Conditional independence concept in various uncertainty calculi

Milan Studený

Institute of Information Theory and Automation of the CAS, Prague, Czech Republic

The 4-th International Conference on *Belief Functions*

BELIEF 2016

Prague, Czech Republic, September 22, 2016, 9:00-10:00

Summary of the talk

1. Introduction: history overview
   - stochastic conditional independence
   - conditional independence beyond probabilistic reasoning
2. Stochastic conditional independence
   - equivalent definitions in discrete case
   - basic formal properties of stochastic conditional independence
3. The concepts of a semi-graphoid and separoid
   - definitions
   - examples from various areas
4. Conditional independence in other uncertainty calculi
   - possibility theory
   - DS theory of evidence
5. Conclusions

History overview: stochastic conditional independence I

Already in the 1950s, Loève in his book on probability theory defined the concept of *conditional independence* (CI) is terms of σ-algebras.

M. Loève (1995). Probability Theory, Foundations, Random Processes. D. van Nostrand, Toronto.

Phil Dawid was probably the first statistician who explicitly formulated certain basic formal properties of stochastic CI.

A. P. Dawid (1979). Conditional independence in statistical theory. Journal of the Royal Statistical Society B 41, 1-31.

He observed that several statistical concepts, e.g. the one of a sufficient statistics, can equivalently be defined in terms of generalized CI.

This observation allows one to derive many results on those statistical concepts in an elegant way, using the formal properties of CI.

History overview: stochastic conditional independence II

These basic formal properties of stochastic CI were independently formulated in the context of philosophical logic by Spohn, who was interested in the interpretation of CI and its relation to causality.

W. Spohn (1980). Stochastic independence, causal independence and shieldability. Journal of Philosophical Logic 9 (1), 73-99.

The same properties, this time formulated in terms of σ-algebras, were also explored by statistician Mouchart and probabilist Rolin.

M. Mouchart and J.-M. Rolin (1984). A note on conditional independence with statistical applications. Statistica 44 (4), 557-584.

Allegedly, the conditional independence symbol $\perp\!\!\!\perp$ was proposed by Dawid and Mouchart during their joint discussion in the end of the 1970s.

History overview: stochastic conditional independence III

The significance of the concept of CI for *probabilistic reasoning* was later recognized by Pearl and Paz, who observed that the above basic formal properties of CI are also valid for certain ternary separation relations induced by undirected graphs.

J. Pearl and A. Paz (1987). Graphoids, graph-based logic for reasoning about relevance relations. In Advances in Artificial Intelligence II. North-Holland, Amsterdam, 357-363.

This led them to the idea describe such formal ternary relations by graphs and introduced an abstract concept of a *semi-graphoid*. Even more abstract concept of a *separoid* was later suggested by Dawid.

A. P. Dawid (2001). Separoids: a mathematical framework for conditional independence and irrelevance. Annals of Mathematics and Artificial Intelligence 32 (1/4), 335-372.

History overview: stochastic conditional independence IV

Pearl and Paz (1987) also raised a conjecture that semi-graphoids coincide with stochastic CI structures, which was later refuted.

M. Studený (1992). Conditional independence relations have no finite complete characterization. In Transactions of 11th Prague Conference B. Kluwer, Dordrecht, 377-396.

A lot of effort and time was devoted to the task to characterize all possible CI structures induced by four discrete random variables. The final solution to that problem was achieved by Matúš.

F. Matúš (1999). Conditional independences among four random variables III., final conclusion. Combinatorics, Probability and Computing 8 (3), 269-276.

P. Šimeček (2007). Independence models (in Czech). PhD thesis, Charles University, Prague.

Šimeček computed that the number of these structures is 18478 and they decompose into 1098 types.

History overview: graphical description

The traditional way of (sketchy) description of (stochastic) CI structures was to use graphs whose nodes correspond to random variables.

This idea had appeared in statistics earlier than Pearl and Paz suggested that in the context of computer science.

One can distinguish two basic trends, namely

- using undirected graphs, and
- using directed (acyclic) graphs.

The theoretical breakthrough leading to (graphical) probabilistic expert systems was the *local computation method*.

S. L. Lauritzen and D. J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society B 50 (2), 157-224.

Overview: non-probabilistic conditional independence I

Nevertheless, the probability theory and statistics is not the only field in which the concept of CI was introduced and examined.

An analogous concept of *embedded multivalued dependency* (EMVD) was studied in theory of relational databases. Sagiv and Walecka showed that there is no finite axiomatic characterization of EMVD structures.

Y. Sagiv and S. F. Walecka (1982). Subset dependencies and completeness result for a subclass of embedded multivalued dependencies. Journal of Association for Computing Machinery 29 (1), 103-117.

Shenoy observed that one can introduce the concept of CI within various calculi for dealing with knowledge and uncertainty in artificial intelligence.

P. P. Shenoy (1994). Conditional independence in valuation-based systems. International Journal of Approximate Reasoning 10 (3), 203-234.

Overview: non-probabilistic conditional independence II

Shenoy's work gave inspiration to several papers on formal properties of CI in various calculi for dealing with knowledge and uncertainty in artificial intelligence.

For example, Vejnarová compared formal properties of CI concepts arising in the frame of *possibility theory*.

J. Vejnarová (2000). Conditional independence in possibility theory. International Journal of Uncertainty and Fuzziness Knowledge-Based Systems 12, 253-269.

As concerns Spohn's calculus of ordinal conditional functions, it was shown that there is no finite axiomatization of CI structures arising in the context of *natural conditional functions*.

M. Studený (1995). Conditional independence and natural conditional functions. International Journal of Approximate Reasoning 12 (1), 43-68.

Overview: non-probabilistic conditional independence III

At least two concepts of CI were proposed in the context of the *Dempster-Shafer theory of evidence*.

B. Ben Yaghlane, P. Smets, K. Mellouli (2002). Belief function independence II. International Journal of Approximate Reasoning 31, 31-75.

R. Jiroušek and J. Vejnarová (2011). Compositional models and conditional independence in evidence theory. International Journal of Approximate Reasoning 52, 316-334.

Various concepts of conditional irrelevance have also been introduced and their formal properties were examined within the theory of *imprecise probabilities*; let us mention the concept of epistemic irrelevance.

