ARTICLE TEMPLATE

Detecting Correlation Between Extreme Probability Events

G. Coletti[a], L.C. van der Gaag[b], D. Petturiti[c], and B. Vantaggi[d]

[a] Dept. Mathematics and Computer Science, University of Perugia, Italy;
[b] Dalle Molle Institute for Artificial Intelligence Research, SUPSI, Lugano, Switzerland;
[c] Dept. Economics, University of Perugia, Italy;
[d] Dept. S.B.A.I., "La Sapienza" University of Rome, Italy

ARTICLE HISTORY
Compiled December 12, 2019

ABSTRACT
As classical definitions of correlation give rise to counterintuitive statements when extreme probability events are involved, we introduce enhanced notions of positive and negative correlation in the general framework of coherent conditional probability. These notions allow to handle extreme probability events in a principled way by accommodating the different levels of strength of the zero probabilities involved (namely, zero layers). Since the detection of correlations by means of zero layers is computationally challenging, we provide a full characterization relying on only conditional probability values.

KEYWORDS
Conditional probability; extreme probability event; coherence; correlation.

1. Introduction

Starting from the seminal paper by de Finetti (1936), the importance of handling *extreme probability events* in a principled way has been stressed in a range of papers (see, for example, de Finetti (1949); Coletti and Scozzafava (1998, 2000, 2002b); Coletti, Scozzafava, and Vantaggi (2001); Scozzafava and Vantaggi (2004); Vantaggi (2001)). By an extreme probability event we mean a highly unexpected event, that is, an event of zero probability, or a nearly sure event, of probability 1. Zero probabilities necessarily arise in uncountable algebras (or σ-algebras) and, hence, prevail in real-world applications involving infinite settings, where the lack of expressive power of the real numbers forces possible events to be assigned zero probability. Although less commonly acknowledged, also in finite settings zero probabilities cannot be ignored. A simple artificial example suffices to demonstrate that even strictly positive probability assessments can lead to zero probability being assigned to possible events (Coletti and Scozzafava 2002a). Consider the three events E, F, G, with $G \subseteq E \wedge F$, and the assessment $P(E) = P(F) = P(G) = \frac{1}{2}$. This assessment may come from a uniform probability distribution on $\Omega = [0,1]^2$, with, for example, $E = [0,1[\, \times\,]0, \frac{1}{2}]$, $F =]0,1] \times [0, \frac{1}{2}[$, and $G = (\,]0,1[\, \times\,]0, \frac{1}{2}[\,) \setminus \{(\frac{1}{4}, y)\colon\ y \in [0,1]\}$. Simple computations suffice to show that the only probability measure on the algebra spanned by the three

CONTACT D. Petturiti. Email: davide.petturiti@unipg.it

events assigns zero probability to the events $E \wedge F \wedge G^c$, $E^c \wedge F \wedge G^c$ and $E \wedge F^c \wedge G^c$.

Zero probability events can arise when constructing real-world probabilistic models by extracting conditional probabilities from (necessarily finite) data sets: in such cases we need to deal with unexpected events and events occurring with negligible frequency. Realistic examples are discussed in Coletti and Scozzafava (1998, 2000); Scozzafava (1984); Scozzafava and Vantaggi (2004) and in Coletti and Scozzafava (2002a), and further concrete examples are found in many real-world applications of probabilistic graphical models.

If an event at hand is not a logical impossibility, for practical applications generally some smoothing technique is employed to forestall the inclusion of zero probabilities in the model (see for example Flach (2012)). Various more or less "ad hoc" techniques are in use for this purpose, such as the well-known Laplace correction and the use of pseudocounts in a Bayesian setting (see, e.g., Niblett and Bratko (1987); Russell and Norvig (2003)). The small probability values thus included in a model, have unexpected and often unwanted effects however, as was demonstrated in van der Gaag and Capotorti (2018); Vantaggi (2002, 2003). Moreover, forcing all events to have positive probability drastically restricts the class of admissible distributions and, hence, the possibilities of extending partial assessments to complete probabilities. The foundational framework of coherent conditional probability offers a principled way of handling zero probability without the adverse effects of current techniques for ignoring extreme probability events.

In many applications of probability theory, stochastic independence and the concepts of positive and negative correlation play an important role, as these concepts often allow for the implementation of efficient computation (see for example Pearl (1988); Lauritzen and Spiegelhalter (1988)). In the context of coherent conditional probability, stochastic independence has been well addressed (Coletti and Scozzafava 1998, 2000, 2002b; Vantaggi 2001, 2002) and has led to an enhanced definition of independence, namely *cs-independence*.

The concept of correlation, on the other hand, has received little to no attention in the setting of coherent probability. In this paper, we demonstrate that the classical definitions of correlation can give rise to highly counterintuitive statements for extreme probability events, such as an event E being uncorrelated with an event H logically implying it. Based on these observations, we introduce an enhanced notion of correlation for coherent conditional probability which takes the different levels of strength of the zero probabilities involved into consideration. More specifically, we develop notions of *positive* and *negative correlation* in a coherent setting, namely *cs-correlation*, referring to the characterization of coherent conditional probabilities represented by their agreeing classes of unconditional probabilities which in turn define the zero layers of the events of interest. Since conditional probability is a primitive notion in our framework, we directly define *conditional* cs-correlation, retrieving unconditional cs-correlation as a particular case.

Although the framework of coherent conditional probability constitutes the basis of our notions of correlation, building on the concept of zero layer for establishing the sign of a correlation does not provide for practicable application in real-world settings, as a consequence of the computational challenges involved. We therefore provide also a full characterization of the correlations relying only on conditional probability values.

The paper is organized as follows. Section 2 presents some preliminaries on coherent conditional probability and thereby introduces our notational conventions. In Section 3, we review the concept of (conditional) stochastic independence in a coherent setting. Given this concept of independence, we develop in Section 4 our concepts of (conditional) positive and negative correlation in the framework of coherent probability, and introduce some of their properties in Section 5. Section 6 then provides a full characterization of all correlations involving extreme probability events. Section 7 concludes the paper with our plans for further research.

2. Preliminaries

Building on conditional probability as a primitive concept, we review coherent conditional probability theory and thereby introduce our notational conventions for the remainder of the paper.

2.1. *Conditional probability as a primitive concept*

We consider an *event* to be any fact described by a Boolean sentence, indicating by Ω the *sure event* and using $\varnothing$ for the *impossible event*. For any event E, we will write E^* to indicate either E itself or its contrary E^c. For an event E, we denote by $|E|$ its *indicator*, which is the random number equal to 1 if E is true and to 0 if E is false. We further take a *conditional event* $E|H$ to be an ordered pair of events E, H with $H \neq \varnothing$. In the pair, the two events E and H have the same type, both being Boolean sentences, yet have different roles in the sense that H has the role of hypothesis.

We recall that an *additive class* of events is a set of events closed under taking disjunction $\vee$. A *Boolean algebra* of events is an additive class which is further closed under taking the contrary $(\cdot)^c$, and hence under conjunction $\wedge$, and is partially ordered by the implication relation $\subseteq$.

In the sequel, we will assume our Boolean algebras to be finite. For any Boolean algebra $\mathcal{A}$, we write $\mathcal{A}^0$ to indicate $\mathcal{A} \setminus \{\varnothing\}$. For any sentence $K \in \mathcal{A}^0$, we denote by $\mathcal{A} \wedge K = \{E \wedge K \colon E \in \mathcal{A}\}$ the Boolean ideal with K for its top element. For an arbitrary family of events $\mathcal{E} = \{E_1, \dots, E_n\}$, we use *algebra*$(\mathcal{E})$ to denote the minimal Boolean algebra of events containing $\mathcal{E}$ and *additive*$(\mathcal{E})$ to denote the minimal additive class of events containing $\mathcal{E}$. By *atoms*$(\mathcal{E})$ we indicate the finest partition of Ω contained in *algebra*$(\mathcal{E})$, in particular, the events in $\mathcal{E}$ are said to be *logically independent* if the cardinality of *atoms*$(\mathcal{E})$ is 2^n.

In this paper, we will build on the following axiomatic definition of *conditional probability*, which dates back to de Finetti (1972) and has been explicitly formulated, with minor differences, by Dubins (1975) and Krauss (1968).

Definition 2.1. Let $\mathcal{A}$ be a Boolean algebra of events and let $\mathcal{H}$ be an additive class with $\mathcal{H} \subseteq \mathcal{A}^0$. A **conditional probability** on $\mathcal{A} \times \mathcal{H}$ is a function $P \colon \mathcal{A} \times \mathcal{H} \to [0, 1]$ that satisfies the following conditions:

(i) $P(E|H) = P(E \wedge H|H)$, for every $E \in \mathcal{A}$ and $H \in \mathcal{H}$;
(ii) $P(\cdot|H)$ is a finitely additive probability on $\mathcal{A}$, for every $H \in \mathcal{H}$;
(iii) $P(E \wedge F|H) = P(E|H) \cdot P(F|E \wedge H)$, for every $H, E \wedge H \in \mathcal{H}$ and $E, F \in \mathcal{A}$.

Whenever $\Omega \in \mathcal{H}$, we will write $P(E) = P(E|\Omega)$ for all $E \in \mathcal{A}$. We note that unconditional probability is thus covered by the above definition as a special case.

Following Dubins (1975), we say that a conditional probability $P(\cdot|\cdot)$ is *full* on the algebra $\mathcal{A}$ if it is defined on $\mathcal{A} \times \mathcal{A}^0$, that is, if it is defined on $\mathcal{A} \times \mathcal{H}$ with $\mathcal{H} = \mathcal{A}^0$. Dubins has shown that every conditional probability on $\mathcal{A} \times \mathcal{H}$ with $\mathcal{H} \subset \mathcal{A}^0$ can be extended to a full conditional probability on $\mathcal{A} \times \mathcal{A}^0$, albeit not necessarily in a unique way.

