Coherent correction of inconsistent conditional probability assessments

A. Capotorti, G.Regoli
Dip. Matematica e Informatica
Università di Perugia - Italy
{capot,regoli}@dipmat.unipg.it

Abstract

In this paper we suggest a procedure to adjust an incoherent conditional probability assessment given on a partial domain. We look for a solution that tries to attain two separate goals: on one hand the solution should be as close as possible to the initial assessments, on the other hand we do not want to insert more information than we had at the beginning. The first goal is achieved by minimizing an appropriately defined distance among assessments, while for the second we look for a "maximum entropy" like solution.

Keywords: Conditional probability coherence, Divergence, Scoring rules.

1 Introduction

In practical applications it is natural to give evaluations of probability only of relevant events; moreover it can happen that these evaluations do not fit well with each other, especially when coming from different sources. Another common feature is that events are judged under specific circumstances, implying a conditional assessment. Often the assessment is intended to be used for inference purposes, i.e. to see how a further (conditional) event can be evaluated consistently with the initial assessment. Of course, the inferential results are meaningful only if the prior information encompassed in the initial assessment is coherent by itself. If not, a modification is required. Usually such a problem is solved with a revision of the initial evaluations. We propose a methodology for choosing an assessment correction automatically. A similar proposal can be found in Kriz [9].

Therefore our input consists of an incoherent conditional probability assessment given on a partial domain. We want to find a coherent assessment on the same domain that will preserve the opinion expressed by the initial assessment as much as possible, without introducing exogenous information. This goal is obtained by minimizing some kind of distance among partial conditional assessments.

(Pseudo)distances among probability distributions are usually measured through divergencies (e.g. Euclidean distance, Kulback-Leibler divergence, Csiszár f-divergences, etc.). Some of them can be applied only among unconditional full probability distributions; others could be applied to our context of partial conditional assessments (see for example [9]), but do not have any probabilistic justification, being purely geometrical tools. Hence, for our purpose, in this paper we introduce an index of "discrepancy" among partial conditional probability assessments which is derived by a particular scoring rule. Such a scoring rule is inspired by the one, introduced by Lad in [11] for unconditional probability distributions, and adapted here to conditional-logical arguments.

Independently of the divergence used to extrapolate the closest coherent assessment, for inference purposes, among all the compatible models, we suggest selecting the one that introduces least additional information. Hence we look for a "maximum entropy" like solution, not far from the suggestion of [7].

Operationally our purpose reduces to a two step optimization procedure. In the first step the discrepancy index is minimized. Peculiar to this part are the incompleteness of the assessment, the non linearity of the objective function and the non convexity of the set of coherent conditional assessments. To avoid the last aspect, we move to a search space formed by unconditional probability distributions. On the other hand, the second step is a canonical maximization of entropy, but with constraints obtained by the first.

The paper is organized as follows: in Section 2, after a brief formalization of the problem, a scoring rule suitable for our framework is proposed. In Section 3, we introduce the discrepancy index and prove its fundamental properties. Then in Subsection 3.1, for comparison purposes, we illustrate other possible divergencies. In Subsection 3.2 numerical examples of discrepancy minimization illustrate the operational effectiveness of our proposal. Finally, in Subsection 3.3 the second step of entropy maximization is formalized and the whole procedure is illustrated by one last numerical example. A short conclusive section closes the paper.

2 A short description of the problem

2.1 Notation

We formalize the problem briefly.

A field expert, in the sequel named the "assessor", elicits a finite family of conditional events $\mathcal{E} = [E_1|H_1, \ldots, E_n|H_n]$ as domain of his/her evaluations. The events E_i's usually represent the situations under consideration, while the H_i's usually represent the different contexts, or scenarios, under which the E_i's are evaluated.

The basic events $E_1, \ldots, E_n, H_1, \ldots, H_n$ can be endowed with logical constraints, that represent dependencies among particular configurations of them.

In the following E_iH_i will denote the logical connection "E_i and H_i", E_i^c will indicate "not E_i" and the event $H^0 = \bigvee_{i=1}^n H_i$ will represent the whole set of contexts.

Starting with the basic events $E_1, \ldots, E_n, H_1, \ldots, H_n$ it is possible, to span a sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$, where ω_j represents a generic atom, in some context named "possible world ". Note that the sample space Ω, together with H^0, are not part of the assessment but only auxiliary tools.