F. G. Cozman and P. Walley (2005). Graphoid properties of epistemic irrelevance and independence. Annals of Mathematics and Artificial Intelligence 45 (1/2), 173-195.

Basic concepts: discrete probability distribution

Definition (probability density)

A *discrete probability measure over* (a finite set) N is defined as follows:

(i) For every $i \in N$ a non-empty finite set X_i is given, which is the *individual sample space* for the variable i. This defines a *joint sample space*, which is the Cartesian product $\mathsf{X}_N := \prod_{i\in N} \mathsf{X}_i$.

(iii) A probability measure P on X_N is given; it is determined by its *probability density*, which is a function $p : \mathsf{X}_N \to [0,1]$ such that $\sum_{\mathbf{x}\in\mathsf{X}_N} p(\mathbf{x}) = 1$. Then $P(\mathbb{A}) = \sum_{\mathbf{x}\in\mathbb{A}} p(\mathbf{x})$ for any $\mathbb{A} \subseteq \mathsf{X}_N$.

Some conventions:

Given $A \subseteq N$, any list of elements $[x_i]_{i\in A}$ such that $x_i \in \mathsf{X}_i$ for $i \in A$ will be named a *configuration* for A. X_A is the set of configurations for A.

Given disjoint $A, B \subseteq N$, the concatenation AB is a shorthand for union $A \cup B$.

In case $A \subseteq B$ and $\mathbf{b} \in \mathsf{X}_B$ the symbol $\mathbf{b}_A$ will denote the *restriction* of the configuration $\mathbf{b}$ for A, that is, the restricted list. ($=$ a *marginal configuration*)

Given $i \in N$ the symbol i will be used as an abbreviation for the singleton $\{i\}$.

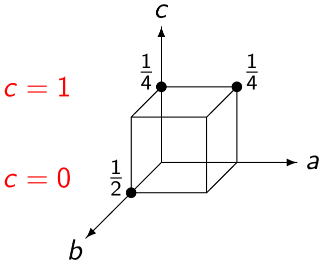
Example: a density of a probability distribution

$N = \{a, b, c\}$ and $\mathsf{X}_i = \{0, 1\}$ for any $i \in N$.

Thus, one has 8 joint configurations: $|\mathsf{X}_N| = 8$.

We put $p(0,0,0) = 0$, $p(0,0,1) = 1/4$, $p(0,1,0) = 1/2$, $p(0,1,1) = 0$,
$p(1,0,0) = 0$, $p(1,0,1) = 1/4$, $p(1,1,0) = 0$, $p(1,1,1) = 0$.

Here, an implicit order of variables is a (= the first one), b and c.



The function p is indeed a density (of a probability distribution).

Basic concepts: marginal density in the discrete case

Definition (marginal density)

Given $A \subseteq N$ and a discrete probability measure P over N, the *marginal measure* P_A is defined by its *marginal density* $p_A : \mathsf{X}_A \to [0,1]$, given by the formula

$$\boxed{p_A(\mathbf{a}) \;:=\; \textstyle\sum_{\mathbf{c}\in \mathsf{X}_{N\setminus A}} p(\mathbf{a},\mathbf{c})} \qquad \text{for } \mathbf{a} \in \mathsf{X}_A,$$

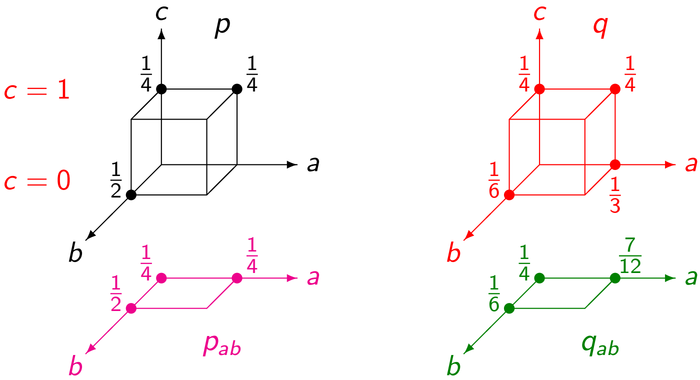
where p is the (joint) density of the probability measure P, whose argument $(\mathbf{a},\mathbf{c})$ is the joint configuration of $\mathbf{a} \in \mathsf{X}_A$ and $\mathbf{c} \in \mathsf{X}_{N\setminus A}$.

Example: marginal density

$N = \{a, b, c\}$ and $\mathsf{X}_i = \{0, 1\}$ for any $i \in N$.

Take $p(0,0,0) = 0$, $p(0,0,1) = 1/4$, $p(0,1,0) = 1/2$, $p(0,1,1) = 0$,
$p(1,0,0) = 0$, $p(1,0,1) = 1/4$, $p(1,1,0) = 0$, $p(1,1,1) = 0$.

For $A = \{a, b\}$, $p_A(0,0) = p(0,0,0) + p(0,0,1) = 1/4$.



Basic concepts: conditional density in the discrete case

Definition (conditional density)

Given disjoint sets $A, C \subseteq N$ of variables and a discrete probability measure P over N, the *conditional density for A given C* is a (partial) function $p_{A|C} : \mathsf{X}_A \times \mathsf{X}_C \to [0,1]$:

$$p_{A|C}(\mathbf{a} \,|\, \mathbf{c}) := \frac{p_{AC}(\mathbf{a}, \mathbf{c})}{p_C(\mathbf{c})} \quad \text{for } \mathbf{a} \in \mathsf{X}_A \text{ and } \mathbf{c} \in \mathsf{X}_C \text{ with } p_C(\mathbf{c}) > 0.$$

Note that the ratio determining the conditional density $p_{A|C}$ is **only defined** for conditioning *positive configurations* $\mathbf{c} \in \mathsf{X}_C$ with $p_C(\mathbf{c}) > 0$!

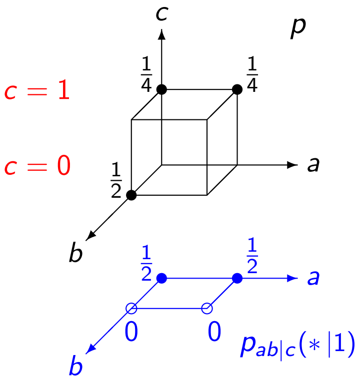
However, sometimes, in *probabilistic reasoning*, a computational convention is accepted that $p_{A|C}(\mathbf{a} \,|\, \mathbf{c}) = 0$ for any *zero configuration* $\mathbf{c} \in \mathsf{X}_C$ with $p_C(\mathbf{c}) = 0$ and $\mathbf{a} \in \mathsf{X}_A$.

