Concepts of Independence for Coherent Probabilities and for Credal Sets

Fabio G. Cozman - University of Sao Paulo, Brazil
fgcozman@usp.br

July 6, 2013

Overview

1. Basic definitions in standard probability theory.
2. Coherent probabilities, and full conditional probabilities.
3. Credal sets.
4. Independence concepts for credal sets, full conditional probabilities, full credal sets.

Standard axioms for probabilities...

Space Ω (FINITE!): subsets are events, functions are random variables.

- **PU1** $P(A) \geq 0$.
- **PU2** $P(\Omega) = 1$.
- **PU3** If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.
- **EXP** $E[X] = \sum X(\omega)P(\omega)$.
- **CP** If $P(B) > 0$, then $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Two problems:

1. Conditional probability is a derived, incomplete concept: may be left undefined even if given event is *possible* (nonempty).
2. Quite a bit of structure is assumed; precise specification of all possible probability values is assumed feasible.

A different scheme: Coherent probabilities

- Basic idea: assessments

$$P(A_1|B_1) = \alpha_1, P(A_2|B_2), \ldots, P(A_m|B_m) = \alpha_m$$

on conditional events must be *coherent*.

- Result: assessments are coherent if and only if they can be extended to a *full conditional probability.*

Assessments are coherent when, for every $\lambda_1, \ldots, \lambda_m$,

$$\sup_{\omega \in \cup_{i=1}^{m} B_i} \sum_{i=1}^{m} \lambda_i (I_{A_i} - \alpha_i) I_{B_i} \geq 0.$$

A full conditional probability is...

...a function $P(\cdot|\cdot)$ on $\mathcal{E} \times \mathcal{E} \backslash \emptyset$ where $\mathcal{E}$ is Boolean algebra of events, such that

- $P(\Omega|C) = 1$;
- $P(A|C) \geq 0$ for all A;
- $P(A \cup B|C) = P(A|C) + P(B|C)$ when $A \cap B = \emptyset$;
- $P(A \cap B|C) = P(A|B \cap C)\, P(B|C)$ when $B \cap C \neq \emptyset$.

- Write the "unconditional" probability $P(A)$ for $P(A|\Omega)$.

- $P(A|B \cap C)$ can be defined even if $P(B|C) = 0$!

Two examples

Two events A and B.

$P(\cdot|\Omega)$

	A	A^c
B	0	α
B^c	0	$1-\alpha$

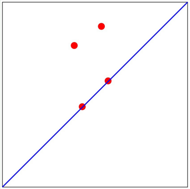
$$P(B|A) = \beta$$
$$P(B^c|A) = 1-\beta$$

Square with uniform distribution (NOTE: infinite space...)

Points a, b, c, d, line e.

$$P(a) = P(b) = P(c) = P(d)\,(= 0...).$$

But $P(a|a \cup b) = 1/2...$
And $P(e|a \cup b) = P(a|a \cup b) = 1/2.$



The Krauss-Dubins representation

- $L_0, \ldots, L_K$ partition Ω, and there is a positive probability P_i over each L_i.
- $P(A|B) = P(A|B \cap L_i)$, where $i = \arg\min_j (P(B|L_j) > 0)$.

Example: two events A and B.

$P(\cdot|\Omega)$

	A	A^c
B	0	α
B^c	0	$1-\alpha$

$$P(B|A) = \beta$$
$$P(B^c|A) = 1-\beta$$

$P(\cdot|A)$

	A	A^c
B	β	
B^c	$1-\beta$	

Example

Two events A and B, three layers

	A	A^c
B	1	0
B^c	0	0

	A	A^c
B		$1-\alpha$
B^c	α	0

	A	A^c
B		
B^c		1

Coletti and Scozzafava's layer numbers

- The layer number of A is the index of the first layer L_i such that $P(A|L_i) > 0$.

Layer numbers

	A	A^c
B	1	0
B^c	0	0

	A	A^c
B		$1-\alpha$
B^c	α	0

	A	A^c
B		
B^c		1

$\circ(A \cap B) = 0, \quad \circ(A \cap B^c) = \circ(A^c \cap B) = 1, \quad \circ(A^c \cap B^c) = 2.$

- Adopt: $\circ(\emptyset) = \infty$.
- Adopt: $\circ(A|B) = \circ(A \cap B) - \circ(B)$ whenever $B \neq \emptyset$.

Layer numbers and Spohn ranking functions

- Given a full conditional probability, the layer numbers induced by it satisfy the properties of Spohn ranking functions (measures of "disbelief").

