Coherent conditional probabilities and proper scoring rules

Angelo Gilio
Dipartimento di Scienze di Base e Applicate per l'Ingegneria,
University of Rome "La Sapienza" (Italy)
gilio@dmmm.uniroma1.it

Giuseppe Sanfilippo
Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli",
University of Palermo (Italy)
sanfilippo@unipa.it

Abstract

In this paper we study the relationship between the notion of coherence for conditional probability assessments on a family of conditional events and the notion of admissibility with respect to scoring rules. By extending a recent result given in literature for unconditional events, we prove, for any given strictly proper scoring rule s, the equivalence between the coherence of a conditional probability assessment and its admissibility with respect to s. In this paper we focus our analysis on the case of continuous bounded scoring rules. In this context a key role is also played by Bregman divergence and by a related theoretical aspect. Finally, we briefly illustrate a possible way of defining (generalized) coherence of interval-valued probability assessments by exploiting the notion of admissibility given for precise probability assessments.

Keywords. Conditional probability assessments, coherence, penalty criterion, proper scoring rules, conditional scoring rules, weak dominance, strong dominance, admissibility, Bregman divergence, g-coherence, total coherence, imprecise probability assessments.

1 Introduction

The theory and the applications of proper scoring rules have a long history in statistical literature (see, e.g., [1, 20, 23, 24, 25, 26, 28, 29, 31, 32, 33, 36, 37]). This theory was central to de Finetti's ideas about assessing the relative values of different subjective probability assessments ([9], see also [12]). A review of the general theory, with applications, has been given in [25] and, more recently, in [20]. A scoring rule for the probability of a given event E is a function of both the observation that comes to be observed, E true, or E false, and of the assessed probability $P(E)$. Assume that you were asked to assert $P(E)$, knowing that your assertion were to be scored according to the rule $s(E, P(E))$; moreover, assume that your degree of belief were $P(E) = p$, while you announced instead some other number $P(E) = x$, in the expectation that you would achieve a better score. The rule is said to be proper if you cannot expect a better score by specifying a value x different from p. Proper scoring rules encourage sincerity, because for you the best decision is to announce probabilities which conform to your beliefs.

The connections between the notions of coherence and of admissibility for probability assessments have been investigated in the work of de Finetti ([9, 10, 11]), by means of a penalty criterion based on the Brier quadratic scoring rule ([5]). A generalization of the work of de Finetti to a broad class of scoring rules has been given by Lindley in [26]. In his paper Lindley assumes suitable properties for the score function and admissibility for the numerical values which describe the uncertainty. Then, he shows that such numerical values can be transformed into numerical values which satisfy the basic properties of conditional probabilities.

The relationship between the notions of coherence and of non-dominance, with respect to strictly proper continuous scoring rules, has been investigated in [27]. In the same paper the connection of coherence and strictly proper scoring rules to Bregman divergence has been clarified.

A rich analysis of scoring rules which extends the results obtained in [27] to conditional probability assessments has been given in [33], where different notions of coherence have been discussed. In the same paper, some conditions are given under which the quadratic scoring rule can be replaced by a general strictly proper scoring rule, preserving the equivalence of the notions of coherence introduced through the gambling and the penalty arguments. In [33] are also examined the cases of scoring rules which are discontinuous and/or not strictly proper. In particular, in Example 8 of the same paper, by using a discontinuous strictly proper scoring rule it is shown that an incoherent probability assessment cannot be weakly

dominated by any coherent probability assessment, while it is dominated by other incoherent assessments. Moreover, in Example 9 of [33], by using a discontinuous merely proper scoring rule it is shown that a coherent probability assessment is weakly dominated by another coherent probability assessment.

In our paper we adopt a notion of coherence for conditional events which is different from that ones given in [33] and is based on the *strengthened* coherence principle of de Finetti ([11], vol. 2, Axiom 3, pag. 339). Such a strengthened principle allows to properly manage *conditioning* events with *zero probability* and, as proved in [14, 15] (see also [18]), is equivalent to the notion of coherence for conditional probability assessments studied by other authors; see e.g. [8, 22, 30, 34, 35]. In order to unify the treatment of unconditional and conditional events, the definition of coherence given by de Finetti with the penalty criterion was suitably modified in [15] (see also [16]).

As it can be shown by suitable examples (see [8, 17]), if a function P defined on a family of conditional events satisfies the axiomatic properties of a conditional probability, but the set of conditioning events doesn't have any structure, it may happen that P is not coherent. On the contrary, if P is coherent, then P satisfies all the properties of conditional probabilities. In particular, (strengthened) coherence requires that $0 \leq P(A|B) \leq 1$, for any given conditional event $A|B$. As another example, let us consider the assessment $P(A_1|B) = 0.9$, $P(A_2|B) = 0.7$, with $A_1 \wedge A_2 = \emptyset$ and $B \neq \Omega$ (see [33], p. 204). Such an assessment, which is coherent based on Definition 1 in [33], is not coherent in our approach.

We observe that the notions of coherence given in [33] and strengthened coherence are equivalent in the case of unconditional probabilities. Moreover, in Example 8 and Example 9 illustrated above only unconditional events are considered; hence, the corresponding results also hold in our approach. Then, in our paper we focus the analysis on continuous strictly proper scoring rules.

In this paper, using the strengthened notion of coherence, we extend the result given in [27] to the case of conditional events. We prove that, for any given (continuous) bounded strictly proper scoring rule s, a probability assessment on an arbitrary family of conditional events is coherent if and only if it is admissible with respect to s.

In ([33], p. 204) the authors leave open the question of whether their results still hold if one restricted the notion of coherence to require that the axioms of probability conditional on events with zero probability be satisfied. Our answer to this open question is that the equivalence between coherence and admissibility still holds with our notion of coherence (which restricts the notions of coherence used in [33]).

In our paper, based on the comments of an anonymous referee, we briefly examine how the notion of admissibility for precise probability assessments can be exploited in the case of interval-valued probability assessments.

The paper is organized as follows: In Section 2 we first give some preliminary notions; then, in Subsection 2.1 we recall the notion of coherence with the betting scheme; in Subsection 2.2 we give the notion of coherence with the penalty criterion of de Finetti; in Subsection 2.3 we illustrate, by a suitable alternative theorem, the equivalence of the betting scheme and the penalty criterion. In Section 3 we recall the notion of strictly proper scoring rule for unconditional events; then, we consider scoring rules for conditional events and we give the notions of weak and strong dominance, and of admissibility, for conditional probability assessments with respect to a scoring rule. We also consider a function $s(p, x)$ connected with the prevision of unconditional and conditional scoring rules. In Section 4 we illustrate some well known properties of $s(p, x)$. In Section 5 we recall the notion of Bregman divergence and a related theoretical aspect. Then, we prove for conditional probability assessments the equivalence between coherence and admissibility with respect to any continuous bounded strictly proper scoring rule. In Section 6 we recall the notions of g-coherence, coherence and total coherence for interval-valued probability assessments and we briefly examine how these notions can be defined by means of the admissibility property. Finally, in Section 7 we give some conclusions.