It is clear that $p_{A|C}(*|*)$ only depends of the marginal density p_{AC}.

Example: conditional density

$N = \{a, b, c\}$ and $\mathsf{X}_i = \{0, 1\}$ for any $i \in N$. The same joint density.

For $A = \{a, b\}$, $C = \{c\}$, $p_{A|C}(0, 0 \,|\, 1) = \frac{p(0,0,1)}{p_C(1)} = \frac{1/4}{1/2} = 1/2$.



Observe that $p_{C|A}(* \,|\, 1, 1)$ is not defined unless a convention is accepted.

Conditional independence in the discrete case

There are several equivalent definitions of stochastic CI in the discrete case. The next one avoids using the concept of a conditional density.

Definition (CI in terms of marginal densities)

Let $A, B, C \subseteq N$ be pairwise disjoint sets of variables and P a discrete probability measure over N given by a joint density p.
We say that *A and B are conditionally independent given C with respect to P* and write $A \perp\!\!\!\perp B \,|\, C \; [P]$ if

$$\forall \mathbf{x} \in \mathsf{X}_N \qquad p_C(\mathbf{x}_C) \cdot p_{ABC}(\mathbf{x}_{ABC}) = p_{AC}(\mathbf{x}_{AC}) \cdot p_{BC}(\mathbf{x}_{BC})\,.$$

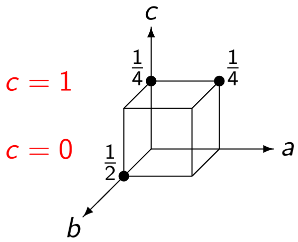
Thus, it follows from this definition that the validity of $A \perp\!\!\!\perp B \,|\, C \; [P]$ only depends on the marginal density p_{ABC}.

Note that in case $C = \emptyset$ the definition of $A \perp\!\!\!\perp B \,|\, \emptyset \; [P]$ reduces to $p_{AB}(\mathbf{x}_{AB}) = p_A(\mathbf{x}_A) \cdot p_B(\mathbf{x}_B)$ for any $\mathbf{x} \in \mathsf{X}_N$ (= stochastic independence).

Example: conditional independence

The definition of CI in terms of marginal densities is awkward to verify directly because it requires to compute marginal densities and to verify $|\mathsf{X}_{ABC}|$ equalities.

$N = \{a, b, c\}$ and $\mathsf{X}_i = \{0, 1\}$ for any $i \in N$. The same joint density.



$$\boxed{p_c \cdot p_{abc} \stackrel{?}{=} p_{ac} \cdot p_{bc}}$$

To verify/disprove $a \perp\!\!\!\perp b \,|\, c$ one needs to check 8 equalities.

For $\mathbf{x} = (0, 0, 0)$ check $p_c(0) \cdot p(0, 0, 0) \stackrel{?}{=} p_{ac}(0, 0) \cdot p_{bc}(0, 0)$,
which is true: $1/2 \cdot 0 = p_c(0) \cdot p(0, 0, 0) = p_{ac}(0, 0) \cdot p_{bc}(0, 0) = 1/2 \cdot 0$,
$\mathbf{x} = (0, 0, 1) : \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = p_c(1) \cdot p(0, 0, 1) = p_{ac}(0, 1) \cdot p_{bc}(0, 1) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$.

Computation for remaining joint configurations confirms $a \perp\!\!\!\perp b \,|\, c$ $[P]$.

Conditional independence in terms of conditional densities

The concept of CI is typically introduced in terms of *conditional densities*.

Observation (CI in terms of conditional densities)

An elegant symmetric description of $A \perp\!\!\!\perp B \,|\, C \; [P]$ in these terms is

$$\forall\, \mathbf{x} \in \mathsf{X}_{ABC} \;\; \text{such that } p_C(\mathbf{x}_C) > 0 \text{ one has}$$

$$p_{AB|C}(\mathbf{x}_{AB}|\mathbf{x}_C) = p_{A|C}(\mathbf{x}_A|\mathbf{x}_C) \cdot p_{B|C}(\mathbf{x}_B|\mathbf{x}_C)\,.$$

However, the most popular equivalent definition in these terms is the next asymmetric one, which basically says that the conditional distribution $P_{A|BC}$ *does not depend on the variables in* B:

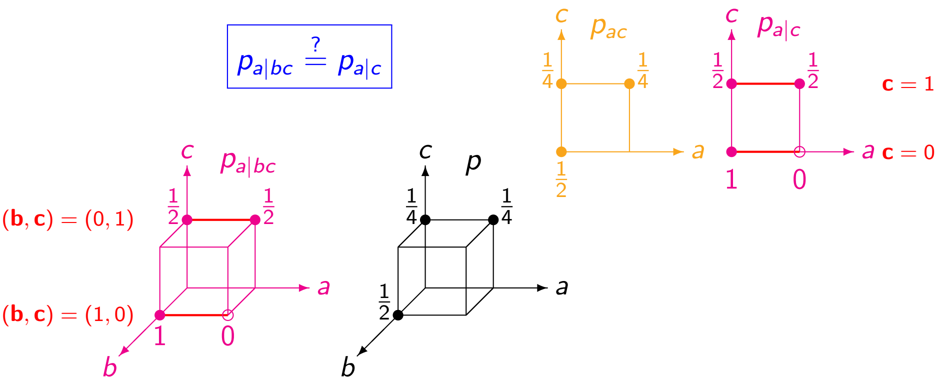
$$\forall\, \mathbf{x} \in \mathsf{X}_{ABC} \text{ with } p_{BC}(\mathbf{x}_{BC}) > 0 \qquad p_{A|BC}(\mathbf{x}_A|\mathbf{x}_{BC}) = p_{A|C}(\mathbf{x}_A|\mathbf{x}_C)\,. \qquad (1)$$

Note that because $p_{BC}(\mathbf{x}_{BC}) > 0 \;\Rightarrow\; p_C(\mathbf{x}_C) > 0$ the RHS in (1) is also defined.

