Coherent conditional previsions and proper scoring rules

Veronica Biazzo[1], Angelo Gilio[2], and Giuseppe Sanfilippo[3]

[1] University of Catania, Italy
vbiazzo@dmi.unict.it
[2] University of Rome "La Sapienza", Italy
angelo.gilio@sbai.uniroma1.it
[3] University of Palermo, Italy
giuseppe.sanfilippo@unipa.it

Abstract. In this paper we study the relationship between the notion of coherence for conditional prevision assessments on a family of finite conditional random quantities and the notion of admissibility with respect to bounded strictly proper scoring rules. Our work extends recent results given by the last two authors of this paper on the equivalence between coherence and admissibility for conditional probability assessments. In order to prove that admissibility implies coherence a key role is played by the notion of Bregman divergence.

Keywords: Conditional prevision assessments, coherence, proper scoring rules, conditional scoring rules, weak dominance, strong dominance, admissibility, Bregman divergence.

1 Introduction

Proper scoring rules have been largely studied in many fields, such as probability, statistics and decision theory. The notion of proper scoring rules was central to de Finetti's ideas about assessing the relative values of different subjective probability assessments ([4]). A review of the general theory, with applications, has been given in [6,7]; an application to sequential forecasting of economic indices has been given in [1]. The connections between the notions of coherence and of admissibility have been investigated in many works (see, e.g., [4,8,9,10,11]). In [5] the last two authors of this paper extended the results given in [9] to the case of conditional probability assessments. In this paper we further extend the work made in [5], by considering the case of conditional prevision assessments on arbitrary families of finite conditional random quantities. We prove the equivalence between the coherence of a conditional prevision assessment on an arbitrary family of finite conditional random quantities and the admissibility of the assessment with respect to any given bounded strictly proper scoring rule. The paper is organized as follows: In Section 2 we give some preliminary notions on conditional prevision assessments; then, we recall some results on the checking of coherence for conditional prevision assessments; in Section 3 we illustrate the notions of strictly proper scoring rules and admissibility for conditional prevision assessments; we also give a list of properties for the conditional prevision of strictly proper scoring rules; finally, in Section 4 we prove that a conditional prevision assessment on an arbitrary family of finite conditional random quantities is coherent if and only if it is admissible with respect to any bounded strictly proper scoring rule.

2 Some preliminary notions

We denote by A^c the negation of A and by $A \vee B$ (resp., AB) the logical union (resp., intersection) of A and B. We use the same symbol to denote an event and its indicator. For each integer n, we set $J_n = \{1, 2, \ldots, n\}$. Given a prevision function $\mathbb{P}$ defined on an arbitrary family $\mathcal{K}$ of finite conditional random quantities, let $\mathcal{F}_n = \{X_i|H_i,\, i \in J_n\}$ be a finite subfamily of $\mathcal{K}$ and $\mathcal{M}_n$ the vector $(\mu_i,\, i \in J_n)$, where $\mu_i = \mathbb{P}(X_i|H_i)$ is the assessed prevision for the conditional random quantity $X_i|H_i$. With the pair $(\mathcal{F}_n, \mathcal{M}_n)$ we associate the random gain $\mathcal{G}_n = \sum_{i \in J_n} s_i H_i (X_i - \mu_i)$, where $s_1, \ldots, s_n$ are arbitrary real numbers and $H_1, \ldots, H_n$ denote the indicators of the corresponding events. We set $\mathcal{H}_n = H_1 \vee \cdots \vee H_n$; moreover, we denote by $\mathcal{G}_n|\mathcal{H}_n$ the restriction of $\mathcal{G}_n$ to $\mathcal{H}_n$. Then, using the *betting scheme* of de Finetti, we have

Definition 1. The function $\mathbb{P}$ is coherent if and only if, $\forall\, n \geq 1, \forall\, \mathcal{F}_n \subseteq \mathcal{K}, \forall\, s_1, \ldots, s_n \in \mathbb{R}$, it holds: $\sup \mathcal{G}_n|\mathcal{H}_n \geq 0$.

Given a family of n conditional random quantities $\mathcal{F}_n = \{X_1|H_1, \ldots, X_n|H_n\}$, for each $i \in J_n$ we assume $X_i \in \{x_{i1}, \ldots, x_{ir_i}\}$; then, for each $i \in J_n$ and $j = 1, \ldots, r_i$, we set $A_{ij} = (X_i = x_{ij})$. Of course, for each $i \in J_n$, the family $\{A_{ij}\,,\ j = 1, \ldots, r_i\}$ is a partition of the sure event Ω. Moreover, for each $i \in J_n$, the family $\{H_i^c, A_{ij}H_i\,,\ j = 1, \ldots, r_i\}$ is a partition of Ω too. Then, the constituents generated by the family $\mathcal{F}_n$ are (the elements of the partition of Ω) obtained by expanding the expression $\bigwedge_{i \in J_n} (A_{i1}H_i \vee \cdots \vee A_{ir_i}H_i \vee H_i^c)$. We set $C_0 = H_1^c \cdots H_n^c$ (it may be $C_0 = \emptyset$); moreover, we denote by $C_1, \ldots, C_m$ the constituents contained in $\mathcal{H}_n = H_1 \vee \cdots \vee H_n$. Hence

$$\bigwedge_{i \in J_n} (A_{i1}H_i \vee \cdots \vee A_{ir_i}H_i \vee H_i^c) = \bigvee_{h=0}^{m} C_h\,.$$

With each C_h, $h \in J_m$, we associate a vector $Q_h = (q_{h1}, \ldots, q_{hn})$, where

$$q_{hi} = \begin{cases} x_{i1}\,, & C_h \subseteq A_{i1}H_i\,, \\ \cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots \\ x_{ir_i}\,, & C_h \subseteq A_{ir_i}H_i\,, \\ \mu_i\,, & C_h \subseteq H_i^c\,. \end{cases} \tag{1}$$

In more explicit terms, for each $j \in \{1, \ldots, r_i\}$ the condition $C_h \subseteq A_{ij}H_i$ amounts to $C_h \subseteq A_{i1}^c \cdots A_{i,j-1}^c A_{ij} A_{i,j+1}^c \cdots A_{ir}^c A_{ir_i}^c H_i$.

