Detecting Correlation Between Extreme Probability Events

Giulianella Coletti
Dept. Mathematics and Computer Science
University of Perugia, Italy
giulianella.coletti@unipg.it

Linda C. van der Gaag
Dept. Information and Computing Sciences
University of Utrecht, NL
l.c.vandergaag@uu.nl

Davide Petturiti
Dept. Economics
University of Perugia, Italy
davide.petturiti@unipg.it

Abstract

Since the classical definitions of correlation give rise to counterintuitive findings for extreme probability events, we build upon the concept of coherent conditional probability to introduce enhanced notions of correlation. Our new notions allow handling extreme events in a principled way by accommodating the different levels of strength of the zero probabilities involved. Where the detection of correlations by means of these levels is computationally challenging, we provide a full characterisation of the correlations between extreme probability events without reference to the complex structure of probability.

1 Introduction

The importance of handling *extreme probability events* in a principled way has been stressed in a range of papers (see for example, [3, 5, 7]); 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 and, hence, in real-world applications involving infinite settings, where the lack of expressive power of the real numbers often forces possible events to be assigned zero probability. Yet, also in finite settings do extreme probabilities arise. When extracting (conditional) probabilities from real-world data, for example, unexpected events and events occurring with negligible frequency will receive zero probabilities. To forestall the inclusion of zero probabilities in probabilistic models, various more or less "ad hoc" solutions are in use, such as the well-known Laplace correction and the use of pseudocounts in a Bayesian setting. Forcing all distinguished events to have positive probability however, drastically restricts the class of admissible distributions and, hence, the possibilities of extending partial assessments to complete probabilities.

In applications of probability theory, stochastic independence and the concepts of positive and negative correlation play an important role. While in the context of extreme probabilities stochastic independence has been well studied (see for example [3, 5, 7]), and has led to an enhanced definition of independence, the concept of correlation has received little to no attention. In this paper, we demonstrate that in the presence of extreme probability events, the classical definition of correlation can give counterintuitive results, such as an event E being uncorrelated with an event H logically implying it. Based on these observations, we introduce enhanced notions of correlation which accommodate the different levels of *strength* of the zero probabilities involved. We develop the notions of positive and negative correlation in a coherent setting, referring to full conditional probabilities represented by their complete agreeing classes which in turn define the zero layers of the events of interest. Although the framework of coherent setting constitutes the principle on which our enhanced notion of correlation is founded, referring to zero layers does not provide for practicable application in real-world settings, as a consequence of the computational challenges involved. We therefore provide also a full characterisation of the correlations involving extreme probability events without reference to the complex structure of probability.

The paper is organised as follows. Section 2 presents some preliminaries on coherent conditional probability and thereby introduces our notational conventions. In Section 3, we present our concepts of positive and negative correlation in a coherent setting and introduce some of their properties. Section 4 then provides the characterisation of all correlations involving extreme probability events. Section 5 concludes the paper with our plans for further research.

2 Preliminaries

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 use E^* to indicate either E itself or its contrary E^c. A *conditional event* $E \mid H$ is 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 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 $\wedge$. For any Boolean algebra $\mathcal{A}$, we use $\mathcal{A}^0$ to denote $\mathcal{A}\backslash\{\varnothing\}$. For an arbitrary family of events $\mathcal{E}$, 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*$(\mathcal{E})$. We will restrict our further discussion to finite Boolean algebras.

In this paper, we build on the following axiomatic definition of *conditional probability* which dates back to de Finetti [8], and has been explicitly formulated, with minor differences, by Dubins [9] and Krauss [10].

Definition 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 write $P(E) = P(E\,|\,\Omega)$, for every $E \in \mathcal{A}$. Following Dubins, we say that a conditional probability $P(\cdot\,|\,\cdot)$ is *full on* $\mathcal{A}$ if it is defined on $\mathcal{A} \times \mathcal{A}^0$, that is, if $\mathcal{H} = \mathcal{A}^0$. 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 $\mathcal{A} \times \mathcal{A}^0$ [9].