Example: CI in terms of conditional densities

$N = \{a, b, c\}$ and $\mathsf{X}_i = \{0, 1\}$ for any $i \in N$. The same joint density.

$$\boxed{p_{a|bc} \stackrel{?}{=} p_{a|c}}$$



Thus, we observe that both $p_{a|bc}(*\,|\,0,1) = \frac{1}{2} \cdot \delta_0 + \frac{1}{2} \cdot \delta_1 = p_{a|c}(*\,|\,1)$ and $p_{a|bc}(*\,|\,1,0) = \delta_0 = p_{a|c}(*\,|\,0)$. Hence, $a \perp\!\!\!\perp b \,|\, c \; [P]$.

Irrelevance interpretation of stochastic CI

The definitions of CI in terms of conditionals offer its natural interpretation.

The asymmetric definition saying that $p_{A|BC}{=}p_{A|C}$ was interpreted in the context of *probabilistic reasoning* as the requirement of *conditional irrelevance of A on B given C*, namely, that values in A depend on the values in B only through the values in C.

This is close to a traditional interpretation of the condition $p_{A|BC}{=}p_{A|C}$ in the theory of Markov processes (so-called *Markovian condition*), which is that the *future* A does depend on the *past* B only through the *present* C.

The symmetric condition $p_{AB|C}(*\,|\,\mathbf{x}_C) = p_{A|C}(*\,|\,\mathbf{x}_C) \cdot p_{B|C}(*\,|\,\mathbf{x}_C)$ for any positive conditioning configuration $\mathbf{x}_C \in \mathsf{X}_C$ leads directly to the next interpretation: *once the configuration for C is fixed/known then the variables is A and B do not influence each other.*

Conditional independence as exchangeability condition

In the case of a discrete probability density, the number of equalities to be checked in order to verify a CI statement can be considerably reduced. One can even avoid computing marginals and conditionals.

An elegant combinatorial characterization of CI in terms of joint densities, already mentioned by Moussouris in the 1970s, can be interpreted as a *cross-exchange condition* for configurations.

J. Moussouris (1974). Gibbs and Markov properties over undirected graphs. Journal of Statistical Physics 10 (1), 11-31.

Observation (CI in the form cross-exchange condition)

Let $A, B, C \subseteq N$ be pairwise disjoint and P a discrete probability measure over N given by a joint density p. Then $A \perp\!\!\!\perp B \,|\, C\ [P]$ if an only if

$$\forall\, \mathbf{a}, \bar{\mathbf{a}} \in \mathsf{X}_A,\ \ \forall\, \mathbf{b}, \bar{\mathbf{b}} \in \mathsf{X}_B,\ \ \forall\, \mathbf{c} \in \mathsf{X}_C \ \text{ one has}$$

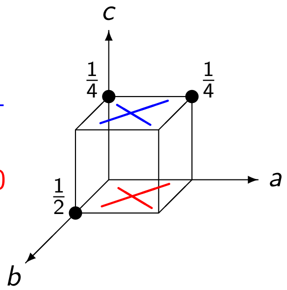
$$p_{ABC}(\mathbf{a}, \mathbf{b}, \mathbf{c}) \cdot p_{ABC}(\bar{\mathbf{a}}, \bar{\mathbf{b}}, \mathbf{c}) = p_{ABC}(\mathbf{a}, \bar{\mathbf{b}}, \mathbf{c}) \cdot p_{ABC}(\bar{\mathbf{a}}, \mathbf{b}, \mathbf{c})\ .$$

Example: CI as the cross-exchange condition

$N = \{a, b, c\}$ and $\mathsf{X}_i = \{0, 1\}$ for any $i \in N$. The same joint density.

$c = 1$

$c = 0$



Only 2 equalities need to be checked for $a \perp\!\!\!\perp b \,|\, c$ using cross-exchange.

For $\mathbf{c} = 1$, $\{\mathbf{a}, \bar{\mathbf{a}}\} = \{0, 1\}$ and $\{\mathbf{b}, \bar{\mathbf{b}}\} = \{0, 1\}$ it means checking

$p(0,0,1) \cdot p(1,1,1) \stackrel{?}{=} p(0,1,1) \cdot p(1,0,1)$, which is true:

$p(0,0,1) \cdot p(1,1,1) = \frac{1}{4} \cdot 0 = 0 = 0 \cdot \frac{1}{4} = p(0,1,1) \cdot p(1,0,1)$.

For $\mathbf{c} = 0$ one has $p(0,0,0) \cdot p(1,1,0) = 0 = p(0,1,0) \cdot p(1,0,0)$.

Conditional independence and factorization

Another characterization of stochastic CI is in term of *factorization*.

Observation (CI in the form of a factorization condition)

Let $A, B, C \subseteq N$ be pairwise disjoint and P a discrete probability measure over N given by a joint density p. Then $A \perp\!\!\!\perp B \,|\, C \; [P]$ if an only if

$\exists\, f^{AC} : \mathsf{X}_{AC} \to \mathbb{R}, \; \exists\, g^{BC} : \mathsf{X}_{BC} \to \mathbb{R}$ such that

$$\forall\, \mathbf{x} \in \mathsf{X}_{ABC} \quad p_{ABC}(\mathbf{x}) = f^{AC}(\mathbf{x}_{AC}) \cdot g^{BC}(\mathbf{x}_{BC}) \,,$$

where the functions f^{AC} and g^{AC} are called *potentials*.

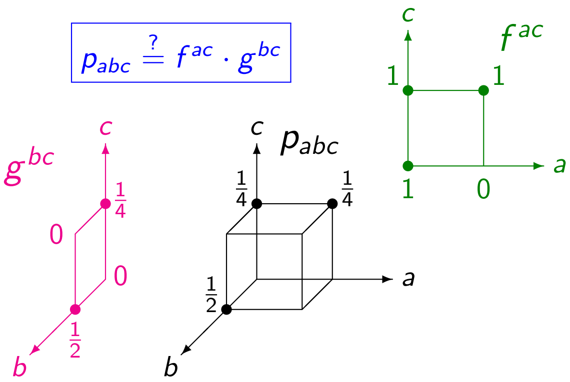
The meaning of this factorization condition is as follows: to record a 3-dimensional density in the memory of a computer it is enough to record only two 2-dimensional potentials, which reduced the memory demands.

This leads to the *second important interpretation* of stochastic CI in terms of *decomposition* of a many-dimensional object into less-dimensional ones.

