Coherent conditional probabilities and proper scoring rules

Angelo Gilio & *Giuseppe Sanfilippo*

Dipartimento di Scienze di Base e Applicate per l'Ingegneria,
University of Rome "La Sapienza", Italy
angelo.gilio@sbai.uniroma1.it

Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli",
University of Palermo, Italy
sanfilippo@unipa.com

Innsbruck
July 28, 2011

Coherence of (precise or imprecise) conditional probability and/or prevision assessments	
Probabilistic aspects in nonmonotonic reasoning	
Compounds of conditionals, probability semantics for categorical syllogisms;	
Generalization of inference rules	Extropy: a complementary of entropy
Coherence based on penalty criterion, equivalence with betting scheme	Scoring analysis of forecasting distributions

Proper scoring rule (p.s.r.)

Let p be the degree of belief that You attribute to an event E and x be the degree of belief on E that You announce publicly. Suppose You are penalized as follows:
pay $f(x) = s(1, x)$ if $E = 1$; or pay $g(x) = s(0, x)$ if $E = 0$.

The rule is said to be proper if you cannot expect a lower penalty by specifying $x \neq p$.
The function $s(E, x) = Es(1, x) + (1 - E)s(0, x)$ is a (strictly) proper scoring rule if (Predd et al. 2009)
(a) for every $x, p \in [0, 1]$, with $x \neq p$, it is

$$p\, s(1, x) + (1 - p)\, s(0, x) > p\, s(1, p) + (1 - p)\, s(0, p)\,;$$

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

By setting $s(p, x) = p\, s(1, x) + (1 - p)\, s(0, x)$ condition (a) amounts to

$$\mathbb{P}(s(E, x)) = s(P(E), x) = s(p, x) > s(p, p),\ \forall x \neq p.$$

Ex.: $s(E, x) = \overbrace{(1 - x)^2}^{proper}$; $s(E, x) = \overbrace{-\log(1 - |E - x|)}^{proper}$; $s(E, x) = \overbrace{x}^{no}$;

Random penalty

Given a scoring rule s and a conditional event $E|H$ we set

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

In such a case $s(p,x)$, where $p = P(E|H)$, represents $\mathbb{P}[s(E|H,x) \,|\, H]$.

Given an assessment $\mathcal{P}$ on an arbitrary family of conditional events $\mathcal{K}$ and a subfamily $\mathcal{F}_n = \{E_1|H_1, E_2|H_2, \ldots, E_n|H_n\} \subseteq \mathcal{K}$, let $\mathcal{P}_n = (p_1, p_2, \ldots, p_n)$, where $p_i = P(E_i|H_i)$, be the restriction of $\mathcal{P}$ to $\mathcal{F}_n$.
Given any p.s.r. s we define the random penalty, or loss function, $\mathcal{L}$ associated with the pair $(\mathcal{F}_n, \mathcal{P}_n)$ as

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

In particular, for $s(E,x) = (E-x)^2$ (Brier quadratic scoring rule) we have

$$\mathcal{L} = \sum_{i=1}^{n} H_i (E_i - p_i)^2.$$

The notion of strengthened coherence

- de Finetti strengthened the notion of coherence for conditional events: "*In order to extend the notions and rules of the calculus of probability to this new case, it is necessary to* strengthen *the condition of coherence*" (de Finetti, 1974, vol.2, p. 339, Axiom 3).

- In (Regazzini, 1985), in agreement with the strengthened coherence principle, a definition of coherence for conditional events based on the betting scheme has been given. Conditioning events with zero probability are properly managed by such a notion of coherence (see also Holzer 1985, Williams 1975).

  Definition 1 A probability assessment $\mathcal{P}$ defined on an arbitrary family of conditional events $\mathcal{K}$ is *coherent* iff, for every finite subfamily $\mathcal{F}_n \subseteq \mathcal{K}$ and for every choice of $s_1, \ldots, s_n$ one has

$$\min \mathcal{G}|\mathcal{H}_n \leq 0 \leq \max \mathcal{G}|\mathcal{H}_n \ \ (\text{ or equiv. } \max \mathcal{G}|\mathcal{H}_n \geq 0),$$

  where $\mathcal{G}|\mathcal{H}_n$ is the gain $\mathcal{G} = \sum_{i=1}^n s_i H_i(E_i - p_i)$, associated with $(\mathcal{F}_n, \mathcal{P}_n)$, restricted to $\mathcal{H}_n = H_1 \vee \cdots \vee H_n$.