For any Boolean algebra of events $\mathcal{A}$, every full conditional probability $P(\cdot\,|\,\cdot)$ on $\mathcal{A}$ has a one-to-one correspondence with a linearly ordered class $\{P_0, \ldots, P_k\}$ of (unconditional) probabilities on $\mathcal{A}$, called its *complete agreeing class*, whose supports form a partition of Ω. For a given full conditional probability $P(\cdot\,|\,\cdot)$, its class $\{P_0, \ldots, P_k\}$ is obtained by setting

- $P_0(\cdot) = P(\cdot\,|\,H_0^0)$, with $H_0^0 = \Omega$;
- for each successive α, $P_\alpha(\cdot) = P(\cdot\,|\,H_0^\alpha)$, with $H_0^\alpha = \bigvee_{H \subseteq H_0^{\alpha-1}, P_{\alpha-1}(H)=0} H \neq \varnothing$;

with the iterative construction halting when $H_0^{k+1} = \varnothing$. We note that for every event $H \in \mathcal{A}^0$, there is an index $\alpha \in \{0, \ldots, k\}$ with $P_\alpha(H) > 0$. Moreover, for every conditional event $E \mid H \in \mathcal{A} \times \mathcal{A}^0$ and α_H being the minimum index in $\{0, \ldots, k\}$ with $P_{\alpha_H}(H) > 0$, we have that

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

Having so far addressed full conditional probabilities on an algebra $\mathcal{A}$, we now consider arbitrary, possibly partially specified, conditional probabilities.

Definition 2. *Let $\mathcal{G} = \{E_j\,|\,H_j\}_{j\in J}$, with J a finite index set, be an arbitrary family of conditional events. A* ***coherent conditional probability*** *on $\mathcal{G}$ is a function $P : \mathcal{G} \to [0,1]$ for which there exists a conditional probability $P'\colon \mathcal{A} \times \mathcal{H} \to [0,1]$, with $\mathcal{A} = algebra(\{E_j, H_j\}_{j\in J})$ and $\mathcal{H} = additive(\{H_j\}_{j\in J})$, such that $P'_{|\mathcal{G}} = P$.*

We note that, since every conditional probability P' on $\mathcal{A}\times\mathcal{H}$ can be extended to a full conditional probability on $\mathcal{A}$, Definition 2 can also be formulated by requiring the existence of a full conditional probability on $\mathcal{A}$ extending the original function P. In the sequel, we will use the phrase *assessment* to denote a function P for which coherence has yet to be established. The following theorem now specifies several characterisations of coherence for such an assessment, relevant to our current context; for proofs of the equivalences stated in the theorem, we refer to [1, 2, 4].

Theorem 1. *Let $\mathcal{G} = \{E_j \mid H_j\}_{j \in J}$, with J a finite index set, be an arbitrary family of conditional events. Then, for any function $P : \mathcal{G} \to [0,1]$, the following statements are equivalent:*

(i) *P is a coherent conditional probability;*

(ii) *There exists a complete agreeing class $\{P_0, \ldots, P_k\}$, $k \geq 0$, of probabilities P_α on algebra$(\{E_j, H_j\}_{j\in J})$ such that, for every $j \in J$, if α_j is the minimum index in $\{0, \ldots, k\}$ with $P_{\alpha_j}(H_j) > 0$, then*

$$P(E_j \,|\, H_j) = \frac{P_{\alpha_j}(E_j \wedge H_j)}{P_{\alpha_j}(H_j)};$$

(iii) *With the atom sets $\mathcal{C}_0 = atoms(\{E_j, H_j\}_{j\in J})$ and, for $\alpha = 1, \ldots, k$, $\mathcal{C}_\alpha = \{C_r \in \mathcal{C}_{\alpha-1} \mid P_{\alpha-1}(C_r) = 0\}$, all systems of equations $\mathcal{S}_\alpha$ in the sequence of systems $\{\mathcal{S}_0, \ldots, \mathcal{S}_k\}$, $k \geq 0$, with non-negative unknowns $x_r^\alpha = P_\alpha(C_r)$ for all $C_r \in \mathcal{C}_\alpha$, are compatible:*

$$\mathcal{S}_\alpha : \begin{cases} \sum\limits_{C_r \in \mathcal{C}_\alpha, C_r \subseteq E_j \wedge H_j} x_r^\alpha = P(E_j \,|\, H_j) \cdot \sum\limits_{C_r \in \mathcal{C}_\alpha, C_r \subseteq H_j} x_r^\alpha, \text{ for all } j \in J \text{ with } P_{\alpha-1}(H_j) = 0 \\ \sum\limits_{C_r \in \mathcal{C}_\alpha} x_r^\alpha = 1. \end{cases}$$