As is well known, a conditional event $E_i|H_i$ is a three-valued logical entity, partitioning Ω in three parts: the atoms satisfying E_iH_i and thus verifying the conditional, those satisfying $E_i^cH_i$, thus falsifying the conditional, and those not fulfilling the context H_i, to which the conditional may not be applied at all. This is usually synthesized with the indicator function

$$|E_i|H_i| = \begin{cases} 1 & \text{if } E_iH_i \text{ occurs} \\ 0 & \text{if } E_i^cH_i \text{ occurs} \\ \text{undetermined} & \text{if } H_i^c \text{ occurs} \end{cases}.$$

We will denote with $\mathcal{A}$ the set of probability distributions $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_k]$ over Ω, with the restriction $\alpha(H_i) = \sum_{\omega_j \in H_i} \alpha_j > 0$, $i = 1, \ldots, n$. In particular, we will mainly use the subset of probability distributions $\mathcal{A}^0 = \{\boldsymbol{\alpha} \in \mathcal{A} \text{ such that } \alpha(H^0) = \alpha(\bigvee H_i) = 1\}$. Any $\boldsymbol{\alpha} \in \mathcal{A}$ induces a coherent conditional assessment on $\mathcal{E}$

$$\mathbf{q}_{\boldsymbol{\alpha}} = [q_{\alpha_i} = \frac{\displaystyle\sum_{\omega_j \subseteq E_iH_i} \alpha_j}{\displaystyle\sum_{\omega_j \subseteq H_i} \alpha_j}, i = 1, \ldots, n].$$

By $\mathcal{Q}_{\mathcal{E}}$ we will denote the set of coherent assessments attainable by some distribution in $\mathcal{A}^0$, i.e. $\mathcal{Q}_{\mathcal{E}} = \{\mathbf{q} \in [0,1]^n : \exists\, \boldsymbol{\alpha} \in \mathcal{A}^0 \text{ s.t. } \mathbf{q} = \mathbf{q}_{\boldsymbol{\alpha}}\}$.

On the other hand, for any $\mathbf{q} \in \mathcal{Q}_{\mathcal{E}}$ it is possible to identify the convex set of probability distributions "compatible" with $\mathbf{q}$

$$\mathcal{A}_{\mathbf{q}} := \{\boldsymbol{\alpha} \in \mathcal{A} | \mathbf{q}_{\boldsymbol{\alpha}} \equiv \mathbf{q}\};$$

and its nonempty convex subset

$$\mathcal{A}^0_{\mathbf{q}} := \mathcal{A}_{\mathbf{q}} \cap \mathcal{A}^0.$$

$\mathcal{A}_{\mathbf{q}}$ can be interpreted as the set of probabilistic models implicitly accepted by an assessor of $\mathbf{q}$.

The input of our problem will consist of an incoherent assessment $\mathbf{p} = [p_1, \ldots, p_n] \in (0,1)^n$, with $\mathbf{p} \notin \mathcal{Q}_{\mathcal{E}}$, and we want to "adjust" it. Note that, at the moment, we exclude "extreme" evaluations $p_i = 0$ or $p_i = 1$.

In Bayesian statistic, to challenge the soundness of any probability assessment, it is usual to introduce a "scoring" rule. Scoring rules can be interpreted in different ways. The simplest one is to think of them as hypothetical penalties suffered by the assessor. The actual loss must depend on which events occur and on what probabilities have been assessed for them. In this way, to minimize the expected loss, an assessor is induced to elicit probability values honestly by following his/her information.

For partial (coherent or not) conditional assessments $\mathbf{v} = [v_1, \ldots, v_n] \in (0,1)^n$ over $\mathcal{E} = [E_1|H_1, \ldots, E_n|H_n]$, we propose as scoring rule the random variable

$$S(\mathbf{v}) := \sum_{i=1}^{n} |E_i H_i| \ln v_i + \sum_{i=1}^{n} |E_i^c H_i| \ln(1 - v_i)$$

with $|\cdot|$ indicator function of unconditional events.

The motivation of such a score is that the assessor "loses less" the higher are the probabilities assessed for events that are verified, and at the same time, the lower are the probabilities assessed for those that are not verified. The values assessed on events that turn out to be undetermined do not influence the score.

Such score $S(\mathbf{v})$ is an "adaptation" of the "proper scoring rule" for probability distributions proposed by Lad in [11](pag. 355). We have extended it to partial and conditional probability assessments.