Example: CI interpreted as factorization

$N = \{a, b, c\}$ and $\mathsf{X}_i = \{0, 1\}$ for any $i \in N$. The same joint density.

$$\boxed{p_{abc} \stackrel{?}{=} f^{ac} \cdot g^{bc}}$$



One can choose the following potentials:

$f^{ac}(0,0) = f^{ac}(0,1) = f^{ac}(1,1) = 1$ and $f^{ac}(1,0) = 0$,

$g^{bc}(0,1) = \frac{1}{4}$, $g^{bc}(1,0) = \frac{1}{2}$, $g^{bc}(0,0) = g^{bc}(1,1) = 0$.

Abstraction: decomposition interpretation of CI

If $A \perp\!\!\!\perp B \,|\, C \; [P]$, the factors in the multiplicative decomposition $p_{ABC} = f^{AC} \cdot g^{BC}$ need not be unique. On the other hand, a unique *standard decomposition* into marginal densities exists:

$$p_{ABC}(\mathbf{x}_{ABC}) = p_{AC}(\mathbf{x}_{AC}) \cdot \underbrace{\frac{1}{p_C(\mathbf{x}_C)} \cdot p_{BC}(\mathbf{x}_{BC})}_{p_{B|C}(\mathbf{x}_B|\mathbf{x}_C)} \quad \text{for any } \mathbf{x} \in \mathsf{X}_N,$$

where a convention $\frac{1}{0} := 0$ is accepted.

This standard decomposition is usually written as $\boxed{p_{ABC} = p_{AC} \times p_{B|C}}$, which emphasizes 3 main operations used to introduce the concept of CI:

- marginalization ($p \mapsto p_{AC}$),
- conditioning ($p \mapsto p_{B|C}$),
- aggregation operation (the multiplication $\times$ is this case).

Analogously, CI can be introduced beyond the probabilistic framework.

Abstraction: towards CI in various uncertainty calculi

This is basically the way to introduce the concept of CI in Shenoy's abstract *valuation-based systems* involving many frames for knowledge representation occurring in artificial intelligence.

A similar idea was utilized by Jiroušek in his *compositional models*.

P. P. Shenoy (1994). Conditional independence in valuation-based systems. International Journal of Approximate Reasoning 10 (3), 203-234.

R. Jiroušek (1997). Composition of probability measures on finite spaces. In Proceedings of UAI, Morgan Kaufmann, San Francisco, 274-281.

In fact, most of the papers on CI within alternative uncertainty calculi introduce the concept of CI by defining these three operations, namely *marginalization*, *conditioning* and some kind of *aggregation* operation.

Note that conditioning and aggregation is often gathered in one *composition operation*.

Basic formal properties of stochastic CI

The next formal properties of CI have been emphasized by many authors.

Observation (formal properties of CI)

Let P be a probability measure over N. Then the following conditions hold for (pairwise disjoint) $A, B, C, D \subseteq N$:

- $\emptyset \perp\!\!\!\perp B \,|\, C \; [P]$
- $A \perp\!\!\!\perp B \,|\, C \; [P] \;\Rightarrow\; B \perp\!\!\!\perp A \,|\, C \; [P]$,
- $A \perp\!\!\!\perp BD \,|\, C \; [P] \;\Rightarrow\; A \perp\!\!\!\perp D \,|\, C \; [P]$,
- $A \perp\!\!\!\perp BD \,|\, C \; [P] \;\Rightarrow\; A \perp\!\!\!\perp B \,|\, DC \; [P]$,
- $A \perp\!\!\!\perp D \,|\, C \; [P] \;\&\; A \perp\!\!\!\perp B \,|\, DC \; [P] \;\Rightarrow\; A \perp\!\!\!\perp BD \,|\, C \; [P]$.

Moreover, if P has a strictly positive density then

- $A \perp\!\!\!\perp B \,|\, DC \; [P] \;\&\; A \perp\!\!\!\perp D \,|\, BC \; [P] \;\Rightarrow\; A \perp\!\!\!\perp BD \,|\, C \; [P]$.

Note that these formal properties are valid far beyond the discrete case.

Formal independence model

These fundamental properties of CI are also valid for certain ternary separation relations induced by graphs, which motivated Pearl and Paz to introduce an abstract concept of a semi-graphoid.

J. Pearl and A. Paz (1987). Graphoids, graph-based logic for reasoning about relevance relations. In Advances in Artificial Intelligence II. North-Holland, Amsterdam, 357-363.

Semi-graphoids are formal independence models in the following sense.

Definition (formal independence model)

A *formal independence model over N* is a set $\mathcal{M}$ of ordered triplets $\langle A, B|C\rangle$ of pairwise disjoint subsets of N, whose elements are interpreted as independence statements.
We write $A \perp\!\!\!\perp B \,|\, C \;[\mathcal{M}]$ to indicate that $\langle A, B|C\rangle \in \mathcal{M}$ is interpreted as an independence statement.

The concept of a semi-graphoid

Definition (disjoint semi-graphoid)

A (disjoint) *semi-graphoid over N* is a formal independence model $\mathcal{M}$ over N satisfying the following conditions/axioms:

$\emptyset \perp\!\!\!\perp B \,|\, C\ [\mathcal{M}]$ triviality,

$A \perp\!\!\!\perp B \,|\, C\ [\mathcal{M}] \ \Rightarrow\ B \perp\!\!\!\perp A \,|\, C\ [\mathcal{M}]$ symmetry,

$A \perp\!\!\!\perp BD \,|\, C\ [\mathcal{M}] \ \Rightarrow\ A \perp\!\!\!\perp D \,|\, C\ [\mathcal{M}]$ decomposition,

$A \perp\!\!\!\perp BD \,|\, C\ [\mathcal{M}] \ \Rightarrow\ A \perp\!\!\!\perp B \,|\, DC\ [\mathcal{M}]$ weak union,

$A \perp\!\!\!\perp D \,|\, C\ [\mathcal{M}]\ \&\ A \perp\!\!\!\perp B \,|\, DC\ [\mathcal{M}] \ \Rightarrow\ A \perp\!\!\!\perp BD \,|\, C\ [\mathcal{M}]$ contraction.