Of the sequence of systems $\mathcal{S}_0, \ldots, \mathcal{S}_k$ introduced in Theorem 1(iii), every sequence of solutions $\{\mathbf{x}^0, \ldots, \mathbf{x}^k\}$ defines a complete agreeing class $\{P_0, \ldots, P_k\}$ on the algebra *algebra*$(\{E_j, H_j\}_{j\in J})$ by setting $P_0(C_r) = x_r^0$ for all $C_r \in \mathcal{C}_0$, and for each successive $\alpha = 1, \ldots, k$, setting

$$P_\alpha(C_r) = 0 \text{ for every } C_r \in \mathcal{C}_0 \setminus \mathcal{C}_\alpha, \quad \text{and} \quad P_\alpha(C_r) = x_r^\alpha \text{ for every } C_r \in \mathcal{C}_\alpha,$$

and then extending each probability P_α by additivity. In turn, the complete agreeing class $\{P_0, \ldots, P_k\}$ described in Theorem 1(ii) has a one-to-one correspondence with a full conditional probability $P'(\cdot \,|\, \cdot)$ on *algebra*$(\{E_j, H_j\}_{j\in J})$ extending P.

To conclude our preliminaries, we recall the concept of *zero layer* [6], which naturally arises from the structure of conditional probability described in Theorem 1.

Definition 3. *Let $\mathcal{A}$ be a Boolean algebra of events and let $P(\cdot \,|\, \cdot)$ be a full conditional probability on $\mathcal{A}$ represented by the complete agreeing class $\{P_0, \ldots, P_k\}$ of probabilities 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*

$$o(H) = \min\{\alpha \in \{0, \ldots, k\} \,:\, P_\alpha(H) > 0\},$$

with the zero layer of the impossible event equal to $o(\varnothing) = +\infty$*. For every conditional event* $E \mid H \in \mathcal{A} \times \mathcal{A}^0$*, the* **zero layer** *of* $E \mid H$ *with respect to* $\{P_0, \ldots, P_k\}$ *is the non-negative number*

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

We note that, for any event E with $P(E) = P(E \mid \Omega) > 0$, we have that $o(E) = 0$. We further note that $P(E \mid H) > 0$ iff $o(E \wedge H) = o(H)$ and hence $o(E \mid H) = 0$.

3 Positive and negative correlation

Before defining our enhanced concept of corrrelation, we review the classical definition of correlation between two events, stated in terms of coherence.

Definition 4. *Let* P *be a coherent conditional probability defined on an arbitrary family of events* $\mathcal{G}$ *with* $E, E \mid H \in \mathcal{G}$*. Then,*

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

Various properties of correlation having been formulated for the classical setting, we review in the following proposition some properties through which we will demonstrate the inadequacy of the classical definitions for describing correlation in the presence of extreme probability events.

Proposition 1. *Let* $\mathcal{G}$ *be an arbitrary family of conditional events including* E^*, H^*, $E^* \mid H^*$*. 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 (or, alternatively: 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* $P(E \mid H) > (<) P(E \mid H^c)$*.*

Proof. The properties (i) and (iii) follow directly from Definition 1. The first part of property (ii) follows from the observation that $E \wedge H = \varnothing$ implies $P(E \mid H) = 0 < P(E)$. As $E^c \wedge H^c = \varnothing$ implies $P(E^c \mid H^c) = 0 < P(E^c)$, we have by property (i) that $P(E \mid H) < P(E)$. 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 1 strictly depends on the premise that the probabilities of E and H are different from 0 and 1. For property (iii), moreover, the implication $P(E \mid H) > P(E) \Rightarrow P(E \mid H) > P(E \mid H^c)$ 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 the presence of extreme probability events, we consider an event H with $P(H) = 1$. By Definition 4, this event is not correlated with any other event, as for any event $E \neq H$ we would find that $P(E \mid H) = P(E)$. We would find the exact same result, in fact, also for an event E which logically contradicts H, 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 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 1 for the basic definition of correlation instead, that is, use Definition 5 below.