Coherence based on the penalty criterion

In order to unify the treatment of unconditional and conditional events, the definition of coherence given by de Finetti with the penalty criterion, based on the Brier scoring rule, was suitably modified in (Gilio 1990, 1992), by avoiding in this way any need for the strengthening of coherence.

Definition 2 A probability assessment $\mathcal{P}$ defined on $\mathcal{K}$ is *coherent* if and only if do not exist a finite subfamily $\mathcal{F}_n \subseteq \mathcal{K}$ and an assessment $\mathcal{P}_n^*$ on $\mathcal{F}_n$ such that $\mathcal{L}^* \leq \mathcal{L}$ and $\mathcal{L}^* \neq \mathcal{L}$, where $\mathcal{L}^* = \sum_{i=1}^n H_i(E_i - p_i^*)^2$ and $\mathcal{L} = \sum_{i=1}^n H_i(E_i - p_i)^2$.

Note that $(\mathcal{L}^* \leq \mathcal{L},\ \mathcal{L}^* \neq \mathcal{L})$ amounts to $L_k^* \leq L_k$ for every k, with $L_k^* < L_k$ in at least one case.

- Definition 1 and Definition 2 are equivalent (Gilio 1990, 1996).
- The two (strengthened) notions of coherence properly manage conditioning events with zero probability.
- If $\mathcal{P}$ is coherent, then $\mathcal{P}$ satisfy the axiomatic properties of a conditional probability. The converse, in general, is not true.

What about a generic proper scoring rule s?

- A generalization of the work of de Finetti to a broad class of scoring rules has been given in (Lindley 1982) where it is shown that the numerical values of the score function, after a suitable transformation, satisfy basic properties of conditional probabilities.

- In (Predd et al., 2009) the relationship between coherence and non-dominance w.r.t. continuous strictly proper scoring rules has been investigated for the case of unconditional events.

- A rich analysis of scoring rules which extends the results obtained in (Predd et al. 2009) to conditional probability assessments has been given in (Schervish et al. 2009) where also the cases of scoring rules which are discontinuous and/or not strictly proper have been examined.

- Moreover, they 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.

Main Result

Definition 3 Let be given a scoring rule s and a probability assessment $\mathcal{P}_n$ on $\mathcal{F}_n$. 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 $\mathcal{L}^* \leq \mathcal{L}$, that is: $L_k^* \leq L_k$, for every k.

Definition 4 We say that $\mathcal{P}_n$ is *admissible w.r.t.* s if $\mathcal{P}_n$ is not weakly dominated by any $\mathcal{P}_n^* \neq \mathcal{P}_n$.

Definition 5 Let be given a scoring rule s and a probability assessment $\mathcal{P}$ on $\mathcal{K}$. We say that $\mathcal{P}$ is admissible w.r.t s if, for every finite subfamily $\mathcal{F}_n \subseteq \mathcal{K}$, the restriction $\mathcal{P}_n$ of $\mathcal{P}$ on $\mathcal{F}_n$ is admissible w.r.t. s.

Our answer to the open question:
coherence and admissibility w.r.t. s are equivalent!

Theorem 1 Let be given a probability assessment $\mathcal{P}$ on a family of conditional events $\mathcal{K}$. The assessment $\mathcal{P}$ is coherent if and only if it is admissible with respect to s, for every bounded (continuous and strictly) proper scoring rule s.

Imprecise Probabilities

We show that the notion of admissibility given for precise assessments can also be trivially exploited in the case of imprecise probabilities.

Definition 6 Let $\mathcal{A}_n = ([l_i, u_i], i = 1, \ldots, n)$ be an interval-valued probability assessment on $\mathcal{F}_n = \{E_i|H_i, i = 1, \ldots, n\}$. 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.

Based on Theorem 1 the Definition 6 can also be given in an equivalent way by replacing the coherence property of $\mathcal{P}_n$ with the property of admissibility w.r.t. a bounded (continuous and strictly) proper scoring rule s.

Proposition 1 Let $\mathcal{A}_n = ([l_i, u_i], i = 1, \ldots, n)$ be an interval-valued probability assessment on $\mathcal{F}_n = \{E_i|H_i, i = 1, \ldots, n\}$. We have that:

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.

Visit our Poster and read our paper.

Thank you for your attention

15 THE FIRST AND SECOND AXIOMS

The entire treatment that we have given was based on a small number of properties, which were justified in the appropriate place in the text as conditions of coherence. In order to develop the theory in an abstract manner, it will now suffice to assume these same properties as axioms.

There will be two axioms (the first and the second) dealing with previsions, and a third dealing with conditional previsions. The third one—which is needed in order to extend the validity of the first two to a special case—will be dealt with later (Section 16): we concentrate for the time being on the first two.