Given any vector $(\lambda_h\,,\,h\in J_m)$ and any event A, we set $\sum_{h:C_h\subseteq A}\lambda_h=\sum_A\lambda_h$. Then, by observing that $H_i=\bigvee_{j=1}^{r_i}A_{ij}H_i$, for each $i\in J_n$ we have

$$\sum_{h\in J_m}\lambda_h q_{hi}=\sum_{H_i}\lambda_h q_{hi}+\sum_{H_i^c}\lambda_h q_{hi}=\sum_{j=1}^{r_i}x_{ij}\sum_{A_{ij}H_i}\lambda_h+\mu_i\sum_{H_i^c}\lambda_h\,. \tag{2}$$

Denoting by $\mathcal{I}_n$ the convex hull of the points $Q_1,\ldots,Q_m$, we examine the satisfiability of the condition $\mathcal{M}_n\in\mathcal{I}_n$; that is we check the existence of a vector $(\lambda_1,\ldots,\lambda_m)$ such that: $\sum_{h\in J_m}\lambda_h Q_h=\mathcal{M}_n\,,\ \sum_{h\in J_m}\lambda_h=1\,,\ \lambda_h\geq 0\,,\ \forall\,h$. More explicitly, we check the solvability of the following system Σ associated with the pair $(\mathcal{F},\mathcal{M})$, in the nonnegative unknowns $\lambda_1,\ldots,\lambda_m$,

$$\Sigma:\qquad \sum_{h\in J_m}\lambda_h q_{hi}=\mu_i\,,\ i\in J_n\,;\ \sum_{h\in J_m}\lambda_h=1\,,\ \lambda_h\geq 0\,,\ \forall\,h\,. \tag{3}$$

We remark that $X_iH_i=\sum_{j=1}^{r_i}x_{ij}A_{ij}H_i$; hence, by interpreting the vector $(\lambda_h,\,h\in J_m)$ as a probability assessment on the family $\{C_1|\mathcal{H}_n,\,\ldots,\,C_m|\mathcal{H}_n\}$, one has: $\mathbb{P}(X_iH_i|\mathcal{H}_n)=\sum_{j=1}^{r_i}x_{ij}\sum_{A_{ij}H_i}\lambda_h=\mathbb{P}(X_i|H_i)P(H_i|\mathcal{H}_n)$, where $P(H_i|\mathcal{H}_n)=\sum_{H_i}\lambda_h$. Then in system (3), by decomposition formula (2), the equality $\sum_{h\in J_m}\lambda_h q_{hi}=\mu_i$ represents the condition $\mathbb{P}(X_iH_i|\mathcal{H}_n)=\mu_i P(H_i|\mathcal{H}_n)$. Given a subset $J\subseteq J_n$, we set $\mathcal{F}_J=\{X_i|H_i\,,\,i\in J\}\,,\ \ \mathcal{M}_J=(\mu_i\,,\,i\in J)\,$; then, we denote by Σ_J, where $\Sigma_{J_n}=\Sigma$, the system like (3) associated with the pair $(\mathcal{F}_J,\mathcal{M}_J)$. Then, it can be proved the following ([2])

Theorem 1. [*Characterization of coherence*]. Given a family of n conditional random quantities $\mathcal{F}=\{X_1|H_1,\ldots,X_n|H_n\}$ and a vector $\mathcal{M}=(\mu_1,\ldots,\mu_n)$, the conditional prevision assessment $\mathbb{P}(X_1|H_1)=\mu_1\,,\,\ldots\,,\,\mathbb{P}(X_n|H_n)=\mu_n$ is coherent if and only if, for every subset $J\subseteq J_n$, defining $\mathcal{F}_J=\{X_i|H_i\,,\,i\in J\}$, $\mathcal{M}_J=(\mu_i\,,\,i\in J)$, the system Σ_J associated with the pair $(\mathcal{F}_J,\mathcal{M}_J)$ is solvable.

3 Scoring rules and admissibility for conditional prevision assessments

In this section we consider scoring rules for conditional prevision assessments and we illustrate the notions of weak and strong dominance, and of admissibility 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. We now extend the notion of strictly proper scoring rule in the following way.

Definition 2. A function $\sigma:(-\infty,+\infty)\times(-\infty,+\infty)\ \rightarrow\ [0,+\infty)$ is said to be a strictly proper scoring rule if the following conditions are satisfied:
(a) given any real numbers $x_1,\ldots,x_r,z,p_1,\ldots,p_r$, with

$$\sum_{i=1}^{r}p_i=1\,,\ \sum_{i=1}^{r}p_ix_i=\mu\neq z\,,\ p_i\geq 0\,,\,\forall\,i\,,$$

it holds

$$\sum_{i=1}^{r}p_i\,\sigma(x_i,z)>\sum_{i=1}^{r}p_i\,\sigma(x_i,\mu)\,; \tag{4}$$

(b) for every real number x, the function $\sigma(x,z)$ is a continuous function of z.

In this paper we focus our attention on strictly proper scoring rules which are bounded. Given a scoring rule σ, with any (finite) conditional random quantity $X|H$, we associate the *conditional scoring rule* $\sigma(X|H,z)$ defined as $\sigma(X|H,z) = H\sigma(X,z)$. Consider any conditional random quantity $X|H$, with $X \in \{x_1,\dots,x_r\}$, and any probability distribution $\mathcal{P} = (p_1,\dots,p_r) \in \mathcal{V}^{r-1}$, where $p_i = P(X = x_i|H)$ and $\mathcal{V}^{r-1} = \{\mathcal{P} = (p_1,\dots,p_r) : \sum_{i=1}^{r} p_i = 1,\ p_i \geq 0\}$. We denote by $\overline{\mathcal{P}}$ the subvector $(p_1,p_2,\dots,p_{r-1})$ of $\mathcal{P}$ and by $\mathcal{S}^{r-1} \subset \mathbb{R}^{r-1}$ the convex set

$$\mathcal{S}^{r-1} = \{\overline{\mathcal{P}} = (p_1,\dots,p_{r-1}) \in \mathbb{R}^{r-1} : \mathcal{P} = (\overline{\mathcal{P}}, 1 - \textstyle\sum_{i=1}^{r-1} p_i) \in \mathcal{V}^{r-1}\}.$$

For any given real number z and for any proper scoring rule σ the conditional prevision of $\sigma(X|H,z)$ w.r.t. $\mathcal{P}$ is given by

$$s(\overline{\mathcal{P}},z) = \mathbb{P}_{\mathcal{P}}(\sigma(X|H,z)|H) = \textstyle\sum_{i=1}^{r-1} p_i\,\sigma(x_i,z) + (1 - \sum_{i=1}^{r-1} p_i)\,\sigma(x_r,z)\,. \quad (5)$$

We give below, without proof, some properties of the function $s(\overline{\mathcal{P}},z)$.