Definition 5. *Let P be a coherent conditional probability defined on an arbitrary family of events $\mathcal{G}$ with $E\,|\,H, E\,|\,H^c \in \mathcal{G}$. 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 and 5 are not equivalent: while Definition 4 implies Definition 5, the reverse does not hold. In fact, by Definition 5, 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 $\alpha_H > 0$ such that $P_{\alpha_H}(H) > 0$ and

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

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 5 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 definitions, 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 1. *Let Ω be the unit square $[0,1]^2$. 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; let the event $H = \{(x,y) \mid x = y,\ x, y \in [0,1]^2\}$ be the diagonal of the unit square. In this setting, we consider the following assessment for a family of four conditional events:*

$$P(E\,|\,H) = P(E\,|\,H^c) = P(H\,|\,E^c) = 0, \quad P(H\,|\,E) = \frac{2}{3}$$

For proving coherence, we consider the set atoms$(\{E, H\}) = \{C_1, C_2, C_3, C_4\}$ *with*

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

and build the sequence of systems $\mathcal{S}_\alpha$ *with non-negative unknowns* $x_i^\alpha = P_\alpha(C_i)$*, as described in Theorem 3. The first system equals*

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

which has $x_1^0 = x_2^0 = x_3^0 = 0$*,* $x_4^0 = 1$*, for its solution. Then, focusing on the zero-probability atoms and writing* x_i^1 *for* x_i^0*, the second system is found to be*

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

which has $x_1^1 = x_2^1 = 0$*,* $x_3^1 = 1$*, for its unique solution. The third system equals*

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

and has $x_1^2 = \frac{2}{3}$*,* $x_2^2 = \frac{1}{3}$ *for its sole solution. Since every constructed system of the sequence has a unique solution, the assessment* P *has a unique complete agreeing class* $\{P_0, P_1, P_2\}$*. 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 taking the event H *for the hypothesis thus is smaller than that of taking* H^c *for the hypothesis. As the conditional event* $E \mid H^c$ *still has zero probability in the structure when* $E \mid H$ *does not, this finding may be naturally construed as a positive correlation of* E *and* H*.*

We further consider the incompatible events R *and* H*. Analogously to the above example, 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 probability than the zero probability of any possible event.*

Based on the considerations in the above example, we now introduce our definition of correlation of extreme probability events in a coherent setting.

Definition 6. *Let P be a coherent conditional probability defined on an arbitrary family of conditional events $\mathcal{G}$ containing $E^* \,|\, H^*, H^* \,|\, E^*$. We say that:*

- *E is* **positively correlated in a coherent setting** *with H, denoted as $E \perp_{cs}^{+} H$, if one of the following conditions holds:*
  - $P(E \,|\, H) > P(E \,|\, H^c)$;
  - $P(E \,|\, H) = P(E \,|\, H^c) = 0$, *and every complete agreeing class $\{P_\alpha\}$ on algebra$(\{E, H\})$ that agrees with P on $\mathcal{D}$, has*

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

  - $P(E \,|\, H) = P(E \,|\, H^c) = 1$, *and every complete agreeing class $\{P_\alpha\}$ on algebra$(\{E, H\})$ that agrees with P on $\mathcal{D}$, has*

$$o(E^c \,|\, H) > o(E^c \,|\, H^c);$$

- *E is* **negatively correlated in a coherent setting** *with H, denoted as $E \perp_{cs}^{-} H$, if one of the following conditions holds:*
  - $P(E \,|\, H) < P(E \,|\, H^c)$;
  - $P(E \,|\, H) = P(E \,|\, H^c) = 0$, *and every complete agreeing class $\{P_\alpha\}$ on algebra$(\{E, H\})$ that agrees with P on $\mathcal{D}$, has*

$$o(E \,|\, H) > o(E \,|\, H^c),$$

  - $P(E \,|\, H) = P(E \,|\, H^c) = 1$, *and every complete agreeing class $\{P_\alpha\}$ on algebra$(\{E, H\})$ that agrees with P on $\mathcal{D}$, has*

$$o(E^c|H) < o(E^c|H^c);$$

- *E is* **not correlated in a coherent setting** *with H, denoted as $E \not\perp_{cs} H$, if it is not positively nor negatively correlated in a coherent setting with H.*

The above definition of 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, the definition allows the identification of a correlation between events which are logically related, as shown in the following theorem.

Theorem 2. *Let P be a coherent conditional probability defined on an arbitrary family of conditional events $\mathcal{G}$ containing $E^* \mid H^*, H^* \mid E^*$. Then, the following properties hold:*

(*i*) *if either* $E \wedge H = \emptyset$ *or* $E^c \wedge H^c = \emptyset$, *then* $E \perp_{cs}^{-} H$;

(*ii*) *if either* $E^c \wedge H = \emptyset$ *or* $E \wedge H^c = \emptyset$ *then* $E \perp_{cs}^{+} H$.

