U, p, \mathfrak{C})$
Output: extr. points W and min. distance δ
begin
    prepare MIP program $\mathcal{P}1$
    solve it and extract the optimal value δ
    if $\delta = 0$ **then**
        return $(\{p\}, 0)$
    else
        prepare MIP program $\mathcal{P}2$
        solve it
        extract the values a_{ij}, r_i, s_i
        $E1$:=columns of matrix a_{ij}
        compute $\bar{p}$ from r_i, s_i
        compute $E2$ with formula 3
        $H1 :=$ FaceEnum($E1$)
        $H2 :=$ FaceEnum($E2$)
        $Q :=$ VertexEnum($H1 \cup H2$)
        return (Q, δ)
    endif
end

3.1 A Simple Numerical Example

Let us illustrate a simplified example that can help one to show the previous procedure step by step.

Example 1. *Consider a statistical analysis of doping in sports and how it improves performance while simultaneously damaging health. So let us consider the binary variables (i.e., events) $X_1 = D \equiv$"the athlete uses banned performance-enhancing drugs" (i.e., "doping"), $X_2 = E \equiv$"the athlete is showing a performance-enhancing in the last period" and $X_3 = H \equiv$"the athlete is showing a significant change in his/her biological profile".*

Hence the domain U of our assessment will be $U = \{D, E, H\}$, while, at the moment, the universal set V can be any, not better specified, superset $V \supseteq U$.

Suppose one obtains the probability values $p(D) = 0.9$, $p(E) = 0.8$ and $p(H) = 0.9$ by collecting information from disparate sources of information (e.g., public health registers, drugs consumption's and physicians' files, trainer interviews, etc.) on athletes showing significant increases in their performances or health alterations. At a first look the numerical evaluation $p = (0.9, 0.8, 0.9)$ on U, except from the extremely high values, could seems "acceptable". On the contrary, since doping causes both an enhancing in the performance and an alteration of the health and furthermore and since information is collected only among athletes already showing at least one of the two "symptoms", the assessment must be endowed with the logical constraints:

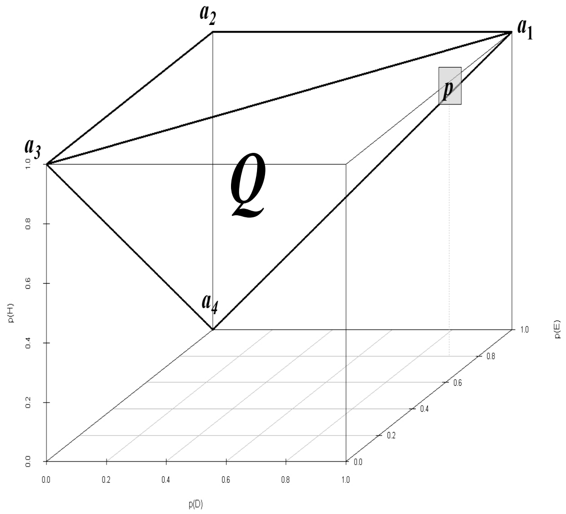

Figure 1: Configuration of the assessment of the doping example: the coherent assessments polytope $\mathcal{Q}$ is delimited by vertexes a_1, a_2, a_3, a_4; the initial incoherent assessment $p = (0.9, 0.8, 0.9)$ is at L_1 distance $\delta = 0.2$ from $\mathcal{Q}$.

$$D \Rightarrow E \wedge H; \quad (4)$$
$$\neg(E \vee H) = \bot, \quad (5)$$

or, equivalently, with the set of clauses

$$\mathfrak{C} = \{E \vee H, \neg D \vee E, \neg D \vee H\}. \quad (6)$$

Consequently the assessment $\pi = (V, U, p, \mathfrak{C})$ is incoherent since the set of coherent values $\mathcal{Q}$ is characterized by the probability inequalities

$$\begin{cases} p'(D) & \leq & p'(E) \\ p'(D) & \leq & p'(H) \\ p'(E) + p'(H) - p'(D) & \geq & 1 \end{cases},$$

and consist of the convex 0-1 polytope with vertexes $a1 = (1, 1, 1)$, $a2 = (0, 1, 1)$, $a3 = (0, 0, 1)$, $a4 = (0, 1, 0)$ and, as it also apparent from Fig. 1, p is outside it and hence incoherent. The first step of the previously described procedure, through MIP program $\mathcal{P}1$, returns that $\delta = 0.2$, while the second MIP program $\mathcal{P}2$ finds $\bar{p} = (0.833, 0.867, 0.967)$ so that one can