In the sequel, we call *assessment* a function $P : \mathcal{G} \to [0,1]$ with $\mathcal{G}$ an arbitrary set of conditional events and, given $\mathcal{D} \subseteq \mathcal{G}$, $P_{|\mathcal{D}}$ stands for the *restriction* of P on $\mathcal{D}$.

2.2. *Coherent conditional probability*

The introduction of conditional probability as a primitive concept paves the way for conditioning on events with zero probability. We review for this purpose the concept of coherent conditional probability.

Definition 2.2. Let $\mathcal{G} = \{E_i|H_i\}_{i\in I}$, with I a finite index set, be an arbitrary family of conditional events. A **coherent conditional probability** on $\mathcal{G}$ is a function $P : \mathcal{G} \to [0,1]$ consistent with a conditional probability, that is, for which there exists a conditional probability $P' : \mathcal{A} \times \mathcal{H} \to [0,1]$, with $\mathcal{A} = \mathit{algebra}(\{E_i, H_i\}_{i\in I})$ and $\mathcal{H} = \mathit{additive}(\{H_i\}_{i\in I})$, such that $P'_{|\mathcal{G}} = P$.

A function $P(\cdot)$ on a family of unconditional events $\mathcal{E} = \{E_1, \dots, E_n\}$ is a *coherent probability* if it is the restriction of an unconditional probability $P'(\cdot)$ on $\mathit{algebra}(\mathcal{E})$. The above definition of coherent conditional probability thus covers coherent unconditional probability as a special case. Since every conditional probability $P(\cdot|\cdot)$ on $\mathcal{A} \times \mathcal{H}$ can be extended to a full conditional probability on $\mathcal{A}$ (Dubins 1975), the above definition can also be formulated by requiring the existence of a full conditional probability on $\mathcal{A}$ extending the original function P.

For every Boolean algebra of events $\mathcal{A}$ and additive class $\mathcal{H}$, every conditional probability $P(\cdot|\cdot)$ on $\mathcal{A} \times \mathcal{H}$ determines a linearly ordered class $\{H_0^0, \dots, H_0^k\}$ of decreasing elements of $\mathcal{H}$, such that

- $H_0^0 = \bigvee_{H\in\mathcal{H}} H$;
- for $\alpha = 1, \dots, k$, $H_0^\alpha = \bigvee\{H \in \mathcal{H} \colon H \subseteq H_0^{\alpha-1}, P(H|H_0^{\alpha-1}) = 0\}$.

Upon constructing the class $\{H_0^0, \dots, H_0^k\}$ iteratively, the value k is found such that $H_0^k \neq H_0^{k+1} = \varnothing$. The constructed class induces a one-to-one correspondence of the conditional probability $P(\cdot|\cdot)$ with a linearly ordered class $\{P_0, \dots, P_k\}$ of coherent unconditional probabilities, each of which is defined on an element $\mathcal{I}_\alpha$ of the class $\{\mathcal{I}_0, \dots, \mathcal{I}_k\}$ of decreasing Boolean ideals of $\mathcal{A}$, with, for $\alpha = 0, \dots, k$, $\mathcal{I}_\alpha = \mathcal{A} \wedge H_0^\alpha$. Each probability P_α of the class thus is the restriction of the conditional probability $P(\cdot|H_0^\alpha)$ on $\mathcal{I}_\alpha$, with $P((H_0^\alpha)^c|H_0^\alpha) = 0$. Noting that every coherent probability P_α on $\mathcal{I}_\alpha$ is completely determined through additivity by its restriction on $\mathcal{I}_\alpha \cap \mathit{atoms}(\mathcal{A})$, we define the *support* of a coherent probability P_α as $\mathbf{supp}(P_\alpha) = \bigvee\{C_r \in \mathcal{I}_\alpha \cap \mathit{atoms}(\mathcal{A}) : P_\alpha(C_r) > 0\}$. We now have that the construction of the ordered class $\{P_0, \dots, P_k\}$ outlined above implies that the supports of its coherent probabilities P_α are pairwise incompatible. For every event $H \in \mathcal{H}$, in fact, there is a unique index $\alpha_H \in \{0, \dots, k\}$ such that $H \in \mathcal{I}_{\alpha_H}$ and $P_{\alpha_H}(H) > 0$. For every conditional event $E|H \in \mathcal{A} \times \mathcal{H}$, we then have that

$$P(E|H) = \frac{P_{\alpha_H}(E \wedge H)}{P_{\alpha_H}(H)}. \tag{1}$$

Given a conditional probability $P(\cdot|\cdot)$ on $\mathcal{A} \times \mathcal{H}$, there can be more than one linearly ordered class $\{P_0, \dots, P_k\}$ of coherent probabilities defined on a decreasing chain of

Boolean ideals of $\mathcal{A}$ and with pairwise incompatible supports assuring the representation of $P(\cdot|\cdot)$ as in (1). Each such class essentially describes a conditional probability $P'(\cdot|\cdot)$ extending the function $P(\cdot|\cdot)$ to $\mathcal{A} \times \mathcal{H}' \supseteq \mathcal{A} \times \mathcal{H}$, and is termed an *agreeing class* corresponding to $P(\cdot|\cdot)$ in Coletti and Scozzafava (2002a). The specific class of coherent probabilities constructed above is minimal given the additive class $\mathcal{H}$, and hence is called the *$\mathcal{H}$-minimal agreeing class* corresponding to $P(\cdot|\cdot)$ on $\mathcal{A} \times \mathcal{H}$. Given a specific $K \in \mathcal{A}^0$, the $(\mathcal{A} \wedge K)^0$-minimal agreeing class corresponding to $P(\cdot|\cdot)$ on $\mathcal{A} \times (\mathcal{A} \wedge K)^0$ is such that the supports of its coherent probabilities form a partition of K. The class is then called a *complete agreeing class* on $\mathcal{A} \wedge K$. If we take $K = \Omega$ and, hence, consider the full conditional probability $P(\cdot|\cdot)$ on $\mathcal{A} \times \mathcal{A}^0$, then the supports of the coherent probabilities of the corresponding $\mathcal{A}^0$-minimal agreeing class form a partition of Ω. This class will be called a *complete agreeing class* on $\mathcal{A}$ in the following and is in fact the only class giving rise to the property stated in equation (1) for a full conditional probability on $\mathcal{A}$.

The following theorem gives some equivalent characterizations of coherence for a given assessment, relevant to the current paper: for proofs of the stated equivalences, we refer to Coletti (1994); Coletti and Scozzafava (1996, 1999).

Theorem 2.3. *Let $\mathcal{G} = \{E_i|H_i\}_{i \in I}$, with I a finite index set, be an arbitrary family of conditional events. Further, let $\mathcal{A} = algebra(\{E_i, H_i\}_{i \in I})$ and let $\mathcal{H}$ be an additive class such that $additive(\{H_i\}_{i \in I}) \subseteq \mathcal{H} \subseteq \mathcal{A}^0$. Then, for any function $P : \mathcal{G} \to [0,1]$, the following statements are equivalent:*

(i) P is a coherent conditional probability on $\mathcal{G}$.

(ii) P is a de Finetti-coherent conditional probability, that is, for every finite index subset $J \subseteq I$, $\{E_j|H_j\}_{j \in J} \subseteq \mathcal{G}$, and $\lambda_j \in \mathbb{R}$, $j \in J$, one has

$$\max_{C_r \subseteq H_0^0} \left[\sum_{j \in J} \lambda_j \cdot (|E_j| - P(E_j|H_j)) \cdot |H_j| \right] \geq 0,$$

where $H_0^0 = \bigvee_{j \in J} H_j$ and $C_r \in atoms(\{E_j|H_j\}_{j \in J})$.

(iii) There exists a $\mathcal{H}$-minimal agreeing class $\{P_0, \ldots, P_k\}$, $k \geq 0$, corresponding to a conditional probability $P'(\cdot|\cdot)$ on $\mathcal{A} \times \mathcal{H}$ extending P, i.e., such that, for every $i \in I$, it holds that

$$P(E_i|H_i) = \frac{P_{\alpha_{H_i}}(E_i \wedge H_i)}{P_{\alpha_{H_i}}(H_i)}.$$

(iv) Denoting $H_0^0 = \bigvee_{H \in \mathcal{H}} H$ and, for $\alpha = 1, \ldots, k$, $H_0^\alpha = \bigvee \{H \in \mathcal{H} : H \subseteq H_0^{\alpha-1}, P_{\alpha-1}(H) = 0\}$, all systems of equations in the sequence $\{\mathcal{S}_0, \ldots, \mathcal{S}_k\}$, $k \geq 0$, with non-negative unknowns $x_r^\alpha = P_\alpha(C_r)$ for all $C_r \in \mathcal{C}_\alpha = \{C_r \in atoms(\{E_i, H_i\}_{i \in I}) : C_r \subseteq H_0^\alpha\}$, are compatible:

$$\mathcal{S}_\alpha : \begin{cases} \sum\limits_{C_r \subseteq E_i \wedge H_i} x_r^\alpha = P(E_i|H_i) \cdot \sum\limits_{C_r \subseteq H_i} x_r^\alpha, & \textit{for all } i \in I \textit{ with } P_{\alpha-1}(H_i) = 0, \\ \sum\limits_{C_r \in \mathcal{C}_\alpha} x_r^\alpha = 1. \end{cases}$$

Following the terminology of Coletti and Scozzafava (2002a), a $\mathcal{H}$-minimal agreeing class $\{P_0, \ldots, P_k\}$ satisfying condition *(iii)* is said to *agree* with the given assessment P. Let us stress that solving the sequence of systems in point *(iv)* of the previous Theorem 2.3 with $\mathcal{H} = additive(\{H_i\}_{i \in I})$ we get a $\mathcal{H}$-minimal agreeing class corresponding to a conditional probability $P'(\cdot|\cdot)$ extending the assessment P on the minimal structured set of conditional events $\mathcal{A} \times \mathcal{H}$ including $\mathcal{G}$. Since, in turn, such $P'(\cdot|\cdot)$ can be always further extended, generally not in a unique way, to a full conditional probability $P''(\cdot|\cdot)$ on the whole $\mathcal{A} \times \mathcal{A}^0$, the previous sequence of systems can be used to find directly a complete agreeing class on $\mathcal{A}$: it will be sufficient to take $\mathcal{H} = \mathcal{A}^0$.

