Independence for decomposable information measures

Giuseppe Busanello
Dip. Matematica e Informatica
Università di Perugia
busanello@dipmat.unipg.it

Giulianella Coletti
Dip. Matematica e Informatica
Università di Perugia
coletti@dipmat.unipg.it

Barbara Vantaggi
Dip. Metodi e Modelli Matematici
Università "La Sapienza" Roma
vantaggi@dmmm.uniroma1.it

Abstract

We deal with information measures and we introduce a suitable notion of conditional independence by studying the main properties.

Keywords: Information measure, unconditional independence.

1 Introduction

In Weiner-Shannon information theory the concept of information is simply derived by the classical concept of stochastic independence for probability by defining independent two events A and B with respect to an information measure I if

$$I(A \wedge B) = I(A) + I(B).$$

This definition induces all the problems of the classical independence in probability theory (see [6, 8, 11]): stochastic independent events with 0 or 1 probability can be logically dependent. This is contrary to the intuitive idea of independence. In fact, by using the above notion we have that if $I(A) = \infty$, then A is independent of every event B, even though the two events are not logically independent (for instance are incompatible or A implies B). A similar problem arises when $I(A \wedge B) = 0$. To avoid these problems a stronger condition of independence has been proposed both in probability theory [6, 8], and in possibility theory [9, 10, 12]. In the generalized information theory, introduced by Kampé de Fériet [13] the problem of introducing a "good" definition of independence has been discussed, starting from the seminal works [15, 16]. Most definitions directly generalize that of Weiner-Shannon information theory, by replacing $+$ with a more general operation $\oplus$, satisfying some "natural" conditions. A different approach has been proposed by Benvenuti in [1], where an axiomatic definition of conditional information measure is introduced and so the independence is defined by the formula:

$$I(A|B) = I(A).$$

Nevertheless, all the definitions present in the literature are not able to avoid the counter-intuitive questions discussed above, related to the connection between independence and logical independence for the events having infinite or null measure of information.

Starting from the general concept of conditional $(\oplus, \odot)$-decomposable information measure given in [4, 5] and the relevant characterization in terms of a class of unconditional information measures, we introduce a new independence definition (inspired to that given in [6] for probabilities and in [10, 12] for possibilities) for such measures, which circumvents the above critical situations. In fact, our definition is such that if A is independent of B, then A and B are logically independent. Moreover, we provide a characterization and we compare our notion with those present in literature.

2 Kampé de Fériet information measures

We recall the definition of (generalized) information measure given by Kampé de Fériet

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 899–906
Torremolinos (Málaga), June 22–27, 2008

and Forte [13].

Definition 1 *A function I on an algebra $\mathcal{B}$ taking values on $R^* = [0, +\infty]$ is an information measure if it is antimonotone,* i.e. *the following condition holds: for every $A, B \in \mathcal{B}$*

$$A \subseteq B \implies I(B) \leq I(A). \tag{1}$$

So, it follows that

$$0 \leq I(\Omega) = \inf_{A \in \mathcal{B}} I(A) \leq \sup_{A \in \mathcal{B}} I(A) = I(\emptyset).$$

Kampé de Fériet[13] claims that the above inequality (1) is necessary and sufficient to build up an information measure; nevertheless, to attribute a universal value to $I(\Omega)$ and $I(\emptyset)$, the further conditions $I(\emptyset) = +\infty$ and $I(\Omega) = 0$ are given. The choice of these two values is obviously aimed at reconciling with the Wiener-Shannon theory. In general, condition (1) implies only that

$$I(A \vee B) \leq \min\{I(A), I(B)\};$$

we can specify the rule of composition by introducing a binary operation $\odot$ to compute $I(A \vee B)$ by means of $I(A)$ and $I(B)$.

Definition 2 *An information measure defined on an additive set of events $\mathcal{A}$ is $\odot$-decomposable if there exists a binary operation $\odot$ on $[0, +\infty] \times [0, +\infty]$ such that, for every $A, B \in \mathcal{A}$ with $A \wedge B = \emptyset$, we have*

$$I(A \vee B) = I(A) \odot I(B).$$

So "min" and the rule of Wiener-Shannon theory[1] are only two of the possible choices of $\odot$.

3 Conditional decomposable information measure

In [4, 5], starting from a general concept of conditional event in which the "third" value is not the same for *all* conditional events, but depends on $E|H$, the axioms defining a generalized $(\odot, \oplus)$-decomposable conditional information measure have been introduced.

Let $\odot$ and $\oplus$ be two commutative, associative, and increasing operations from $[0, +\infty]^2$ to $[0, +\infty]$, having respectively $+\infty$ and 0 as neutral elements and with $\oplus$ distributive with respect to $\odot$ (for instance min and +).