A semi-graphoid $\mathcal{M}$ will be called a *graphoid* (over N) if it satisfies

$A \perp\!\!\!\perp B \,|\, DC\ [\mathcal{M}]\ \&\ A \perp\!\!\!\perp D \,|\, BC\ [\mathcal{M}] \ \Rightarrow\ A \perp\!\!\!\perp BD \,|\, C\ [\mathcal{M}]$ intersection.

Thus, the observation about basic properties of stochastic CI can be formulated as a statement that $\mathcal{M}_P := \{\, \langle A, B | C \rangle \ :\ A \perp\!\!\!\perp B \,|\, C\ [P] \,\}$ is a semi-graphoid for any (discrete) probability measure P over N.

Remarks on semi-graphoids

Pearl and Paz considered the semi-graphoid implications to be natural properties of **conditional irrelevance relations**. They indeed occur in various areas (see later examples).

Introducing a semi-graphoid in the form of an abstract mathematical concept inspired several theoretical papers on this topic.

The methods for efficient representing (disjoint) semi-graphoids in the memory of a computer have been proposed using so-called *elementary CI statements*, which are statements of the form $a \perp\!\!\!\perp b \mid C$.

Also, natural dominance ordering between triplets $\langle A, B|C \rangle$ leads to a dual concept of a *dominant CI statement* and to the research on (implementing) the *semi-graphoid/graphoid closure*. This further led to some papers on *complexity of semi-graphoid inference*.

Morton (2009) in his thesis even revealed surprising occurrence of semi-graphoids in the context of *polyhedral geometry*.

The concept of a separoid

Some authors studied *general semi-graphoids* over N, in which case the requirement of pairwise disjointness of sets A, B and C is omitted. This allows one to model *functional dependency* relations.

Dawid took very abstract point of view and introduced the next notion.

Definition (separoid)

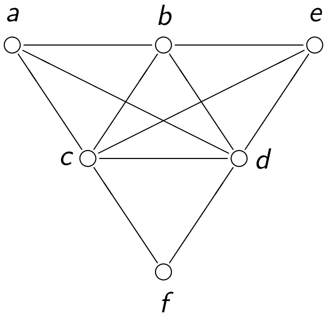
Let $\mathbf{S}$ be a joint semi-lattice, that is, a partially ordered set in which every two elements a, b have a supremum (= joint), denoted by $a \vee b$. A set of ordered triplets $a \perp\!\!\!\perp b \,|\, c$ of elements of $\mathbf{S}$ will be named a *separoid* if

- $b \vee c = c \;\Rightarrow\; a \perp\!\!\!\perp b \,|\, c$,
- $a \perp\!\!\!\perp b \,|\, c \;\Leftrightarrow\; b \perp\!\!\!\perp a \,|\, c$,
- $a \perp\!\!\!\perp b \vee d \,|\, c \;\Leftrightarrow\; \{\, a \perp\!\!\!\perp d \,|\, c \;\;\&\;\; a \perp\!\!\!\perp b \,|\, d \vee c \,\}$.

A. P. Dawid (2001). Separoids: a mathematical framework for conditional independence and irrelevance. Annals of Mathematics and Artificial Intelligence 32 (1/4), 335-372.

Examples of semi-graphoids: undirected graph

Given an undirected graph G whose set of nodes is N and a triplet $\langle A, B|C\rangle$ of pairwise disjoint subsets of N we say that *A and B are separated by C in G* and write $A \perp\!\!\!\perp B \,|\, C \; [G]$ if every path in G from a node in A to a node in B must contain a node in C.



one has $b \perp\!\!\!\perp f \,|\, \{c, d\} \; [G]$

$a \perp\!\!\!\perp e \,|\, \{b, c, d\} \; [G]$

but $a \not\perp\!\!\!\perp e \,|\, \{b, c\} \; [G]$

It is always a graphoid. Every such graphoid is induced by a discrete probability distribution. Nevertheless, in case $|N| = 4$ one has 64 such graphoids in comparison 18478 probabilistic semi-graphoids.

Examples of semi-graphoids: a class of subsets

Let $\mathcal{P}(N) := \{A : \ A \subseteq N\}$ denote the power set of N.

Given a class $\mathcal{T} \subseteq \mathcal{P}(N)$ of subsets on N and a triplet $\langle A, B|C\rangle$ of disjoint subsets of N we write $A \perp\!\!\!\perp B \,|\, C \; [\mathcal{T}]$ if

$$\forall\ T \in \mathcal{T} \qquad T \subseteq ABC \ \Rightarrow\ [\, T \subseteq AC \text{ or } T \subseteq BC\,].$$

One can verify easily that it is always a graphoid.

Note that every graphical graphoid is induced by a class of sets: take the collection of complete sets in the graph G in place of $\mathcal{T}$.

Examples of semi-graphoids: supermodular function

A set function $m : \mathcal{P}(N) \to \mathbb{R}$ is called *supermodular* if

$$m(D \cup E) + m(D \cap E) \geq m(D) + m(E) \qquad \text{for any } D, E \subseteq N.$$

Given a supermodular function $m : \mathcal{P}(N) \to \mathbb{R}$ and a triplet $\langle A, B|C \rangle$ of pairwise disjoint subsets of N we write $A \perp\!\!\!\perp B \,|\, C \; [m]$ if

$$m(C) + m(ABC) = m(AC) + m(BC)\,.$$

This is always a semi-graphoid and semi-graphoids defined in this way are called *structural* for some deeper reasons. In fact, every (discrete) probabilistic semi-graphoid is structural.

In comparison with graphical semi-graphoids, structural semi-graphoids are much more common. If $|N| = 4$ one has 22108 structural semi-graphoids in comparison with all 26424 semi-graphoids.

Examples of semi-graphoids: relational database

Theory of *relational databases* is an area of computer science, in which a concept analogous to CI was studied even earlier than in probabilistic reasoning.