2.3. *Zero layers*

To conclude our review of the setting of coherent conditional probability, we recall the concept of *zero layer* from Coletti and Scozzafava (2000); Coletti, Scozzafava, and Vantaggi (2001); Coletti and Scozzafava (2002a), which naturally arises from the structure of conditional probability described in Theorem 2.3.

Definition 2.4. Let $\mathcal{A}$ be a Boolean algebra of events and let $\{P_0, \ldots, P_k\}$ be a complete agreeing class on $\mathcal{A}$. For every event $H \in \mathcal{A}^0$, the **zero layer** of H with respect to $\{P_0, \ldots, P_k\}$ is the non-negative number

$$o(H) = \alpha_H,$$

where $\alpha_H \in \{0, \ldots, k\}$ is the unique index such that $H \in \mathcal{I}_{\alpha_H}$ and $P_{\alpha_H}(H) > 0$. The zero layer of the impossible event is set to $o(\emptyset) = +\infty$. For every conditional event $E|H \in \mathcal{A} \times \mathcal{A}^0$, the **zero layer** of $E|H$ with respect to $\{P_0, \ldots, P_k\}$ is the non-negative number

$$o(E|H) = o(E \wedge H) - o(H).$$

We note that for any event E with $P(E) = P(E|\Omega) > 0$, we have that $o(E) = 0$. We further note that $P(E|H) > 0$ iff $o(E \wedge H) = o(H)$ and hence $o(E|H) = 0$. For further properties of zero layers and their role, we refer to, for example, Coletti, Scozzafava, and Vantaggi (2001).

For a given event $K \in \mathcal{A}^0$, the concept of zero layer is readily generalized to the events is $\mathcal{A} \wedge H$ by referring to a complete agreeing class $\{P_0, \ldots, P_k\}$ on $\mathcal{A} \wedge K$ and, hence, to the conditional events in $\mathcal{A} \times (\mathcal{A} \wedge K)^0$ since, for every $E|H \in \mathcal{A} \times (\mathcal{A} \wedge K)^0$, it holds $E \wedge H \in \mathcal{A} \wedge K$ and $H \in (\mathcal{A} \wedge K)^0$.

Example 2.5. Let $\mathcal{A}$ be a Boolean algebra with $atoms(\mathcal{A}) = \{C_1, C_2, C_3, C_4\}$ and let $K = C_1 \vee C_2 \vee C_3$. We consider the complete agreeing class of coherent probabilities $\{P_0, P_1\}$ on $\mathcal{A} \wedge K$ with

$atoms(\mathcal{A})$	C_1	C_2	C_3	C_4
P_0	1	0	0	$\bullet$
P_1	$\bullet$	$\frac{1}{3}$	$\frac{2}{3}$	$\bullet$

The class $\{P_0, P_1\}$ determines a unique conditional probability $P(\cdot|\cdot)$ on $\mathcal{A} \times (\mathcal{A} \wedge K)^0$, along with the zero layers for the corresponding conditional events. We have for example

- $P(C_2|C_1 \vee C_2) = \dfrac{P_0(C_2)}{P_0(C_1 \vee C_2)} = 0$, with the zero layer $o(C_2|C_1 \vee C_2) = o(C_2) -$

$o(C_1 \vee C_2) = 1 - 0 = 1$;

- $P(C_2|C_2 \vee C_3) = \dfrac{P_1(C_2)}{P_1(C_2 \vee C_3)} = \frac{1}{3}$, with the zero layer $o(C_2|C_2 \vee C_3) = o(C_2) - o(C_2 \vee C_3) = 1 - 1 = 0$;
- $P(C_4 \vee C_2|C_1 \vee C_2) = P(C_2|C_1 \vee C_2)$ and $o(C_4 \vee C_2|C_1 \vee C_2) = o(C_2|C_1 \vee C_2)$.

□

3. On conditional cs-independence

Alternative classical definitions of stochastic independence have been studied in detail under de Finetti conditional probabilities (see, for example, Vantaggi (2001, 2002)). In the setting of coherent conditional probability, these alternative definitions are not equivalent (Coletti and Scozzafava 2000; Vantaggi 2001, 2002). The alternative concepts have moreover been shown to lead to counterintuitive statements in the presence of zero probability and have as a consequence been enhanced to the concept of *cs-independence*[1]. This concept of independence was further generalized to *conditional cs-independence* in Vantaggi (2001).

Definition 3.1. Let $\mathcal{G}$ be a family of events containing $\mathcal{D} = \{E^*|H^* \wedge K, H^*|E^* \wedge K\}$ and let P be a coherent conditional probability on $\mathcal{G}$. Then, E is **cs-independent of** H **conditionally on** K, denoted as $E \perp\!\!\!\perp_{cs} H|K$, if the following two conditions hold:

(i) $P(E|H \wedge K) = P(E|H^c \wedge K)$;

(ii) there exists a complete agreeing class $\{P_\alpha\}$ on *algebra*$(\{E, H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$ such that

$$o(E|H \wedge K) = o(E|H^c \wedge K) \quad \text{and} \quad o(E^c|H \wedge K) = o(E^c|H^c \wedge K).$$

Note that for each triple of events E, H, K such that $K \subseteq E$, in particular for $E = \Omega$ or $E = K$, we have $E \perp\!\!\!\perp_{cs} H|K$ whenever $\emptyset \neq H \wedge K \neq K$, but the statement $H \perp\!\!\!\perp_{cs} E|K$ does not make sense since $E^c \wedge K = \emptyset$, so, $P(H|E^c \wedge K)$ is without any meaning. An analogous consideration holds for $E^c \subseteq K$. This shows that the conditional cs-independence relation is generally not symmetric.

Here, we report the characterization of conditional cs-independence in terms of conditional probabilities avoiding to refer to zero layers (Coletti and Scozzafava 2002a; Vantaggi 2001).

In the following we say that two events E, H are logically independent with respect to $K \neq \emptyset$ if all the events $E^* \wedge H^* \wedge K$ differ from the impossible event. Obviously if E, H are logically independent with respect to K then also E, H are logically independent.

Theorem 3.2. *Let E, H be two events logically independent with respect to K. If P is a coherent conditional probability such that $P(E|H \wedge K) = P(E|H^c \wedge K)$, then $E \perp\!\!\!\perp_{cs} H|K$ if and only if one (and only one) of the following conditions holds:*

(a) $0 < P(E|H \wedge K) < 1$;

(b) $P(E|H \wedge K) = 0$ *and the extension of P to $H|K$ and $H|E \wedge K$ satisfies one of the following conditions*

(1) $P(H|K) = 0$, $P(H|E \wedge K) = 0$,

(2) $P(H|K) = 1$, $P(H|E \wedge K) = 1$,

[1] *cs* is used as an abbreviation of *coherent setting.*

(3) $0 < P(H|K) < 1$, $0 < P(H|E \wedge K) < 1$;

(c) $P(E|H \wedge K) = 1$ *and the extension of* P *to* $H|K$ *and* $H|E^c \wedge K$ *satisfies one of the following conditions*

(1) $P(H|K) = 0$, $P(H|E^c \wedge K) = 0$,

(2) $P(H|K) = 1$, $P(H|E^c \wedge K) = 1$,

(3) $0 < P(H|K) < 1$, $0 < P(H|E^c \wedge K) < 1$.

From the previous result a theorem characterizing the symmetric property for conditional cs-independence follows Coletti and Scozzafava (2002a); Vantaggi (2001):

Theorem 3.3. *Let* E, H *be two events logically independent with respect to* K. *If* P *is a coherent conditional probability, then* $E \perp\!\!\!\perp_{cs} H|K$ *and* $H \perp\!\!\!\perp_{cs} E|K$ *if and only if*

$$P(E|H \wedge K) = P(E|H^c \wedge K) \quad \textit{and} \quad P(H|E \wedge K) = P(H|E^c \wedge K).$$

4. Defining positive and negative correlation

Starting from the concept of cs-independence, we now study the *sign* (positive or negative) of a correlation between dependent events. Before defining our concepts of positive and negative correlation for coherent conditional probability, we first address the impact of using classical definitions of correlation between two events in the setting of coherence.

Definition 4.1. Let $\mathcal{G}$ be an arbitrary family of events with $E, E|H \in \mathcal{G}$ and let P be a coherent conditional probability on $\mathcal{G}$. Then,

- E is **positively correlated** with H iff $P(E|H) > P(E)$;
- E is **negatively correlated** with H iff $P(E|H) < P(E)$;
- E and H are **not correlated** iff $P(E|H) = P(E)$.

We note that, with the above definition, if two events E and H are not correlated, they are independent in the classical sense.

Based on the classical definition of correlation, various properties of correlation between two events have been formulated. We review some of these properties in the following proposition, tailored to the coherent setting. We will subsequently demonstrate the inadequacy of the classical definitions for describing correlation for coherent conditional probability.