Finally we need to express the expected value of a score $S(\mathbf{v})$ with respect to a probability distribution $\boldsymbol{\alpha} \in \mathcal{A}$:

$$E_{\boldsymbol{\alpha}}(S(\mathbf{v})) := \sum_{j=1}^{k} \alpha_j \left(\sum_{i: E_i H_i \supseteq \{\omega_j\}} \ln v_i + \sum_{i: E_i^c H_i \supseteq \{\omega_j\}} \ln(1 - v_i) \right).$$

3 Our correction procedure

The aim of our proposal is to "adjust" $\mathbf{p}$ by a coherent assessment $\mathbf{q_p} \in \mathcal{Q}_{\mathcal{E}}$ in such a way that the difference between the corresponding expected scores is minimal.

Hence, we can introduce the "discrepancy" between an assessment $\mathbf{p}$ over $\mathcal{E}$ and a distribution $\boldsymbol{\alpha} \in \mathcal{A}$, with respect to its conditional coherent assessment $\mathbf{q}_{\boldsymbol{\alpha}}$, as

$$\Delta(\mathbf{p}, \boldsymbol{\alpha}) := E_{\boldsymbol{\alpha}}(S(\mathbf{q}_{\boldsymbol{\alpha}}) - S(\mathbf{p})). \qquad (1)$$

This discrepancy $\Delta(\mathbf{p}, \boldsymbol{\alpha})$ behaves like other common divergences. In fact the following theorem holds.

Theorem 1 *Let $\Delta(\mathbf{p}, \boldsymbol{\alpha})$ be defined as in (1). Then*

i) $\Delta(\mathbf{p}, \boldsymbol{\alpha}) \geq 0 \quad \forall \boldsymbol{\alpha} \in \mathcal{A}$;
ii) $\Delta(\mathbf{p}, \boldsymbol{\alpha}) = 0$ *iff* $\mathbf{p} \equiv \mathbf{q}_{\boldsymbol{\alpha}}$;
iii) $\Delta(\mathbf{p}, \boldsymbol{\alpha})$ *is convex on* $\boldsymbol{\alpha}$.

Proof: Recall that, by definition,

$$q_{\alpha i} = \alpha(E_i|H_i) = \frac{\alpha(E_i H_i)}{\alpha(H_i)}; \qquad (2)$$

$$(1 - q_{\alpha i}) = (1 - \alpha(E_i|H_i)) = \frac{\alpha(H_i) - \alpha(E_i H_i)}{\alpha(H_i)} = \frac{\alpha(E_i^c H_i)}{\alpha(H_i)}. \qquad (3)$$

Hence we have

$$\Delta(\mathbf{p}, \boldsymbol{\alpha}) = \sum_{j=1}^{k} \alpha_j \left(\sum_{i: E_i H_i \supseteq \{\omega_j\}} \ln \frac{q_{\alpha i}}{p_i} + \sum_{i: E_i^c H_i \supseteq \{\omega_j\}} \ln \frac{1 - q_{\alpha i}}{1 - p_i} \right) =$$

$$= \sum_{j=1}^{k} \sum_{i: E_i H_i \supseteq \{\omega_j\}} \alpha_j \ln \frac{q_{\alpha i}}{p_i} + \sum_{j=1}^{k} \sum_{i: E_i^c H_i \supseteq \{\omega_j\}} \alpha_j \ln \frac{1 - q_{\alpha i}}{1 - p_i} =$$

(by distributivity)

$$\begin{array}{rcr} = & \sum_{i=1}^n \ln \frac{q_{\alpha i}}{p_i} \sum_{j:\{\omega_j\}\subseteq E_i H_i} \alpha_j & + \\ + & \sum_{i=1}^n \ln \frac{1-q_{\alpha i}}{1-p_i} \sum_{j:\{\omega_j\}\subseteq E_i^c H_i} \alpha_j & = \\ = & \sum_{i=1}^n \ln \frac{q_{\alpha i}}{p_i} \alpha(E_i H_i) & + \\ + & \sum_{i=1}^n \ln \frac{1-q_{\alpha i}}{1-p_i} \alpha(E_i^c H_i) & = \\ = & \sum_{i=1}^n \ln \frac{q_{\alpha i}}{p_i} q_{\alpha i} \alpha(H_i) & + \\ + & \sum_{i=1}^n \ln \frac{1-q_{\alpha i}}{1-p_i} (1-q_{\alpha i}) \alpha(H_i) & = (4) \end{array}$$