A *relational database over N* is simply a set of configurations over N: $\mathbb{D} \subseteq \mathsf{X}_N \equiv \prod_{i\in N} \mathsf{X}_i$. Operations of *marginalization* and *combination* for relational databases:

- given $A \subseteq N$ and $\mathbb{D} \subseteq \mathsf{X}_N$, put $\mathbb{D}_A := \{\mathbf{b}_A \,:\, \mathbf{b} \in \mathbb{D}\}$,
- given a disjoint triplet $\langle A, B|C\rangle$ over N and databases $\mathbb{D}^1 \subseteq \mathsf{X}_{AC}$, $\mathbb{D}^2 \subseteq \mathsf{X}_{BC}$ its *combination* is as follows:

$$\mathbb{D}^1 \bowtie \mathbb{D}^2 \;:=\; \{\, (\mathbf{a},\mathbf{b},\mathbf{c}) \in \mathsf{X}_{ABC} \,:\, (\mathbf{a},\mathbf{c}) \in \mathbb{D}^1 \;\&\; (\mathbf{b},\mathbf{c}) \in \mathbb{D}^2 \,\}\,.$$

Given a database $\mathbb{D} \subseteq \mathsf{X}_N$ and a triplet $\langle A, B|C\rangle$ of disjoint subsets of N we write $A \perp\!\!\!\perp B \,|\, C\; [\mathbb{D}]$ if $\mathbb{D}_{ABC} = \mathbb{D}_{AC} \bowtie \mathbb{D}_{BC}$.

It is always a semi-graphoid. These semi-graphoid are analogous to probabilistic ones, but they are different.

Examples of semi-graphoids: ordinal conditional functions

Spohn (1988) in his epistemic-belief theory used special *ordinal conditional functions* to model the state of knowledge.

Following Shenoy (1994) we call a function $\kappa : \mathsf{X}_N \to \mathbb{N}$ such that $\min \{ \kappa(x) \, : \, x \in \mathsf{X}_N \} = 0$ a *disbelief function*.

One defines a special *marginalization* operation: given $A \subseteq N$ put

$$\kappa_A(\mathbf{a}) \;:=\; \min \{ \kappa(\mathbf{a}, \mathbf{c}) \; : \; \mathbf{c} \in \mathsf{X}_{N \setminus A} \} \quad \text{for } \mathbf{a} \in \mathsf{X}_A .$$

Given a disbelief function $\kappa : \mathsf{X}_N \to \mathbb{N}$ and a triplet $\langle A, B | C \rangle$ of disjoint subsets of N we write $A \perp\!\!\!\perp B \,|\, C \; [\kappa]$ if

$$\forall \, \mathbf{x} \in \mathsf{X}_N \qquad \kappa_C(\mathbf{x}_C) + \kappa_{ABC}(\mathbf{x}_{ABC}) = \kappa_{AC}(\mathbf{x}_{AC}) + \kappa_{BC}(\mathbf{x}_{BC}) \, .$$

Possibility theory

Possibility theory was proposed as a model for quantification of judgements on basis of fuzzy theory.

D. Dubois and H. Prade (1988). Possibility Theory, an Approach to Computerized Processing of Uncertainty. Plenum Press, New York.

Definition (possibility distribution)

Possibility distribution over N is a function $\pi : \mathsf{X}_N \to [0,1]$ with $\max\{\pi(\mathbf{a}) : \mathbf{a} \in \mathsf{X}_N\} = 1$. Then $\Pi(\mathbb{A}) = \max_{\mathbf{x}\in\mathbb{A}} \pi(\mathbf{x})$ for any $\mathbb{A} \subseteq \mathsf{X}_N$ is the respective *possibility measure*.

Given $A \subseteq N$, the *marginal possibility distribution* is defined by

$$\pi_A(\mathbf{a}) = \max\{\pi(\mathbf{a},\mathbf{b}) : \mathbf{b} \in \mathsf{X}_{N\setminus A}\}.$$

The above concept is a counterpart of the concept of a *probability density*; the main point is that the operation of summation is replaced by the operation of maximization.

CI in possibility theory I

There is a bunch of different CI concepts introduced in this framework.

The following one was pinpointed by Shenoy (1994) in his work on valuation-based systems.

Definition (multiplicative possibilistic CI)

Given a possibility distribution $\pi : \mathsf{X}_N \to [0,1]$ and a triplet $\langle A, B|C\rangle$ of pairwise disjoint subsets of N we write $A \perp\!\!\!\perp B \mid C \; [\pi]$ if

$$\pi_C(\mathbf{c}) \cdot \pi_{ABC}(\mathbf{a},\mathbf{b},\mathbf{c}) = \pi_{AC}(\mathbf{a},\mathbf{c}) \cdot \pi_{BC}(\mathbf{b},\mathbf{c})\,.$$

for any $\mathbf{a} \in \mathsf{X}_A$, $\mathbf{b} \in \mathsf{X}_B$ and $\mathbf{c} \in \mathsf{X}_C$.

The induced formal independence model is always a semigraphoid.

This class of possibilistic semi-graphoids involves semi-graphoids induced by relational databases and by Spohn's disbelief functions; use the transformation $\pi(\mathbf{x}) := \exp^{-\kappa(\mathbf{x})}$ for $\mathbf{x} \in \mathsf{X}_N$.

CI in possibility theory II

A lot of papers in possibility theory was devoted to the ways to introduce various concepts of *conditional possibilistic densities* and to define the concept of (conditional) independence on basis of that.

For example, Fonck (1994) combined Zadeh's conditioning rule in which $\pi_{A|C} = \pi_{AC}$ with miminization aggregation operation, which lead to the concept of *conditional non-interactivity of A and B given C*:

$$\pi_{ABC}(\mathbf{a},\mathbf{b},\mathbf{c}) = \min\{\, \pi_{AC}(\mathbf{a},\mathbf{c}), \pi_{BC}(\mathbf{b},\mathbf{c})\,\}$$

for any $\mathbf{a} \in \mathsf{X}_A$, $\mathbf{b} \in \mathsf{X}_B$ and $\mathbf{c} \in \mathsf{X}_C$.

Vejnarová (2000) made an overview of those various conditioning concepts and offered a unifying point using a measure-theoretical approach. In this approach, every *continuous triangular norm defines* the respective conditioning operation and a concept of CI in possibilitic framework.

Every possibilistic CI concept defined through such triangular norm then leads to semi-graphoids.

Dempster-Shafer theory of evidence I

This is one of most popular approaches to deal with uncertain knowledge in AI. It offers a far-reaching generalization of the probabilistic approach.

There is even a special regular conference on belief functions!

A. P. Dempster (1967). Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 11, 325-339.

G. Shafer (1976). A Mathematical Theory of Evidence. Princeton Univerity Press, Princeton.

Belief functions are defined by means the following elementary concept.