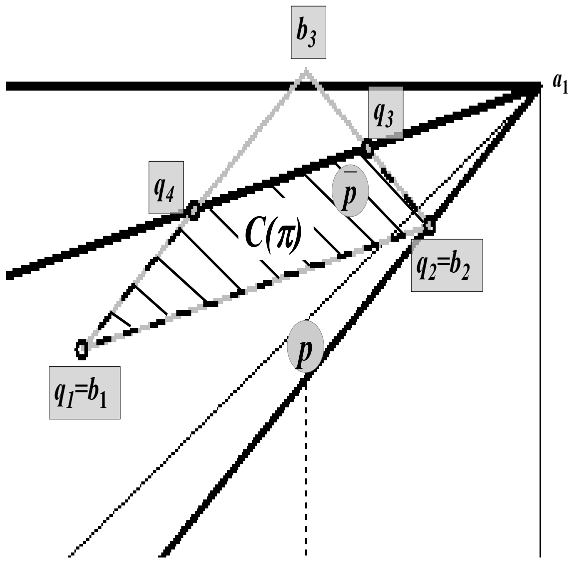

Figure 2: Zooming on the doping example assessment: facet F2 of $\mathcal{B}_\pi(\delta)$ is the grey triangle (b_1, b_2, b_3), the set of corrected assessment $\mathcal{C}(\pi)$ is the dashed polygon with vertexes $q_1 = b1$, $q_2 = b2$, q_3 and q_4.

derive the V-representations of the two facets as

$$E1 = \left\{ \begin{array}{l} a_{1\cdot} = a_1, \\ a_{2\cdot} = a_3, \\ a_{3\cdot} = a_4 \end{array} \right\} \quad (7)$$

$$E2 = \left\{ \begin{array}{l} b_1 = (0.7, 0.8, 0.9), \\ b_2 = (0.9, 1, 0.9), \\ b_3 = (0.9, 0.8, 1.1) \end{array} \right\}. \quad (8)$$

With them the intermediate calls to FaceEnum produce the H-representations (not reported here because of an hard reading) of the facets F1 and F2. The final call to the VertexEnum returns the V-representation of the whole set $C(\pi)$ of corrected values (the dashed polygon in Fig. 2) that is

$$Q = \left\{ \begin{array}{l} q_1 = b1 = (0.7, 0.8, 0.9), \\ q_2 = b2 = (0.9, 1, 0.9), \\ q_3 = (0.9, 0.9, 1), \\ q_4 = (0.8, 0.8, 1) \end{array} \right\}. \quad (9)$$

□

4 Merging Probability Assessments

Given two coherent probability assessments $\pi_1 = (V, U, p, \mathfrak{C})$ and $\pi_2 = (V, W, q, \mathfrak{D})$, on the same propositional variables V, one can say that π_1 and π_2 are compatible if for each variable $x \in U \cap W$, $p(x) = q(x)$. In other words, p and q coincide on the variables in common among π_1 and π_2. The compatibility of two probability assessments means that they do not assign, in an apparent way, different probability values to the same variable. Nevertheless, they can be contradictory by assigning different values to same proposition in an *implicit way*, thus the assessment formed by joining together π_1 and π_2 could be incoherent.