Proposition 4.2. *Let* $\mathcal{G}$ *be an arbitrary family of events including* $E^*, H^*, E^*|H^*$. *Let* P *be a coherent conditional probability on* $\mathcal{G}$ *such that* $P(E), P(H) \in]0,1[$. *Then, the following properties hold:*

(i) if E *is positively [negatively] correlated with* H, *then* E^c *is positively [negatively] correlated with* H^c;

(ii)
- ∘ *if either* $E \wedge H = \varnothing$ *or* $E^c \wedge H^c = \varnothing$, *then* E *is negatively correlated with* H;
- ∘ *if either* $E^c \wedge H = \varnothing$ *or* $E \wedge H^c = \varnothing$, *then* E *is positively correlated with* H;

(iii) E *is positively [negatively] correlated with* H *iff* $P(E|H) > [<] P(E|H^c)$.

Proof. Properties *(i)* and *(iii)* follow directly from Definition 4.1. The first part of property *(ii)* follows by observing that $E \wedge H = \varnothing$ implies $P(E|H) = 0 < P(E)$. As $E^c \wedge H^c = \varnothing$ implies $P(E^c|H^c) = 0 < P(E^c)$, we have by property *(i)* that

$P(E|H) < P(E)$. In both cases, therefore, E is negatively correlated with H. The second part of property *(ii)* follows analogously. □

We note that property *(i)* of Proposition 4.2 strictly depends on the premise that the probabilities of E and H are different from 0 and 1 and such premise also underlies the cases of $E^c \wedge H^c = \varnothing$ and $E \wedge H^c = \varnothing$. For property *(iii)*, moreover, the implication $P(E|H) > P(E) \Longrightarrow P(E|H) > P(E|H^c)$ holds only when $P(E), P(H) \in]0, 1[$, while the reversed implication is universally valid.

To illustrate the inadequacy of the classical definition above for describing correlation in a coherent setting, we consider an event H with $P(H) = 1$. By Definition 4.1, this event is not correlated with any other event, as for any event $E \neq H$ we would find that $P(E|H) = P(E)$. We would find the same result, in fact, also for an event E which logically contradicts H, that is, for which $E \wedge H = \varnothing$, as we would then have $P(E|H) = P(\varnothing|H) = P(E) = 0$. Yet, E could clearly not be considered uncorrelated with H. Similarly counterintuitive conclusions are found for an event E which is logically implied by H and for an event E with $P(E) = 0$.

Not all researchers accept Definition 4.1 as the basic definition of correlation, however, and may argue that the above observations are due to using an inappropriate definition. They may use property *(iii)* of Proposition 4.2 for the basic definition of correlation instead, that is, they would use the following definition for coherent conditional probability.

Definition 4.3. Let $\mathcal{G}$ be an arbitrary family of events with $E|H, E|H^c \in \mathcal{G}$ and let P be a coherent conditional probability on $\mathcal{G}$. Then,

- E is **positively correlated** with H iff $P(E|H) > P(E|H^c)$;
- E is **negatively correlated** with H iff $P(E|H) < P(E|H^c)$;
- E and H are **not correlated** iff $P(E|H) = P(E|H^c)$.

We note that Definitions 4.1 and 4.3 are not equivalent: while Definition 4.1 implies Definition 4.3, the reverse does not hold. In fact, by Definition 4.3, a conditioning event H with $P(H) = 1$ is not necessarily uncorrelated with an event E. Since $P(H^c) = 0$, there is an index $\alpha_{H^c} > 0$ such that $P_{\alpha_{H^c}}(H) > 0$ and

$$P(E|H^c) = \frac{P_{\alpha_{H^c}}(E \wedge H^c)}{P_{\alpha_{H^c}}(H^c)},$$

which, without any further information, can assume any value in $[0, 1]$ and hence also values larger, or smaller, than $P(E)$. Yet, also Definition 4.3 does not capture the full impact of the hypothesis H on the degree of belief in E when $P(E|H) = P(E|H^c) = 0$ or $P(E|H) = P(E|H^c) = 1$.

From the above considerations, we conclude that, with both classical definitions of correlation, we need to distinguish between different zeroes, depending on their strengths, before concluding that two extreme probability events are uncorrelated. We provide an example to illustrate our conclusion.

Example 4.4. Let Ω be the unit square $[0, 1]^2$, as shown in Figure 1. Let the event E be the Boolean sentence $E = P \vee Q \vee R$ where P, Q, R are the points $P = \left(\frac{3}{4}, \frac{3}{4}\right)$, $Q = \left(\frac{1}{2}, \frac{1}{2}\right)$, $R = \left(\frac{3}{4}, \frac{1}{4}\right)$ in Ω. The event H further is the diagonal of the unit square, that is, $H = \{(x, y)|x = y,\ x, y \in [0, 1]\}$.

In this setting, we consider the assessment $P(E|H) = P(E|H^c) = P(H|E^c) = 0$,

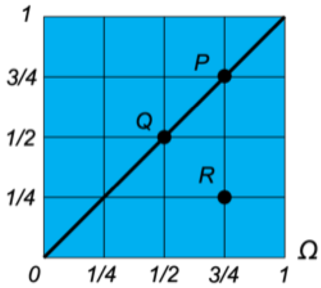

Figure 1. Representation of events P, Q, R and H in Ω

$P(H|E) = \frac{2}{3}$ and check its coherence. To this end, we consider the set $atoms(\{E, H\}) = \{C_1, C_2, C_3, C_4\}$ with

$$C_1 = E \wedge H = P \vee Q, \quad C_2 = E \wedge H^c = R, \quad C_3 = E^c \wedge H, \quad C_4 = E^c \wedge H^c,$$

To search for a complete agreeing class on $algebra(\{E, H\})$, we build the sequence of systems $\mathcal{S}_\alpha$ with non-negative unknowns $x_r^\alpha = P_\alpha(C_r)$, as described in Theorem 2.3. The first system equals

$$\mathcal{S}_0 : \begin{cases} x_1^0 = 0 \cdot (x_1^0 + x_3^0) \\ x_2^0 = 0 \cdot (x_2^0 + x_4^0) \\ x_3^0 = 0 \cdot (x_3^0 + x_4^0) \\ x_1^0 = \frac{2}{3} \cdot (x_1^0 + x_2^0) \\ x_1^0 + x_2^0 + x_3^0 + x_4^0 = 1 \end{cases}$$

whose unique solution is $x_1^0 = x_2^0 = x_3^0 = 0$, $x_4^0 = 1$. We note that this solution implies through additivity $P_0(H) = 0 = P_0(E)$. Focusing on the zero-probability atoms and writing x_r^1 for x_r^0, the second system in the sequence is

$$\mathcal{S}_1 : \begin{cases} x_1^1 = 0 \cdot (x_1^1 + x_3^1) \\ x_1^1 = \frac{2}{3} \cdot (x_1^1 + x_2^1) \\ x_1^1 + x_2^1 + x_3^1 = 1 \end{cases}$$

whose unique solution is $x_1^1 = x_2^1 = 0$, $x_3^1 = 1$. This solution implies through additivity $P_1(E) = 0$, yet no longer has zero for $P_1(H)$. The third, and final, system equals

$$\mathcal{S}_2 : \begin{cases} x_1^2 = \frac{2}{3} \cdot (x_1^2 + x_2^2) \\ x_1^2 + x_2^2 = 1 \end{cases}$$

whose unique solution is $x_1^2 = \frac{2}{3}$, $x_2^2 = \frac{1}{3}$. Since every system of the constructed sequence has a unique solution, the assessment P has a unique complete agreeing class $\{P_0, P_1, P_2\}$ on $algebra(\{E, H\})$, that is equal to

$atoms(\{E, H\})$	C_1	C_2	C_3	C_4
P_0	0	0	0	1
P_1	0	0	1	$\bullet$
P_2	$\frac{2}{3}$	$\frac{1}{3}$	$\bullet$	$\bullet$

This class implies that

$$o(E|H) = o(E \wedge H) - o(H) = 2 - 1 \ < \ 2 - 0 = o(E \wedge H^c) - o(H^c) = o(E|H^c).$$

The zero layer of E under the hypothesis H thus is smaller than that of E under the hypothesis H^c. As the conditional event $E|H^c$ still has zero probability in the structure when $E|H$ does not, this finding may be naturally construed as a positive correlation of E with H.

We further consider the event R that is incompatible with H. Analogously to the above approach, we find for the conditional events $R|H$ and $R|H^c$ that

$$o(R|H) = o(R \wedge H) - o(H) = +\infty - 1 > 2 - 0 = o(R \wedge H^c) - o(H^c) = o(R|H^c),$$

which demonstrates that the logical impossibility of R under the hypothesis H results in a zero layer which is infinitely larger than that of R under the hypothesis H^c. The zero probability resulting from a logical impossibility will thus always be deeper in the complex structure of conditional probability than the zero probability of any possible event. □

Based on the considerations in the above example, we now introduce our enhanced definition of correlation in a coherent setting. This definition builds on a previously detailed notion of correlation in a context of coherence (Coletti, van der Gaag, and Petturiti 2018), extending it to conditional correlation.