Definition (basic probability assignment)

A *basic probability assignment* (BPA) over N is a function $\mathbf{m} : \mathcal{P}(N) \to [0, \infty]$ satisfying

$$\sum \{ \mathbf{m}(\mathbb{D}) \ : \ \mathbb{D} \subseteq \mathsf{X}_N \} = 1 \quad \text{and} \quad \mathbf{m}(\emptyset) = 0 \,.$$

Dempster-Shafer theory of evidence II

Shenoy (1994) introduced the concept of CI also within this framework. Two further concepts needed to recall his definition.

Definition (marginal BPA)

Given a BPA $\mathbf{m} : \mathcal{P}(N) \to [0, \infty]$ over N and $A \subseteq N$ the respective *marginal BPA* is defined as follows

$$\mathbf{m}_A(\mathbb{A}) = \sum \{ \, \mathbf{m}(\mathbb{D}) \; : \; \mathbb{D} \subseteq \mathsf{X}_N \;\&\; \mathbb{D}_A = \mathbb{A} \, \} \quad \text{for } \mathbb{A} \subseteq \mathsf{X}_A \,,$$

where $\mathbb{D}_A$ denotes the database projection ($=$ marginalization).

Definition (commonality function)

Given a BPA $\mathbf{m}_A : \mathcal{P}(A) \to [0, \infty]$ over $A \subseteq N$ the respective *commonality function* $Q_A^{\mathbf{m}}$ is defined by means of superset summing:

$$Q_A^{\mathbf{m}}(\mathbb{E}) = \sum \{ \, \mathbf{m}_A(\mathbb{A}) \; : \; \mathbb{E} \subseteq \mathbb{A} \subseteq \mathsf{X}_A \, \} \quad \text{for } \mathbb{E} \subseteq \mathsf{X}_A \,.$$

CI in DS theory of evidence

This is how Shenoy (1994) introduced the concept of CI in DS theory.

Definition (commonalistic CI)

Given a BPA $\mathbf{m} : \mathcal{P}(N) \to [0, \infty]$ over N and a triplet $\langle A, B|C\rangle$ of pairwise disjoint subsets of N we write $A \perp\!\!\!\perp B \,|\, C \; [Q^{\mathbf{m}}]$ if

$$Q^{\mathbf{m}}_{C}(\mathbb{E}_C) \cdot Q^{\mathbf{m}}_{ABC}(\mathbb{E}) = Q^{\mathbf{m}}_{AC}(\mathbb{E}_{AC}) \cdot Q^{\mathbf{m}}_{BC}(\mathbb{E}_{BC}) \qquad \text{for every } \mathbb{E} \subseteq \mathsf{X}_{ABC}.$$

Note that in the unconditional case ($C = \emptyset$) this coincides the the definition proposed (independently) by de Campos and Huete (1993).

One can observe that this CI concept generalizes the CI concept introduced for database relations and probability distributions.

On the other hand, neither of two possibilistic CI concepts is generalized by the commonalistic CI in the DS theory (provided a standard embedding of possibility distributions into DS frame is considered).

Inconsistence with marginalization

Nevertheless, commonalistic CI has one disadvantage in comparison with standard probabilistic CI, which was mentioned in (Studený, 1993) and also discussed by Ben Yaghlane, P. Smets and K. Mellouli (2002).

M. Studený (1993). Formal properties of conditional independence in different calculi of AI. In Symbolic and Quantitative Approaches to Reasoning and Uncertainty, LNCS 747, Springer, Berlin, pp. 341-348.

B. Ben Yaghlane, P. Smets, K. Mellouli (2002). Belief function independence II. International Journal of Approximate Reasoning 31, 31-75.

The problem with commonalistic CI is that it is *not consistent with marginalization* in the sense that

- BPAs $\mathbf{m}^1$ over AC and $\mathbf{m}^2$ over BC may exist such $\mathbf{m}^1_C = \mathbf{m}^2_C$,
- but no BPA $\mathbf{m}$ over ABC exists such that $\mathbf{m}_{AC} = \mathbf{m}^1$, $\mathbf{m}_{BC} = \mathbf{m}^2$ and $A \perp\!\!\!\perp B \,|\, C\ [Q^{\mathbf{m}}]$.

Consistence with marginalization

A solution to the problem of inconsistency of CI definition with marginalization was suggested by Jiroušek and Vejnarová (2011).

R. Jiroušek and J. Vejnarová (2011). Compositional models and conditional independence in evidence theory. International Journal of Approximate Reasoning 52, 316-334.

They proposed a modified definition of a CI concept in DS theory which removes the above mentioned disadvantage.

Definition (BPA-listic CI)

Given a BPA $\mathbf{m} : \mathcal{P}(N) \to [0, \infty]$ over N and a triplet $\langle A, B|C\rangle$ of pairwise disjoint subsets of N we write $A \perp\!\!\!\perp B \,|\, C \;[\mathbf{m}]$ if

$$\mathbf{m}_C(\mathbb{D}_C) \cdot \mathbf{m}_{ABC}(\mathbb{D}_{ABC}) = \mathbf{m}_{AC}(\mathbb{D}_{AC}) \cdot \mathbf{m}_{BC}(\mathbb{D}_{BC})$$

for every $\mathbb{D} \subseteq \mathsf{X}_{ABC}$ such that $\mathbb{D} = \mathbb{D}_{AC} \bowtie \mathbb{D}_{BC}$ and $\mathbf{m}(\mathbb{E}) = 0$ for any other $\mathbb{E} \subseteq \mathsf{X}_{ABC}$.

Conclusions

The concept of CI has been introduced and became the topic of research interest in several areas of AI. The shared phenomenon of those CI concepts is the *conditional irrelevance interpretation*.

In the basic probabilistic framework, there is only one concept of stochastic CI with several equivalent definitions. The analogues of these definitions in other uncertainty calculi however leads to *different CI concepts*.

There are other frameworks, besides the calculi mentioned in the talk, in which a concept analogous to CI was introduced. Let us mention the *epistemic irrelevance* studied in the context of *imprecise probabilities* (Cozman, 2005) and the *cs-independence* coming from de Finetti's conditional-event approach (Vantaggi, 2002).

Most of the CI concepts mentioned in the talk exhibit the *semi-graphoid properties*. The abstract mathematical concept of a semi-graphoid, respectively of a *separoid*, has a potential to enrich the lattice theory.