Coherent conditional probabilities and proper scoring rules

Angelo Gilio♣ and Giuseppe Sanfilippo♠

♣ Dipartimento di Scienze di Base e Applicate per l'Ingegneria, University of Rome "La Sapienza", Italy

♠ Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli", University of Palermo, Italy

angelo.gilio@sbai.uniroma1.it, sanfilippo@unipa.it

de Finetti's notions of coherence

- Bruno de Finetti introduced two operational definitions of probabilities:
  (1^{st} criterion) coherence based on the betting scheme ;
  (2^{nd} criterion) coherence based on the penalty criterion.
- de Finetti proved the equivalence between the two criterions for unconditional events and random quantities.
- The penalty criterion, based on the Brier quadratic scoring rule, was adopted as the main criterion in case of conditional events.
  "*In order to give definitions of conditional probability and conditional prevision, and as a foundation for rigorous proofs, we choose to base ourselves on the second criterion*" (de Finetti, 1974, vol.1, p. 135).

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

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

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 $\mathcal{H}_n = H_1 \vee \cdots \vee H_n$.

Proper scoring rules

Let p be your degree of belief on the event E and x be the degree of belief on E that You announce publicly. Suppose You are penalized as follows:
if $E=1$, you have to pay $f(x)$; otherwise, if $E=0$ you have to pay $g(x)$. By setting $s(1,x)=f(x)$ and $s(0,x)=g(x)$ your random penalty is

$$s(E,x)=Es(1,x)+(1-E)s(0,x).$$

The function $s(E,x)$ is a (strictly) proper scoring rule if
(a) for every $x,p\in[0,1]$, with $x\neq p$, it holds

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

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

Condition (a) means that your expected penalty is minimized only at $x=p$.
Thus, You are strictly incentivized to tell the truth.
Given a scoring rule s and a conditional event $E|H$ we set

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

Dominance and admissibility w.r.t. s

Given an assessment $\mathcal{P}$ on $\mathcal{K}$ and a subfamily $\mathcal{F}_n = \{E_1|H_1, E_2|H_2,\ldots,E_n|H_n\}\subseteq\mathcal{K}$, let $\mathcal{P}_n=(p_1,p_2,\ldots,p_n)$ be the restriction of $\mathcal{P}$ to $\mathcal{F}_n$. Given any proper scoring rule 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

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

Given the pair $(\mathcal{F}_n,\mathcal{P}_n)$ and a scoring rule s, we denote by L_k the value of $\mathcal{L}$ associated with the constituent C_k, $k=0,1,\ldots,m$; if $\mathcal{H}_n=H_1\vee\cdots\vee H_n\neq\Omega$, we denote by L_0 the loss associated with $C_0=H_1^cH_2^c\cdots H_n^c$. Of course, $L_0=0$.

Definition 2. 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 3. 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. By Definition 3, 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.

Definition 4. 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.

Coherence and admissibility

- 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 5. 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 $\mathcal{L}=\sum_{i=1}^n H_i(E_i-p_i)^2$.
- The equivalence between Definition 1 and Definition 5 has been proved in (Gilio 1990, 1996).
- 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. Notice that in such a paper it is not used the *strengthened* coherence of de Finetti which allows to properly manage *conditioning* events with *zero* probability.

Main result

We prove the equivalence between the notion of (*strengthened*) coherence based on the betting scheme and the notion of coherence based on the penalty criterion in which the Brier quadratic scoring rule is replaced by a generic strictly (bounded) continuous proper scoring rule.

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

Sketch of the proof

Geometrical approach to coherence.