Example 2. *Take for example as $\pi_1 = (V, U, \overline{p}, \mathfrak{C})$ the "barycentric" correction of π in Ex.1 given by the first $\mathcal{P}1$ MIP program, and let $\pi_2 = (V, W, q, \mathfrak{D})$ be a further investigation over official training data that, agreeing with the percentages of enhancing performers and of biological perturbations, claims moreover that the percentage of athletes that naturally, i.e., without doping, are able to enhance significantly their performance showing biological modifications is of the 1%. Hence one has*

$$W = \{X_2 = E, X_3 = H, X_4 = (\neg D \wedge E \wedge H)\}; \quad (10)$$

$$q = \left(\begin{array}{l} q(E) = \overline{p}(E) = 0.867, \\ q(H) = \overline{p}(H) = 0.967, \\ q(X_4) = 0.01 \end{array} \right); \quad (11)$$

$$\mathfrak{D} \equiv \mathfrak{C} \cup \{\neg D \vee \neg X_4, E \vee \neg X_4, H \vee \neg X_4\}. \quad (12)$$

By construction π_1 and π_2 are "compatible" since they give the same probabilities to the common subdomain $U \cap W = \{E, H\}$, but they disagree on X_4 since the all coherent extensions of $\overline{p}$ give zero probability to X_4.□

If two probability assessments $\pi_1 = (V, U, p, \mathfrak{C})$ and $\pi_2 = (V, W, q, \mathfrak{D})$ are compatible, one can denote $\pi_1 + \pi_2$ as the probability assessment $(V, U \cup W, r, \mathfrak{C} \cup \mathfrak{D})$, where $r : U \cup W \to [0, 1]$ is defined by joining together p and q, i.e.,

$$r(x) = \left\{ \begin{array}{ll} p(x) & \text{if } x \in U \\ q(x) & \text{if } x \in W \end{array} \right.$$

The compatibility condition assures that the value of $r(x)$, when $x \in U \cap W$, is uniquely defined.

Given two compatible probability assessments $\pi_1 = (V, U, p, \mathfrak{C})$ and $\pi_2 = (V, W, q, \mathfrak{D})$, the merging operation of π_1 and π_2 is defined by

$$\pi_1 \oplus \pi_2 = \text{Correct}(\pi_1 + \pi_2)$$

Example 2. *(continues) If one takes the juxtaposition of the two assessments π_1 and π_2 one gets an assessment $\pi_1 + \pi_2$ with components V, $U \cup W = (D, E, H, X_4)$, $r = (0.8333, 0.8667, 0.9667, 0.01)$ and logical constraints $\mathfrak{D}$. It is, as explained before, incoherent, with an L_1 minimal distance of $\delta = 0.01$ and its correction $\pi_1 \oplus \pi_2$ is the credal set with extremal* *numerical values*

$$
\begin{aligned}
q_1 &= (0.8333, 0.8767, 0.9667, 0.01)\\
q_2 &= (0.8333, 0.8667, 0.9767, 0.01)\\
q_3 &= (0.8233, 0.8667, 0.9667, 0.01)\\
q_4 &= (0.8333, 0.8667, 0.9667, 0.00). \qquad \square
\end{aligned}
$$

Compatibility is not always present between different assessments, especially if the two sources of information stem from disparate contexts. When the probability assessments to be merged are non compatible, i.e., they assign different probability values to some common variables, it is not possible to join directly them into a unique assessment. Anyhow two different solutions are possible: a "weighted combination" of the two assessments, or a "assignment to duplicates", as detailed in the next paragraphs.

The first approach requires one to create a non contradictory probability assessment from π_1 and π_2, by choosing a unique probability value for each variable in $U \cap W$. A possible solution is to use a weighted average of p and q, i.e., chosen a weighting coefficient $\omega \in [0,1]$, where $\pi_1 +_\omega \pi_2$ defines the probability assessment $(V, U \cup W, r, \mathfrak{C} \cup \mathfrak{D})$, where $r : U \cup W \to [0,1]$ is now defined

$$
r(x) = \begin{cases} p(x) & \text{if } x \in U \setminus W \\ q(x) & \text{if } x \in W \setminus U \\ \omega p(x) + (1-\omega) q(x) & \text{if } x \in U \cap W \end{cases}
$$

Finally, the merging operation of π_1 and π_2 is

$$
\pi_1 \oplus_\omega \pi_2 = \mathrm{Correct}(\pi_1 +_\omega \pi_2)
$$

When $\omega = \frac{1}{2}$, equal importance is given to π_1 and π_2 and $\oplus_{\frac{1}{2}}$ becomes commutative. While the extreme values $\omega = 0$ and $\omega = 1$ correspond to the cases where the values of π_2 (or π_1, respectively), are used for contradictory situations. In some sense $\frac{\omega}{1-\omega}$ is a measure of the relative reliability of π_1 with respect to π_2.