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

We note that the correlations $\perp_{cs}^{+}$ and $\perp_{cs}^{-}$ introduced above generally are not symmetric, as demonstrated by the following example.

Example 2. *We address two events* E, H *which are logically independent, and consider the following coherent probability assessment for these events:*

$$P(E \mid H) = \frac{3}{4},\ P(E \mid H^c) = \frac{1}{4},\ P(E) = \frac{1}{4},\ P(H) = 0.$$

By definition, we have that E *is positively correlated with* H *and, hence,* $E \perp_{cs}^{+} H$.

We now address the way in which H *is correlated with* E. *Building upon the set of atoms atoms*$(\{E, H\}) = \{C_1, C_2, C_3, C_4\}$ *with*

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

we consider the sequence of systems $\mathcal{S}_\alpha$ *with non-negative unknowns* $x_r^\alpha = P_\alpha(C_r)$ *as before. The first system of equations equals*

$$\mathcal{S}_0 : \begin{cases} x_1^0 = \frac{3}{4} \cdot (x_1^0 + x_3^0) \\ x_2^0 = \frac{1}{4} \cdot (x_2^0 + x_4^0) \\ x_1^0 + x_2^0 = \frac{1}{4} \cdot (x_1^0 + x_2^0 + x_3^0 + x_4^0) \\ x_1^0 + x_3^0 = 0 \cdot (x_1^0 + x_2^0 + x_3^0 + x_4^0) \\ x_1^0 + x_2^0 + x_3^0 + x_4^0 = 1 \end{cases}$$

having $x_1^0 = x_3^0 = 0$, $x_2^0 = \frac{1}{4}$, $x_4^0 = \frac{3}{4}$ *for its solution. The second system then is*

$$\mathcal{S}_1 : \begin{cases} x_1^1 = \frac{3}{4} \cdot (x_1^1 + x_3^1) \\ x_1^1 + x_3^1 = 1 \end{cases}$$

whose unique solution is $x_1^1 = \frac{3}{4}$, $x_3^1 = \frac{1}{4}$. *These solutions determine the unique agreeing class* $\{P_0, P_1\}$ *on the algebra algebra*$(\{E, H\})$, *which, in turn, has a one-to-one correspondence with a full conditional probability* P' *on algebra*$(\{E, H\})$ *extending* P. *This probability* P' *has*

$$P'(H \mid E) = \frac{P_0(H \wedge E)}{P_0(E)} = 0 = \frac{P_0(H \wedge E^c)}{P_0(E^c)} = P'(H \mid E^c),$$

and, hence,

$$o(H \,|\, E) = o(H \wedge E) - o(E) = 1 = o(H \wedge E^c) - o(E^c) = o(H \,|\, E^c),$$

from which we find that $H \not\perp_{cs}^{+} E$*. We conclude that, while* E *is positively correlated with* H*, the reverse does not hold.*

For symmetric concepts of positive and negative correlation in a coherent setting, the definitions of $\perp_{cs}^{+}$ and $\perp_{cs}^{-}$ need be further enhanced, by setting

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

Because of space limitations, we do not further elaborate on this enhancement.

4 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, in terms of just the probabilities $P(H), P(E^* \mid H^*)$ and $P(H^* \mid E^*)$. The theorem thereby provides for detecting all correlations between the two events without the need to explicitly identify the zero layers for the conditional events involved.

Theorem 3. *Let* E, H *be logically independent events and let* P *be a coherent conditional probability on a family of conditional events* $\mathcal{G}$ *containing the subset* $\mathcal{D} = \{E^* \,|\, H^*, H^* \,|\, E^*\}$*, with* $P(E \,|\, H) = P(E \,|\, H^c)$*. Then, the following properties hold:*

- (i) $E \perp_{cs}^{+} H$ *if and only if one of the following conditions holds:*
  - (a) $P(E \mid H) = 0$ *and all extensions of* P *to* $H, H \mid E$ *meet either of the following conditions:*
    1. $P(H) = 0$ *and* $P(H \,|\, E) > 0$*;*
    2. $0 < P(H) < 1$ *and* $P(H \,|\, E) = 1$*;*
  - (b) $P(E \,|\, H) = 1$ *and all extensions of* P *to* H *and* $H|E$ *meet either of the following conditions:*
    1. $P(H) = 0$ *and* $P(H \,|\, E^c) > 0$*;*
    2. $0 < P(H) < 1$ *and* $P(H \,|\, E^c) = 1$*;*
- (ii) $E \perp_{cs}^{-} H$ *if and only if one of the following conditions holds:*
  - (c) $P(E \mid H) = 0$ *and all extensions of* P *to* $H, H \mid E$ *meet either of the following conditions:*