Given the pair $(\mathcal{F}_n,\mathcal{P}_n)$ with each constituent C_k, $k=1,2,\ldots,m$ we associated the point

$$Q_k=(q_{h1},\ldots,q_{kn})\,,\ q_{kj}=\begin{cases}1, & \text{if } C_h\subseteq E_jH_j,\\ 0, & \text{if } C_h\subseteq E_j^cH_j,\\ p_j=P(E_j|H_j), & \text{if } C_k\subseteq H_j^c.\end{cases}$$

Denoting by $\mathcal{I}$ the convex hull of the points $Q_1,\ldots,Q_m$ the following characterization of the notion of coherence for conditional events can be proved (Gilio 1990, 1995)

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

For every $\mathcal{F}_n\subseteq\mathcal{K}$ coherence of $\mathcal{P}$ requires that $\mathcal{P}_n\in\mathcal{I}$, that is $\mathcal{P}_n=\sum_k\lambda_kQ_k$, with $\lambda_k\geq0$ and $\sum_k\lambda_k=1$. Then for any $\mathcal{P}_n^*\neq\mathcal{P}_n$ it can be proved that $\sum_k\lambda_kL_k<\sum_k\lambda_kL_k^*$, which implies $L_k<L_k^*$ for at least an index k.
Then coherence of $\mathcal{P}$ implies that $\mathcal{P}$ is admissible.

The function $s(p,x)$

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

$$s(p,x)=p\,s(1,x)+(1-p)\,s(0,x)\,.$$

We have

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

In case $p=P(E|H)$ it is $s(p,x)=\mathbb{P}[s(E|H,x)\,|\,H]$.

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 holds

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

Bregman Divergence

Given two vectors

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

we set $S(V_n,\mathcal{P}_n)=\sum_{i=1}^n s(v_i,p_i)$. By exploiting the properties of $s(p,x)$ we have:

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

where $\Phi(\mathcal{P}_n)=-S(\mathcal{P}_n,\mathcal{P}_n)$, is a strictly convex function, differentiable in $(0,1)^n$. The Bregman divergence (see e.g. Censor and Zenios, 1997) corresponding to Φ defined on $[0,1]^n\times[0,1]^n$ is given by

$$\begin{aligned} d_\Phi(V_n,\mathcal{P}_n) &= \Phi(V_n)-\Phi(\mathcal{P}_n)-\nabla\Phi(\mathcal{P}_n)\cdot(V_n-\mathcal{P}_n) = \\ &= S(V_n,\mathcal{P}_n)-S(V_n,V_n)\,. \end{aligned} \qquad (1)$$

It is $d_\Phi(V_n,\mathcal{P}_n)\geq0$ and $d_\Phi(V_n,\mathcal{P}_n)=0$ if and only if $V_n=\mathcal{P}_n$.

Given $\mathcal{I}\subseteq[0,1]^n$, for each $\mathcal{P}_n\in[0,1]^n\setminus\mathcal{I}$, 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, for all $V_n\in\mathcal{I}$, $\mathcal{P}_n\in[0,1]^n\setminus\mathcal{I}$ it is

$$d_\Phi(V_n,\mathcal{P}_n^*)+d_\Phi(\mathcal{P}_n^*,\mathcal{P}_n)\ \leq\ d_\Phi(V_n,\mathcal{P}_n)\,, \qquad (2)$$

By recalling (1), the value L_k of the penalty $\mathcal{L}$ associated with $(\mathcal{F}_n,\mathcal{P}_n)$ is given by

$$L_k=S(Q_k,\mathcal{P}_n)-S(Q_k,Q_k)+\alpha_k=d_\Phi(Q_k,\mathcal{P}_n)+\alpha_k\,,$$

where α_k depends only on $\mathcal{F}_n$ (and not on $\mathcal{P}_n$).
If $\mathcal{P}$ is not coherent, there exists $\mathcal{F}_n$ such that $\mathcal{P}_n\notin\mathcal{I}$; then considering the projection $\mathcal{P}_n^*$ of $\mathcal{P}_n$ onto $\mathcal{I}$ and its associated points Q_k^*, for every $k=1,2\ldots,m$, it holds

$$L_k^*-\alpha_k=d_\Phi(Q_k^*,\mathcal{P}_n^*)\leq d_\Phi(Q_k,\mathcal{P}_n^*)<d_\Phi(Q_k,\mathcal{P}_n)=L_k-\alpha_k.$$

Hence $\mathcal{P}$ would be not admissible.
Therefore, if $\mathcal{P}$ is admissible then $\mathcal{P}$ is coherent.

References

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

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

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

[9] 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.

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

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

[12] 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.