Example 3. *If one renders explicit the contradiction on X_4 of the two assessment π_1 and π_2 of Ex.2, i.e., by considering $p(X_4) = 0$, and one chooses $\omega = \frac{1}{2}$, one has the starting weighted assessment $\pi_1 +_{\frac{1}{2}} \pi_2$ with components V, $U \cup W = (D, E, H, X_4)$, $r = (0.8333, 0.8667, 0.9667, 0.005)$ and logical constraints again expressed through the same set of clauses $\mathfrak{D}$. It is anyhow incoherent with an L_1 minimal distance of $\delta = 0.01$ but its correction $\pi_1 \oplus_{\frac{1}{2}} \pi_2$ is now the credal set with extremal values*

$$
\begin{aligned}
q_1 &= (0.8333, 0.8742, 0.9667, 0.0075)\\
q_2 &= (0.8308, 0.8642, 0.9667, 0.00)\\
q_3 &= (0.8333, 0.8667, 0.9742, 0.0075)\\
q_4 &= (0.8308, 0.8667, 0.9642, 0.00)\\
q_5 &= (0.8358, 0.8692, 0.9667, 0.00)\\
q_6 &= (0.8258, 0.8667, 0.9667, 0.0075). \qquad \square
\end{aligned}
$$

Of course, by varying the weight ω in $\pi_1 \oplus_\omega \pi_2$, one obtains a class of new coherent (imprecise) assessments over the domain $U \cup W$ that have the peculiarity of being "compromises" of the two original π_1 and π_2, but with the same weight for each event with associated different values.

A different approach is to create a probability assessment which maintains both numerical values and to solve the apparent contradiction by adding a new logical variable X_i', for each event $X_i \in U \cap W$ such that $p(X_i) \neq q(X_i)$, and assigning the values $r(X_i) = p(X_i)$ and $r(X_i') = q(X_i)$. Moreover, the logical constraint $X_i = X_i'$ is added to $\mathfrak{C} \cup \mathfrak{D}$.

Indeed, the assessment so obtained $\pi_1 + \pi_2$ is obviously incoherent and the merging operation of π_1 and π_2 is computed as

$$
\pi_1 \oplus_I \pi_2 = \mathrm{Correct}(\pi_1 + \pi_2).
$$

Note that, whenever the two assessments π_1 and π_2 are compatible, this merging operator $\pi_1 \oplus_I \pi_2$ coincides with the previous $\pi_1 \oplus \pi_2$ since no duplication of variables is needed in such a case.

The main difference between the two approaches is that the latter $\oplus_I$ tries to automatically solve the contradiction, while the operator $\oplus_\omega$ needs an explicit way of solving it. The approach of $\oplus_\omega$ is in some sense a supervised one, because the user must explicitly provide a weight ω, while $\oplus_I$ adopts an unsupervised approach, and these difference can leads to very different final results, as the following example shows.

Example 4. *Let us proceed as in Ex.3 but maintaining the two distinct values associated to X_4, i.e., let us start with the assessment $\pi_1 + \pi_2$ with components V, $U' = (D, E, H, X_4, X_4')$, $r = (0.8333, 0.8667, 0.9667, 0.00, 0.01)$ and with logical constraints augmented to $\mathfrak{D} \cup \{\neg X_4 \vee X_4', X_4 \vee \neg X_4'\}$. This further assessment has again a minimal L_1 distance of $\delta = 0.01$ from the polytope $\mathcal{Q}$ of coherent assessments (note anyhow the different cardinality of the space $n = 5$), but whose correction leads now to a precise assessment with numerical values*