Proposition 1. The function $s(\overline{\mathcal{P}},z) : \mathcal{S}^{r-1} \times (-\infty,+\infty) \to [0,+\infty)$ satisfies the following properties:

1. $s(\alpha\overline{\mathcal{P}}' + (1-\alpha)\overline{\mathcal{P}}'', z) = \alpha\, s(\overline{\mathcal{P}}', z) + (1-\alpha)\, s(\overline{\mathcal{P}}'', z)$ for every $\alpha \in [0,1]$;
2. we have $s(\overline{\mathcal{P}},z) \geq s(\overline{\mathcal{P}},\mu)$, where $\mu = \sum_{i=1}^{r-1} p_i x_i + (1 - \sum_{i=1}^{r-1} p_i)x_r$, with $s(\overline{\mathcal{P}},z) = s(\overline{\mathcal{P}},\mu)$ if and only if $z = \mu$;
3. $s(\overline{\mathcal{P}},\mu)$, with $\mu = \sum_{i=1}^{r-1} p_i x_i + (1 - \sum_{i=1}^{r-1} p_i)x_r$, is a strictly concave function of $\overline{\mathcal{P}}$;
4. given any $\overline{\mathcal{P}} = (p_1,\dots,p_{r-1})$, with $\sum_{i=1}^{r-1} p_i x_i + (1 - \sum_{i=1}^{r-1} p_i)x_r = \mu$, $s(\overline{\mathcal{P}},z)$ is partially derivable with respect to z, $\forall\, z$, and it holds $\frac{\partial s(\overline{\mathcal{P}},z)}{\partial z}|_{z=\mu} = 0$;
5. given any interior point $\overline{\mathcal{P}}$ of $\mathcal{S}^{r-1}$, with $\sum_{i=1}^{r-1} p_i x_i + (1 - \sum_{i=1}^{r-1} p_i)x_r = \mu$, for each $j = 1,\dots,r-1$, we have $\frac{\partial s(\overline{\mathcal{P}},\mu)}{\partial p_j} = \sigma(x_j,\mu) - \sigma(x_r,\mu)$. Moreover, $s(\overline{\mathcal{P}},\mu)$ is differentiable in the interior of $\mathcal{S}^{r-1}$;
6. for any interior point $\overline{\mathcal{P}}$ of $\mathcal{S}^{r-1}$, with $\sum_{i=1}^{r-1} p_i x_i + (1 - \sum_{i=1}^{r-1} p_i)x_r = \mu$, and for every $\overline{\mathcal{P}}' \in \mathcal{S}^{r-1}$, we have
$$s(\overline{\mathcal{P}}',\mu) = s(\overline{\mathcal{P}},\mu) + \nabla s(\overline{\mathcal{P}},\mu) \cdot (\overline{\mathcal{P}}' - \overline{\mathcal{P}})\,.$$

Given a prevision assessment $\mathcal{M}_n = (\mu_1,\mu_2,\dots,\mu_n)$ on a family of conditional random quantities $\mathcal{F}_n = \{X_1|H_1, X_2|H_2,\dots,X_n|H_n\}$, where $\mu_i = \mathbb{P}(X_i|H_i)$, and a proper scoring rule σ, let $C_0,C_1,\dots,C_m$ be the constituents generated by $\mathcal{F}_n$ and $Q_1,\dots,Q_m$ the points associated with the pair $(\mathcal{F}_n,\mathcal{M}_n)$, as defined by formula (1). The penalty $\mathcal{L}$ associated with the pair $(\mathcal{F}_n,\mathcal{M}_n)$ is given by

$$\mathcal{L} = \sum_{i=1}^{n} \sigma(X_i|H_i,\mu_i) = \sum_{i=1}^{n} H_i\sigma(X_i,\mu_i)\,.$$

We denote by L_k the value of $\mathcal{L}$ associated with C_k, $k = 0,1,\dots,m$. Of course, $L_0 = 0$; moreover, by defining the quantities

$$h_{ki} = \begin{cases} 1\,, & C_k \subseteq H_i\,, \\ 0\,, & C_k \subseteq H_i^c\,, \end{cases} \qquad e_{kij} = \begin{cases} 1\,, & C_k \subseteq A_{ij}\,, \\ 0\,, & C_k \subseteq A_{ij}^c\,, \end{cases}$$

we have

$$L_k = \sum_{i=1}^{n} h_{ki} \sum_{j=1}^{r_i} e_{kij}\sigma(x_{ij}, \mu_i)\,,\; k = 1, \ldots, m\,. \tag{6}$$

We give below the notions of weak and strong dominance and admissibility with respect to scoring rules.

Definition 3. Let σ be a scoring rule and $\mathcal{M}_n$ be a prevision assessment on a family $\mathcal{F}_n$ of n conditional random quantities. Given any assessment $\mathcal{M}_n^*$ on $\mathcal{F}_n$, with $\mathcal{M}_n^* \neq \mathcal{M}_n$, we say that $\mathcal{M}_n$ is *weakly dominated* by $\mathcal{M}_n^*$, with respect to σ, if denoting by $\mathcal{L}$ (resp., $\mathcal{L}^*$) the penalty associated with the pair $(\mathcal{F}_n, \mathcal{M}_n)$ (resp., $(\mathcal{F}_n, \mathcal{M}_n^*)$), it holds $\mathcal{L}^* \leq \mathcal{L}$, that is: $L_k^* \leq L_k$, for every $k = 0, 1, \ldots, m$. Moreover, by observing that $L_0 = L_0^* = 0$, we say that $\mathcal{M}_n$ is *strongly dominated* by $\mathcal{M}_n^*$, with respect to σ, if $L_k^* < L_k$, for every $k = 1, \ldots, m$.