Definition 3 *A function $I : \mathcal{C} \to [0, +\infty]$, with $\mathcal{C} = \mathcal{B} \times \mathcal{H}$, $\mathcal{B}$ a Boolean algebra, $\mathcal{H} \subseteq \mathcal{B}$ an additive set not containing $\{\emptyset\}$, is a* conditional *$(\odot, \oplus)$*-decomposable information measure *if it satisfies the following conditions*

(I1) *$I(E|H) = I(E \wedge H|H)$, for every $E \in \mathcal{B}$ and $H \in \mathcal{H}$,*

(I2) *for any given $H \in \mathcal{H}$ $I(\cdot|H)$ is a $\odot$-decomposable information measure, i.e.*

$$I(\Omega|H) = 0 \ , \ I(\emptyset|H) = +\infty,$$

and, for any E, $A \in \mathcal{B}$ with $A \wedge E \wedge H = \emptyset$, one has

$$I((E \vee A)|H) = I(E|H) \odot I(A|H)$$

(I3) *for every $A \in \mathcal{B}$ and E, H, $E \wedge H \in \mathcal{H}$,*

$$I((E \wedge A)|H) = I(E|H) \oplus I(A|(E \wedge H)).$$

Remark 1 *We note that, for a given $\odot$, the choice of operation $\oplus$ is not free (for instance, in the case of Wiener-Shannon information measure, we can prove that* the only possible *choice for $\oplus$ is +). The constraints are given by the requirement of distributivity and, obviously, by the axioms. Nevertheless, if for instance we choose $\odot$ = min, then we have many different possible choices for $\oplus$.*

In the rest of the paper we assume that $\oplus$ is strictly increasing.

As proved for conditional decomposable uncertainty measures [3, 7, 8, 10], conditional information measures can be characterized in terms of a nested class of information measures [4, 5]: any nested class of information measures induces a conditional information measure and any conditional $(\odot, \oplus)$-decomposable information measure gives rise to a nested class of information measures.

Definition 4 *Let $\mathcal{B}$ be a finite algebra and $\mathcal{C}_o$ the set of atoms in $\mathcal{B}$. The class $\mathcal{I} = \{I_o, ..., I_k\}$ of information measures, defined on $\mathcal{B}$ is nested if (for $j = 0, ..., k$) the following conditions hold:*

[1] $x \odot y = -c \log[e^{-x/c} + e^{-y/c})]$

- $I_j(H^j) = 0$;
- $I_j(A) = +\infty$ *for* $A \wedge H^j = \emptyset$;
- $I_j(A) = I_j(A \wedge H^j)$ *for* $A \wedge H^j \neq \emptyset$;
- *for any* $C \in \mathcal{C}_o$ *there exists a (unique)* j *such that* $I_j(C) < +\infty$,

where $H^0 = \Omega$ *and, for* $j = 1, ..., k$, $\mathcal{C}_j = \{C \in \mathcal{C}_{j-1} : I_{j-1}(C) = +\infty\}$ *and* $H^j = \bigvee_{C \in \mathcal{C}_j} C$.

The concept of coherence, developed for uncertainty measures (starting from de Finetti [11] relatively to probability), is the tool to manage partial assessments, which are defined in arbitrary sets of (conditional) events (i.e. sets without particular Boolean structure). Coherence requires that some numbers associated to some events are, in fact, the restriction of a specific uncertainty measure on a Boolean algebra or (for conditional events) on a product of an algebra and an additive set. In [4, 5] for conditional $(\odot, \oplus)$-decomposable information measures, coherence has been introduced and characterized in terms of the existence of a class of unconditional $\odot$-decomposable information measures and in terms of a solvability of a sequence of systems.

Definition 5 *Let* $\mathcal{E}$ *be an arbitrary set of conditional events and* $I : \mathcal{E} \rightarrow [0, +\infty]$. *The function* I *is a* coherent *conditional* $(\odot, \oplus)$*-decomposable information assessment iff it can be extended on* $\mathcal{B} \times \mathcal{H} \supset \mathcal{E}$ *(with* $\mathcal{B}$ *a Boolean algebra,* $\mathcal{H} \subseteq \mathcal{B}$ *additive set not containing* $\emptyset$*) as a conditional* $(\odot, \oplus)$*-decomposable information measure.*