(by (2) and (3))

$$\begin{array}{rl} = & \sum_{i=1}^n \alpha(H_i)[q_{\alpha i} \ln \frac{q_{\alpha i}}{p_i} - q_{\alpha i} + p_i + \\ + & (1-q_{\alpha i}) \ln \frac{1-q_{\alpha i}}{1-p_i} - (1-q_{\alpha i}) + (1-p_i)] \end{array}$$

$$= \sum_{i=1}^n \alpha(H_i)[\beta_i^1 + \beta_i^2] \qquad (5)$$

where each β_i^ϵ is of the form $x_i \ln \frac{x_i}{y_i} - x_i + y_i$, $x_i, y_i \in (0,1)$, $\epsilon = 1, 2$, and therefore it is easy to verify that $\beta_i^\epsilon \geq 0$, $i = 1, \ldots, n$, with $\beta_i^\epsilon = 0$ if and only if $x_i = y_i$. This proves *i*) and *ii*).

On the other hand, by applying (2) and (3) to (4) it follows that

$$\begin{array}{rcl} \Delta(\mathbf{p}, \boldsymbol{\alpha}) & = & \sum_{i=1}^n (\alpha(E_i H_i) \ln \frac{\alpha(E_i H_i)}{\alpha(H_i)} \\ & + & \alpha(E_i^c H_i) \ln \frac{\alpha(E_i^c H_i)}{\alpha(H_i)} \\ & - & \alpha(E_i H_i) \ln p_i - \alpha(E_i^c H_i) \ln(1-p_i)) \end{array}$$

that is Δ is of the form $\sum \gamma_i \ln \frac{\gamma_i}{\varsigma_i} + \gamma_i \psi_i$ which is convex on the pairs (γ_i, ς_i). Therefore Δ is convex on $\boldsymbol{\alpha}$, being γ_i and ς_i sums of α_j's. $\square$

The remark about expression (5) reveals that the discrepancy $\Delta(\mathbf{p}, \boldsymbol{\alpha})$ turns out to be a generalization of the sum of two different "Bregman divergences" [2]

$$D(\mathbf{x}, \mathbf{y}) := \phi(\mathbf{x}) - \phi(\mathbf{y}) - < \mathbf{x} - \mathbf{y}, \nabla\phi(\mathbf{y}) > \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

In fact if $D_1(\mathbf{x}, \mathbf{y})$ is obtained by $\phi_1(\mathbf{x}) = \sum_{i=1}^n x_i \ln x_i$ and $D_2(\mathbf{x}, \mathbf{y})$ by $\phi_2(\mathbf{x}) = \sum_{i=1}^n (1 - x_i) \ln(1 - x_i)$ we would have $D_1(\mathbf{x}, \mathbf{y}) + D_2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n [\beta_i^1 + \beta_i^2]$. In $\Delta(\mathbf{p}, \boldsymbol{\alpha})$ each term is weighted by $\alpha(H_i)$, which reflects the "relevance" of each context H_i with respect to all the assessments.

Now we suggest using as coherent correction of $\mathbf{p}$ the assessment $\mathbf{q_p}$ generated by the distribution $\widetilde{\boldsymbol{\alpha}}$ solution of the nonlinear optimization program

$$\min_{\boldsymbol{\alpha} \in \mathcal{A}^0} \Delta(\mathbf{p}, \boldsymbol{\alpha}). \qquad (6)$$

The motivation for this choice is that (intuitively) the assessor of $\mathbf{p}$ would expect to suffer the penalty $S(\mathbf{p})$, hence we select the coherent assessment $\mathbf{q_p}$ that has a (probabilistic) expected score as close as possible.

Note that we restrict ourself to evaluate the discrepancy only over $\mathcal{A}^0$ because it is easy to see that, by making $\alpha(H^0)$ smaller and smaller, we get

$$\inf_{\boldsymbol{\alpha} \in \mathcal{A}} \{\Delta(\mathbf{p}, \boldsymbol{\alpha})\} = 0.$$

Hence we would be induced to look for the closest adjustment of $\mathbf{p}$ among those generated by distributions $\boldsymbol{\alpha}$ that weakly support the whole set of scenarios H^0. This would be meaningless because, relatively to $\mathcal{E}$, the only relevant atoms of Ω are those in H^0.

The optimal solution $\widetilde{\boldsymbol{\alpha}}$ of the optimization problem (6) is just an auxiliary component, while our attention is focused on its extension $\mathbf{q_p}$. In fact, adopting $\mathbf{q_p}$ as "best" approximation of $\mathbf{p}$ implicitly admits as reasonable models not only $\widetilde{\boldsymbol{\alpha}}$ but also all others agreeing distributions in $\mathcal{A}^0_{\mathbf{q_p}}$.