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

Definition 4. Let σ be a scoring rule and $\mathcal{M}_n$ be a prevision assessment on a family $\mathcal{F}_n$ of n conditional random quantities. We say that $\mathcal{M}_n$ is *admissible w.r.t.* σ if $\mathcal{M}_n$ is not weakly dominated by any $\mathcal{M}_n^* \neq \mathcal{M}_n$. Moreover, given a prevision assessment $\mathcal{M}$ on an arbitrary family of conditional random quantities $\mathcal{K}$, we say that $\mathcal{M}$ is admissible w.r.t. σ if, for every finite subfamily $\mathcal{F}_n \subseteq \mathcal{K}$, the restriction $\mathcal{M}_n$ of $\mathcal{M}$ to $\mathcal{F}_n$ is admissible w.r.t. σ.

Remark 1. We observe that, by Definition 4, it follows:
- If the assessment $\mathcal{M}_n$ on $\mathcal{F}_n$ is admissible, then for every subfamily $\mathcal{F}_J \subset \mathcal{F}_n$ the sub-assessment $\mathcal{M}_J$ associated with $\mathcal{F}_J$ is admissible.

4 Coherence and admissibility of conditional prevision assessments

In this section we give the main result of the paper, by showing the equivalence between the coherence of conditional prevision assessments and admissibility with respect to proper scoring rules. Given the assessment $\mathcal{M}_n = (\mu_1, \ldots, \mu_n)$ on $\mathcal{F}_n = \{X_1|H_1, X_2|H_2, \ldots, X_n|H_n\}$ and a bounded strictly proper scoring rule σ, we set $S(\overline{P}, \mathcal{Z}_n) = S(\overline{\mathcal{P}}_1, \ldots, \overline{\mathcal{P}}_n, \mathcal{Z}_n) = \sum_{i=1}^{n} s(\overline{\mathcal{P}}_i, z_i)$, where $\overline{P} = (\overline{\mathcal{P}}_1, \ldots, \overline{\mathcal{P}}_n), \overline{\mathcal{P}}_i = (p_{i1}, \ldots, p_{ir_i-1})$, $\sum_{j=1}^{r_i-1} p_{ij}x_{ij} + (1 - \sum_{j=1}^{r_i-1} p_{ij})x_{ir_i} = \mu_i$ and $\mathcal{Z}_n = (z_1, \ldots, z_n)$. Given any vector $\overline{P}' = (\overline{\mathcal{P}}'_1, \ldots, \overline{\mathcal{P}}'_n) \in \Pi$, with $\overline{\mathcal{P}}'_i = (p'_{i1}, \ldots, p'_{ir_i-1})$ and $\Pi = \prod_{i=1}^{n} \mathcal{S}^{r_i-1} \subset \mathbb{R}^{r-n}$, $r = \sum_{i=1}^{n} r_i$, from the properties 5 and 6 in Proposition 1 we have $S(\overline{P}', \mathcal{M}_n) = S(\overline{P}, \mathcal{M}_n) + \nabla S(\overline{P}, \mathcal{M}_n) \cdot (\overline{P}' - \overline{P})$. We set $\Phi(\overline{P}) = -S(\overline{P}, \mathcal{M}_n) = -\sum_{i=1}^{n} s(\overline{\mathcal{P}}_i, \mu_i)$. Then, we have $S(\overline{P}', \mathcal{M}_n) = -\Phi(\overline{P}) - \nabla\Phi(\overline{P}) \cdot (\overline{P}' - \overline{P})$, that is

$$S(\overline{P}', \mathcal{M}_n) - S(\overline{P}, \mathcal{M}_n) = -\nabla\Phi(\overline{P}) \cdot (\overline{P}' - \overline{P})\,. \tag{7}$$

We observe that the function $\Phi(\overline{P})$ is continuous on Π and strictly convex in the interior of Π. Moreover, $\Phi(\overline{P})$ has continuous partial derivatives on the interior of Π, so that $\Phi(\overline{P})$ is differentiable in the interior of Π and the gradient $\nabla\Phi(\overline{P})$ is a continuous function on the interior of Π. As the functions $\sigma(x_{ij}, z_i)$ are assumed bounded, then $\nabla\Phi(\overline{P})$ extends to a bounded continuous function on Π. The Bregman divergence ([3,9]) associated with the function Φ is given by

$$d_{\Phi}(\overline{P}', \overline{P}) = \Phi(\overline{P}') - \Phi(\overline{P}) - \nabla\Phi(\overline{P}) \cdot (\overline{P}' - \overline{P}) .$$

Then, from (7) it follows

$$d_{\Phi}(\overline{P}', \overline{P}) = \Phi(\overline{P}') - \Phi(\overline{P}) + S(\overline{P}', \mathcal{M}_n) + \Phi(\overline{P}) = S(\overline{P}', \mathcal{M}_n) - S(\overline{P}', \mathcal{M}'_n) . \quad (8)$$

We illustrate now the relationship between the notion of coherence and the property of non dominance, by first examining a single assessment $\mathbb{P}(X|H) = \mu$.

Lemma 1. Given any event $H \neq \emptyset$, any finite random quantity $X \in \{x_1, \ldots, x_r\}$ and any strictly proper continuous and bounded scoring rule σ, the assessment $\mathbb{P}(X|H) = \mu$ is coherent if and only if μ is admissible with respect to σ.