Theorem 1 *Let* $\mathcal{F} = \{E_1|H_1, ..., E_n|H_n\}$ *be a finite set of conditional events,* $\mathcal{C}_o$ *and* $\mathcal{B}$ *denote, respectively, the set of atoms and the algebra generated by* $\{E_1, H_1, ..., E_n, H_n\}$.
For a real function $I : \mathcal{F} \rightarrow [0, +\infty]$ *the following two statements are equivalent:*

a) I *is a coherent conditional* $(\odot, \oplus)$*-decomposable information assessment on* $\mathcal{F}$*;*

b) there exists (at least) a nested class $\mathcal{I} = \{I_o, ..., I_k\}$ *of information measures on* $\mathcal{B}$*, such that for any* $E_i|H_i \in \mathcal{F}$, $I(E_i|H_i)$ *is solution of the equations*

$$I_\alpha(E_i \wedge H_i) = x \oplus I_\alpha(H_i) \qquad (2)$$

for $\alpha = 0, ..., j_{H_i}$ *and* $I(E_i|H_i)$ *is the* unique *solution of equation (2) for* $\alpha = j_{H_i}$ *with* $I_{j_{H_i}}(H_i) < +\infty$*;*

c) there exists a sequence of compatible systems S_α *(with* $\alpha = 0, ..., k$*), with unknowns* $x_r^\alpha = I_\alpha(C_r)$

$$S_\alpha = \begin{cases} \bigodot_{C_r \subseteq E_i \wedge H_i} x_r^\alpha = I(E_i|H_i) \oplus \bigodot_{C_r \subseteq H_i} x_r^\alpha \\ \qquad \text{if } \bigodot_{C_r \subseteq H_i} \mathbf{x}_r^{\alpha-1} = +\infty \\ \bigodot_{C_r \subseteq H_0^\alpha} x_r^\alpha = 0 \\ x_r^\alpha \geq 0 \end{cases}$$

where $\mathbf{x}^{\alpha-1}$ *(with* r*-th component* $\mathbf{x}_r^{\alpha-1}$*) is solution of the system* $S_{\alpha-1}$ *and* $\mathbf{x}_r^{-1} = +\infty$ *for any* $C_r \in \mathcal{C}_o$*. Moreover*

$$H_0^\alpha = \bigvee_{H_i : \bigodot_{C_r \subseteq H_i} \mathbf{x}_r^{\alpha-1} = +\infty} H_i.$$

The nested class in condition *b)* of Theorem 1 is said to agree with the coherent conditional information assessment.

In the sequel of the paper we call a coherent $(\odot, \oplus)$-decomposable information assessment briefly conditional information assessment.

4 Independence

We propose a new independence definition for coherent conditional information assessments, inspired to that given in [6] for conditional probabilities and in [9, 10, 12] for some T-conditional possibilities, which circumvents the above critical situations.

Definition 6 *Let* $\mathcal{F} = \{E_1|H_1, ...E_n|H_n\}$ *be an arbitrary family of conditional events and let* $\mathcal{B}$ *be the algebra spanned by* $\mathcal{U}_\mathcal{F} = \{E_1, H_1, ...E_n, H_n\}$ *and* $\mathcal{B}^0 = \mathcal{B} \setminus \{\emptyset\}$. *Let* I *be a coherent conditional information assessment on* $\mathcal{F}$*, and* $\mathcal{I}$ *be a nested class agreeing with* I*, then, for every event* $E \in \mathcal{B}^0$*, the significant-layer of* E *(denoted as* $\circ(E)$*) related to* $\mathcal{I}$ *is defined as the minimum number* α *such that* $I_\alpha(E) < +\infty$.

Moreover, define $\circ(\emptyset) = \infty$.

We can define also the significant-layer of a conditional event $E|H$ (denoted as $\circ(E|H)$), related to an agreeing class $\mathcal{I}$, as the (nonnegative) number

$$\circ(E|H) = \circ(E \wedge H) - \circ(H). \quad (3)$$

Note that $\circ(E|H) = \infty$ iff $E \wedge H = \emptyset$.

Significant-layers single-out a partition of the algebra $\mathcal{B}$, in particular it follows that the significant-layer of any event E with finite information measure (e.g. Ω) is zero (i.e. $\circ(E) = 0$). It is immediate to prove that the significant-layers, related to an agreeing class $\mathcal{I} = \{I_0, ..., I_k\}$ of I, satisfy the following formal properties: for every $E, F \in \mathcal{B}$

$$\circ(E \vee F) = \min\{\circ(E), \circ(F)\}$$

$$\circ(E \wedge F) \geq \max\{\circ(E), \circ(F)\}.$$

These properties recall those related to the notion of zero-layer given in de Finetti conditional probability framework [8], the significant-layer given for T-conditional possibility [10, 12] and the κ functions of Spohn [17].

Moreover, by properties of conditional information measure, for any conditioning event H there is at least one atom $C \subseteq H$ such that $\circ(C|H) = 0$.

We introduce now a definition of independence, directly for coherent partial information assessment defined on an arbitrary set of conditional events.