$$
(0.8333, 0.8667, 0.9667, 0.00, 0.00). \qquad \square
$$

Anyway, the idea behind these two definitions is the same, i.e., the merging of two information sources can be performed in two steps. First, put together all the information $\mathcal{I}$, and then find the smallest number of corrections on $\mathcal{I}$ such that the new information $\mathcal{I}'$ is consistent. The choice of which merging operator adopt should be based on the availability or not of relevance, or better of the reliability, of the sources of information. If a reliability grade is available, or reasonably assessed, the $\oplus_\omega$ should be preferred, if not the $\oplus_I$ operator avoids the use of unrealistic assumptions.

Thinking the probability assessments as belief states, the merging operators are a belief merging functions (see, e.g., [22]).

Our approach is different from usual imprecise probability technique "a la Walley" (see in particular [34]), where usually the convex hull of incompatible assessments is considered. This is a so called "least commitment" procedure, while our proposal can be dually thought as "maximal commitment". In fact, in our merging operators, values which are exogenous to the initial assessments (like those appearing by doing the convex hull) are avoided as much as possible, and original opinions are maintained fixed and crisp as much as possible. Moreover the convex hull of initial assessments is not guaranteed to at least "avoid sure loss", so that the Walley's "natural extension" procedure is not always applicable. On the contrary, our approach is always applicable.

5 Revising Probability Assessments

In this section we propose how the correction procedure can be used to revise a probability assessment.

Suppose that the coherent probability assessment $\pi_1 = (V, U, p, \mathfrak{C})$ represents our current belief state and a new reliable information arrives, represented by the probability assessment $\pi_2 = (V, W, q, \mathfrak{D})$.

One could merge π_1 and π_2 as described in the previous section, but suppose that one would rather update our belief state with the new available information, with the idea that

- one assumes that the new information is correct
- one allows to revise, as less as possible, our current state in order to adapt it to the new information

The revision can be performed as follows. First, π_1 and π_2 are merged together with the operator $+_0$, thus in the case of contradiction, the values from π_2 are used. Second, the resulting assessment is corrected by forbidding any change the probabilities of the variables in W. This can be achieved with the procedure Correct2 which is a small modification of the procedure Correct. Correct2 has a further parameter, the set T of the variables whose probability value cannot be corrected, and when the MIP systems $\mathcal{P}1$ and $\mathcal{P}2$ are built, the constraint 1 for the variables of T reduces to

$$\sum_{j=1}^{n+1} b_{ij} = p(X_i)$$

and their corresponding variable r_i and s_i are not created.

The revision of π_1 with π_2 is then computed as

$$\pi_1 \star \pi_2 = Correct2(\pi_1 +_0 \pi_2, W)$$

Note that any probability assessment $(V, U \cup W, r', \mathfrak{C} \cup \mathfrak{D})$ resulting from $\pi_1 \star \pi_2$ is such that it agrees with q, i.e., $r'(x) = q(x)$ for all $x \in W$.

Example 5. *If in Ex.2 one wants to inevitably maintain as valid the latter investigation π_2 one starts with an adjoined initial assessment $\pi_1 +_0 \pi_2$ with components V, $U \cup W = (D, E, H, X_4)$, $W = (E, H, X_4)$, $r = (0.8333, 0.8667, 0.9667, 0.01)$ and logical constraints $\mathfrak{D}$. The only possibility to correct it is to reduce the numerical evaluation $r(D) = 0.8333$ to $r'(D) = 0.823$, so that the result of the revision is the precise assessment $\pi_1 \star \pi_2$ with components V , $U \cup W = (D, E, H, X_4)$, $r' = (0.8233, 0.8667, 0.9667, 0.01)$ and the same logical constraints $\mathfrak{D}$.* □

Such revising methodology, that in general leads to an imprecise model, could be thought as an analogous of the famous Jeffrey's rule of combination [26]. The main difference between the two is that our proposal minimize the probability mass dislocation from the original assessment, maintaining as much as possible the magnitude of the values, hence working in an "additive" way, while Jeffrey's rule maintains as much as possible the odds ratios, hence working in a "multiplicative" way.