Proof. ($\Rightarrow$) Assume that μ is coherent. With no loss of generality, we can suppose $x_1 < \cdots < x_r$ and $A_i H \neq \emptyset$ where $A_i = (X = x_i)$, $i = 1, \ldots, r$; then coherence of μ amounts to $x_1 \leq \mu \leq x_r$, so that there exist $p_1, \ldots, p_r$ such that $\sum_i p_i x_i = \mu$, with $\sum_i p_i = 1$, $p_i \geq 0$, for every i. Then, for any given $\mu^* \neq \mu$, by recalling (4) we have $\sum_i p_i \sigma(x_i, \mu) < \sum_i p_i \sigma(x_i, \mu^*)$, so that $\sigma(x_k, \mu) < \sigma(x_k, \mu^*)$ for at least an index k; hence, μ is not weakly dominated by μ^*.
($\Leftarrow$) Assume that μ is not coherent; that is $\mu \notin [x_1, x_r]$. We consider the random quantity $Y = XH + \mu H^c$ with possible values: $x_1, \ldots, x_r, \mu$, which are associated with the $r+1$ constituents: $A_1 H, \ldots, A_r H, H^c$. Let $P = (p_1, p_2, \ldots, p_{r+1})$ be a probability distribution on Y; we set $\overline{P} = (p_1, p_2, \ldots, p_r)$. Then, the prevision of the score $\sigma(Y, \mu)$ is $s(\overline{P}, \mu)$. We observe that particular choices of P are the vectors $W_k = (w_{k1}, \ldots, w_{kr}, w_{k\,r+1})$, $k = 1, 2, \ldots, r+1$, with $W_1 = (1, 0, \ldots, 0)$, $W_2 = (0, 1, 0, \ldots, 0)$, $\ldots, W_{r+1} = (0, \ldots, 0, 1)$. We set $\overline{W}_k = (w_{k1}, \ldots, w_{kr})$; then $s(\overline{W}_k, \mu) = \sum_{j=1}^{r} w_{kj} \sigma(x_j, \mu) + (1 - \sum_{j=1}^{r} w_{kj}) \sigma(\mu, \mu) = \sigma(x_k, \mu)$, $k = 1, \ldots, r$, with $s(\overline{W}_{r+1}, \mu) = \sigma(\mu, \mu)$. As $s(\overline{W}_k, x_k) = \sigma(x_k, x_k)$, $k = 1, \ldots, r$, we obtain $L_k = \sigma(x_k, \mu) = s(\overline{W}_k, \mu) - s(\overline{W}_k, x_k) + \alpha_k$, $k = 1, \ldots, r$, with $\alpha_k = \sigma(x_k, x_k)$ and $L_{r+1} = \sigma(\mu, \mu)$. We set $\mathcal{C} = [0, 1]^r$; then, we consider the function $\Phi(\overline{P}) : \mathcal{C} \to R$, defined as $\Phi(\overline{P}) = -s(\overline{P}, \mu(\overline{P}))$, with $\overline{P} = (p_1, \ldots, p_r)$, $\mu(\overline{P}) = p_1 x_1 + \ldots + p_r x_r + p_{r+1}\mu$, $p_{r+1} = 1 - \sum_{j=1}^{r} p_j$. Based on (8), we have $d_{\Phi}(\overline{P}', \overline{P}) = s(\overline{P}', \mu(\overline{P})) - s(\overline{P}', \mu(\overline{P}'))$ and, by observing that $\mu(\overline{W}_k) = x_k$ and $\mu(\overline{W}_{r+1}) = \mu$, we obtain

$$L_k = s(\overline{W}_k, \mu) - s(\overline{W}_k, x_k) + \alpha_k = d_{\Phi}(\overline{W}_k, \overline{W}_{r+1}) + \alpha_k, \;\; k = 1, \ldots, r .$$

Denoting by $\mathcal{I}_W$ the convex hull of $\overline{W}_1, \ldots, \overline{W}_r$, for each $\overline{P} = (p_1, \ldots, p_r) \in \mathcal{I}_W$ we have $\overline{P} = \sum_{i=1}^{r} p_i \overline{W}_i$, with $\sum_{i=1}^{r} p_i = 1$, so that $p_{r+1} = 0$. Then $\mu(\overline{P}) \in [x_1, x_r]$ and, as $\mu \notin [x_1, x_r]$, we have $\mu(\overline{P}) \neq \mu$, $\forall\, \overline{P} \in \mathcal{I}_W$. Then, for every $\overline{P}_{\mu} \in \mathcal{C}$ such that $\mu(\overline{P}_{\mu}) = \mu$, it holds that $\overline{P} \notin \mathcal{I}_W$; thus, there exists a projection point

$\overline{P}^*_\mu \in \mathcal{I}_W$, with $\mu(\overline{P}^*_\mu) = \mu^*$, such that: $d_\Phi(\overline{W}_k, \overline{P}^*_\mu) + d_\Phi(\overline{P}^*_\mu, \overline{P}_\mu) \leq d_\Phi(\overline{W}_k, \overline{P}_\mu)$, and, as $\overline{P}^*_\mu \neq \overline{P}_\mu$, one has: $d_\Phi(\overline{W}_k, \overline{P}^*_\mu) < d_\Phi(\overline{W}_k, \overline{P}_\mu)$, $k = 1, \ldots, r$, that is $\sigma(x_k, \mu^*) < \sigma(x_k, \mu)$, $k = 1, \ldots, r$. Now, by considering the alternative assessments μ^* and μ for the prevision of $X|H$, we have

$$L^*_0 = L_0 = 0 \text{ and } L^*_k = \sigma(x_k, \mu^*) < \sigma(x_k, \mu) = L_k\,,\ k = 1, \ldots, r\,;$$

thus, μ is (strictly) dominated by μ^* with respect to σ. □

Based on the previous Lemma, given any prevision assessment $\mathcal{M}_n = (\mu_1, \ldots, \mu_n)$, in what follows we can assume $\mu_i \in [min\, X_i|H_i, max\, X_i|H_i,]$ for every i. We have

Theorem 2. Let $\mathcal{M}$ be a prevision assessment on a family $\mathcal{K}$ of conditional random quantities, with $\mu_{X|H} = \mathbb{P}(X|H) \in [minX|H, maxX|H]$ for every $X|H \in \mathcal{K}$; moreover, let σ be any bounded strictly proper scoring rule. $\mathcal{M}$ is coherent if and only if $\mathcal{M}$ is admissible with respect to σ.