2 Some preliminary notions

Given a real function $P : \mathcal{K} \to \mathbb{R}$, where $\mathcal{K}$ is an arbitrary family of conditional events, let us consider a sub-family $\mathcal{F}_n = \{E_1|H_1, \ldots, E_n|H_n\} \subseteq \mathcal{K}$, and the vector $\mathcal{P}_n = (p_1, \ldots, p_n)$, where $p_i = P(E_i|H_i)\,, \;\; i = 1, \ldots, n$. The vector $\mathcal{P}_n$ represents the restriction of the function P to $\mathcal{F}_n$. We denote by $\mathcal{H}_n$ the disjunction $H_1 \vee \cdots \vee H_n$. Since

$$E_i H_i \vee E_i^c H_i \vee H_i^c = \Omega\,, \;\; i = 1, \ldots, n\,,$$

where Ω is the sure event, by expanding the expression $\bigwedge_{i=1}^n (E_i H_i \vee E_i^c H_i \vee H_i^c)$, we can represent Ω as the disjunction of 3^n logical conjunctions, some of which may be impossible. The remaining ones are the constituents generated by the family $\mathcal{F}$. We denote by $C_1, \ldots, C_m$ the constituents contained in $\mathcal{H}_n$ and (if $\mathcal{H}_n \neq \Omega$) by C_0 the further constituent $\mathcal{H}_n^c = H_1^c \cdots H_n^c$, so that

$$\mathcal{H}_n = C_1 \vee \cdots \vee C_m\,,$$
$$\Omega = \mathcal{H}_n^c \vee \mathcal{H}_n = C_0 \vee C_1 \vee \cdots \vee C_m\,, \quad m + 1 \leq 3^n\,.$$

2.1 Coherence with betting scheme

Using the same symbols for the events and their indicators, with the pair $(\mathcal{F}_n, \mathcal{P}_n)$ we associate the random gain

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

where $s_1, \ldots, s_n$ are n arbitrary real numbers. Let g_h be the value of $\mathcal{G}$ when C_h is true. Of course $g_0 = 0$ (notice that g_0 will not play any role in the definition of coherence). Denoting by $\mathcal{G}|\mathcal{H}_n$ the restriction of $\mathcal{G}$ to $\mathcal{H}_n$, it is $\mathcal{G}|\mathcal{H}_n \in \{g_1, \ldots, g_m\}$. Then, the function P defined on $\mathcal{K}$ is said *coherent* if and only if, for every integer n, for every finite sub-family $\mathcal{F}_n \subseteq \mathcal{K}$ and for every $s_1, \ldots, s_n$, one has

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

Remark 1. If the function P is coherent, then it is called a *conditional probability* on $\mathcal{K}$. Notice that, if P is coherent, then P satisfies all the well known properties of conditional probabilities (while the converse is not true; see [8], Example 13; or [17], Example 8).

2.2 Coherence with penalty criterion

Another operational definition of probabilities based on the quadratic scoring rule has been proposed by de Finetti ([10, 11]). This definition has been extended to the case of conditional events in [15].
With the pair $(\mathcal{F}_n, \mathcal{P}_n)$ we associate the loss $\mathcal{L} = \sum_{i=1}^{n} H_i (E_i - p_i)^2$; we denote by L_h the value of $\mathcal{L}$ if C_h is true. If You specify the assessment $\mathcal{P}_n$ on $\mathcal{F}_n$ as representing your belief's degrees, You are required to pay a penalty L_h when C_h is true. Then, the function P is said *coherent* if and only if do not exist an integer n, a finite sub-family $\mathcal{F}_n \subseteq \mathcal{K}$, and an assessment ${\mathcal{P}_n}^* = (p_1^*, \ldots, p_n^*)$ on $\mathcal{F}_n$ such that, for the loss $\mathcal{L}^* = \sum_{i=1}^{n} H_i (E_i - p_i^*)^2$, associated with $(\mathcal{F}_n, \mathcal{P}_n^*)$, it results $\mathcal{L}^* \leq \mathcal{L}$ and $\mathcal{L}^* \neq \mathcal{L}$; that is $L_h^* \leq L_h \, , \quad h = 1, \ldots, m \, ,$ with $L_h^* < L_h$ in at least one case.
We can develop a geometrical approach to coherence by associating, with each constituent C_h contained in $\mathcal{H}_n$, a point $Q_h = (q_{h1}, \ldots, q_{hn})$, where

$$q_{hj} = \begin{cases} 1, & \text{if } C_h \subseteq E_j H_j, \\ 0, & \text{if } C_h \subseteq E_j^c H_j, \\ p_j, & \text{if } C_h \subseteq H_j^c. \end{cases} \tag{2}$$

Denoting by $\mathcal{I}$ the convex hull of the points $Q_1, \ldots, Q_m$, based on the penalty criterion, the following result can be proved ([15], see also [17])

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

2.3 Equivalence between betting scheme and penalty criterion

The betting scheme and the penalty criterion are *equivalent* ([14, 15]). This equivalence can also be proved by the following steps ([18]):
1. The condition $\mathcal{P}_n \in \mathcal{I}$ amounts to solvability of the following system Σ in the unknowns $\lambda_1, \ldots, \lambda_m$

$$(\Sigma) \qquad \begin{cases} \sum_{h=1}^{m} q_{hj} \lambda_h = p_j \ , \ \ j = 1, \ldots, n \, ; \\ \sum_{h=1}^{m} \lambda_h = 1 \ , \ \ \lambda_h \geq 0 \, , \ h = 1, \ldots, m. \end{cases}$$

2. Let $\mathbf{x} = (x_1, \ldots, x_m)$, $\mathbf{y} = (y_1, \ldots, y_n)^t$ and $A = (a_{ij})$ be, respectively, a row $m-$vector, a column $n-$vector and a $m \times n-$matrix. The vector $\mathbf{x}$ is said *semipositive* if $x_i \geq 0, \forall \, i,$ and $x_1 + \cdots + x_m > 0$. Then, we have (cf. [13], Theorem 2.9)

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

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

3 Scoring rules and admissibility for conditional probability assessments

In this section we recall the notion of (strictly) proper scoring rule for unconditional events; then, based on this notion, we consider scoring rules for conditional events, called *conditional scoring rules*. Then, we illustrate the notions of weak and strong dominance, and of admissibility, for a probability assessment with respect to a scoring rule.
A score may represent a reward or a penalty; we think of scores as penalties, so that to improve the score means to reduce it. To introduce strictly proper scoring rules, we use the definition given in [27].

Definition 1. A function $s : \{0, 1\} \times [0, 1] \to [0, +\infty]$ is said to be a strictly proper scoring rule if the following conditions are satisfied:
(a) for every $x, p \in [0, 1]$, with $x \neq p$, it is

$$p \, s(1, x) + (1 - p) \, s(0, x) > p \, s(1, p) + (1 - p) \, s(0, p) \, ; \tag{3}$$

(b) the functions $s(1, x)$ and $s(0, x)$ are continuous.