3.1 Other divergencies

For the purpose of comparison, we also considered other divergencies among assessments:

1. $L_1(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n |q_i - p_i|$;
2. $L_2(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n (q_i - p_i)^2$;
3. $CD(\mathbf{p}, \mathbf{q}, \boldsymbol{\alpha}) = \sum_{i=1}^n (\sum_{\omega_j \subseteq H_i} \alpha_j)(q_i \ln(q_i/p_i) - q_i + p_i)$ with $\boldsymbol{\alpha} \in \mathcal{A}_{\mathbf{q}}$;

4. $R(\mathbf{p},\mathbf{q}) = \sum_{i \neq j} |\ln(p_i/p_j) - \ln(q_i/q_j)|$;

5. $O(\mathbf{p},\mathbf{q}) = \sum_{i \neq j} |\ln(p_i/p_j) - \ln(q_i/q_j)| + \sum_i |\ln(p_i/(1-p_i)) - \ln(q_i/(1-q_i))|$.

L_1 and L_2 are usual metric distances, endowed with all their geometric properties, but without either an intuitive or a probabilistic interpretation to be used as distances between conditional assessments. CD is a direct adaptation of the logarithmic Bregman divergence to conditional probabilities. On a non conditional framework, such divergence is the most frequently adopted, because of its information theoretic properties. In fact it generalizes the well known Kulback-Leibler divergence [10] to partial assessments. Anyhow, it is known that this Bregman divergence is generated by a logarithmic scoring rule (see e.g. [12]). This logarithmic scoring rule evaluates only the events that occur, without considering those that turn out to be false. On the contrary, our score $S(\mathbf{p})$ evaluates all the events, both those that occur and those that do not, since it encompasses the implicit assessments $(1-p_i)$'s.

We have considered the last two divergences R and O just to see if it was possible to find a correction $\mathbf{q}$ that could maintain the relative proportions among the components p_i's, and among components p_i's and their complements $1-p_i$, respectively. However, we have not studied their theoretical properties because, as we will see in the following numerical examples, both R and O present computational drawbacks due to the presence of local minima.

3.2 Numerical results

We describe now some examples that will show the effectiveness of our procedure. The numerical results have been obtained trough the nonlinear optimization software CONOPT of the package General Algebraic Modeling System (GAMS) [3].

Example 1

By borrowing the framework from [1], we take $\mathcal{E} = [A|H, B|AH, AB|H]$ without any logical relation among A, B, H. Hence the sample space is composed by 8 atoms, 4 of them inside $H^0 \equiv H$. The set of coherent assessments $\mathcal{Q}_{\mathcal{E}}$ is made by the triples $[q_1, q_2, q_3] \in [0,1]^3$ with $q_3 = q_1 q_2$ (Fig.1). Note that the set $\mathcal{Q}_{\mathcal{E}}$ is evidently non-convex.

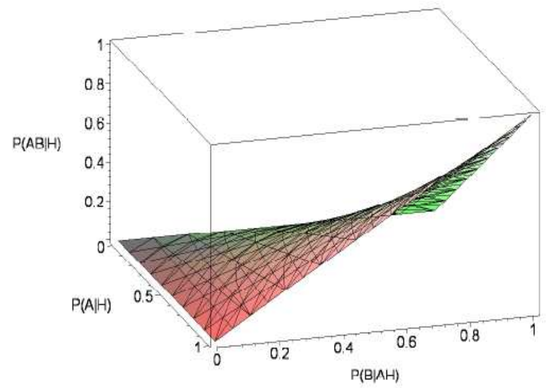

Figure 1: The set of coherent assessments $\mathcal{Q}_{\mathcal{E}}$

Starting from different initial incoherent assessments $\mathbf{p} = [p(A|H), p(B|AH), p(AB|H)]$ we got the coherent corrections $\mathbf{q_p}$ indicated in Table 1. Note that when the initial assessment is "weakly" inconsistent (first and third case) all the corrections are similar and meaningful. On the contrary, when $\mathbf{p}$ is "heavily" inconsistent (second case) only those based on Δ and L_2 are reasonable adjustments. The others are less meaningful or absolutely biased by the presence of local minima in the associated divergence. In particular those marked with * are affected by the choice of optimization starting points

Example 2

By borrowing a different framework from [6], we take three conditional events $\mathcal{E}_1 = [C|A, C|B, C|A \vee B]$ built by the three basic unconditional logically independent events A, B, C. Hence the whole sample space would again consist of 8 atoms, with 6 of them inside $H^0 \equiv A \vee B$. The set of coherent assessments $\mathcal{Q}_{\mathcal{E}_1}$ is made now by the triples $[q_1, q_2, q_3] \in (0,1)^3$ with the last component q_3 necessarily in the range $[\frac{q_1 q_2}{q_1+q_2-q_1 q_2}, \frac{q_1+q_2-2q_1 q_2}{1-q_1 q_2}]$ (the

Table 1: Coherent corrections $\mathbf{q_p}$ for different $\mathbf{p}$.