Proof. ($\Rightarrow$) Assuming $\mathcal{M}$ coherent, let σ be a bounded proper scoring rule. Given any subfamily $\mathcal{F}_n = \{X_1|H_1, \ldots, X_n|H_n\}$ of $\mathcal{K}$, let $\mathcal{M}_n = (\mu_1, \ldots, \mu_n)$ be the restriction to $\mathcal{F}_n$ of $\mathcal{M}$. Now, given any $\mathcal{M}^*_n = (\mu^*_1, \ldots, \mu^*_n) \neq \mathcal{M}_n$, we distinguish two cases:
(a) $\mu^*_i \neq \mu_i$, for every $i = 1, \ldots, n$; **(b)** $\mu^*_i = \mu_i$, for at least one index i.
Case **(a)**. We still denote by $C_0, C_1, \ldots, C_m$, where $C_0 = H^c_1 \wedge \cdots \wedge H^c_n$, the constituents generated by $\mathcal{F}_n$ and by $Q_k = (q_{k1}, \ldots, q_{kn})$ the point associated with $C_k, k = 1, \ldots, m$. With the assessment $\mathcal{M}_n$ we associate the loss

$$\mathcal{L} = \textstyle\sum_{i=1}^n \sigma(X_i|H_i, \mu_i) = \sum_{i=1}^n H_i \sigma(X_i, \mu_i)\,,$$

with $L_0 = 0$ and, recalling (6), $L_k = \sum_{i=1}^n h_{ki} \sum_{j=1}^{r_i} e_{kij} \sigma(x_{ij}, \mu_i)\,,\ k = 1, \ldots, m$. Of course, with any other assessment $\mathcal{M}^*_n$ on $\mathcal{F}_n$ we associate the loss

$$\mathcal{L}^* = \textstyle\sum_{i=1}^n \sigma(X_i|H_i, \mu^*_i) = \sum_{i=1}^n H_i \sigma(X_i, \mu^*_i)\,,$$

with $L^*_0 = 0$ and $L^*_k = \sum_{i=1}^n h_{ki} \sum_{j=1}^{r_i} e_{kij} \sigma(x_{ij}, \mu^*_i)\,,\ k = 1, \ldots, m$.
As $L_0 = L^*_0 = 0$, in what follows we will only refer to the values L_k, L^*_k, $k = 1, \ldots, m$. As $\mathcal{M}_n$ is coherent, there exists a vector $(\lambda_1, \ldots, \lambda_m)$, with $\lambda_k \geq 0$ and $\sum_k \lambda_k = 1$, which is a coherent extension of $\mathcal{M}_n$ on the family of conditional events $\{C_1|\mathcal{H}_n, \ldots, C_m|\mathcal{H}_n\}$, with $\lambda_h = P(C_h|\mathcal{H}_n)$. We have

$$P(H_i|\mathcal{H}_n) = \textstyle\sum_{C_k \subseteq H_i} P(C_k|\mathcal{H}_n) = \sum_k \lambda_k h_{ki}\,,$$

with $\sum_{i=1}^n P(H_i|\mathcal{H}_n) \geq P(\mathcal{H}_n|\mathcal{H}_n) = 1$, so that $P(H_i|\mathcal{H}_n) > 0$ for at least an index i. Moreover

$$P(A_{ij} H_i|\mathcal{H}_n) = \textstyle\sum_{C_k \subseteq A_{ij} H_i} P(C_k|\mathcal{H}_n) = \sum_k \lambda_k h_{ki} e_{kij} = P(A_{ij}|H_i) P(H_i|\mathcal{H}_n)\,.$$

We set: $I' = \{i : \sum_k \lambda_k h_{ki} > 0\} \subseteq \{1, 2, \ldots, n\}$. Of course, $I' \neq \emptyset$. We set $P(A_{ij}|H_i) = p_{ij}$; then, by observing that

$$\begin{aligned} &\textstyle\sum_{j=1}^{r_i} (\sum_k \lambda_k h_{ki} e_{kij}) x_{ij} = \sum_{j=1}^{r_i} P(A_{ij} H_i|\mathcal{H}_n) x_{ij} = \\ &= \textstyle\sum_{j=1}^{r_i} P(A_{ij}|H_i) P(H_i|\mathcal{H}_n) x_{ij} = P(H_i|\mathcal{H}_n) \sum_{j=1}^{r_i} p_{ij}\, x_{ij} = \mu_i P(H_i|\mathcal{H}_n)\,, \end{aligned}$$

for each $i \in I'$ it holds

$$\sum_{j=1}^{r_i}(\sum_k \lambda_k h_{ki} e_{kij})\sigma(x_{ij},\mu_i) = P(H_i|\mathcal{H}_n)\sum_{j=1}^{r_i} p_{ij}\sigma(x_{ij},\mu_i) < \\ < P(H_i|\mathcal{H}_n)\sum_{j=1}^{r_i} p_{ij}\sigma(x_{ij},\mu_i^*)\,.$$

It follows:

$$\sum_k \lambda_k L_k = \sum_k \lambda_k \sum_{i=1}^{n} h_{ki} \sum_{j=1}^{r_i} e_{kij}\sigma(x_{ij},\mu_i) = \sum_{i\in I'}\sum_{j=1}^{r_i}(\sum_k \lambda_k h_{ki} e_{kij})\sigma(x_{ij},\mu_i) = \\ \sum_{i\in I'} P(H_i|\mathcal{H}_n)\sum_{j=1}^{r_i} p_{ij}\sigma(x_{ij},\mu_i) < \sum_{i\in I'} P(H_i|\mathcal{H}_n)\sum_{j=1}^{r_i} p_{ij}\sigma(x_{ij},\mu_i^*) = \\ = \sum_k \lambda_k \sum_{i=1}^{n} h_{ki}\sum_{j=1}^{r_i} e_{kij}\sigma(x_{ij},\mu_i^*) = \sum_k \lambda_k L_k^*\,.$$