We observe that, if x is your announced probability for the event E, while p represents your degree of

belief on E, then the quantity $ps(1,x)+(1-p)s(0,x)$ is nothing but your expected score.
For brevity, a *strictly proper scoring rule* will be called *proper scoring rule*.
We indicate by the same symbol the events and their indicators. Then, given any event E, we have

$$s(E,x)=\begin{cases} s(1,x), & E, \\ s(0,x), & E^c. \end{cases}$$

Given a scoring rule s, with any conditional event $E|H$ we associate the conditional scoring rule $s(E|H,x):\{0,1\}\times[0,1]\to[0,+\infty]$ defined as

$$s(E|H,x)=Hs(E,x)=\begin{cases} s(1,x), & EH, \\ s(0,x), & E^cH, \\ 0, & H^c. \end{cases}$$

We consider, for any given proper scoring rule s defined on the set $\{0,1\}\times[0,1]$, the extension of s to the set $[0,1]\times[0,1]$, defined as

$$s(p,x)=p\,s(1,x)+(1-p)\,s(0,x)\,. \tag{4}$$

We remark that, if x is your announced probability for the conditional event $E|H$, while p numerically represents your degree of belief on $E|H$, then the quantity $s(p,x)$ in (4) represents the *conditional prevision*

$$\mathbb{P}[s(E|H,x)\,|\,H]=\mathbb{P}[Hs(E,x)\,|\,H]=\mathbb{P}[s(E,x)\,|\,H]\,.$$

Moreover,

$$\mathbb{P}[s(E|H,x)]=s(1,x)P(EH)+s(0,x)P(E^cH)\,;$$

of course, $s(p,x)\neq s(1,x)P(EH)+s(0,x)P(E^cH)$.
Given a probability assessment $\mathcal{P}_n=(p_1,p_2,\ldots,p_n)$, with $p_i\in[0,1]$, on a family of conditional event $\mathcal{F}_n=\{E_1|H_1,E_2|H_2,\ldots,E_n|H_n\}$, where $p_i=P(E_i|H_i)$, and a proper scoring rule s, assuming that the scores are additive, we define the random penalty, or loss function, $\mathcal{L}$ associated with the pair $(\mathcal{F}_n,\mathcal{P}_n)$ as

$$\mathcal{L}=\sum_{i=1}^{n}s(E_i|H_i,p_i)=\sum_{i=1}^{n}H_is(E_i,p_i).$$

For the Brier quadratic scoring rule $s(E,x)=(E-x)^2$ it is $s(E|H,x)=H(E-x)^2$. The loss function associated with this conditional scoring rule was used in [15] (see also [18]), in the framework of the penalty criterion of de Finetti, to give a unified definition of the notion of coherence for conditional and unconditional events.
For the (unbounded and proper) logarithmic scoring rule ([21]) $s(E,x)=-\log(1-|E-x|)$, we have

$$s(E|H,x)=-H\log(1-|E-x|)\,.$$

The associated random penalty is

$$\mathcal{L}=-\sum_{i=1}^{n}[E_iH_i\log p_i+E_i^cH_i\log(1-p_i)]\,,$$

which was proposed in ([25], p. 355) for the case of unconditional events, with $\{E_1,\ldots,E_n\}$ a partition of Ω. The above random penalty was used in [7] to introduce a suitable discrepancy measure with the aim of correcting incoherent conditional probability assessments.
Given the constituents $C_0,C_1,\ldots,C_m$ generated by $\mathcal{F}_n$, we denote by L_k the value of $\mathcal{L}$ associated with C_k, $k=0,1,\ldots,m$. Of course, $L_0=0$.

Definition 2. Let be given a scoring rule s and a probability assessment $\mathcal{P}_n$ on a family of n conditional events $\mathcal{F}_n$. Given any assessment $\mathcal{P}_n^*$ on $\mathcal{F}_n$, with $\mathcal{P}_n^*\neq\mathcal{P}_n$, we say that $\mathcal{P}_n$ is *weakly dominated* by $\mathcal{P}_n^*$, with respect to s, if denoting by $\mathcal{L}$ (resp., $\mathcal{L}^*$) the penalty associated with the pair $(\mathcal{F}_n,\mathcal{P}_n)$ (resp., $(\mathcal{F}_n,\mathcal{P}_n^*)$), it is $\mathcal{L}^*\leq\mathcal{L}$, that is: $L_k^*\leq L_k$, for every $k=0,1,\ldots,m$.

We observe that $\mathcal{P}_n$ is not weakly dominated by $\mathcal{P}_n^*$ if and only if $L_k^*>L_k$ for at least a subscript k.

Definition 3. Let be given a scoring rule s and a probability assessment $\mathcal{P}_n$ on a family of n conditional events $\mathcal{F}_n$. We say that $\mathcal{P}_n$ is *admissible w.r.t.* s if $\mathcal{P}_n$ is not weakly dominated by any $\mathcal{P}_n^*\neq\mathcal{P}_n$.

Remark 2. We observe that, by Definition 3, it follows:
- If the assessment $\mathcal{P}_n$ on $\mathcal{F}_n$ is admissible, then for every subfamily $\mathcal{F}_J\subset\mathcal{F}_n$ the sub-assessment $\mathcal{P}_J$ associated with $\mathcal{F}_J$ is admissible.

In order to manage infinite families of conditional events we give the following

Definition 4. Let be given a scoring rule s and a probability assessment $\mathcal{P}$ on an arbitrary family of conditional events $\mathcal{K}$. We say that $\mathcal{P}$ is admissible with respect to s if, for every finite subfamily $\mathcal{F}_n\subseteq\mathcal{K}$, the restriction of $\mathcal{P}$ on $\mathcal{F}_n$ is admissible w.r.t. s.

By observing that $L_0=L_0^*=0$, we give the following

Definition 5. Let be given a scoring rule s and a probability assessment $\mathcal{P}_n$ on a family of n conditional events $\mathcal{F}_n$. Given any assessment $\mathcal{P}_n^*$ on $\mathcal{F}_n$, we say that $\mathcal{P}_n$ is *strongly dominated* by $\mathcal{P}_n^*$, with respect to s, if $L_k^*<L_k$, for every $k=1,\ldots,m$.

4 Properties of the function $s(p,x)$

For the convenience of the reader and to make our exposition self-contained, in the Proposition below we illustrate some well known properties of the function $s(p,x)$ defined in (4).

Proposition 1. Given a proper scoring rule s, the function $s(p,x)$ satisfies the following properties:

1. $s(\alpha p'+(1-\alpha)p'',x) = \alpha\, s(p',x)+(1-\alpha)\, s(p'',x)$;
2. $s(p,x) \geq s(p,p)$, with $s(p,x) = s(p,p)$ if and only if $x = p$;
3. $s(p,p)$ is strictly concave on $(0,1)$;
4. $s(p,x)$ is partially derivable with respect to x at (p,p), for every $p \in (0,1)$, and it is
$$\frac{\partial s(p,x)}{\partial x}|_{(p,p)} = 0\,;$$
5. for every $p \in (0,1)$, $s(p,p)$ is differentiable, with a continuous decreasing derivative
$$s'(p,p) = a(p) = s(1,p) - s(0,p)\,;$$
6. for every $p \in [0,1]$, $x \in (0,1)$, it is
$$s(p,x) = s(x,x) + s'(x,x)(p-x)\,.$$

Proof. 1. We have $s(p,x) = a(x)p + b(x)$, where
$$a(x) = s(1,x) - s(0,x)\,, \quad b(x) = s(0,x)\,,$$
so that
$$\begin{aligned}&s(\alpha p' + (1-\alpha)p'',x) = \\ &a(x)[\alpha p' + (1-\alpha)p''] + b(x)[\alpha + (1-\alpha)] = \\ &\alpha\, s(p',x) + (1-\alpha)\, s(p'',x).\end{aligned}$$

2. The property immediately follows by observing that the restriction of the function $s(p,x)$ to the set $\{0,1\} \times [0,1]$ is a proper scoring rule.
3. For every $x, y, \alpha \in (0,1)$, by setting $z = \alpha x + (1-\alpha)y$, we have $s(x,x) < s(x,z)$, $s(y,y) < s(y,z)$; then
$$\begin{aligned}&s(z,z) = s(\alpha x + (1-\alpha)y, \alpha x + (1-\alpha)y) = \\ &\alpha\, s(x,z) + (1-\alpha)\, s(y,z) > \alpha\, s(x,x) + (1-\alpha)\, s(y,y)\end{aligned}$$