		$A \mid H$	$B \mid AH$	$AB \mid H$
	$\mathbf{p}$	.2	.3	.05
$\mathbf{q_p}$	Δ	.196	.28	.0549
	L_1	.2	.3	.06
	L_2	.197	.298	.059
	CD	.195	.277	.054
	R^*	.167	.25	.042
	O^*	.183	.274	.05
	$\mathbf{p}$	.8	.9	.02
$\mathbf{q_p}$	Δ	.419	.445	.186
	L_1	.799	.9	.719
	L_2	.545	.702	.383
	CD	.402	.212	.085
	R^*	.046	.028	.001
	O^*	.133	.15	.02
	$\mathbf{p}$	.8	.9	.7
$\mathbf{q_p}$	Δ	.793	.896	.711
	L_1	.794	.9	.714
	L_2	.793	.893	.708
	CD	.793	.891	.707
	R^*	.778	.875	.681
	O^*	.798	.9	.718

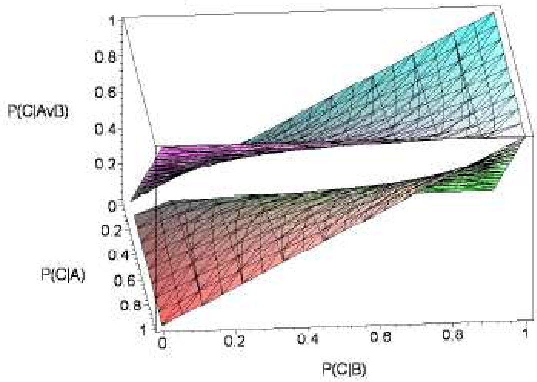

Figure 2: Lower and upper bounds of $\mathcal{Q}_{\mathcal{E}1}$

Table 2: Non-associativity of the procedure

$\mathcal{E}$	$C \mid A$	$C \mid B$	$C \mid A \vee B$	$D \mid B$	$E \mid B$
$\mathbf{p}_1$	.2	.3	.5		
$\mathbf{q}_{\mathbf{p}1}$	.226	.339	.446		
$\mathbf{p}_2$	.226	.339	.446	.7	
$\mathbf{q}_{\mathbf{p}2}$	.231	.323	.437	.677	
$\mathbf{p}'_2$	.2	.3	.5	.7	
$\mathbf{q}_{\mathbf{p}'2}$	.231	.323	.437	.677	
$\mathbf{p}_3$	.231	.321	.437	.677	.4
$\mathbf{q}_{\mathbf{p}3}$	.237	.350	.429	.650	.350
$\mathbf{p}'_3$	.2	.3	.5	.7	.4
$\mathbf{q}_{\mathbf{p}'3}$	.227	.344	.450	.656	.344

lower and upper bounds of $\mathcal{Q}_{\mathcal{E}1}$ are shown in Fig.2). Note also in this case the evident non-convexity of such coherent set. This example focuses on the iterative behavior of the procedure by adding events and assessments. More precisely, by considering two more conditional events $D|B$ and $E|B$, endowed with logical constraints $D \subset (A \wedge B) \vee (A \vee B \vee C)^c$ and $E \subset A^c \wedge B \wedge C$, the sample space Ω refines to 12 atoms. We can now see the difference between adding to the assessments one further evaluation per time, and making at once a full correction. Results are reported in Table 2 where $\mathbf{p}_2$ is obtained by extending $\mathbf{q}_{\mathbf{p}1}$ with a new assessment over $D|B$, while $\mathbf{p}'_2$ is obtained by extending $\mathbf{p}_1$ with the same assessment. Analogously, $\mathbf{p}_3$ is obtained by extending $\mathbf{q}_{\mathbf{p}2}$ with an assessment over $E|B$, while $\mathbf{p}'_3$ is the full assessment over the entire domain $\mathcal{E}_3 = [C|A, C|B, C|A \vee B, D|B, E|B]$ obtained by joining the various partial assessments. The coherent corrections $\mathbf{q}_{\mathbf{p}(\cdot)}$ of the corresponding incoherent assessments $\mathbf{p}_{(\cdot)}$ are reported under each input. While, at a first glance, equivalence between $\mathbf{q}_{\mathbf{p}2}$ and $\mathbf{q}_{\mathbf{p}'2}$ would suggest that the procedure is independent of the order, the differences between $\mathbf{q}_{\mathbf{p}3}$ and $\mathbf{q}_{\mathbf{p}'3}$ indicate the non-associativity of our correction process. This agrees with the similar behavior of the maximum entropy "fusion" operator (for further details refer to Kern-Isberner& Rodder[8]).