Definition 4.5. Let $\mathcal{G}$ be an arbitrary family of conditional events containing the set $\mathcal{D} = \{E^*|H^* \wedge K, H^*|E^* \wedge K\}$, and let P be a coherent conditional probability on $\mathcal{G}$. We say that:

- E is **positively cs-correlated with** H **conditionally on** K, denoted as $E \perp_{cs}^{+} H|K$, if one of the following conditions holds:
  - ◦ $P(E|H \wedge K) > P(E|H^c \wedge K)$;
  - ◦ $P(E|H \wedge K) = P(E|H^c \wedge K) = 0$, and there exists a complete agreeing class $\{P_\alpha\}$ on *algebra*$(\{E, H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$ such that
  $$o(E|H \wedge K) < o(E|H^c \wedge K);$$
  - ◦ $P(E|H \wedge K) = P(E|H^c \wedge K) = 1$, and there exists a complete agreeing class $\{P_\alpha\}$ on *algebra*$(\{E, H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$ such that
  $$o(E^c|H \wedge K) > o(E^c|H^c \wedge K);$$
- E is **negatively cs-correlated with** H **conditionally on** K, denoted as $E \perp_{cs}^{-} H|K$, if one of the following conditions holds:
  - ◦ $P(E|H \wedge K) < P(E|H^c \wedge K)$;
  - ◦ $P(E|H \wedge K) = P(E|H^c \wedge K) = 0$, and there exists a complete agreeing class $\{P_\alpha\}$ on *algebra*$(\{E, H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$ such that
  $$o(E|H \wedge K) > o(E|H^c \wedge K),$$
  - ◦ $P(E|H \wedge K) = P(E|H^c \wedge K) = 1$, and there exists a complete agreeing

class $\{P_\alpha\}$ on $algebra(\{E,H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$ such that

$$o(E^c|H \wedge K) < o(E^c|H^c \wedge K);$$

- E is **not cs-correlated with H conditionally on** K, denoted as $E \not\perp_{cs}^{+/-} H|K$, if it is not positively nor negatively cs-correlated with H conditionally on K.

The definition of unconditional positive and negative cs-correlation is covered by the above definition as a special case with $K = \Omega$: in case of unconditional cs-correlation, we will simply write $E \perp_{cs}^{+} H$, $E \perp_{cs}^{-} H$ and $E \not\perp_{cs}^{+/-} H$, respectively.

The following proposition shows that the notion of conditional non-cs-correlation coincides with the notion of conditional cs-independence recalled in Section 3.

Proposition 4.6. *For every E, H, K it holds*

$$E \not\perp_{cs}^{+/-} H|K \iff E \perp\!\!\!\perp_{cs} H|K.$$

Proof. If $E \perp\!\!\!\perp_{cs} H|K$ then it trivially holds that $E \not\perp_{cs}^{+/-} H|K$. Thus, suppose $E \not\perp_{cs}^{+/-} H|K$ which implies that $P(E|H \wedge K) = P(E|H^c \wedge K) = p$ and $P(E^c|H \wedge K) = P(E^c|H^c \wedge K) = 1-p$. If $p \in]0,1[$ then $o(E|H \wedge K) = o(E|H^c \wedge K)$ and $o(E^c|H \wedge K) = o(E^c|H^c \wedge K)$ and the conclusion follows. If $p = 0$, then $o(E^c|H \wedge K) = o(E^c|H^c \wedge K)$, while it cannot be $o(E|H \wedge K) < o(E|H^c \wedge K)$ nor $o(E|H \wedge K) > o(E|H^c \wedge K)$, so, the conclusion follows. If $p = 1$, then $o(E|H \wedge K) = o(E|H^c \wedge K)$, while it cannot be $o(E^c|H \wedge K) < o(E^c|H^c \wedge K)$ nor $o(E^c|H \wedge K) > o(E^c|H^c \wedge K)$, so, the conclusion follows. □

The above definition of conditional positive and negative correlation in a coherent setting avoids the counterintuitive findings from the classic definitions of correlation which were illustrated above. In fact, in the presence of extreme probability events, this definition allows the identification of a correlation between events which are logically related, as shown in the following theorem.

Theorem 4.7. *Let P be a coherent conditional probability defined on an arbitrary family of conditional events $\mathcal{G}$ containing the set $\mathcal{D} = \{E^*|H^* \wedge K, H^*|E^* \wedge K\}$. Then, the following properties hold:*

(i) if either $E \wedge H \wedge K = \varnothing$ or $E^c \wedge H^c \wedge K = \varnothing$, then $E \perp_{cs}^{-} H|K$;
(ii) if either $E^c \wedge H \wedge K = \varnothing$ or $E \wedge H^c \wedge K = \varnothing$ then $E \perp_{cs}^{+} H|K$.

Proof. We prove property *(i)* as the proof of property *(ii)* is analogous. If $E \wedge H \wedge K = \varnothing$, we just have to consider the case where $P(E|H \wedge K) = 0 = P(E|H^c \wedge K)$: in this case we have that $o(E|H \wedge K) = +\infty > o(E|H^c \wedge K)$ and the negative correlation follows. Similarly, if $E^c \wedge H^c \wedge K = \varnothing$, we address just the case where $P(E|H \wedge K) = 1 = P(E|H^c \wedge K)$; since then $o(E^c|H \wedge K) < +\infty = o(E^c|H^c \wedge K)$, the negative correlation equally follows. □

Condition *(i)* emphasizes, for example, that when $E \wedge H \wedge K = \varnothing$ obviously supposing that $H \wedge K$ occurs implies that E cannot be true and so gives rise to a negative correlation between the two events. Analogously, if $E^c \wedge H \wedge K = \varnothing$, then supposing that $H \wedge K$ occurs implies E occurs and so gives rise to a positive correlation.

Apparently, Definition 4.5 is tied to the choice of a complete agreeing class $\{P_\alpha\}$ on $algebra(\{E,H\}) \wedge K$ guaranteeing a particular ordering of zero layers in case of extreme probability conditional events. This choice could seem arbitrary, nevertheless, the following theorem shows that if the ordering of zero layers holds for a choice of $\{P_\alpha\}$ on *algebra*$(\{E,H\}) \wedge K$ than it holds for all the possible choices. This means that positive/negative cs-stochastic correlation is invariant with respect to the choice of the complete agreeing class $\{P_\alpha\}$ on *algebra*$(\{E,H\}) \wedge K$.

Theorem 4.8. *Let E, H be logically independent events with respect to K and let P be a coherent conditional probability on a family of conditional events $\mathcal{G}$ containing the subset $\mathcal{D} = \{E^*|H^* \wedge K, H^*|E^* \wedge K\}$. If there exists a complete agreeing class $\{P_\alpha\}$ on algebra$(\{E,H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$, such that one of the following conditions holds*

- $P(E|H \wedge K) = P(E|H^c \wedge K) = 0$ *and* $o(E|H \wedge K) < [>] o(E|H^c \wedge K)$*;*
- $P(E|H \wedge K) = P(E|H^c \wedge K) = 1$ *and* $o(E^c|H \wedge K) > [<] o(E^c|H^c \wedge K)$*;*

then this holds true for any other complete agreeing class on algebra$(\{E,H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$.

Proof. Under the hypotheses on E, H, K, the elements of *atoms*$(\{E,H,K\})$ implying K are

$$C_1 = E \wedge H \wedge K,\ C_2 = E \wedge H^c \wedge K,\ C_3 = E^c \wedge H \wedge K,\ C_4 = E^c \wedge H^c \wedge K.$$

Suppose $P(E|H \wedge K) = P(E|H^c \wedge K) = 0$ and let $P(H|E \wedge K) = p$ and $P(H|E^c \wedge K) = q$. We find all the possible complete agreeing classes on *algebra*$(\{E,H\}) \wedge K$ that agree with $P_{|\mathcal{D}}$, by solving the sequence of systems in Theorem 2.3.

If $p = q = 0$, then the possible complete agreeing classes on *algebra*$(\{E,H\}) \wedge K$ that agree with $P_{|\mathcal{D}}$ are

	C_1	C_2	C_3	C_4
P_0	0	0	0	1
P_1	0	δ	$1-\delta$	$\bullet$
P_2	1	$\bullet$	$\bullet$	$\bullet$

$\delta \in]0,1[$

	C_1	C_2	C_3	C_4
P_0	0	0	0	1
P_1	0	1	0	$\bullet$
P_2	0	$\bullet$	1	$\bullet$
P_3	1	$\bullet$	$\bullet$	$\bullet$

	C_1	C_2	C_3	C_4
P_0	0	0	0	1
P_1	0	0	1	$\bullet$
P_2	0	1	$\bullet$	$\bullet$
P_3	1	$\bullet$	$\bullet$	$\bullet$

and in any case $o(E|H \wedge K) = o(E|H^c \wedge K)$, thus the hypothesis is never satisfied.

If $p = q = 1$, then the possible complete agreeing classes on *algebra*$(\{E,H\}) \wedge K$ that agree with $P_{|\mathcal{D}}$ are

	C_1	C_2	C_3	C_4
P_0	0	0	1	0
P_1	δ	0	$\bullet$	$1-\delta$
P_2	$\bullet$	1	$\bullet$	$\bullet$

$\delta \in]0,1[$

	C_1	C_2	C_3	C_4
P_0	0	0	1	0
P_1	1	0	$\bullet$	0
P_2	$\bullet$	0	$\bullet$	1
P_3	$\bullet$	1	$\bullet$	$\bullet$

	C_1	C_2	C_3	C_4
P_0	0	0	1	0
P_1	0	0	$\bullet$	1
P_2	1	0	$\bullet$	$\bullet$
P_3	$\bullet$	1	$\bullet$	$\bullet$

and in any case $o(E|H \wedge K) = o(E|H^c \wedge K)$, thus the hypothesis is never satisfied.

If $p, q \in]0,1[$, the unique complete agreeing class on *algebra*$(\{E,H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$ is

	C_1	C_2	C_3	C_4
P_0	0	0	q	$1-q$
P_1	p	$1-p$	$\bullet$	$\bullet$

and it holds $o(E|H \wedge K) = o(E|H^c \wedge K)$, thus the hypothesis is not satisfied.

If $p > q$ with $p = 1$ or $q = 0$, then we always have a unique complete agreeing class on $algebra(\{E, H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$ and it holds

$p = 1, q = 0$

	C_1	C_2	C_3	C_4
P_0	0	0	0	1
P_1	0	0	1	•
P_2	1	0	•	•
P_3	•	1	•	•

$p \in]0, 1[, q = 0$

	C_1	C_2	C_3	C_4
P_0	0	0	0	1
P_1	0	0	1	•
P_2	p	$1 - p$	•	•

$p = 1, q \in]0, 1[$

	C_1	C_2	C_3	C_4
P_0	0	0	q	$1 - q$
P_1	1	0	•	•
P_2	•	1	•	•

and in any case $o(E|H \wedge K) < o(E|H^c \wedge K)$, thus the hypothesis is always satisfied and the conclusion follows.