4. Given any $p \in (0,1)$ and $0 < \varepsilon < 1-p$, by property 2 we have
$$\begin{aligned}&\frac{s(p,p+\varepsilon) - s(p,p)}{\varepsilon} > 0\,, \\ &\frac{s(p+\varepsilon,p+\varepsilon) - s(p+\varepsilon,p)}{\varepsilon} < 0\,.\end{aligned} \tag{5}$$

Moreover
$$\begin{aligned}&\tfrac{s(p+\varepsilon,p+\varepsilon)-s(p+\varepsilon,p)}{\varepsilon} = \\ &= \tfrac{s(p+\varepsilon,p+\varepsilon)-s(p,p+\varepsilon)}{\varepsilon} - \tfrac{s(p+\varepsilon,p)-s(p,p)}{\varepsilon} + \\ &+ \tfrac{s(p,p+\varepsilon)-s(p,p)}{\varepsilon} = \\ &= \tfrac{\varepsilon[s(1,p+\varepsilon)-s(0,p+\varepsilon)]}{\varepsilon} - \tfrac{\varepsilon[s(1,p)-s(0,p)]}{\varepsilon} + \\ &+ \tfrac{s(p,p+\varepsilon)-s(p,p)}{\varepsilon} = \\ &= \tfrac{\varepsilon[s(1,p+\varepsilon)-s(0,p+\varepsilon)]}{\varepsilon} - \tfrac{\varepsilon[s(1,p)-s(0,p)]}{\varepsilon} + \\ &+ \tfrac{s(p,p+\varepsilon)-s(p,p)}{\varepsilon} = \\ &= [s(1,p+\varepsilon)-s(0,p+\varepsilon)] - [s(1,p)-s(0,p)] + \tfrac{s(p,p+\varepsilon)-s(p,p)}{\varepsilon}\,.\end{aligned}$$

Then, by (5), it follows
$$\begin{aligned}&0 < \frac{s(p,p+\varepsilon) - s(p,p)}{\varepsilon} < \\ &< [s(1,p) - s(0,p)] - [s(1,p+\varepsilon) - s(0,p+\varepsilon)]\,,\end{aligned} \tag{6}$$
and by continuity of the function $s(1,x) - s(0,x)$ it follows
$$\lim_{\varepsilon\to 0^+} \frac{s(p,p+\varepsilon) - s(p,p)}{\varepsilon} = 0\,.$$
Analogously, given any $p \in (0,1)$ and $0 < \varepsilon < p$, by property 2 we have
$$\begin{aligned}&\frac{s(p,p-\varepsilon) - s(p,p)}{\varepsilon} > 0\,, \\ &\frac{s(p-\varepsilon,p-\varepsilon) - s(p-\varepsilon,p)}{\varepsilon} < 0\,.\end{aligned} \tag{7}$$

Moreover
$$\begin{aligned}&\tfrac{s(p-\varepsilon,p-\varepsilon)-s(p-\varepsilon,p)}{\varepsilon} = \\ &= \tfrac{s(p-\varepsilon,p-\varepsilon)-s(p,p-\varepsilon)}{\varepsilon} - \tfrac{s(p-\varepsilon,p)-s(p,p)}{\varepsilon} + \\ &+ \tfrac{s(p,p-\varepsilon)-s(p,p)}{\varepsilon} = \\ &= \tfrac{-\varepsilon[s(1,p-\varepsilon)-s(0,p-\varepsilon)]}{\varepsilon} - \tfrac{-\varepsilon[s(1,p)-s(0,p)]}{\varepsilon} + \\ &+ \tfrac{s(p,p-\varepsilon)-s(p,p)}{\varepsilon} = \\ &= -[s(1,p-\varepsilon)-s(0,p-\varepsilon)] + [s(1,p)-s(0,p)] + \tfrac{s(p,p-\varepsilon)-s(p,p)}{\varepsilon}\,.\end{aligned}$$

Then, by (7), it follows
$$\begin{aligned}&0 < \frac{s(p,p-\varepsilon) - s(p,p)}{\varepsilon} < \\ &< [s(1,p-\varepsilon) - s(0,p-\varepsilon)] - [s(1,p) - s(0,p)]\,,\end{aligned}$$
and by continuity of the function $s(1,x) - s(0,x)$ it follows
$$\begin{aligned}&\lim_{\varepsilon\to 0^+} \frac{s(p,p-\varepsilon) - s(p,p)}{-\varepsilon} = \\ &= -\lim_{\varepsilon\to 0^+} \frac{s(p,p-\varepsilon) - s(p,p)}{\varepsilon} = 0\,.\end{aligned}$$

Therefore, for every $p \in (0,1)$, there exists the partial derivative of $s(p,x)$ with respect to x at (p,p) and it is zero.
5. Given any $p \in (0,1)$ and $-p < \varepsilon < 1-p$, $\varepsilon \neq 0$, we have

$$\begin{aligned}
&\tfrac{s(p+\varepsilon,p+\varepsilon)-s(p,p)}{\varepsilon} = \\
&= \tfrac{s(p+\varepsilon,p+\varepsilon)-s(p,p+\varepsilon)}{\varepsilon} + \tfrac{s(p,p+\varepsilon)-s(p,p)}{\varepsilon} = \\
&= \tfrac{\varepsilon[s(1,p+\varepsilon)-s(0,p+\varepsilon)]}{\varepsilon} + \tfrac{s(p,p+\varepsilon)-s(p,p)}{\varepsilon} = \\
&= [s(1,p+\varepsilon)-s(0,p+\varepsilon)] + \tfrac{s(p,p+\varepsilon)-s(p,p)}{\varepsilon}\,;
\end{aligned}$$

then, by continuity of the function $s(1,x)-s(0,x)$ and by property 4, it follows

$$\begin{aligned}
&s'(p,p) = \lim_{\varepsilon\to 0} \frac{s(p+\varepsilon,p+\varepsilon)-s(p,p)}{\varepsilon} = \\
&= a(p) = s(1,p) - s(0,p)\,.
\end{aligned}$$

We observe that, in agreement with the strict concavity of $s(p,p)$ and as shown in (6), $a(p)$ is decreasing.
6. For every $p \in [0,1]$, $x \in (0,1)$, we have

$$\begin{aligned}
&s(p,x) - s(x,x) = [a(x)p + b(x)] - [a(x)x + b(x)] = \\
&= s'(x,x)(p-x)\,;
\end{aligned}$$

hence $s(p,x) = s(x,x) + s'(x,x)(p-x)$. □

5 Coherence and admissibility

In this section we recall the notion of Bregman divergence and a related theoretical aspect. Then, we prove the main result of the paper, by showing the equivalence between the coherence of conditional probability assessments and admissibility with respect to any bounded (strictly) proper scoring rule s.
Given two vectors

$$V_n = (v_1,\ldots,v_n),\ \ \mathcal{P}_n = (p_1,\ldots,p_n) \in [0,1]^n,$$

we set

$$S(V_n,\mathcal{P}_n) = \sum_{i=1}^{n} s(v_i,p_i)\,. \tag{8}$$

By property 3, the function S is strictly concave; moreover, by property 5, S is differentiable in $(0,1)^n$. By property 6, given any $\mathcal{P}_n \in (0,1)^n$ we have

$$\begin{aligned}
&S(V_n,\mathcal{P}_n) = \textstyle\sum_{i=1}^{n}[s(p_i,p_i) + s'(p_i,p_i)(v_i-p_i)] = \\
&= S(\mathcal{P}_n,\mathcal{P}_n) + \nabla S(\mathcal{P}_n,\mathcal{P}_n)\cdot(V_n-\mathcal{P}_n)\,;
\end{aligned} \tag{9}$$

then, by setting

$$\Phi(\mathcal{P}_n) = -S(\mathcal{P}_n,\mathcal{P}_n)\,,$$

we have

$$S(V_n,\mathcal{P}_n) = -\Phi(\mathcal{P}_n) - \nabla\Phi(\mathcal{P}_n)\cdot(V_n-\mathcal{P}_n)\,. \tag{10}$$