Definition 7 *Given two events A and B (with $\emptyset \neq B \neq \Omega$) and a coherent conditional information assessment I on a set $\mathcal{G}$ containing $\mathcal{D} = \{A^*|B^*, B^*|A^*\}$, then A is independent of B under I (in symbol $A \perp\!\!\!\perp B\,[I]$) if both the following conditions hold:*
(i) $I(A|B) = I(A|B^c)$, $I(A^c|B) = I(A^c|B^c)$;
(ii) for the restriction of I to $\mathcal{D}$ there exists an agreeing class $\mathcal{I} = \{I_\alpha\}$ such that

$$\circ(A|B) = \circ(A|B^c) \text{ and } \circ(A^c|B) = \circ(A^c|B^c).$$

Remark 2 *Definition 7 requires for the statement $A \perp\!\!\!\perp B\,[I]$ that $B \neq \emptyset$ and $B \neq \Omega$ (since conditioning events cannot be impossible). This syntactical constraint has also a semantical counterpart: Ω and $\emptyset$ correspond to a situation of complete information (since the former is always true and the latter always false), and so it does not make sense to ask whether they could "influence" the information measure of another event.*

Conversely, by definition it follows that, under any coherent conditional information assessment, the events Ω and $\emptyset$ are independent of every possible (i.e. different from $\emptyset$ and Ω) event B. In fact, condition (i) holds and, for any agreeing class, one has $\circ(\Omega|B) = \circ(\Omega|B^c) = 0$ and $\circ(\emptyset|B) = \circ(\emptyset|B^c) = +\infty$.

This conclusion is natural, since the information measure (0 and $+\infty$, respectively) of Ω and $\emptyset$ cannot be changed by assuming the occurrence of any other possible event B. This is the first instance of a lack of symmetry in this concept of independence.

In condition (i) of Definition 7 we require both the equalities $I(A|B) = I(A|B^c)$ and $I(A^c|B) = I(A^c|B^c)$ and not just one as for instance in probability theory. This is due to the fact that for the information measures $I(A)$ and $I(A^c)$ are not strictly linked (as, e.g., by additivity) and so we can have, for instance: $I(A|B) = I(A|B^c) = 0$, but $I(A^c|B) \neq I(A^c|B^c)$.

When $I(A|B) = I(A^c|B) < +\infty$, condition (i) of Definition 7 assures that A is independent of B, in fact in this case all the significant-layers in condition (ii) are equal to 0 and so condition *(ii)* is trivially satisfied.

On the contrary when $I(A|B) = +\infty$, then necessarily $I(A^c|B) = 0$, and so if condition (i) holds, the independence statement $A \perp\!\!\!\perp B$ under I is ruled by the first equality under condition (ii) (analogously $I(A^c|B) = +\infty$).

We finally note that the statement $A \perp\!\!\!\perp B\ [I]$ depends only on the restriction of the assessment I on $\mathcal{D}$, hence the statement is not effected by the values of the assessment I on $\mathcal{G} \setminus \mathcal{D}$. Since (ii) depends on an agreeing class for the coherent conditional information assessment I, and since this class is in general not unique, it is necessary to prove that independence is well-defined by Definition 7, in other words that it is invariant with respect to the choice of the nested class. This is proved by the following theorems.

Theorem 2 *Let A and B be two logically independent events, and let I be a coherent conditional information assessment defined on $\mathcal{G}$ containing $\mathcal{D} = \{A^*|B^*, B^*|A^*\}$, such that*

$$I(A|B) = I(A|B^c) \text{ and } I(A^c|B) = I(A^c|B^c).$$

If there exists an agreeing class for $I_{|\mathcal{D}}$ such that

$$\circ(A|B) = \circ(A|B^c),\ \circ(A^c|B) = \circ(A^c|B^c) \quad (4)$$

then condition (4) holds for any other agreeing class for $I_{|\mathcal{D}}$.

Proof: Let $C_1 = (A \wedge B)$, $C_2 = (A \wedge B^c)$, $C_3 = (A^c \wedge B)$ and $C_4 = (A^c \wedge B^c)$ be the atoms generated by the events A and B, we put $x_j^i = I_i(C_j)$ with $I_i \in \mathcal{I}$. One has the following situations:

a. If $I(A|B) = I(A^c|B) < +\infty$, then, as discussed above condition (4), $\circ(A|B) = \circ(A|B^c) = \circ(A^c|B) = \circ(A^c|B^c) = 0$, for any agreeing class for $I_{|\mathcal{D}}$.