1. $P(H) > 0$ *and* $P(H\,|\,E) = 0$*;*

2. $P(H) = 1$ *and* $0 < P(H\,|\,E) < 1$*;*

(d) $P(E \mid H) = 1$ *and all extensions of* P *to* $H, H \mid E$ *meet either of the following conditions:*

1. $P(H) = 0$ *and* $P(H\,|\,E^c) > 0$*;*

2. $P(H) = 1$ *and* $0 < P(H\,|\,E^c) < 1$*.*

Proof. For proving the theorem, we take *atoms*$(\{E, H\}) = \{C_1, C_2, C_3, C_4\}$ with

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

We further consider a complete class $\{P_\alpha\}$ on *algebra*$(\{E, H\})$ agreeing with the restriction of P to $\mathcal{D}$, obtained by solving a sequence of systems $\mathcal{S}_\alpha$ as in Theorem 1.

We first prove that condition (a)*1.* implies property (i); proofs of the conditions (a)*2.* and (b) implying (i) are analogous. We assume that $P(E\,|\,H) = P(E\,|\,H^c) = 0$ and, moreover, that $P(H) = 0$ and $P(H\,|\,E) > 0$; we take $P(H\,|\,E) = p \in\ ]0, 1]$. Under these conditions, every complete agreeing class $\{P_\alpha\}$ that agrees with P on $\mathcal{D}$, 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 $o(E \wedge H) = 2$ and $o(E \wedge H^c) \geq 2$ while $o(H) = 1$ and $o(H^c) = 0$. We conclude that $o(E\,|\,H) < o(E\,|\,H^c)$ and, hence, that $E \perp^+_{cs} H$.

We now prove that condition (a) suffices for concluding $E \perp^+_{cs} H$; the proof involving condition (b) is analogous. We assume $P(E\,|\,H) = P(E\,|\,H^c) = 0$, $o(E\,|\,H) < o(E \mid H^c)$, and take $P(H \mid E) = q \in\ ]0, 1]$. We now distinguish between three cases:

- We suppose that $P(H) = \delta \in]0, 1[$. The only complete agreeing class satisfying $o(E \mid H) < o(E \mid H^c)$ 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 $P(H \mid E) = 1$.
- We suppose that $P(H) = 0$. 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, $P_2(C_1) = P(H\,|\,E) = q$ and $P_2(C_2) = 1 - q$, it follows that $o(E\,|\,H) = o(E\,|\,H^c)$, which contradicts our assumption;
  - if $P_1(C_2) = 1$, we must have that $P(H\,|\,E) = 0$ and, as $P_2(C_3) = 1$ and $P_3(C_1) = 1$, also $o(E\,|\,H) = o(E\,|\,H^c)$, contradicting our assumption.
- We cannot have $P(H) = 1$, as this would contradict $o(E\,|\,H) < o(E\,|\,H^c)$.

□

5 Concluding observations

Based on the observation that the classical definitions of correlation can give 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 applications, we gave a full characterisation of correlations involving extreme probability events without referring to the underlying complex structure of the probability involved. We noted that our definition of correlation in a coherent setting is not symmetric; as a next step in our research, we will address this asymmetry by studying the conditions under which it occurs. 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.

References

[1] G. Coletti. Coherent numerical and ordinal probabilistic assessments. *IEEE Transactions on Systems, Man and Cybernetics*, 24(0):1747–1754, 1994.

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

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

[4] G. Coletti and R. Scozzafava. Conditioning and inference in intelligent systems. *Soft Computing*, 3:118–130, 1999.

[5] G. Coletti and R. Scozzafava. *Zero probabilities in stochastic independence*, volume 516 of *The Springer International Series in Engineering and Computer Science*, pages 185–196. Springer, Boston, 2000.

[6] G. Coletti and R. Scozzafava. Probabilistic logic in a coherent setting. volume 15 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht, 2002.

[7] G. Coletti and R. Scozzafava. Stochastic independence in a coherent setting. *Annals of Mathematics and Artificial Intelligence*, 35:151–176, 2002.

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

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

[10] P. Krauss. Representation of conditional probability measures on boolean algebras. *Acta Mathematica Academiae Scientiarum Hungarica*, 19(3-4):229–241, 1968.