Axiom 1: *Non-negativity*: if we *certainly* have $X \geqslant 0$, we must have $\mathbf{P}(X) \geqslant 0$;

Axiom 2: *Additivity* (finite):

$$\mathbf{P}(X + Y) = \mathbf{P}(X) + \mathbf{P}(Y).$$

From these it also follows that

$$\mathbf{P}(aX) = a\mathbf{P}(X), \qquad \inf X \leqslant \mathbf{P}(X) \leqslant \sup X,$$

as well as the (Convexity) condition, which includes Axioms 1 and 2:

(C) *any linear equation (or inequality) between random quantities* X_i *must be satisfied by the respective previsions* $\mathbf{P}(X_i)$; in other words,

$$\text{if we } \textit{certainly} \text{ have } c_1X_1 + c_2X_2 + \ldots + c_nX_n = c \text{ (or } \geqslant c)$$

$$\text{then } \textit{necessarily } c_1\mathbf{P}(X_1) + c_2\mathbf{P}(X_2) + \ldots + c_n\mathbf{P}(X_n) = c \text{ (or } \geqslant c).$$

By taking differences, (C) can be written in an alternative form:

(C') *No linear combination of (fair!) random quantities can be uniformly positive*; in other words, the $\mathbf{P}(X_h)$ must be chosen in such a way that whatever be the given $c_1, c_2, \ldots, c_n$, *there does not exist a* $c > 0$ *such that*

$$c_1(X_1 - \mathbf{P}(X_1)) + c_2(X_2 - \mathbf{P}(X_2)) + \ldots + c_n(X_n - \mathbf{P}(X_n)) \geqslant c$$

certainly holds.

We could put forward as a further (possible) axiom one which consists in excluding the addition of other axioms; i.e. one which considers *admissible*, way—for example (as was done in Chapter 4, in line with the subjectivistic point of view), by means of conditional bets—then the meaning would be retained.

But the theorem which expresses coherence, connecting it to the non-conditional **P** (the theorem of compound probabilities), no longer holds (and neither does the criterion of coherence) if its formulation (Chapter 4, Section 4.2) has to be in terms of the existence of a 'certainly smaller' loss. In order to extend the notions and rules of the calculus of probability to this new case, it is necessary to strengthen the condition of coherence by saying that *the evaluations conditional on H must turn out to be coherent conditional on H* (i.e. under the hypothesis that H turns out to be true). This is automatic if one evaluates $\mathbf{P}(H) \neq 0$, in which case we reduce to the certainty of a loss in the case of incoherence. The loss for $\tilde{H}$ (Chapter 4, Section 4.3) is, in fact, the sum of the squares of $\mathbf{P}(H)$ and $\mathbf{P}(EH)$; but if $\mathbf{P}(H)$, and therefore $\mathbf{P}(EH)$, are zero, this loss is also zero in the case $\tilde{H}$ (which has probability $= 1$, and is, in any case, possible).

Although this strengthening of the condition of coherence might seem obvious, we had better be careful with it. There are several other forms of strengthening of conditions, often considered as 'obvious', which have consequences that lead us to regard them as inadmissible. In this case, however, there do not seem to be any drawbacks of this kind; moreover, the 'nature' of the strengthening of the condition seems more firmly based on fundamental arguments (rather than for conventional or formal reasons, or for 'mathematical convenience') than others we have come across, and to which we shall return later. In any case, we propose to accept the given extension of the notion of coherence, and to base upon it the theory of conditional probability, without excluding, or treating as special in any way, the case in which one makes the evaluation $\mathbf{P}(H) = 0$.

If we wish to base ourselves upon a new axiom, we could express it in the following way:

Axiom 3: *The conditions of coherence (Axioms* 1 *and* 2) *must be satisfied, also, by the* $\mathbf{P}_H$ *conditional on a possible* H, where

$$\mathbf{P}_H(E) = \mathbf{P}(E|H), \qquad \mathbf{P}_H(E|A) = \mathbf{P}(E|AH)$$

is to be understood.

This means that $\mathbf{P}_H$ is the prevision function that we may have ready for the case in which H turns out to be true, and the axioms oblige us to make this possible evaluation in such a way that if it is to have any effect it must be coherent. This is implicit in the previous definition if one makes the evaluation $\mathbf{P}(H) \neq 0$. Axiom 3 obliges us to behave in the same way, simply on the grounds that H is possible and we might find ourselves actually having to behave according to the choice of $\mathbf{P}_H$—even if, in the case in which