b. Suppose that $I(A^c|B) = I(A^c|B^c) = +\infty$ (so $I(A|B) = I(A|B^c) = 0$), then

$$S_0 = \begin{cases} x_1^0 = 0 \oplus (x_1^0 \odot x_2^0) \\ x_1^0 = I(B|A) \oplus (x_1^0 \odot x_2^0) \\ x_2^0 = 0 \oplus (x_2^0 \odot x_4^0) \\ x_2^0 = I(B^c|A) \oplus (x_1^0 \odot x_2^0) \\ x_3^0 = +\infty \oplus (x_1^0 \odot x_3^0) \\ x_3^0 = I(B|A^c) \oplus (x_3^0 \odot x_4^0) \\ x_4^0 = +\infty \oplus (x_2^0 \odot x_4^0) \\ x_4^0 = I(B^c|A^c) \oplus (x_3^0 \odot x_4^0) \\ x_1^0 \odot x_2^0 \odot x_3^0 \odot x_4^0 = 0 \end{cases}$$

Therefore, $x_1^0 \odot x_2^0 = 0$ and $x_3^0 = x_4^0 = +\infty$. Then, $\circ(A|B) = \circ(A|B^c) = 0$

1. If $x_1^0 < +\infty$, $x_2^0 < +\infty$, then

$$S_1 = \begin{cases} x_3^1 = I(B|A^c) \oplus (x_3^1 \odot x_4^1) \\ x_4^1 = I(B^c|A^c) \oplus (x_3^1 \odot x_4^1) \\ x_3^1 \odot x_4^1 = 0 \end{cases}$$

Therefore, $x_3^1 = I(B|A^c)$ and $x_4^1 = I(B^c|A^c)$.

   - If $x_3^1 = I(B|A^c) = +\infty$, then $x_4^1 = I(B^c|A^c) = 0$ and $\circ(A^c|B) \geq 2 \neq \circ(A^c|B^c) = 1$. Hence, for $I(B|A^c) = +\infty$ (so $I(B^c|A^c) = 0$) there is no agreeing class for $I_{|\mathcal{D}}$ satisfying (4).
   - If $x_4^1 = I(B^c|A^c) = +\infty$, then $\circ(A^c|B^c) \geq 2 \neq \circ(A^c|B) = 1$; and for $I(B^c|A^c) = +\infty$ there is no agreeing class for $I_{|\mathcal{D}}$ satisfying (4).
   - If $x_3^1 = I(B|A^c) < +\infty$ and $x_4^1 = I(B^c|A^c) < +\infty$, then $\circ(A^c|B) = \circ(A^c|B^c) = 1$, so any agreeing class for $I_{|\mathcal{D}}$ satisfies (4).

2. If $x_1^0 = +\infty$ then $x_2^0 = 0$, $I(B|A) = +\infty$ (so $I(B^c|A) = 0$) and then

$$S_1 = \begin{cases} x_3^1 = +\infty \oplus (x_1^1 \odot x_3^1) \\ x_3^1 = I(B|A^c) \oplus (x_3^1 \odot x_4^1) \\ x_4^1 = I(B^c|A^c) \oplus (x_3^1 \odot x_4^1) \\ x_1^1 \odot x_3^1 \odot x_4^1 = 0 \end{cases}$$

Then $x_1^1 \odot x_4^1 = 0$ and $x_3^1 = +\infty$.

   - If $x_1^1 < +\infty$ and $x_4^1 < +\infty$ then $I(B|A^c) = +\infty$ and $I(B^c|A^c) < +\infty$. Hence, $\circ(A^c|B) = \circ(A^c|B^c) = 1$.
   - If $x_1^1 = 0$ and $x_4^1 = +\infty$ then

$$S_2 = \begin{cases} x_3^2 = I(B|A^c) \oplus (x_3^2 \odot x_4^2) \\ x_4^2 = I(B^c|A^c) \oplus (x_3^2 \odot x_4^2) \\ x_3^2 \odot x_4^2 = 0 \end{cases}$$

   Therefore, $x_3^2 = I(B|A^c)$ and $x_4^2 = I(B^c|A^c)$.

     * If $x_3^2 = I(B|A^c) = +\infty$, then $x_4^2 = I(B^c|A^c) = 0$ and $\circ(A^c|B) = \circ(A^c|B^c) = 2$. Hence, for $I(B|A^c) = +\infty$ under all the agreeing class for $I_{|\mathcal{D}}$ the condition (4) holds.
     * If $x_4^2 = I(B^c|A^c) = +\infty$, then $x_3^2 = I(B|A^c) = 0$ and $\circ(A^c|B^c) \geq 3 \neq \circ(A^c|B) = 1$. Hence, for $I(B^c|A^c) = +\infty$ there is no agreeing class for $I_{|\mathcal{D}}$ satisfying (4).
     * If $x_3^2 = I(B|A^c) < +\infty$ and $x_4^2 = I(B^c|A^c) < +\infty$ then $\circ(A^c|B) = 1 \neq \circ(A^c|B^c) = 2$. Hence, for $I(B|A^c) < +\infty$ and

$I(B^c|A^c) < +\infty$ there is no agreeing class for $I_{|\mathcal{D}}$ satisfying (4).