We recall that the function $s(p,p)$ is continuous on $[0,1]$ and strictly concave on $(0,1)$; then $\Phi(\mathcal{P}_n)$ is continuous on $[0,1]^n$ and strictly convex on $(0,1)^n$. Moreover, $s(p,p)$ has a continuous first derivative on $(0,1)$; then, the function $\Phi(\mathcal{P}_n)$ has continuous partial derivatives on $(0,1)^n$. Hence, $\Phi(\mathcal{P}_n)$ is differentiable on $(0,1)^n$ and its gradient $\nabla\Phi(\mathcal{P}_n)$ is a continuous function on $(0,1)^n$. If s is bounded, then $\nabla\Phi(\mathcal{P}_n)$ extends to a bounded continuous function on $[0,1]^n$.
In the definition below we recall the notion of Bregman divergence (see e.g. [6]).

Definition 6. Let $\mathcal{C}$ be a convex subset of $\mathbb{R}^n$ with nonempty interior. Let $\Phi : \mathcal{C} \to \mathbb{R}$ be a strictly convex function, differentiable in the interior of $\mathcal{C}$, whose gradient $\nabla\Phi$ extends to a bounded, continuous function on $\mathcal{C}$. For $V_n, \mathcal{P}_n \in \mathcal{C}$ the **Bregman divergence** $d_\Phi : \mathcal{C}\times\mathcal{C} \to \mathbb{R}$ corresponding to Φ is given by

$$d_\Phi(V_n,\mathcal{P}_n) = \Phi(V_n) - \Phi(\mathcal{P}_n) - \nabla\Phi(\mathcal{P}_n)\cdot(V_n-\mathcal{P}_n)\,.$$

It is $d_\Phi(V_n,\mathcal{P}_n) \geq 0$ and, as Φ is strictly convex, $d_\Phi(V_n,\mathcal{P}_n) = 0$ if and only if $V_n = \mathcal{P}_n$.
We remark that, assuming s bounded, $\mathcal{C} = [0,1]^n$ and $\Phi(\mathcal{X}) = -S(\mathcal{X},\mathcal{X})$, by (10) and Definition 6 it follows

$$d_\Phi(V_n,\mathcal{P}_n) = S(V_n,\mathcal{P}_n) - S(V_n,V_n)\,. \tag{11}$$

We observe that, for $s(E,x) = -\log(1-|E-x|)$, we have

$$S(V_n,\mathcal{P}_n) = -\sum_{i=1}^{n}[v_i\log p_i + (1-v_i)\log(1-p_i)]\,; \tag{12}$$

then, formula (11) becomes

$$d_\Phi(V_n,\mathcal{P}_n) = \sum_{i=1}^{n}\left[v_i\log\left(\frac{v_i}{p_i}\right) + (1-v_i)\log\left(\frac{1-v_i}{1-p_i}\right)\right].$$

This logarithmic Bregman divergence is connected with the discrepancy measure proposed in [7] to correct incoherent conditional probability assessments.
Now, we recall the following result given in [27]; see also [6].

Proposition 2. Let $d_\Phi : \mathcal{C}\times\mathcal{C} \to \mathbb{R}$ be a Bregman divergence and let $\mathcal{I} \subseteq \mathcal{C}$ be a closed convex subset of $\mathbb{R}^n$. For each $\mathcal{P}_n \in \mathcal{C}\setminus\mathcal{I}$, there exists a unique $\mathcal{P}_n^* \in \mathcal{I}$, called the **projection** of $\mathcal{P}_n$ onto $\mathcal{I}$, such that

$$d_\Phi(\mathcal{P}_n^*,\mathcal{P}_n)\ \leq\ d_\Phi(V_n,\mathcal{P}_n)\,,\ \forall\, V_n \in \mathcal{I}\,.$$

Moreover

$$\begin{aligned}
&d_\Phi(V_n,\mathcal{P}_n^*) + d_\Phi(\mathcal{P}_n^*,\mathcal{P}_n)\ \leq\ d_\Phi(V_n,\mathcal{P}_n)\,, \\
&\forall\, V_n \in \mathcal{I}\,,\ \mathcal{P}_n \in \mathcal{C}\setminus\mathcal{I}\,.
\end{aligned} \tag{13}$$

In the next result we illustrate the relationship between the notion of coherence and the property of non dominance.

Theorem 3. Let be given a probability assessment $\mathcal{P}$ on a family of conditional events $\mathcal{K}$; moreover, let be given any bounded (strictly) proper scoring rule s. The assessment $\mathcal{P}$ is coherent if and only if it is admissible with respect to s.

Proof. ($\Rightarrow$) Assuming $\mathcal{P}$ coherent, let s be any bounded proper scoring rule. Given any subfamily $\mathcal{F}_n = \{E_1|H_1, \ldots, E_n|H_n\}$ of $\mathcal{K}$, let $\mathcal{P}_n = (p_1, \ldots, p_n)$ be the restriction to $\mathcal{F}_n$ of $\mathcal{P}$. Now, given any $\mathcal{P}_n^* = (p_1^*, \ldots, p_n^*) \neq \mathcal{P}_n$, we distinguish two cases:
(a) $p_i^* \neq p_i$, for every $i = 1, \ldots, n$;
(b) $p_i^* = p_i$, for at least one index i.
Case **(a)**. We still denote by $C_0, C_1, \ldots, C_m$, where $C_0 = H_1^c \wedge \cdots \wedge H_n^c$, the constituents generated by $\mathcal{F}_n$ and by $Q_k = (q_{k1}, \ldots, q_{kn})$ the point associated with $C_k, k = 1, \ldots, m$.