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 $\mathcal{M}_n$ is admissible. Since $\mathcal{F}_n$ is arbitrary, it follows that $\mathcal{M}$ is admissible.
Case **(b)**. Let $\mathcal{M}_n^* \neq \mathcal{M}_n$, with $\mu_i^* = \mu_i$, for at least one index i. We set $J = \{i : \mu_i^* \neq \mu_i\} \subset J_n = \{1,\ldots,n\}$. We denote by $\mathcal{M}_J$ (resp., $\mathcal{M}_{J_n\setminus J}$) the subvector of $\mathcal{M}_n$ associated with J (resp., $J_n \setminus J$). Analogously, we can consider the subvectors $\mathcal{M}_J^*$ and $\mathcal{M}_{J_n\setminus J}^*$ of $\mathcal{M}_n^*$. 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 $\mathcal{M}_n$ is admissible. Since $\mathcal{F}_n$ is arbitrary, $\mathcal{M}$ is admissible.
($\Leftarrow$). We will prove that, given any bounded proper scoring rule σ, if $\mathcal{M}$ is not coherent, then $\mathcal{M}$ is not admissible with respect to σ. Assume that $\mathcal{M}$ is not coherent. Then, there exists a subfamily $\mathcal{F}_n = \{X_1|H_1,\ldots,X_n|H_n\} \subseteq \mathcal{K}$ such that, for the restriction $\mathcal{M}_n = (\mu_1,\ldots,\mu_n)$ of $\mathcal{M}$ to $\mathcal{F}_n$, denoting by $\mathcal{I}_n$ the associated convex hull, we have $\mathcal{M}_n \notin \mathcal{I}_n$. For each constituent C_k we set $\Gamma_k = \{i : C_k \subseteq H_i\}$, $I_k = \{i : C_k \subseteq H_i^c\} = J_n \setminus \Gamma_k$. As for each i it holds $minX_i|H_i \leq \mu_i \leq maxX_i|H_i$, with each quantity q_{ki}, defined as in (1), we associate a vector $W_{ki} = (w_{ki1},\ldots,w_{kir_i}) \in \mathcal{V}^{r_i-1}$, with

$$w_{kij} = \begin{cases} 1\,, & C_k \subseteq A_{ij}H_i\,, \\ 0\,, & C_k \subseteq A_{ij}^c H_i\,, \\ p_{ij}\,, & C_k \subseteq H_i^c\,, \end{cases} \tag{9}$$

where $\sum_{j=1}^{r_i} p_{ij}x_{ij} = \mu_i$, $\sum_{j=1}^{r_i} p_{ij} = 1$, $p_{ij} \geq 0$. We denote by $\overline{W}_{ki}$ the subvector $(w_{ki1},\ldots,w_{ki(r_i-1)})$ of W_{ki}. As $W_{ki} \in \mathcal{V}^{r-1}$ it follows $\overline{W}_{ki} \in \mathcal{S}^{r-1}$. We observe that, if $C_k \subseteq H_i$, that is $i \in \Gamma_k$, then

$$s(\overline{W}_{ki},q_{ki}) = \sum_{j=1}^{r_i-1} w_{kij}\sigma(x_{ij},q_{ki}) + (1-\sum_{j=1}^{r_i-1} w_{kij})\sigma(x_{ir},q_{ki}) = \sigma(q_{ki},q_{ki})\,; \\ s(\overline{W}_{ki},\mu_i) = \cdots = \sigma(q_{ki},\mu_i)\,.$$

If $C_k \subseteq H_i^c$, that is $i \in I_k$, then:

$$s(\overline{W}_{ki}, q_{ki}) = s(\overline{W}_{ki}, \mu_i) = \sum_{j=1}^{r_i-1} p_{ij}\sigma(x_{ij}, \mu_i) + (1 - \sum_{j=1}^{r_i-1} p_{ij})\sigma(x_{ir}, \mu_i)\,.$$

Then, by taking into account that $\sum_{i\in I_k}[s(\overline{W}_{ki}, \mu_i) - s(\overline{W}_{ki}, q_{ki})] = 0$ and defining $\overline{W}_k = (\overline{W}_{k1}, \dots, \overline{W}_{kn})$, for the value L_k of the penalty $\mathcal{L}$, we obtain

$$\begin{aligned}
&L_k = \sum_{i=1}^{n} h_{ki} \sum_{j=1}^{r_i} e_{kij}\sigma(x_{ij}, \mu_i) = \sum_{i\in\Gamma_k} \sigma(q_{ki}, \mu_i) = \\
&= \sum_{i\in\Gamma_k} \sigma(q_{ki}, \mu_i) - \sum_{i\in\Gamma_k} \sigma(q_{ki}, q_{ki}) + \sum_{i\in\Gamma_k} \sigma(q_{ki}, q_{ki}) = \\
&= \sum_{i\in\Gamma_k}[\sigma(q_{ki}, \mu_i) - \sigma(q_{ki}, q_{ki})] + \alpha_k = \sum_{i\in\Gamma_k}[s(\overline{W}_{ki}, \mu_i) - s(\overline{W}_{ki}, q_{ki})] + \alpha_k = \\
&= \sum_{i\in\Gamma_k}[s(\overline{W}_{ki}, \mu_i) - s(\overline{W}_{ki}, q_{ki})] + \sum_{i\in I_k}[s(\overline{W}_{ki}, \mu_i) - s(\overline{W}_{ki}, q_{ki})] + \alpha_k = \\
&= \sum_{i=1}^{n}[s(\overline{W}_{ki}, \mu_i) - s(\overline{W}_{ki}, q_{ki})] + \alpha_k = S(\overline{W}_k, \mathcal{M}_n) - S(\overline{W}_k, Q_k) + \alpha_k,
\end{aligned}$$

where $\alpha_k = \sum_{i\in\Gamma_k} \sigma(q_{ki}, q_{ki})$. Then, by applying (8) with $\overline{P}' = \overline{W}_k$, so that $\mathcal{M}'_n = Q_k$, we have

$$L_k = S(\overline{W}_k, \mathcal{M}_n) - S(\overline{W}_k, Q_k) + \alpha_k = d_\Phi(\overline{W}_k, \overline{P}) + \alpha_k\,. \tag{10}$$