- If $x_1^1 = +\infty$, then $x_4^1 = 0$, $I(B|A^c) = +\infty$, $I(B^c|A^c) = 0$.

$$S_2 = \begin{cases} x_1^2 = 0 \oplus (x_1^2 \odot x_3^2) \\ x_3^2 = +\infty \oplus (x_1^2 \odot x_3^2) \\ x_1^2 \odot x_3^2 = 0 \end{cases}$$

Therefore, $x_1^2 = 0$ and $x_3^2 = +\infty$. So, $\circ(A^c|B) = \circ(A^c|B^c) = 1$. Thus when $I(B|A) = I(B|A^c)$ all the agreeing classes for $I_{|\mathcal{D}}$ satisfy the condition (4).

3. If $x_2^0 = +\infty$, then $x_1^0 = 0$.

$$S_1 = \begin{cases} x_2^1 = 0 \oplus (x_2^1 \odot x_4^1) \\ x_3^1 = I(B|A^c) \oplus (x_3^1 \odot x_4^1) \\ x_4^1 = +\infty \oplus (x_2^1 \odot x_4^1) \\ x_4^1 = I(B^c|A^c) \oplus (x_3^1 \odot x_4^1) \\ x_2^1 \odot x_3^1 \odot x_4^1 = 0 \end{cases}$$

Therefore, $x_2^1 \odot x_3^1 = 0$ and $x_4^1 = +\infty$.

- If $x_2^1 = +\infty$, then $x_3^1 = 0$, $I(B^c|A^c) = +\infty$, $I(B|A^c) = 0$.

$$S_2 = \begin{cases} x_2^2 = 0 \oplus (x_2^2 \odot x_4^2) \\ x_4^2 = +\infty \oplus (x_2^2 \odot x_4^2) \\ x_2^2 \odot x_4^2 = 0 \end{cases}$$

Then, $x_2^2 = 0$, $x_4^2 = +\infty$ and $\circ(A^c|B) = \circ(A^c|B^c) = 1$.

- If $x_3^1 = +\infty$ then $x_2^1 = 0$.

$$S_2 = \begin{cases} x_3^2 = I(B|A^c) \oplus (x_3^2 \odot x_4^2) \\ x_4^2 = I(B^c|A^c) \oplus (x_3^2 \odot x_4^2) \\ x_3^2 \odot x_4^2 = 0 \end{cases}$$

Therefore $x_3^2 = I(B|A^c)$ and $x_4^2 = I(B^c|A^c)$.

  * If $x_3^2 = +\infty$ then $x_4^2 = 0$ and $\circ(A^c|B) = 3 \neq \circ(A^c|B^c) = 1$. Hence, for $I(B|A^c) = +\infty$ there is no agreeing class for $I_{|\mathcal{D}}$ satisfying (4).
  * If $x_4^2 = +\infty$ then $x_3^2 = 0$ and $\circ(A^c|B) = \circ(A^c|B^c) = 2$.
  * If $x_3^2 < +\infty$ and $x_4^2 < +\infty$ then $\circ(A^c|B) = 2 \neq \circ(A^c|B^c) = 1$. Hence, for $I(B|A^c) < +\infty$ and $I(B^c|A^c) < +\infty$ there is no agreeing class for $I_{|\mathcal{D}}$ satisfying the condition (4).

- If $x_2^1 < +\infty$ and $x_3^1 < +\infty$ then $I(B^c|A^c) = +\infty$ and $\circ(A^c|B) = \circ(A^c|B^c) = 1$.

Hence, when $I(B^c|A^c) = I(B^c|A) = +\infty$ under any agreeing class the condition (4) holds.

c. In the case $I(A|B) = I(A|B^c) = +\infty$ and $I(A^c|B) = I(A^c|B^c) = 0$ the proof goes in the same line of case **b.** by changing the role of the event A with that of A^c.

Theorem 2 proves the invariance of independence with respect to the choice of the agreeing class under the hypothesis that A and B are logically independent. The case of $A = \emptyset$ (or $A = \Omega$) have been discussed in Remark 2. Theorem 3 proves that events not logically independent cannot be independent for any choice of I. This result is important per se and represents the main goal of the reinforcement of independence by condition *(ii)* of Definition 7.

Theorem 3 *Let A and B be two possible events, if A and B are not logically independent, then under any given conditional information measure I the event A is not independent of B.*

Proof: Since A and B are not logically independent, at least one of the following events $A \wedge B$, $A \wedge B^c$, $A^c \wedge B$ and $A^c \wedge B^c$ is impossible. Suppose without loss of generality $A \wedge B = \emptyset$ $(A \subseteq B^c)$, then for every coherent conditional information assessment on $\mathcal{G}$ and for every agreeing class $\mathcal{I}$, it follows

$$\circ(A|B) = \circ(A \wedge B) - \circ(B) = \infty$$

while $\circ(A|B^c) = \circ(B^c) - \circ(B^c) = 0$. Then condition (ii) of Definition 7 does not hold.