Ranking functions satisfy:

- $\circ(A) = 0$ or $\circ(A^c) = 0$ (or both).
- $\circ(A \cup B) = \min(\circ(A), \circ(B))$.
- $\circ(\emptyset) = \infty$.
- $\circ(A|B) = \circ(A \cap B) - \circ(B)$ if $A \cap B \neq \emptyset$.

Coherent and full conditional probabilities offer...

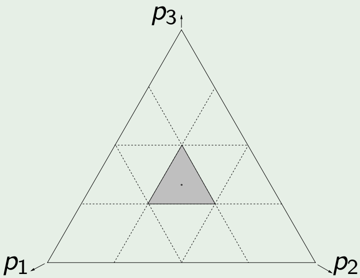
- a principled approach to conditional probability (and conditional events).
- a flexible framework for assessments, that does not require much structure on events, and does not assume complete specification.
- a unifying language that can express a variety of formalisms (even Spohn-like measures, probabilistic logics, default reasoning).

Credal sets

- A *credal set* is a set of probability measures (distributions).
- A credal set is usually defined by a set of *assessments*.

Possibility space with three states: $\Omega = \{\omega_1, \omega_2, \omega_3\}$

Assessments:
$P(\omega_1) \in [1/3, 2/3]$,
$P(\omega_2) \in [1/3, 2/3]$,
$P(\omega_3) \in [1/3, 2/3]$,



A few points...

- Credal set with distributions for X is denoted $K(X)$.
- Given credal set $K(X)$:
  - $\underline{P}(A) = \inf_{P \in K(X)} P(A)$.
  - $\underline{E}[X] = \inf_{P \in K(X)} E_P[X]$.
- Consider the lower expectation *functional* that associates every random variable X with its lower expectation.
  - There is a one-to-one correspondence between closed convex credal sets and lower expectation functionals.

Justifying credal sets

1. From partially ordered (binary) preferences, closed convex credal sets:

$$X \succ Y \quad \text{iff} \quad E_P[X] > E_P[Y] \text{ for all } P \in K.$$

   From more general preferences, more general credal sets.

2. From coherence: a set of assessments $\{\underline{E}[X_i] = \alpha_i\}_{i=1}^{m}$ is *coherent* if and only if it can be extended to a credal set.

> For every $X_0, X_1, \ldots, X_n$, any $m > 0$, there is $\omega \in \Omega$ such that
>
> $$\sum_{i=1}^{n} (X_i(\omega) - \underline{E}[X_i]) \geq m \times (X_0(\omega) - \underline{E}[X_0]).$$

Credal sets offer...

- a unifying language to express assessments:

$$P(A) = 1/2; \quad E[X] = 10;$$

$$P(B) \in [1/2, 3/4]; \quad \underline{E}[X + Y] = 1; \quad P(A) \leq P(B \cap C);\ldots$$

  and belief functions, Choquet capacities, p-boxes, probability intervals, possibility measures...
- a framework for robustness analysis.
- a model for ambiguity aversion and risk assessment in decision-making, economics, and finance.
- a scheme for aggregation of beliefs within an agent, or a community of agents.

Conditioning

1. One option:

$$K(X|B) = \{P(\cdot|B) : P \in K(X)\} \text{ if } \underline{P}(B) > 0.$$

2. Another option:

$$K^{>}(X|B) = \{P(\cdot|B) : P \in K(X) \text{ and } P(B) > 0\} \text{ if } \overline{P}(B) > 0.$$

Problem:

- Conditional probability is a derived, incomplete concept: may be left undefined even if given event is *possible* (nonempty).

Full credal sets

- A set of full conditional probabilities (used by Levi, Williams, Walley).
- Now: a set of assessments of lower/upper probabilities/expectations
  - is *coherent* if and only if
  - it can be extended to a set of full conditional probabilities

Williams coherence (refined Pelessoni-Vicig version, finite Ω):
For every $X_0|B_0, X_1|B_1, \ldots, X_n|B_n$, any $s_0 \geq 0, s_1 \geq 0, \ldots, s_n \geq 0$, there is $\omega \in \cup_{i=0}^{n} B_i$ such that

$$\sum_{i=1}^{n} s_i I_{B_i}(\omega)(X_i(\omega) - \underline{E}[X_i]) \geq s_0 I_{B_0}(\omega)(X_0(\omega) - \underline{E}[X_0]).$$

Independence: first, for standard probabilities

- Variables $X_1, \ldots, X_n$ are *stochastically independent* when
  - $P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \times \cdots \times P(X_n = x_n)$.
  - $P(X_i = x_i | \cap_{j \neq i} X_j = x_j) = P(X_i = x_i)$,
    whenever $P(\cap_{j \neq i} X_j = x_j) > 0$.
- Conditional independence: independence given every $\{Z = z\}$.