We introduce the following binary quantities

$$e_{ki} = \begin{cases} 1, & C_k \subseteq E_i, \\ 0, & C_k \subseteq E_i^c, \end{cases} \quad , \quad h_{ki} = \begin{cases} 1, & C_k \subseteq H_i, \\ 0, & C_k \subseteq H_i^c. \end{cases}$$

Then, by recalling (2), for every $i = 1, \ldots, n$, $k = 1, \ldots, m$ it is

$$q_{ki} = e_{ki} h_{ki} + (1 - h_{ki}) p_i \,. \tag{14}$$

With the assessment $\mathcal{P}_n$ it is associated the loss

$$\mathcal{L} = \sum_{i=1}^{n} [E_i H_i s(1, p_i) + E_i^c H_i s(0, p_i)] = \sum_{i=1}^{n} H_i s(E_i, p_i) \,;$$

of course, with any other assessment $\mathcal{P}_n^*$ on $\mathcal{F}_n$ it associated the loss

$$\mathcal{L}^* = \sum_{i=1}^{n} H_i [E_i s(1, p_i^*) + E_i^c s(0, p_i^*)] = \sum_{i=1}^{n} H_i s(E_i, p_i^*) \,.$$

For each constituent $C_k, k = 0, 1, \ldots, m$, the values of $\mathcal{L}$ and $\mathcal{L}^*$ are, respectively

$$L_k = \sum_{i=1}^{n} [e_{ki} h_{ki} s(1, p_i) + (1 - e_{ki}) h_{ki} s(0, p_i)] \,,$$

$$L_k^* = \sum_{i=1}^{n} [e_{ki} h_{ki} s(1, p_i^*) + (1 - e_{ki}) h_{ki} s(0, p_i^*)] \,.$$

By recalling that $L_0 = L_0^* = 0$, in what follows we will only refer to the values L_k, L_k^*, $k = 1, \ldots, m$.
As $\mathcal{P}_n$ is coherent, there exists a vector $(\lambda_1, \ldots, \lambda_m)$, with $\lambda_k \geq 0$ and $\sum_k \lambda_k = 1$, such that $\mathcal{P}_n = \sum_k \lambda_k Q_k$; that is, by (14)

$$p_i = \textstyle\sum_k \lambda_k q_{ki} = \sum_k \lambda_k e_{ki} h_{ki} + p_i - p_i \sum_k \lambda_k h_{ki} \,,$$

for every $i = 1, \ldots, n$; so that

$$\sum_k \lambda_k e_{ki} h_{ki} = p_i \sum_k \lambda_k h_{ki} \,, \; i = 1, \ldots, n \,,$$

or equivalently

$$\sum_k \lambda_k (1 - e_{ki}) h_{ki} = (1 - p_i) \sum_k \lambda_k h_{ki} \,, \; i = 1, \ldots, n \,.$$

Then,

$$\begin{aligned} &\textstyle\sum_k \lambda_k L_k = \\ &= \textstyle\sum_k \lambda_k \sum_{i=1}^{n} [e_{ki} h_{ki} s(1, p_i) + (1 - e_{ki}) h_{ki} s(0, p_i)] = \\ &\qquad = \textstyle\sum_i \left(\sum_k \lambda_k e_{ki} h_{ki}\right) s(1, p_i) + \\ &\qquad + \textstyle\sum_i \left(\sum_k \lambda_k (1 - e_{ki}) h_{ki}\right) s(0, p_i) = \\ &= \textstyle\sum_i [p_i \sum_k \lambda_k h_{ki}) s(1, p_i) + (1 - p_i) \sum_k \lambda_k h_{ki}) s(0, p_i)] \\ &\quad = \textstyle\sum_i (\sum_k \lambda_k h_{ki}) \left[p_i \, s(1, p_i) + (1 - p_i) \, s(0, p_i)\right] \,. \end{aligned}$$

We set $I' = \{i : \sum_k \lambda_k h_{ki} > 0\} \subseteq \{1, 2, \ldots, n\}$. We observe that I' is not empty. In fact, for each $i = 1, \ldots, n$, there exists a constituent C_k such that $C_k \subseteq H_i$ and then $\sum_i h_{ki} \geq 1$. Moreover, as

$$\sum_i \sum_k \lambda_k h_{ki} = \sum_k \lambda_k \sum_i h_{ki} \geq \sum_k \lambda_k = 1 \,,$$

there exists an index i such that $\sum_k \lambda_k h_{ki} > 0$; i.e. $I' \neq \emptyset$.

Then, by recalling that for each $i = 1, \ldots, n$ it is

$$p_i \, s(1, p_i) + (1 - p_i) s(0, p_i) < p_i \, s(1, p_i^*) + (1 - p_i) s(0, p_i^*),$$

we have

$$\begin{aligned} &\textstyle\sum_k \lambda_k L_k = \\ &= \textstyle\sum_{i \in I'} (\sum_k \lambda_k h_{ki}) [p_i \, s(1, p_i) + (1 - p_i) \, s(0, p_i)] < \\ &< \textstyle\sum_{i \in I'} (\sum_k \lambda_k h_{ki}) [p_i \, s(1, p_i^*) + (1 - p_i) \, s(0, p_i^*)] = \\ &= \textstyle\sum_i (\sum_k \lambda_k h_{ki}) [p_i \, s(1, p_i^*) + (1 - p_i) \, s(0, p_i^*)] = \\ &= \textstyle\sum_k \lambda_k L_k^* \,. \end{aligned}$$

The inequality $\sum_k \lambda_k L_k < \sum_k \lambda_k L_k^*$ implies that there exists an index k such that $L_k < L_k^*$; that is $\mathcal{L}^* > \mathcal{L}$ in at least one case. Hence $\mathcal{P}_n$ is admissible. Since $\mathcal{F}_n$ is arbitrary, it follows that $\mathcal{P}$ is admissible.
Case **(b)**. Let be given any $\mathcal{P}_n^* \neq \mathcal{P}_n$, with $p_i^* = p_i$, for at least one index i. We set $J = \{i : p_i^* \neq p_i\} \subset J_n = \{1, \ldots, n\}$. We denote by $\mathcal{P}_J$ (resp., $\mathcal{P}_{J_n \setminus J}$) the subvector of $\mathcal{P}_n$ associated with J (resp., $J_n \setminus J$).

Analogously, we can consider the subvectors $\mathcal{P}^*_J$ and $\mathcal{P}^*_{J_n\setminus J}$ of $\mathcal{P}^*_n$. Then, we have

$$\mathcal{L} = \mathcal{L}_J + \mathcal{L}_{J_n\setminus J}\,,\ \ \mathcal{L}^* = \mathcal{L}^*_J + \mathcal{L}^*_{J_n\setminus J}\,,\ \ \mathcal{L}_{J_n\setminus J} = \mathcal{L}^*_{J_n\setminus J}\,.$$

By the same reasoning as in case (a), it holds that $\mathcal{L}^*_J > \mathcal{L}_J$ in at least one case. Then, by observing that $\mathcal{L} - \mathcal{L}^* = \mathcal{L}_J - \mathcal{L}^*_J$, it is $\mathcal{L}^* > \mathcal{L}$ in at least one case; hence $\mathcal{P}_n$ is admissible. Since $\mathcal{F}_n$ is arbitrary, $\mathcal{P}$ is admissible.

($\Leftarrow$). We will prove that, given any bounded proper scoring rule s, if $\mathcal{P}$ is not coherent, then $\mathcal{P}$ is not admissible with respect to s. Assume that $\mathcal{P}$ is not coherent. Then, there exists a subfamily $\mathcal{F}_n = \{E_1|H_1, \ldots, E_n|H_n\} \subseteq \mathcal{K}$ such that, for the restriction $\mathcal{P}_n = (p_1, \ldots, p_n)$ of $\mathcal{P}$ to $\mathcal{F}_n$, denoting by $\mathcal{I}_n \subseteq [0,1]^n$ the associated convex hull, it is $\mathcal{P}_n \notin \mathcal{I}_n$. For each constituent C_k we set $I_k = \{i : C_k \subseteq H_i^c\}$, $J_k = \{i : C_k \subseteq H_i\}$; then, by recalling (11), the value L_k of the penalty $\mathcal{L}$ is given by