If $p < q$ with $p = 0$ or $q = 1$, then we always have a unique complete agreeing class on $algebra(\{E, H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$ and it holds

$p = 0, q = 1$

	C_1	C_2	C_3	C_4
P_0	0	0	1	0
P_1	0	0	•	1
P_2	0	1	•	•
P_3	1	•	•	•

$p = 0, q \in]0, 1[$

	C_1	C_2	C_3	C_4
P_0	0	0	q	$1 - q$
P_1	0	1	•	•
P_2	1	•	•	•

$p \in]0, 1[, q = 1$

	C_1	C_2	C_3	C_4
P_0	0	0	1	0
P_1	0	0	•	1
P_2	p	$1 - p$	•	•

and in any case $o(E|H \wedge K) > o(E|H^c \wedge K)$, thus the hypothesis is always satisfied and the conclusion follows.

The proof when $P(E|H \wedge K) = P(E|H^c \wedge K) = 1$ goes along the same line. □

5. Properties of cs-correlation

In view of classical probability, the concept of independence lies at the core of modern modelling approaches, such as probabilistic graphical models, where it has proven to be the key to efficient probability computation. The concept of correlation lies at the heart of comparative probability approaches, such as those found in qualitative probabilistic networks (Wellman 1990; Druzdzel and Henrion 1993). Probability computation in the latter type of approach is shaped by exploiting properties such as symmetry and transitivity of correlations. In this section we study these properties in view of our enhanced concepts of positive and negative correlation.

Positive correlation fulfils the property of *reflexivity*, under any hypothesis, as stated in the following Proposition.

Proposition 5.1. *Every possible event E is positively cs-correlated with itself under any hypothesis K compatible with E.*

Proof. The proof is straightforward. □

Positive as well as negative cs-correlation are not necessarily symmetric, as the following example shows.

Example 5.2. Let E, H be two logically independent events and let us consider the following coherent conditional probability $P(E|H) = \frac{1}{2} > \frac{1}{4} = P(E|H^c)$ and $P(H|E) = P(H|E^c) = 0$. Obviously, $E \perp^+_{cs} H$, but not $H \perp^+_{cs} E$, since by a trivial computation $o(E|H) = o(E|H^c)$. □

For a symmetric concept of both positive and negative cs-correlation, the definitions of $\perp_{cs}^{+}$ and $\perp_{cs}^{-}$ need to be further strengthened by setting

$$E \perp_{S-cs}^{+} H|K \quad \text{iff} \quad E \perp_{cs}^{+} H|K \text{ and } H \perp_{cs}^{+} E|K,$$
$$E \perp_{S-cs}^{-} H|K \quad \text{iff} \quad E \perp_{cs}^{-} H|K \text{ and } H \perp_{cs}^{-} E|K.$$

The following theorem now characterizes symmetric positive and negative cs-correlation.

Theorem 5.3. *Let E, H be logically independent events with respect to K and let P be a coherent conditional probability on a family of conditional events $\mathcal{G}$ containing the set $\mathcal{D} = \{E^*|H^* \wedge K, H^*|E^* \wedge K\}$. Then $E \perp_{S-cs}^{+} H|K$ if and only if one of the following conditions holds:*

(a) $P(E|H \wedge K) > P(E|H^c \wedge K)$ *and* $P(H|E \wedge K) > P(H|E^c \wedge K)$*;*
(b) *either* $1 = P(E|H \wedge K) > P(E|H^c \wedge K)$ *or* $P(E|H \wedge K) > P(E|H^c \wedge K) = 0$, *and* $P(H|E \wedge K) = P(H|E^c \wedge K) \in \{0, 1\}$*;*
(b') *either* $1 = P(H|E \wedge K) > P(H|E^c \wedge K)$ *or* $P(H|E \wedge K) > P(H|E^c \wedge K) = 0$, *and* $P(E|H \wedge K) = P(E|H^c \wedge K) \in \{0, 1\}$.

Proof. We limit to proof conditions *(a)* and *(b)* since *(b')* is the symmetric version of *(b)*.

We first prove the sufficiency of the above conditions. Condition *(a)* trivially implies $E \perp_{S-cs}^{+} H|K$.

By a direct computation, it is easy to prove that in condition *(b)*, for every complete agreeing class $\{P_\alpha\}$ on *algebra*$(\{E, H\}) \wedge K$ one has $o(E|H \wedge K) > o(E|H^c \wedge K)$, in the case $P(H|E \wedge K) = 0$, and $o(E^c|H \wedge K) < o(E^c|H^c \wedge K)$, in the case $P(H|E \wedge K) = 1$. Therefore, under condition *(b)*, one obtains $H \perp_{cs}^{+} E \wedge K$ and, so, $E \perp_{S-cs}^{+} H|K$.

To prove the necessity of the above conditions we need to prove that in all the other cases symmetry fails. For that, we only have to consider the case $1 \neq P(E|H \wedge K) > P(E|H^c \wedge K) \neq 0$ and $P(H|E \wedge K) = P(H|E^c \wedge K) \in \{0, 1\}$, since in the remaining cases we obtain conditional cs-independence (see Vantaggi (2001)). Again by a direct computation, in the above conditions, we obtain $o(H|E \wedge K) = o(H|E^c \wedge K)$ in the case $P(H|E \wedge K) = 0$, and $o(H^c|E \wedge K) = o(H^c|E^c \wedge K)$ in the case $P(H|E \wedge K) = 1$. □

A completely analogous result holds for symmetric negative cs-independence.

Theorem 5.4. *Let E, H be logically independent events with respect to K and let P be a coherent conditional probability on a family of conditional events $\mathcal{G}$ containing the set $\mathcal{D} = \{E^*|H^* \wedge K, H^*|E^* \wedge K\}$. Then $E \perp_{S-cs}^{-} H|K$ if and only if one of the following conditions holds:*

(a) $P(E|H \wedge K) < P(E|H^c \wedge K)$ *and* $P(H|E \wedge K) < P(H|E^c \wedge K)$*;*
(b) *either* $0 = P(E|H \wedge K) < P(E|H^c \wedge K)$ *or* $P(E|H \wedge K) < P(E|H^c \wedge K) = 1$, *and* $P(H|E \wedge K) = P(H|E^c \wedge K) \in \{0, 1\}$*;*
(b') *either* $0 = P(H|E \wedge K) < P(H|E^c \wedge K)$ *or* $P(H|E \wedge K) < P(H|E^c \wedge K) = 1$, *and* $P(E|H \wedge K) = P(E|H^c \wedge K) \in \{0, 1\}$.

Proof. The proof goes along the same line of the proof of Theorem 5.3. □

The following theorem allows to easily check symmetry of conditional positive/negative correlation and non-correlation (independence) of E and H conditionally on K, whenever the probabilities of $E|K$ and $H|K$ are fixed and non-extreme.

Theorem 5.5. *Let P be a coherent conditional probability on a family of conditional events $\mathcal{G}$ containing the set $\mathcal{D} = \{E^*|K, H^*|K, E^*|H^* \wedge K, H^*|E^* \wedge K\}$. If $P(E|K), P(H|K) \in]0,1[$ then*

$$P(E|H \wedge K) \leq P(E|H^c \wedge K) \Longleftrightarrow P(H|E \wedge K) \leq P(H|E^c \wedge K).$$

Proof. Since $P(E|K), P(H|K) \in]0,1[$, we can write

$$\begin{aligned} P(E|H \wedge K) &= \frac{P(E \wedge H|K)}{P(H|K)} = \frac{P(E|K)P(H|E \wedge K)}{P(E|K)P(H|E \wedge K) + P(E^c|K)P(H|E^c \wedge K)}, \\ P(E|H^c \wedge K) &= \frac{P(E \wedge H^c|K)}{P(H^c|K)} = \frac{P(E|K)P(H^c|E \wedge K)}{P(E|K)P(H^c|E \wedge K) + P(E^c|K)P(H^c|E^c \wedge K)}. \end{aligned}$$

Hence, we have that $P(E|H \wedge K) \leq P(E|H^c \wedge K)$ is equivalent to

$$\frac{P(E|K)P(H|E \wedge K)}{P(E|K)P(H|E \wedge K) + P(E^c|K)P(H|E^c \wedge K)} \leq \frac{P(E|K)P(H^c|E \wedge K)}{P(E|K)P(H^c|E \wedge K) + P(E^c|K)P(H^c|E^c \wedge K)},$$

which, in turn, holds if and only if

$$P(H|E \wedge K)P(H^c|E^c \wedge K) \leq P(H^c|E \wedge K)P(H|E^c \wedge K).$$

The previous inequality can be rewritten as

$$P(H|E \wedge K)(1 - P(H|E^c \wedge K)) \leq (1 - P(H|E \wedge K))P(H|E^c \wedge K),$$

that holds if and only if $P(H|E \wedge K) \leq P(H|E^c \wedge K)$. □

Since *transitivity* is an important property in the context of comparative probability for computing indirect correlations, we address it in our context of coherent probability. In particular, given events E, F, G and K taken as hypothesis, by transitivity we mean

$$E \perp_{cs}^{+} H|K \text{ and } H \perp_{cs}^{+} G|K \Longrightarrow E \perp_{cs}^{+} G|K,$$

with an analogous formulation for $\perp_{cs}^{-}$.

The following example shows that transitivity is generally violated, even when the symmetric definition of cs-correlation is adopted.