3.3 Selecting a specific coherent distribution

In each correction step of the previous examples the probability distribution $\boldsymbol{\alpha}$ compatible with the coherent corrections is unique, i.e. $\mathcal{A}^0_{\mathbf{q_p}} \equiv \{\boldsymbol{\alpha}\}$. Hence in these cases, once the initial assessment $\mathbf{p}$ has been corrected to $\mathbf{q_p}$, one can directly use $\boldsymbol{\alpha}$ for inference purposes, e.g. to extend the assessment to other relevant conditional events. But this is not the general situation. In fact it can happen that the set $\mathcal{A}^0_{\mathbf{q_p}}$ contains several distributions. In this case one can either continue to use the whole set $\mathcal{A}^0_{\mathbf{q_p}}$ by adopting imprecise probability techniques, or make a choice by "capturing" a specific distribution $\boldsymbol{\alpha}$, e.g. the one with maximum entropy $\boldsymbol{\alpha}_{ME}$ as prescribed by Kern-Isberner in [7] or by Kriz in [9]. In the latter case, our whole procedure would consist of two main steps: derive the coherent correction $\mathbf{q_p}$ by solving the nonlinear optimization problem (6), and then select inside the agreeing set of models $\mathcal{A}^0_{\mathbf{q_p}}$ the maximum entropy distribution $\boldsymbol{\alpha}_{ME}$, solving the nonlinear optimization program

$$\begin{array}{ll} \text{maximize} & -\sum_{\omega_j \in \Omega} \alpha_j \ln \alpha_j \\ \text{s.t.} & \boldsymbol{\alpha} \in \mathcal{A}^0_{\mathbf{q_p}}. \end{array} \quad (7)$$

Example 3

To describe the potential of our procedure in its two basic steps we have borrowed a medical application example from [4] and adapted it to our needs. There are 12 events involved, 4 unconditional and 8 conditional. We start with five basic events H_1, H_2, H_3, E, F, that have the following interpretations

- H_1 = cardiac insufficiency;
- H_2 = asthma attack;
- H_3 = asthma attack and a cardiac lesion;
- E = taking the medicine M for asthma does not reduce choking symptoms;
- F = taking the medicine M for asthma increases tachycardia,

and have the logical constraints

$$\begin{aligned} H_3 &\subseteq H_1H_2; \\ EH_1^cH_2 &\equiv \phi; \\ FH_1^cH_2 &\equiv \phi. \end{aligned}$$

We have given an incoherent assessment $\mathbf{p}$ over the events $\mathcal{E}$ listed in the first column of Table 3. The coherent corrections produced through the discrepancy index Δ and through the quadratic divergence L_2 are compared in the same Table 3. One can see some significant discrepancy. The correction obtained through Δ, apart from numerical considerations, has the advantage of having a probabilistic interpretation.

Table 3: The incoherent assessment $\mathbf{p}$ and its coherent corrections

		$\mathbf{q_p}$	
$\mathcal{E}$	$\mathbf{p}$	Δ	L_2
H_1	0.5	0.476	0.474
H_2	0.333	0.319	0.309
H_3	0.2	0.2	0.190
$H_1 \vee H_2$	0.6	0.595	0.593
$E\|H_1$	0.9	0.840	0.814
$E\|H_1^c$	0.7	0.571	0.601
$E\|H_2$	0.45	0.502	0.472
$E\|H_2^c$	0.75	0.791	0.805
$E\|H_3$	0.75	0.8	0.768
$E\|H_3^c$	0.6	0.674	0.686
$F\|EH_1$	0.875	0.855	0.852
$F\|EH_2$	0.6	0.637	0.610

We can describe now the second step of our procedure for the correction Δ. In fact the five basic events H_1, H_2, H_3, E, F span a sample space Ω with 17 atoms, as detailed in Table 4. By coherent extension of Δ to the sample space Ω we obtain a set $\mathcal{A}$ of probability distributions. Among all the admissible distributions $\boldsymbol{\alpha} \in \mathcal{A}$ we can select the one with maximum entropy by solving the optimization program (7). The solution $\boldsymbol{\alpha}_{ME}$ is reported in the last column of Table 4 and one can see that its components coincide with mid points of the coherent lower and upper bounds[1] $\underline{\boldsymbol{\alpha}}$ and $\overline{\boldsymbol{\alpha}}$.