$$\begin{aligned}
&L_k = \textstyle\sum_{i=1}^n s(e_{ki}, p_i) h_{ki} = \\
&= \textstyle\sum_{i=1}^n s(q_{ki}, p_i) - \sum_{i\in I_k} s(p_i, p_i) = \\
&= \textstyle\sum_{i=1}^n s(q_{ki}, p_i) - \sum_{i=1}^n s(q_{ki}, q_{ki}) + \\
&+ \textstyle\sum_{i=1}^n s(q_{ki}, q_{ki}) - \sum_{i\in I_k} s(p_i, p_i) = \\
&= \textstyle\sum_{i=1}^n s(q_{ki}, p_i) - \sum_{i=1}^n s(q_{ki}, q_{ki}) + \\
&+ \textstyle\sum_{i\in J_k}^n s(e_{ki}, e_{ki}) = \\
&= S(Q_k, \mathcal{P}_n) - S(Q_k, Q_k) + \alpha_k = \\
&= d_\Phi(Q_k, \mathcal{P}_n) + \alpha_k\,,
\end{aligned} \tag{15}$$

where $\alpha_k = \sum_{i\in J_k}^n s(e_{ki}, e_{ki})$ and $\Phi(\mathcal{X}) = -S(\mathcal{X}, \mathcal{X})$. By applying Proposition 2 with $\mathcal{C} = [0,1]^n$ and $\mathcal{I} = \mathcal{I}_n$, by (13) we have

$$d_\Phi(Q_k, \mathcal{P}^*_n) + d_\Phi(\mathcal{P}^*_n, \mathcal{P}_n) \le d_\Phi(Q_k, \mathcal{P}_n)\,,$$

where $\mathcal{P}^*_n = (p^*_1, \ldots, p^*_n)$ is the projection of $\mathcal{P}_n$ onto $\mathcal{I}_n$. Moreover, as $\mathcal{P}^*_n \neq \mathcal{P}_n$ it is $d_\Phi(\mathcal{P}^*_n, \mathcal{P}_n) > 0$ and hence

$$d_\Phi(Q_k, \mathcal{P}^*_n) < d_\Phi(Q_k, \mathcal{P}_n)\,,\ k = 1, \ldots, m\,.$$

Now, denoting by $Q^*_1 = (q^*_{11}, \ldots, q^*_{1n}), \ldots, Q^*_m = (q^*_{m1}, \ldots, q^*_{mn})$ the points associated with the pair $(\mathcal{F}_n, \mathcal{P}^*_n)$, recalling property 2, for each $k = 1, \ldots, m$ we have

$$\begin{aligned}
&d_\Phi(Q_k, \mathcal{P}^*_n) - d_\Phi(Q^*_k, \mathcal{P}^*_n) = \\
&= S(Q_k, \mathcal{P}^*_n) - S(Q_k, Q_k) - S(Q^*_k, \mathcal{P}^*_n) + S(Q^*_k, Q^*_k) = \\
&= \sum_{i=1}^n [s(q_{ki}, p^*_i) - s(q_{ki}, q_{ki}) - s(q^*_{ki}, p^*_i) + s(q^*_{ki}, q^*_{ki})] = \\
&= \textstyle\sum_{i=1}^n [s(q_{ki}, p^*_i) - s(q^*_{ki}, p^*_i)] + \\
&- \textstyle\sum_{i=1}^n [s(q_{ki}, q_{ki}) - s(q^*_{ki}, q^*_{ki})] = \\
&= \textstyle\sum_{i\in I_k} [s(p_i, p^*_i) - s(p^*_i, p^*_i)] + \\
&- \textstyle\sum_{i\in I_k} [s(p_i, p_i) - s(p^*_i, p^*_i)] = \\
&= \textstyle\sum_{i\in I_k} [s(p_i, p^*_i) - s(p_i, p_i)] \ge 0\,.
\end{aligned}$$

Therefore, for each $k = 1, \ldots, m$, it is

$$d_\Phi(Q^*_k, \mathcal{P}^*_n) \ \le\ d_\Phi(Q_k, \mathcal{P}^*_n) \ <\ d_\Phi(Q_k, \mathcal{P}_n)\,.$$

Then, by (15), for each $k = 1, \ldots, m$ it follows

$$L^*_k = d_\Phi(Q^*_k, P^*_n) + \alpha_k \ <\ d_\Phi(Q_k, P_n) + \alpha_k = L_k\,;$$

that is, $\mathcal{P}_n$ is strongly dominated (and hence weakly dominated) by $\mathcal{P}^*_n$; hence $\mathcal{P}_n$ is not admissible. This implies that $\mathcal{P}$ is not admissible. □

We remark that in the first part of the proof of the previous theorem it has not been necessary to use the Bregman divergence .

We observe that Theorem 3 can be formulated in the following equivalent way.

Theorem 4. Given an arbitrary family of conditional events $\mathcal{K}$, let Π_c the set of coherent conditional probability assessments $\mathcal{P}$ on $\mathcal{K}$. Moreover, denoting by Σ the class of bounded (continuous strictly) proper scoring rules, let be given any $s \in \Sigma$. Then, let Π_s be the set of conditional probability assessments $\mathcal{P}$ on $\mathcal{K}$ which are admissible with respect to s. We have

$$\Pi_s = \Pi_c\,,\ \forall\, s \in \Sigma\,. \tag{16}$$

Remark 3. The equality (16) in the case $s(E, x) = (E - x)^2$ has been proved in [15] (see also [18]).

6 The case of imprecise probability assessments

In this section we illustrate a possible way of studying the relationship between coherence and admissibility with respect to scoring rules in the case of interval-valued conditional probability assessments. An anonymous referee observed that "there is an *impossibility* result due to the authors of [33] (probably still unpublished) showing that there does not exist a real-valued proper IP-scoring rule".

Moreover, the referee claims that in the same paper it is shown that "there is a lexicographic, i.e. non-standard valued, proper scoring rule for eliciting probability intervals".

Here, we just show that the notion of admissibility given for precise assessments can also be exploited in the case of imprecise probabilities.

We recall below the notions of generalized coherence (g-coherence, [2, 3, 4]), coherence and total coherence ([19]) for interval-valued conditional probability assessments.

Definition 7. Let be given an interval-valued probability assessment $\mathcal{A}_n = ([l_i, u_i], i = 1, \ldots, n)$, defined on a family of n conditional events $\mathcal{F}_n = \{E_i|H_i, i = 1, \ldots, n\}$. We say that:

a) $\mathcal{A}_n$ is g-coherent if there exists a coherent precise probability assessment $\mathcal{P}_n = (p_i, i = 1, \ldots, n)$ on $\mathcal{F}_n$, with $p_i = P(E_i|H_i)$, which is consistent with $\mathcal{A}_n$, that is such that $l_i \leq p_i \leq u_i$ for each $i = 1, \ldots, n$;
b) $\mathcal{A}_n$ is coherent if, given any $j \in \{1, \ldots, n\}$ and any $x_j \in [l_j, u_j]$, there exists a coherent precise probability assessment $\mathcal{P}_n = (p_i, i = 1, \ldots, n)$ on $\mathcal{F}_n$, which is consistent with $\mathcal{A}_n$ and is such that $p_j = x_j$;
c) $\mathcal{A}_n$ is totally coherent if every precise probability assessment $\mathcal{P}_n = (p_i, i = 1, \ldots, n)$ on $\mathcal{F}_n$, consistent with $\mathcal{A}_n$, is coherent.

We observe that the notions of g-coherence and coherence above amount to the well known notions of *avoiding uniform loss* and *coherence*, respectively, used in the literature on imprecise probabilities (see, e.g., [34]). Based on Definition 7 we can give the following versions of our main result in the case of interval-valued probability assessments.