Remark 3 *As a particular case of Theorem 3 we have that, for every I, the statement $A \perp\!\!\!\perp A$ $[I]$ does not hold. This irreflexivity is a goal, in fact* any event must be dependent on itself. *Nevertheless, the classical independence, i.e.*

$$I(A \wedge B) = I(A) \oplus I(B) \qquad (5)$$

implies that any event A with $I(A) = 0$ is independent of itself under I.

Theorem 3 proves that our independence definition (Definition 7) implies logical independence, while classical independence (5) does not imply logical independence. Actually this implication is guaranteed by the requirement of *(ii)* in Definition 7.

In the sequel we study some relevant properties of our independence definition: in particular, the following result characterizes independence of two events in terms of the values of $I(B|A)$, $I(B^c|A)$, $I(B|A^c)$, $I(B^c|A^c)$, giving up any direct reference to significant-layers.

Theorem 4 *Let A and B be two logically independent events. If a coherent conditional information assessment I is such that*

$$I(A|B) = I(A|B^c) = \alpha$$

$$I(A^c|B) = I(A^c|B^c) = \beta,$$

then $A \perp\!\!\!\perp B\,[I]$ if and only if one (and only one) of the following conditions holds:

(a) $\alpha < +\infty$, $\beta < +\infty$;

(b) $\alpha = 0$, $\beta = +\infty$ *and the extension of I to $I(B|A)$, $I(B^c|A)$, $I(B|A^c)$, $I(B^c|A^c)$ satisfies one of the following conditions:*

- *(1)* $I(B|A^c) \oplus I(B^c|A^c) \oplus I(B|A) \oplus I(B^c|A) < +\infty$.
- *(2)* $I(B|A) = I(B|A^c) = +\infty$.
- *(3)* $I(B^c|A) = I(B^c|A^c) = +\infty$.

(c) $\alpha = +\infty$, $\beta = 0$ *and the extension of I to $I(B|A)$, $I(B^c|A)$, $I(B|A^c)$, $I(B^c|A^c)$ satisfies one of the following conditions:*

- *(1)* $I(B|A^c) \oplus I(B^c|A^c) \oplus I(B|A) \oplus I(B^c|A) < +\infty$.
- *(2)* $I(B|A^c) = I(B|A) = +\infty$.
- *(3)* $I(B^c|A) = I(B^c|A^c) = +\infty$.

Proof: The proof is essentially contained in that of Theorem 2.

By the following result we show the connections with the independence given in [1].

Proposition 1 $A \perp\!\!\!\perp B\,[I] \Rightarrow I(A|B) = I(A)$.

Proof: Let $I(A|B) = I(A|B^c) = \alpha$ and by distributivity $I(A) = I(A \wedge B) \odot I(A \wedge B^c) = (I(A|B) \oplus I(B)) \odot (I(A|B^c) \oplus I(B^c)) = \alpha \oplus (I(B) \odot I(B^c)) = \alpha$.

The validity of the independence statement $A \perp\!\!\!\perp B$ $[I]$, under a given conditional information measure I, does not guarantee the symmetric independence relation $B \perp\!\!\!\perp A$ $[I]$. However, concerning the symmetry of the independence we have the following result:

Theorem 5 *Let A and B be two possible events. Given a coherent conditional information assessment I such that $A \perp\!\!\!\perp B$ $[I]$, we have:*

1. *if $I(A) = I(A^c) < +\infty$, then $B \perp\!\!\!\perp A$ $[I]$;*
2. *if $I(A) \oplus I(A^c) = +\infty$ and either $I(B|A) = +\infty$ or $I(B^c|A) = +\infty$, then $B \perp\!\!\!\perp A$ $[I]$.*

Proof: If $I(A|B) < +\infty$ then

$I(A \wedge B) = I(A|B) \oplus I(B) = I(B|A) \oplus I(A)$.

Since $\oplus$ is strictly increasing the cancellation law also, so from the previous equation $I(B|A) = I(B)$ and $I(B|A^c) = I(B)$, then $B \perp\!\!\!\perp A$ by Theorem 4.

If $A \perp\!\!\!\perp B$ and $I(A|B) = +\infty = I(B|A)$, then by Theorem 4 it follows that $I(B|A^c) = +\infty$, so $I(B^c|A) = I(B^c|A^c) = 0$. Again, by Theorem 4 (condition *(c)*) it follows $B \perp\!\!\!\perp A$.