The semi-graphoid properties

Proposed as a way to encode the intuitive meaning of "conditional independence".

Symmetry: $(X \perp\!\!\!\perp Y \mid Z) \Rightarrow (Y \perp\!\!\!\perp X \mid Z)$

Redundancy: $(X \perp\!\!\!\perp Y \mid X)$

Decomposition: $(X \perp\!\!\!\perp (W, Y) \mid Z) \Rightarrow (X \perp\!\!\!\perp Y \mid Z)$

Weak union: $(X \perp\!\!\!\perp (W, Y) \mid Z) \Rightarrow (X \perp\!\!\!\perp W \mid (Y, Z))$

Contraction:

$$(X \perp\!\!\!\perp Y \mid Z) \ \& \ (X \perp\!\!\!\perp W \mid (Y, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y) \mid Z)$$

Independence for full conditional probabilities

Example: two events A and B

	A	A^c
B	1	0
B^c	0	0

	A	A^c
B		$1/10$
B^c	$1/10$	$4/5$

- Note: $P(A^c \cap B^c) = P(A^c)\,P(B^c)$.
- However, $P(A^c|B^c) = \frac{4/5}{1/10+4/5} = 8/9$, while $P(A^c) = 0$!

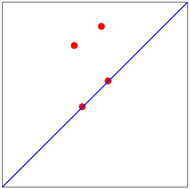
Another example, as a digression (infinite space)

Square with uniform distribution (NOTE: infinite space...)

Points a, b, c, d, line e.

$$P(a) = P(b) = P(c) = P(d) \, (= 0...).$$

But $P(a|a \cup b) = 1/2...$
And $P(e|a \cup b) = P(a|a \cup b) = 1/2$.



- Note: $P(e \cap (a \cup b)) = P(a) = P(a)\, P(b) = 0$.
- However, $P(e|a \cup b) = 1/2!$

Failure of symmetry

Example: two events A and B

	A	A^c
B	$1/2$	0
B^c	$1/2$	0

	A	A^c
B		$1/3$
B^c		$2/3$

- Note: $P(A^c|B) = P(A^c)$.
- However, $P(B|A^c) = 1/3$, while $P(B) = 1/2$!

Epistemic and cs-independence (for variables)

Idea: for independence of X and Y, require

$$P(X = x|Y = y, Z = z) = P(X = x|Z = z)$$

and

$$P(Y = y|X = x, Z = z) = P(Y = y|Z = z)$$

and

$$\circ(X = x, Y = y|Z = z) = \circ(X = x|Z = z) + \circ(Y = y|Z = z)\,.$$

- This definition fails the following property:

  Weak union: $(X \perp\!\!\!\perp (W, Y)\,|\,Z) \Rightarrow (X \perp\!\!\!\perp W\,|\,(Y, Z))$

Other concepts of independence

1. Hammond's concept of independence (fails Contraction).
2. Blume et al's concept of preference independence (fails Contraction).
3. Kohlberg and Reny's concept of "strong" independence (fails Contraction).
4. Layer independence.

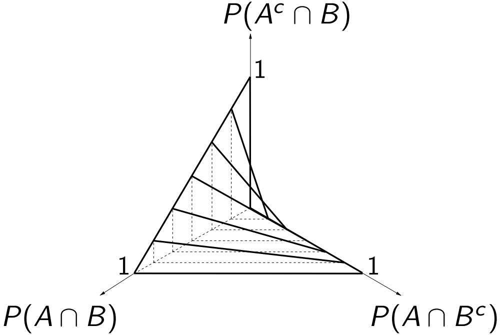
Back to credal sets: Complete independence

X and Y are *completely independent* when

for all $P \in K(X, Y)$,

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)\,.$$

- That is, elementwise stochastic independence.
    - Example: several experts agree on stochastic independence.
- This concept violates convexity.



Strong independence

X and Y are *strongly independent* when

$K(X, Y)$ is the convex hull of a set of distributions satisfying complete independence.

- Variants:
  - Walley and Fine's independent products;
  - Walley's type-1 and type-2 products;
  - Weichselberger's mutual independence;
  - Campos and Moral's type-2 and type-3 independence;
  - Couso et al.'s repetition independence.
- Justification by extendibility (Moral and Cano 2002), by exchangeability (Cozman 2012, De Bock and De Cooman 2012).