Proposition 3. Let be given an interval-valued probability assessment $\mathcal{A}_n = ([l_i, u_i], i = 1, \ldots, n)$, defined on $\mathcal{F}_n = \{E_i|H_i, i = 1, \ldots, n\}$. Moreover, let be given any bounded (continuous and strictly) proper scoring rule s. We have:
a) $\mathcal{A}_n$ is g-coherent if and only if there exists a precise probability assessment $\mathcal{P}_n = (p_i, i = 1, \ldots, n)$ on $\mathcal{F}_n$, consistent with $\mathcal{A}_n$, which is admissible w.r.t. s;
b) $\mathcal{A}_n$ is coherent if, given any $j \in \{1, \ldots, n\}$ and any $x_j \in [l_j, u_j]$, there exists a precise probability assessment $\mathcal{P}_n = (p_i, i = 1, \ldots, n)$ on $\mathcal{F}_n$, with $p_j = x_j$, consistent with $\mathcal{A}_n$, which is admissible w.r.t. s;
c) $\mathcal{A}_n$ is totally coherent if every precise probability assessment $\mathcal{P}_n = (p_i, i = 1, \ldots, n)$ on $\mathcal{F}_n$, consistent with $\mathcal{A}_n$, is admissible w.r.t. s.

7 Conclusions

In this paper we have studied the relationship between the notion of (strengthened) coherence for conditional probability assessments and the property of admissibility with respect to scoring rules. We have extended to the case of conditional events a result given in [27] for unconditional events. We have shown that, given any bounded (continuous and strictly) proper scoring rule s, a conditional probability assessment on an arbitrary family of conditional events is coherent if and only if it is admissible with respect to s. To obtain our main result a key role has also been played by Bregman divergence. Finally, we have shown that the property of admissibility can be exploited to characterize the notions of g-coherence, coherence and total coherence for interval-valued conditional probability assessments.

Acknowledgments

The authors thank the anonymous referees for their very useful comments and suggestions.

References

[1] G. Agró, F. Lad, G. Sanfilippo. Sequentially forecasting economic indices using mixture linear combinations of EP distributions. *Journal of Data Science*, 8(1):101–126, 2010.

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

[3] V. Biazzo, A. Gilio, G. Sanfilippo. Coherence Checking and Propagation of Lower Probability Bounds, *Soft Computing* 7, 310-320, 2003.

[4] V. Biazzo, A. Gilio, G. Sanfilippo. Generalized coherence and connection property of imprecise conditional previsions. In: *Proceedings of 12th Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference*, IPMU 2008, Malaga, Spain, June 22 - 27, 907–914, 2008.

[5] G. W. Brier. Verification of forecasts expressed in terms of probability, *Monthly Weather Review*, 78:1–3, 1950.

[6] Y. Censor, S.A. Zenios. *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford Univ. Press, Oxford, 1997.

[7] A. Capotorti, G. Regoli, F. Vattari. Correction of incoherent conditional probability assessments. *International Journal of Approximate Reasoning*, 51:718-727, 2010.

[8] G. Coletti, R. Scozzafava. *Probabilistic logic in a coherent setting*. Kluwer, Dordrecht, 2002.

[9] B. de Finetti. Does it make sense to speak of 'good probability appraisers'? In: *The scientist speculates: an anthology of of partly-baked ideas*, I. J. Good (ed.), Heinemann, London, 357–364, 1962.

[10] B. de Finetti. Probabilità composte e teoria delle decisioni. *Rendiconti di Matematica*, 23:128-134, 1964.

[11] B. de Finetti. *Teoria delle probabilità*, voll. 1-2. Einaudi, Torino, 1970, (Engl. transl.: Theory of Probability, voll. 1-2, Wiley, Chichester, 1974, 1975).

[12] B. de Finetti. The Role of 'Dutch Books' and of 'Proper Scoring Rules'. *The British Journal for the Philosophy of Science*, 32(1):55-56, 1981.

[13] D. Gale. *The theory of linear economic models.* McGraw-Hill, New York, 1960.

[14] A. Gilio. Probabilità condizionate $\mathcal{C}_0$-coerenti. *Rendiconti di Matematica*, Serie VII, 9:277–295, 1989.

[15] A. Gilio. Criterio di penalizzazione e condizioni di coerenza nella valutazione soggettiva della probabilità. *Boll. Un. Mat. Ital.*, [7a] 4-B(3): 645–660, 1990.

[16] A. Gilio. C_0-Coherence and Extension of Conditional Probabilities. In: *Bayesian Statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (eds.), Oxford University Press, 633-640, 1992.

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

[18] A. Gilio. Algorithms for conditional probability assessments. In: *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, D. A. Berry, K. M. Chaloner, J. K. Geweke (eds.), John Wiley, 29–39, 1996.

[19] A. Gilio, S. Ingrassia. Totally coherent set-valued probability assessments. *Kybernetika* 34(1): 3–15, 1998.

[20] T. Gneiting, A. Raftery. Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[21] I. J. Good. Rational decisions. *J. Roy. Statist. Soc.*, 14:107-114, 1952.

[22] S. Holzer. On coherence and conditional prevision. *Boll. Un. Mat. Ital.*, 4(6):441–460, 1985.

[23] A. D. Hendrickson, R. J. Buehler. Proper scores for probability forecasters. *Ann. Math. Statist.* 42:1916-1921, 1971.

[24] V. R. R. Jose, R. F. Nau, R. L. Winkler. Scoring Rules, Generalized Entropy, and Utility Maximization, *Operations Research*, 56(5):1146-1157, 2008.

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

[26] D. V. Lindley. Scoring rules and the inevitability of probability. *Int. Statist. Rev.*, 50:1-11, 1982.

[27] J. B. Predd, R. Seiringer, E. H. Lieb, D. N. Osherson, H. V. Poor, S. R. Kulkarni. Probabilistic Coherence and Proper Scoring Rules. *IEEE T. Inform. Theory*, 55:4786–4792, 2009.

[28] J. McCarthy. Measures of the value of information. *Proceedings of the National Academy of Sciences*, 42:654–655, 1956.

[29] R. F. Nau, V. R. R. Jose, R. L. Winkler. Scoring Rules, Entropy, and Imprecise Probabilities. In: *Proceedings of the 5th International Symposium on Imprecise Probabilities and their Applications, ISIPTA07*, Prague, Czech Republic, 307–316, 2007.

[30] E. Regazzini. Finitely additive conditional probabilities. *Rend. Sem. Mat. Fis. Milano*, 55:69–89, 1985.

[31] G. Sanfilippo, G. Agró, F. Lad. Assessing fat-tailed sequential forecast distributions for the Dow-Jones index with logarithmic scoring rules. In: *Proc. of 56th Session of the International Statistical Institute*, Lisbon, 22-29, August, 2007.

[32] L. J. Savage. Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.*, 66:783-801, 1971

[33] M. J. Schervish, T. Seidenfeld, J. B. Kadane. Proper Scoring Rules, Dominated Forecasts, and Coherence. *Decision Analysis*, 6(4):202–221, 2009.

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

[35] P. M. Williams. Notes on conditional previsions, Technical report, University of Sussex, 1975. Reprinted in a revised form in: *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.

[36] R. L. Winkler. Scoring rules and the evaluation of probability assessors. *J. Amer. Statist. Assoc.*, 64:1073-1078, 1969.

[37] R. L. Winkler. Scoring rules and the evaluation of probabilities. *Test* 5(1):1-60, 1996.