Example 5.6. Let E, H, G be three logically independent events where *atoms*$(\{E, H, G\}) = \{C_1, \ldots, C_8\}$ with

$$\begin{array}{llll} C_1 = E \wedge H \wedge G, & C_2 = E \wedge H \wedge G^c, & C_3 = E \wedge H^c \wedge G, & C_4 = E \wedge H^c \wedge G^c, \\ C_5 = E^c \wedge H \wedge G, & C_6 = E^c \wedge H \wedge G^c, & C_7 = E^c \wedge H^c \wedge G, & C_8 = E^c \wedge H^c \wedge G^c. \end{array}$$

Consider the coherent assessment $P(E) = \frac{9}{10}$, $P(H) = \frac{9}{20}$, $P(G) = \frac{3}{10}$, $P(H|E) = \frac{5}{10}$ and $P(G|H) = \frac{2}{5}$. Applying axioms of Definition 2.1, from this coherent assessment we derive:

- $P(E \wedge H) = P(H|E) \cdot P(E) = \frac{9}{20}$,

- $P(E \wedge H^c) = P(E) - P(E \wedge H) = \frac{9}{20}$,
- $P(E^c \wedge H) = P(H) - P(E \wedge H) = 0$,
- $P(H \wedge G) = P(G|H) \cdot P(H) = \frac{9}{50}$,
- $P(H \wedge G^c) = P(H) - P(H \wedge G) = \frac{27}{100}$,
- $P(H^c \wedge G) = P(G) - P(H \wedge G) = \frac{6}{50}$,
- $P(E|H) = \frac{P(E \wedge H)}{P(H)} = 1$,
- $P(E|H^c) = \frac{P(E \wedge H^c)}{P(H^c)} = \frac{9}{11}$,
- $P(H|E^c) = \frac{P(E^c \wedge H)}{P(E^c)} = 0$,
- $P(H|G) = \frac{P(H \wedge G)}{P(G)} = \frac{6}{10}$,
- $P(H|G^c) = \frac{P(H \wedge G^c)}{P(G^c)} = \frac{27}{70}$,
- $P(G|H^c) = \frac{P(H^c \wedge G)}{P(H^c)} = \frac{12}{55}$,

so, all conditional probabilities extending this assessment satisfy the above equations. In particular, by Theorem 5.3, the given assessment determines $E \perp^+_{S-cs} H$ and $H \perp^+_{S-cs} G$ but these two statements do not imply any form of correlation between E and G.

Indeed, considering the complete agreeing class $\{P_0, P_1\}$ on *algebra*$(\{E, H, G\})$ that agrees with P, where

atoms$(\{E,H,G\})$	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
P_0	$\frac{9}{50}$	$\frac{27}{100}$	$\frac{6}{50}$	$\frac{33}{100}$	0	0	0	$\frac{1}{10}$
P_1	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\bullet$

we have

$$P(E|G) = \frac{P_0(C_1 \vee C_3)}{P_0(G)} = 1 > \frac{6}{7} = \frac{P_0(C_2 \vee C_4)}{P_0(G^c)} = P(E|G^c),$$
$$P(G|E) = \frac{P_0(C_1 \vee C_3)}{P_0(E)} = \frac{1}{3} > 0 = \frac{P_0(C_5 \vee C_7)}{P_0(E^c)} = P(G|E^c),$$

that by Theorem 5.3 imply $E \perp^+_{S-cs} G$.

On the other hand, taking the complete agreeing class $\{P'_0, P'_1\}$ on *algebra*$(\{E, H, G\})$ that agrees with P, where

atoms$(\{E,H,G\})$	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
P'_0	$\frac{9}{50}$	$\frac{27}{100}$	$\frac{1}{50}$	$\frac{43}{100}$	0	0	$\frac{1}{10}$	0
P'_1	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\frac{1}{3}$	$\frac{1}{3}$	$\bullet$	$\frac{1}{3}$

we have

$$P(E|G) = \frac{P'_0(C_1 \vee C_3)}{P'_0(G)} = \frac{2}{3} \quad < \quad 1 = \frac{P'_0(C_2 \vee C_4)}{P'_0(G^c)} = P(E|G^c),$$
$$P(G|E) = \frac{P'_0(C_1 \vee C_3)}{P'_0(E)} = \frac{2}{9} \quad < \quad 1 = \frac{P'_0(C_5 \vee C_7)}{P'_0(E^c)} = P(G|E^c),$$

that by Theorem 5.3 imply $E \perp^-_{S-cs} G$.

Finally, taking the complete agreeing class $\{P''_0, P''_1\}$ on *algebra*$(\{E, H, G\})$ that agrees with P, where

$atoms(\{E,H,G\})$	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
P'_0	$\frac{9}{50}$	$\frac{27}{100}$	$\frac{9}{100}$	$\frac{9}{25}$	0	0	$\frac{3}{100}$	$\frac{7}{100}$
P'_1	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\frac{1}{2}$	$\frac{1}{2}$	$\bullet$	$\bullet$

we have

$$P(E|G) = \frac{P''_0(C_1 \vee C_3)}{P''_0(G)} = \frac{9}{10} = \frac{9}{10} = \frac{P''_0(C_2 \vee C_4)}{P''_0(G^c)} = P(E|G^c),$$
$$P(G|E) = \frac{P''_0(C_1 \vee C_3)}{P''_0(E)} = \frac{3}{10} = \frac{3}{10} = \frac{P''_0(C_5 \vee C_7)}{P''_0(E^c)} = P(G|E^c),$$

that by Theorem 3.3 imply $E \perp\!\!\!\perp_{cs} G$ and $G \perp\!\!\!\perp_{cs} E$ or, equivalently, by Proposition 4.6 $E \not\perp_{cs}^{+/-} G$ and $G \not\perp_{cs}^{+/-} E$.

Even though transitivity does not hold in general neither for $\perp_{cs}^{+}$ nor for $\perp_{cs}^{-}$, the following weak form of transitivity involving conditional probability values holds.

Theorem 5.7. *Let P be a coherent conditional probability on a family of conditional events $\mathcal{G}$ containing the set $\mathcal{D} = \{E^*|H^* \wedge G^* \wedge K, E^*|H^* \wedge K, H^*|G^* \wedge K, E^*|G^* \wedge K\}$. Assuming that $P(E|H \wedge G \wedge K) = P(E|H \wedge G^c \wedge K)$ and $P(E|H^c \wedge G \wedge K) = P(E|H^c \wedge G^c \wedge K)$ it holds*

$$\left.\begin{array}{l} P(E|H \wedge K) \geq P(E|H^c \wedge K) \\ P(H|G \wedge K) \geq P(H|G^c \wedge K) \end{array}\right\} \Longrightarrow P(E|G \wedge K) \geq P(E|G^c \wedge K).$$

Proof. First notice that

$$\begin{aligned} P(E|H \wedge K) &= P(E|H \wedge G \wedge K) = P(E|H \wedge G^c \wedge K), \\ P(E|H^c \wedge K) &= P(E|H^c \wedge G \wedge K) = P(E|H^c \wedge G^c \wedge K), \end{aligned}$$

moreover, it holds

$$\begin{aligned} P(E|G \wedge K) &= P(H|G \wedge K)P(E|H \wedge G \wedge K) + P(H^c|G \wedge K)P(E|H^c \wedge G \wedge K), \\ P(E|G^c \wedge K) &= P(H|G^c \wedge K)P(E|H \wedge G^c \wedge K) + P(H^c|G^c \wedge K)P(E|H^c \wedge G^c \wedge K). \end{aligned}$$

If $P(E|H \wedge K) = P(E|H^c \wedge K)$ or $P(H|G \wedge K) = P(H|G^c \wedge K)$, then $P(E|G \wedge K) = P(E|G^c \wedge K)$, and the conclusion trivially follows. Hence, assume $P(E|H \wedge K) > P(E|H^c \wedge K)$ and $P(H|G \wedge K) > P(H|G^c \wedge K)$ and suppose by contradiction

$$P(E|G \wedge K) < P(E|G^c \wedge K).$$

Substituting the expressions of $P(E|G \wedge K)$ and $P(E|G^c \wedge K)$ in the previous inequality and simplifying we reach the conclusion

$$P(E|H \wedge K) < P(E|H^c \wedge K)$$

which is a contradiction. □

6. Detecting correlations in a coherent setting

Detecting correlations in the presence of extreme probability events by means of the definitions introduced in the previous section, involves the construction of a sequence of systems of equations to determine the zero layers of the conditional probabilities involved. The next theorem now characterizes the possible correlations between two logically independent events E and H under the hypothesis K, in terms of just the probabilities $P(H|K), P(E^*|H^* \wedge K)$ and $P(H^*|E^* \wedge K)$. The theorem thereby provides a procedure for detecting all correlations between the two events without the need to explicitly identify the zero layers for the conditional events involved.