The case $I(A|B) = +\infty = I(B^c|A)$ follows analogously.

The proof of the case $I(A^c|B) = +\infty$ is similar to the previous one.

From Theorem 4 and Proposition 1 it follows that $A \perp\!\!\!\perp B$ with $I(A) = +\infty$ cannot necessary imply $B \perp\!\!\!\perp A$ if $I(B|A) \oplus I(B^c|A) < +\infty$. For instance, if we have $A \perp\!\!\!\perp B$ with $I(A|B) = +\infty$, $I(B|A) \oplus I(B^c|A) < +\infty$ and $I(B) \neq I(B|A)$ then by Proposition 1 $B \perp\!\!\!\perp A$ does not hold. In fact, $I(A^c|B) = 0 = I(A^c)$ and $I(A^c \wedge B) = I(A^c|B) \oplus I(B) = I(B|A^c) \oplus I(A^c)$, therefore $I(B|A^c) = I(B) \neq I(B|A)$.

By the following result we show the connection with the classical independence definition.

Proposition 2 *If* $A \perp\!\!\!\perp B$ *under* I, *then* $I(A \wedge B) = I(A) \oplus I(B)$.

Proof: From Proposition 1 one has that $A \perp\!\!\!\perp B\,[I]$ implies that $I(A \wedge B) = I(A|B) \oplus I(B) = I(A) \oplus I(B)$.

References

[1] P. Benvenuti (1977). Sur l'indépendance dans l'information. *Colloques internationaux du C.N.R.S.*, N.276 - Théorie de l'Information, pages 44-55.

[2] B. Bouchon-Meunier, G. Coletti, C. Marsala (2001). Conditional possibility and necessity. In *Technologies for Constructing Intelligent Systems (B. Bouchon-Meunier, J. Gutiérrez-Rios, L. Magdalena, R.R. Yager eds.)* volume 1, Springer, Berlin, 2001.

[3] B. Bouchon-Meunier, G. Coletti, C. Marsala (2002). Independence and possibilistic conditioning. *Annals of Math. and Artif. Intelligence*, 35, pages 107-124.

[4] B. Bouchon-Meunier, G. Coletti, C. Marsala (2006). A general theory of conditional decomposable information measures, *Proc. of the Conf. IPMU'2006*, Paris, pages 97-104.

[5] B. Bouchon-Meunier, G. Coletti, C. Marsala (2008). Towards a general theory of conditional decomposable information measures, in *Uncertainty and Intel. Information Syst.*, World Scientific, (B. Bouchon-Meunier, et al. Eds.) to appear.

[6] G. Coletti, R. Scozzafava (2000). Zero probabilities in stochastical independence, in: *Information , Uncertainty, Fusion* (Bouchon-Meunier, Yager, Zadeh Eds.), Kluwer, pages 185–196.

[7] G. Coletti, R. Scozzafava (2001). From conditional events to conditional measures: a new axiomatic approach, *Annals of Math. and Artif. Intelligence*, 32, pages 373–392.

[8] G. Coletti, R. Scozzafava (2002). *Probabilistic logic in a coherent setting*, trends in Logic n. 15, Kluwer, Dorbdrecht, Boston, London.

[9] G. Coletti, B. Vantaggi (2004). Independence in conditional possibility theory, *Proc. of the Conf. IPMU'2004*, Perugia, Italy, pages 849–856.

[10] G. Coletti, B. Vantaggi (2006). Possibility theory: conditional independence *Fuzzy Sets and Systems*, 157, pages 1491-1513.

[11] B. de Finetti (1970). *Theory of Probability.* vol. I,II, Willey & Sons, London.

[12] L. Ferracuti and B. Vantaggi (2006). Independence and conditional possibilities for strictly monotone triangular norms. *Inter. Jour. of Intel. Systems*, 21, pages 299-323.

[13] J. Kampé de Fériet (1969). Measure de l'information fournie par un événement. In: *Colloques Internationaux C.N.R.S.* 186, pages 191–221.

[14] J. Kampé de Fériet (1975). L'indépendance des événements dans la théorie géneralisée de l'information *Journées Lyonnaises des questionnaires. C.N.R.S. Groupe de recherche 22*, n.1, pages 1–30.

[15] J. Kampé de Fériet, B. Forte (1967). Information et probabilité. *Comp. Rendus Acad. Sci. Paris*, 265 A, pages 350–353.

[16] J. Kampé de Fériet, B. Forte, P. Benvenuti (1969). Forme général de l'opération de composition continue d'une information. *CRAS*, 269, pages 529–534.

[17] W. Spohn (1994). On the Properties of Conditional Independence, in: *Scientific Phil. 1: Probability and Probabilistic Causality*, Kluwer, Dordrecht (Humphreys, Suppes, Eds), pages 173–194.