4 Conclusions

We conclude by saying that this is a preliminary study about the adoption of the discrepancy Δ in revision problems. The numerical examples and the theoretical properties re-

[1] Computed by "Check Coherence Interface" software of the *PAID (PArtial Information and Decision)* Research Group `http://www.dipmat.unipg.it/~upkd/paid/software.html`

Table 4: The sample space, the optimal solution $\boldsymbol{\alpha}_I$, the lower $\underline{\boldsymbol{\alpha}}$ and upper $\overline{\boldsymbol{\alpha}}$ bounds and the maximum entropy solution $\boldsymbol{\alpha}_{ME}$ of Example 3

	$H_1H_2H_3EF$	$\boldsymbol{\alpha}_I$	$\underline{\boldsymbol{\alpha}}$	$\overline{\boldsymbol{\alpha}}$	$\boldsymbol{\alpha}_{ME}$
ω_1	1 1 1 0 0	.032	.000	.040	.020
ω_2	1 1 0 0 0	.000	.000	.000	.000
ω_3	0 1 0 0 0	.119	.119	.119	.119
ω_4	1 0 0 0 0	.018	.000	.036	.018
ω_5	0 0 0 0 0	.053	.000	.106	.053
ω_6	1 1 1 1 0	.058	.058	.058	.058
ω_7	1 1 0 1 0	.000	.000	.000	.000
ω_8	1 0 0 1 0	.000	.000	.000	.000
ω_9	0 0 0 1 0	.059	.000	.299	.150
ω_{10}	1 1 1 0 1	.008	.000	.040	.020
ω_{11}	1 1 0 0 1	.000	.000	.000	.000
ω_{12}	1 0 0 0 1	.018	.000	.036	.018
ω_{13}	0 0 0 0 1	.053	.000	.106	.053
ω_{14}	1 1 1 1 1	.102	.102	.102	.102
ω_{15}	1 1 0 1 1	.000	.000	.000	.000
ω_{16}	1 0 0 1 1	.240	.240	.240	.240
ω_{17}	0 0 0 1 1	.240	.000	.299	.150
	entropy	**2.108**			**2.174**

ported in this paper encourage us to complete the study.

In particular, at the moment we have restricted the conditional events H_i to have probabilities bounded away from zero. To remove this limitation we should profit from the cornerstone representation of coherent conditional partial assessments through different zero "layers" (see e.g. [5]). This method could also allow us to split the problem on different subproblems and to localize the optimization goals.

References

[1] V. Biazzo, A. Gilio. Some Theoretical Properties of Conditional Probability Assessments. In *LNAI (Lecture Notes in Computer Science)* 3571: 775–787, (2005).

[2] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Physics*, 7:200–217, (1967).

[3] A. Brooke , D. Kendrick , A. Meeraus and R. Raman. GAMS: a Users Guide (2003), Washington, D.C.: GAMS Development Corp.

[4] G. Coletti, R. Scozzafava. The role of coherence in eliciting and handling imprecise probabilities and its application to medical diagnosis. *Information Science*, 13:41–65 (2000).

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

[6] A. Gilio. Probabilistic Relations Among Logically Dependent Conditional Events, *Soft Computing*,3:154–161, (1999).

[7] G. Kern-Isberner, Characterizing the Principle of Minimum Cross-Entropy within a Conditional-Logical Framework, *Artificial Intelligence*, 98:169–208, (1998).

[8] G. Kern-Isberner, W. Rödder, Belief Revision and Information Fusion on Optimum Entropy, *Int. J. of Intelligent Systems*, 19:837–857 (2004).

[9] O. Kriz, Conditional problem for Objective Probability, *Kybernetica*, 34(1), 27–40 (1998)

[10] S. Kulback,*Information Theory and Statistics*, New York: John Wiley, (1957)

[11] F. Lad, *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, New York: John Wiley, (1996)

[12] R. Nau, V. R. Jose and R. Winkler: Scoring Rules, Entropy, and Imprecise Probabilities, in *Proc. 5th Int. Symp. on Imprecise Probability : Theorie and Applications. ISIPTA'07 - Prague (CZ)*, 307–315, (2007).