Moreover the Jeffrey's rule produces a final probability assessment which could be too different from π since it inevitably alters all the values of p on $U \setminus W$, while our approach tries to modify p as less as possible, in line with the belief revision methodology [22].

6 Conclusions

In this article, a preliminary proposal for a correction of incoherent probability assessments on finite domains through L_1 distance minimization was presented. The proposal's novelty is reflected in a new procedure that uses mixed integer programming while profiting from geometrical properties of the convex sets involved, makes such a method easily applicable.

Apart from the applicability of the direct incoherent correction "per se," we have stressed that such a method can be tailored to naturally implement the merging of disparate assessments or reasonable belief revisions. We focused on the combination of two different assessments, but the generalization to the merging or revision of several assessments is straightforward: it simply requires the generalization of the weighted combination $+_\omega$ by allowing convex combination of several values and to iterate the juxtaposition + which duplicates several times.

Our procedure can be seen as a reasonable way to generate lower-upper probability models from precise, but incoherent, probabilities estimates.

Future research should systematically analyze the procedure through simulation studies and investigate formal properties of the correction operator. We are confident that the revising $\star$ operator satisfies some properties which are the probabilistic counter-parts of the Katsuno-Mendelzon axiom for belief revision operators.

Acknowledgments

This work has been partially supported by Italian Ministry of Education, University and Research funding of Research Projects of National Interest (PRIN 2010-11) under grant 2010FP79LR_003 *"Logical methods of information management"* and by Italian Ministry of Health under grant J52I14001640001 *"Sistemi intelligenti di ausilio alle decisioni per l'identificazione precoce e la dissuasione all'utilizzo del doping"*.

References

[1] M. Baioletti, A. Capotorti, S. Tulipani, B. Vantaggi. Elimination of Boolean variables for probabilistic coherence. *Soft Computing*, 4(2):81–88, 2000.

[2] M. Baioletti, A. Capotorti, S. Tulipani, B. Vantaggi. Simplification Rules for the Coherent Probability Assessment Problem. *Ann. of Math. and Artif. Intell.*, 35:11–28, 2002.

[3] M. Baioletti, A. Capotorti, S. Tulipani. An empirical complexity study for a 2CPA solver. In: *Modern Information Processing: From Theory to Applications*. B. Bouchon-Meunier, G. Coletti and R.R. Yager, Eds. 1–12, 2005.

[4] C.E. Bonferroni. Teorie e probabilità. Discorso Inaugurale annuario 1925-1926. *Regio Istituto Superiore di Scienze Economiche e Commerciali*, Bari, Casa Editrice Cressati.

[5] A. Capotorti. Benefits of embedding structural constraints in coherent diagnostic processes. *Int. J. of Approx. Reasoning*, 39(2-3):211–233, 2005.

[6] A. Capotorti. A Further Empirical Study on the Over-Performance of Estimate Correction in Statistical Matching. In: *Advances in Computational Intelligence*. S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, R.R. Yager Eds. Springer Berlin Heidelberg. 124–133, 2012.

[7] A. Capotorti, G. Regoli, Coherent correction of inconsistent conditional probability assessments, in: L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds), Proceeding of IPMU'08, Malaga (ES), 891–898, 2008.

[8] A. Capotorti, G. Regoli, F. Vattari, Correction of incoherent conditional probability assessments, *Int. J. of Approx. Reasoning* 51(6): 718–727, 2010.

[9] A. Capotorti, B. Vantaggi, Incoherence correction strategies in statistical matching, In: *Proceedings of ISIPTA 2011*, Innsbruck (Austria), 109–118, 2011.

[10] A. Capotorti, B. Vantaggi. Correction of Incoherences in Statistical Matching. In: *Statistical Models for Data Analysis*. P. Giudici, and S. Ingrassia, and M. Vichi Eds. Springer International Publishing, 73–80, 2013.