Theorem 6.1. *Let E, H be logically independent events with respect to K and let P be a coherent conditional probability on a family of conditional events $\mathcal{G}$ containing the set $\mathcal{D} = \{E^*|H^* \wedge K, H^*|E^* \wedge K\}$, with $P(E|H \wedge K) = P(E|H^c \wedge K)$. Then, the following properties hold:*

(i) $E \perp_{cs}^{+} H|K$ if and only if one of the following conditions holds:
- *(a) $P(E|H \wedge K) = 0$ and the extension of P to $H|K$ and $H|E \wedge K$ meets either of the following conditions:*
  - *(1) $P(H|K) = 0$ and $P(H|E \wedge K) > 0$;*
  - *(2) $0 < P(H|K) < 1$ and $P(H|E \wedge K) = 1$;*
- *(b) $P(E|H \wedge K) = 1$ and the extension of P to $H|K$ and $H|E \wedge K$ meets either of the following conditions:*
  - *(1) $P(H|K) = 0$ and $P(H|E^c \wedge K) > 0$;*
  - *(2) $0 < P(H|K) < 1$ and $P(H|E^c \wedge K) = 1$;*

(ii) $E \perp_{cs}^{-} H|K$ if and only if one of the following conditions holds:
- *(c) $P(E|H \wedge K) = 0$ and the extension of P to $H|K$ and $H|E \wedge K$ meets either of the following conditions:*
  - *(1) $P(H|K) > 0$ and $P(H|E|K) = 0$;*
  - *(2) $P(H|K) = 1$ and $0 < P(H|E \wedge K) < 1$;*
- *(d) $P(E|H \wedge K) = 1$ and the extensions of P to $H|K$ and $H|E \wedge K$ meets either of the following conditions:*
  - *(1) $P(H|K) = 0$ and $P(H|E^c \wedge K) > 0$;*
  - *(2) $P(H|K) = 1$ and $0 < P(H|E^c \wedge K) < 1$.*

Proof. For proving the theorem, we take the elements of $atoms(\{E, H, K\})$ implying K which are

$$C_1 = E \wedge H \wedge K,\ C_2 = E \wedge H^c \wedge K,\ C_3 = E^c \wedge H \wedge K,\ C_4 = E^c \wedge H^c \wedge K.$$

We further consider a complete agreeing class $\{P_\alpha\}$ on $algebra(\{E, H\}) \wedge K$ that agrees with $P_{|\mathcal{D}}$, obtained by solving a sequence of systems $\mathcal{S}_\alpha$ as in Theorem 2.3.

We first prove that condition *(a)1.* implies property *(i)*: proofs of conditions *(a)2.* and *(b)* implying *(i)* are analogous. We assume that

$$P(E|H \wedge K) = P(E|H^c \wedge K) = 0$$

and, moreover, that $P(H|K) = 0$, $P(H|E \wedge K) > 0$, and we take $P(H|E \wedge K) = p \in]0, 1]$. Under these conditions, every complete agreeing class $\{P_\alpha\}$ that agrees with $P_{|\mathcal{D}}$, has $P_0(C_4) = 1$, $P_1(C_3) = 1$, $P_2(C_1) = p$ and $P_2(C_2) = 1 - p$, which implies that $o(E \wedge H \wedge K) = 2$ and $o(E \wedge H^c \wedge K) \geq 2$ while $o(H \wedge K) = 1$ and $o(H^c \wedge K) = 0$.

We conclude that $o(E|H \wedge K) < o(E|H^c \wedge K)$ and, hence, that $E \perp_{cs}^{+} H|K$.

We now prove that condition *(a)* suffices for concluding $E \perp_{cs}^{+} H|K$; the proof involving condition *(b)* is analogous. We assume

$$P(E|H \wedge K) = P(E|H^c \wedge K) = 0$$

$o(E|H \wedge K) < o(E|H^c \wedge K)$, and take $P(H|E \wedge K) = q \in\ ]0, 1]$.

We now distinguish between three cases:

- We suppose that $P(H|K) = \delta \in]0, 1[$. The only complete agreeing class satisfying $o(E|H \wedge K) < o(E|H^c \wedge K)$ is the class $\{P_0, P_1, P_2\}$ with $P_0(C_3) = \delta$, $P_0(C_4) = 1 - \delta$, $P_1(C_1) = 1$ and $P_2(C_2) = 1$, which implies that $P(H|E \wedge K) = 1$.
- We suppose that $P(H|K) = 0$. Every complete agreeing class having $P_0(C_4) = 1$, $P_1(C_1) = 0$, we consider the following possibilities for the remaining atoms:
  - if $P_1(C_2) = \delta \in\ ]0, 1[$ and $P_1(C_3) = 1 - \delta$, we must have that $P(H|E)$ equals zero and $o(E|H) = o(E|H^c)$, which contradicts our assumption;
  - if $P_1(C_3) = 1$ and, hence, $P_2(C_1) = P(H|E \wedge K) = q$ and $P_2(C_2) = 1 - q$, it follows that $o(E|H \wedge K) = o(E|H^c \wedge K)$, which contradicts our assumption;
  - if $P_1(C_2) = 1$, we must have that $P(H|E \wedge K) = 0$ and, as $P_2(C_3) = 1$ and $P_3(C_1) = 1$, also $o(E|H \wedge K) = o(E|H^c \wedge K)$, contradicting our assumption.
- We cannot have $P(H|K) = 1$, as this would contradict $o(E|H \wedge K) < o(E|H^c \wedge K)$.

□

7. Concluding observations

Based on the observation that the classical definitions of correlation can give rise to counterintuitive results in the presence of extreme probability events, we provided an enhanced definition of correlation in a coherent setting. To allow ready applicability of our definition to real-world problems, we gave a full characterization of correlations involving extreme probability events without referring to the underlying complex structure of the probability involved. Noting that our definition of correlation in a coherent setting is not symmetric, we identified a condition for symmetry to hold for cs-correlation. In the future, we will investigate how our enhanced definition of correlation can be embedded in the framework of qualitative probabilistic influence, to render this framework suitable to real-world applications involving extreme probability events. We also aim to extend the definition of of positive and negative correlation proposed here to imprecise probabilities.

Acknowledgments

The authors are members of the GNAMPA-INdAM research group. This work was partially supported by the University of Perugia, funding of 2015 Research Projects under grant "Inviluppi di probabilità condizionate coerenti e loro applicazioni statistiche".

References

Coletti, G. 1994. "Coherent numerical and ordinal probabilistic assessments." *IEEE Transactions on Systems, Man, and Cybernetics* 24 (12): 1747–1754.

Coletti, G., and R. Scozzafava. 1996. "Characterization of coherent conditional probabilities as a tool for their assessment and extension." *Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 44: 101–132.

Coletti, G., and R. Scozzafava. 1998. "Null events and stochastical independence." *Kybernetika* 34: 69–78.

Coletti, G., and R. Scozzafava. 1999. "Conditioning and inference in intelligent systems." *Soft Computing* 3 (3): 118–130.

Coletti, G., and R. Scozzafava. 2000. *Zero probabilities in stochastic independence*, 185–196. Kluwer.

Coletti, G., and R. Scozzafava. 2002a. *Probabilistic Logic in a Coherent Setting.* Vol. 15 of *Trends in Logic.* Dordrecht/Boston/London: Kluwer Academic Publisher.

Coletti, G., and R. Scozzafava. 2002b. "Stochastic Independence in a Coherent Setting." *Annals of Mathematics and Artificial Intelligence* 35: 151–176.

Coletti, G., R. Scozzafava, and B. Vantaggi. 2001. "Probabilistic reasoning as a general unifying tool." *Lecture Notes* 2143: 120–131.

Coletti, G., L.C. van der Gaag, and D. Petturiti. 2018. "Detecting Correlation Between Extreme Probability Events." In *Proceedings of the 11th Workshop on Uncertainty Processing (WUPES'18) Třeboň, Czech Republic*, edited by J. Vejnarov'à V. Kratochvìl, 37–48.

de Finetti, B. 1936. "Les probabilités nulles." *Bulletin des Sciences Mathematiques* 40: 275–288.

de Finetti, B. 1949. "Sull'impostazione assiomatica del calcolo delle probabilità." *Annali Triestini* 19: 29–81.

de Finetti, B. 1972. *Probability, Induction and Statistics: The art of guessing.* John Wiley & Sons, London, New York, Sydney, Toronto.

Druzdzel, M.J., and M. Henrion. 1993. "Efficient reasoning in qualitative probabilistic networks." In *Proceedings of the Eleventh National Conference on Artificial Intelligence, AAAI Press, Menlo Park*, 548 – 553.

Dubins, L.E. 1975. "Finitely Additive Conditional Probabilities, Conglomerability and Disintegrations." *The Annals of Probability* 3 (1): 89–99.

Flach, P. 2012. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data.* Cambridge University Press.

Krauss, P.H. 1968. "Representation of conditional probability measures on Boolean algebras." *Acta Mathematica Hungarica* 19 (3): 229–241.

Lauritzen, S.L., and D.J. Spiegelhalter. 1988. "Local computations with probabilities on graphical structures and their application to expert systems." *Journal of the Royal Statistical Society, Series B* 50: 157 – 224.

Niblett, T., and I. Bratko. 1987. "Learning Decision Rules in Noisy Domains." In *Proceedings of Expert Systems '86, The 6Th Annual Technical Conference on Research and Development in Expert Systems III*, 25–34. Cambridge University Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference.* Morgan Kaufmann.

Russell, S., and P. Norvig. 2003. *Artificial Intelligence: A Modern Approach.* 2nd ed. Upper Saddle River: Prentice Hal.

Scozzafava, R. 1984. "A survey of some common misunderstandings concerning the role and meaning of finitely additive probabilities in statistical inference." *Statistica* 44: 21–45.

Scozzafava, R., and B. Vantaggi. 2004. "The role of zero probabilities in dealing with zero frequency problems." In *Soft Methodology and Random Information Systems*, 265–272.

van der Gaag, L.C., and A. Capotorti. 2018. "Naive Bayesian classifiers with extreme probability features." In *Journal of Machine Learning Research, Proceedings Series*, Vol. 72, 499–510.

Vantaggi, B. 2001. "Conditional independence in a finite additive setting." *Annals of Mathematics and Artificial Intelligence* 32: 287–314.

Vantaggi, B. 2002. "The L-separation criterion for description of cs-independence models." *International Journal of Approximate Reasoning* 29: 291–316.

Vantaggi, B. 2003. *Qualitative Bayesian Networks with Logical Constraints*, Vol. 2711 LNAI, 100–112.

Wellman, M.P. 1990. "Fundamental concepts of qualitative probabilistic networks." *Artificial Intelligence* 44: 257 – 303.