We recall that for the probability assessment $P = (\mathcal{P}_1, \dots, \mathcal{P}_n)$ on the family $\mathcal{A} = \{A_{ij}|H_i\,,\ j = 1, \dots, r_i;\ i \in J_n\}$ it holds that $\sum_{j=1}^{r_i} p_{ij}x_{ij} = \mu_i$, $i = 1, \dots, n$. We recall that $\mathcal{M}_n \notin \mathcal{I}_n$; then, denoting by $\mathcal{I}_{\overline{P}}$ the convex hull associated with the pair $(\overline{\mathcal{A}}, \overline{P})$, where $\overline{\mathcal{A}} = \{A_{ij}|H_i\,,\ j = 1, \dots, (r_i - 1);\ i \in J_n\}$, $\overline{P} = (\overline{\mathcal{P}}_1, \dots, \overline{\mathcal{P}}_n)$ and $\overline{\mathcal{P}}_i = (p_{i1}, \dots, p_{i(r_i-1)})$, we have $\overline{P} \notin \mathcal{I}_{\overline{P}}$. Then, by recalling the projection lemma associated with Bregman divergences ([9], see also [5], Proposition 2), for the projection $\overline{P}^*$ of $\overline{P}$ on $\mathcal{I}_{\overline{P}}$ we have

$$d_\Phi(\overline{W}_k, P^*) + d_\Phi(\overline{P}^*, \overline{P}) \leq d_\Phi(\overline{W}_k, \overline{P})\,.$$

Moreover, as $\overline{P}^* \neq \overline{P}$, we have $d_\Phi(\overline{P}^*, \overline{P}) > 0$; therefore

$$d_\Phi(\overline{W}_k, \overline{P}^*) < d_\Phi(\overline{W}_k, \overline{P})\,,\ k = 1, \dots, m\,.$$

Now, with the point $\overline{P}^*$ we associate the probability assessment $P^* = (\mathcal{P}_1^*, \dots, \mathcal{P}_n^*)$, where $\mathcal{P}_i^* = (p_{i1}^*, \dots, p_{i(r_i-1)}^*, 1 - \sum_{j=1}^{r-1} p_{ij}^*)$, and the (possibly not coherent) prevision assessment $\mathcal{M}_n^* = (\mu_1^*, \dots, \mu_n^*)$, with $\sum_{j=1}^{r_i-1} p_{ij}^* x_{ij} + (1 - \sum_{j=1}^{r_j-1} p_{ij}^*)x_{ir_i} = \mu_i^*$, $i = 1, \dots, n$. For each constituent C_k we consider the vector Q_k^* associated with the pair $(\mathcal{F}_n, \mathcal{M}_n^*)$; moreover, based on (9), with the pair (P^*, Q_k^*) we associate the vector W_k^*. Then, for the values of the penalty $\mathcal{L}^*$, we have

$$L_k^* = S(\overline{W}_k^*, \mathcal{M}_n^*) - S(\overline{W}_k^*, Q_k^*) + \alpha_k^* = d_\Phi(\overline{W}_k^*, \overline{P}^*) + \alpha_k^*\,,\ k = 1, \dots, m\,, \tag{11}$$

with $L_0^* = 0$ and $\alpha_k^* = \alpha_k = \sum_{i\in\Gamma_k} \sigma(q_{ki}, q_{ki})$. We observe that

$$\overline{W}_{ki} = \overline{W}_{ki}^*\,,\ \ q_{ki} = q_{ki}^*\,,\ \forall\, i \in \Gamma_k\,,\ \forall\, k = 1, \dots, m\,.$$

Then, by virtue of the property 2 of the function $s(\overline{\mathcal{P}}, z)$ (which is connected with condition (a) in Definition 2), for each $k = 1, \dots, m$ we have

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

Therefore, $d_\Phi(\overline{W}^*_k, \overline{P}^*) \leq d_\Phi(\overline{W}_k, \overline{P}^*) < d_\Phi(\overline{W}_k, \overline{P})$, for each $j = 1, 2, \dots, m$. Then, recalling (10) and (11), for each $k = 1, \dots, m$ we obtain

$$L^*_k = d_\Phi(\overline{W}^*_k, \overline{P}^*) + \alpha_k < d_\Phi(\overline{W}_k, \overline{P}) + \alpha_k = L_k \,;$$

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

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

References

1. Agró, G., Lad, F., Sanfilippo, G.: Sequentially forecasting economic indices using mixture linear combinations of EP distributions. Journal of Data Science 8(1), 101–126, (2010)
2. Biazzo, V., Gilio, A., Sanfilippo, G.: Generalized coherence and connection property of imprecise conditional previsions. In: Proceedings of IPMU'08, Malaga, Spain, June 22-27, pp. 907–914 (2008)
3. Censor, Y., Zenios, S.A.: Parallel Optimization: Theory, Algorithms, and Applications. Oxford Univ. Press, Oxford (1997)
4. de Finetti, B.: Does it make sense to speak of 'good probability appraisers'? In: Good, I.J. (ed.) The scientist speculates: an anthology of partly-baked ideas, pp. 357–364, Heinemann, London (1962)
5. Gilio, A., Sanfilippo, G.: Coherent Conditional Probabilities and Proper Scoring Rules. In Proc. of ISIPTA'11, Innsbruck, Austria, July 25-28, pp. 189–198 (2011)
6. Gneiting, T., Raftery, A.: Strictly proper scoring rules, prediction and estimation. J. Amer. Statist. Assoc. 102(477), 359–378, (2007)
7. Lad, F.: Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction. John Wiley, New York (1996)
8. Lindley, D.V.: Scoring rules and the inevitability of probability. Int. Statist. Rev. 50, 1–11 (1982)
9. Predd, J.B., Seiringer, R., Lieb, E.H., Osherson, D.N., Poor, H.V., Kulkarni, S.R.: Probabilistic Coherence and Proper Scoring Rules. IEEE T. Inform. Theory 55, 4786–4792 (2009)
10. Savage, L.J.: Elicitation of personal probabilities and expectations. J. Amer. Statist. Assoc. 66, 783–801 (1971)
11. Schervish, M.J., Seidenfeld T., Kadane, J. B.: Proper Scoring Rules, Dominated Forecasts, and Coherence. Decision Analysis 6(4), 202–221 (2009)