[11] G. Coletti. Numerical and Qualitative Judgments in Probabilistic Expert Systems. In *Proc. of the International Workshop on Probabilistic Methods in Expert Systems*, Romano Scozzafava Ed. Roma: SIS, 37–55, 1993.

[12] G. Coletti. Coherent numerical and Ordinal probabilistic assessments. *IEEE Transaction on Systems, Man, and Cybernetics*, 24:1747–1754, 1994.

[13] G. Coletti, R. Scozzafava, *Probabilistic Logic in a Coherent Setting*, Dordrecht: Kluwer, Series "Trends in Logic", 2002.

[14] F.G. Cozman, L. Fargoni di Ianni. Probabilistic Satisfiability and Coherence Checking through Integer Programming. *Lecture Notes in Computer Science*, 7958, 145–156, 2013.

[15] A. P. Dawid. The well-calibrated Bayesian. *Journal of the American Statistical Association*, 77:605-613, 1982.

[16] A. P. Dawid. Calibration-Based Empirical Probability. *Ann. Statist.* 13(4): 1251–1274, 1985.

[17] B. de Finetti. La prevision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré. Section B*, 7:l–68, 1937.

[18] B. de Finetti. *Teoria della Probabilità.* Torino: Einaudi, 1970 (Engl. transl. *Theory of probability*, London: Wiley & Sons, 1974).

[19] M. Finger, G. De Bona. Probabilistic satisfiability: Logic based algorithms and phase transition. *IJCAI*, 528-533, 2011.

[20] M. Fréchet. *Les Mathématiques et le Concret*, Paris. P.U.F.

[21] K. Fukuda. Lecture: Polyhedral Computation, Spring 2011. Institute for Operations Research and Institute of Theoretical Computer Science. ETH Zurich. Available online at `http://stat.ethz.ch/ifor/teaching/lectures/poly_comp_ss11/lecture_notes`.

[22] P. Gärdenfors and H. Rott. Belief revision. In Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 4, pages 35–132. Oxford University Press. (1995)

[23] G.Georgakopoulos, D.Kavvadias and C.H.Papadimitriou. Probabilistic Satisfiability. *Journal of Complexity.* 4:1–11, 1988.

[24] A. Gilio. Probabilistic Consistency of Knowledge Bases in Inference Systems. *Lecture Notes in Computer Science*, M. Clarke, R. Kruse, S. Moral Eds. Springer–Verlag, 747:160–167, 1993.

[25] B. Jaumard, P. Hansen, M. Poggi de Aragão. Column Generation Methods for Probabilistic Logic. *ORSA Journal on Computing*, 3(2): 135-148, 1991.

[26] R.C. Jeffrey. *The logic of decision.* McGraw-Hill, 1965, 2nd Ed. Univ. Chicago Press, Chicago, 1983.

[27] F. Lad, Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction, Wiley, New York, 1996.

[28] D. V. Lindley, A. Tversky And R. V. Brown. On the Reconciliation of Probability Assessments. *J. R. Statist. Soc. A*, 142(2):146–180, 1979.

[29] K.G. Murty, K.S. Al-Sultan. On Relationship between L_1, L_2, L_∞ Minima. *Technical Report 89-26*, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, USA, 1989.

[30] D. Osherson, D. Lane, P. Hartley, and R. R. Batsell. Coherent Probability from Incoherent Judgment. *Journal of Experimental Psychology: Applied*, 7(1):3–12, 2001.

[31] J. Pratt. Must subjective probabilities be realized as relative frequencies? *Unpublished seminar paper*, Harvard University Graduate School of Business Administration, *Technical report* 1962.

[32] D. Pretolani. Probability Logic and Optimization SAT: the PSAT and CPA Models. *Ann. of Math. and Artif. Intell.*, 43:211–221, 2005.

[33] B. Vantaggi. Statistical matching of multiple sources: A look through coherence. *Int. J. of Approx. Reasoning* 49:701–711, 2008.

[34] P. Walley, The elicitation and aggregation of beliefs. *Technical Report University of Warwick*, 1982.

[35] P. Walley, *Statistical reasoning with Imprecise Probabilities.* Chapman and Hall, London 1991.