Confirmational and epistemic irrelevance

Levi's proposal: Y is *confirmationally irrelevant* to X when

$$K(X|Y=y)=K(X)\,.$$

Walley's proposal: Y is *epistemically irrelevant* to X when for any function $f(X)$,

$$\underline{E}[f(X)|Y=y]=\underline{E}[f(X)]\,.$$

Failure of symmetry (Couso et al 1999, Example 3)

- Consider three urns (with red and white balls):

Urn	Red	White	Unknown
A	5	2	3
B	3	3	4
C	3	3	4

- Take ball X from A, then
  - if red, take ball Y from B,
  - otherwise take ball Y from C.
- What are $P(Y = R|X = R)$, $P(Y = R|X = W)$, $P(Y = R)$?
- But notice: $P(X = R|Y = R) \in [3/10, 28/31]$ (symmetry fails).

Epistemic independence

- Walley's clever idea: "symmetrize" irrelevance.

X and Y are *epistemically independent* when

- Y is epistemically irrelevant to X and
- X is epistemically irrelevant to Y.

- Quite an intuitive concept.

The zoo, for credal sets...

- Complete independence.
  - Elementwise stochastic independence.
- Strong independence and its variants.
- Confirmational irrelevance.
- Epistemic irrelevance and independence.
  - $\underline{E}[f(X)|Y] = \underline{E}[f(X)]$ and $\underline{E}[g(Y)|X] = \underline{E}[g(Y)]$.
- *Cognitive* independence.
- *Kuznetsov* independence.
- *Type-5* irrelevance.

By conditioning on every value of a given variable Z, we obtain concepts of *conditional* independence...

Comparing concepts

- All concepts satisfy forms of laws of large numbers (results by De Cooman and Miranda).
- Complete independence satisfies all semi-graphoid properties.
- When lower probabilities are positive, epistemic independence satisfies Symmetry, Redundancy, Decomposition, Weak Union, but fails Contraction.
- Other concepts fail various properties; usually Contraction is violated.

Independence and zero probabilities

- So far we have avoided zero probabilities in our discussion of independence concepts for credal sets.

- But consider conditional epistemic independence of X and Y given Z:

$$\underline{E}[f(X)|Y = y, Z = z] = \underline{E}[f(X)|Z = z] \,.$$

What happens if $\underline{P}(Z = z) = 0$?

Conditioning and independence

- Two options:

  1. $\underline{E}[f(X)|Y = y, Z = z] = \underline{E}[f(X)|Z = z]$
     whenever $\underline{P}(Y = y, Z = z) > 0$.
     - TOO WEAK!!!

  2. $\underline{E}^{>}[f(X)|Y = y, Z = z] = \underline{E}^{>}[f(X)|Z = z]$
     whenever $\overline{P}(Y = y, Z = z) > 0$.
     - De Campos and Moral's type-5 independence; perhaps the best idea if standard conditioning is adopted.

- However, we can resort to full conditional probabilities here.

Now, full credal sets

- Complete independence satisfies all semi-graphoid properties, but too weak for full conditional probabilities.
- Is there some appropriate form of "elementwise" independence?
- Epistemic independence fails Decomposition, Weak Union and Contraction (!).

Hammond's independence for full credal sets

Say that Y is h-irrelevant to X given Z when

$$\underline{E}[f(X)|A(X), B(Y), Z = z] = \underline{E}[f(X)|A(X), Z = z]\,.$$

- Say that X and Y are h-independent given Z if X is h-irrelevant to Y given Z and Y is h-irrelevant to X given Z.
- This concept satisfies Symmetry, Redundancy, Decomposition and Weak Union, but fails Contraction.

Full Bayesian and Credal networks

- Consider an extension of Bayesian networks, where each node is associated with a full conditional probability. Or perhaps where each node is associated with a credal set.
- Example:

$$X \longrightarrow Y \longrightarrow Z.$$

  Presumably, Z is independent of X given Y (Markov condition).
- Obviously, the properties of the joint model will depend on the sort of independence that is adopted (factorization, d-separation).

Conclusion

- Standard probability theory offers a simple and flexible framework, but it has some drawbacks.
  1. Conditional probability does not get the right treatment.
  2. Precise specification of probability values is assumed.
- This talk considered some ways to bypass these difficulties:
  1. Full conditional probabilities.
  2. Sets of standard probabilities.
  3. Sets of full conditional probabilities ("full" credal sets).
- We examined concepts of independence for them.
  - A single best concept has not emerged: many options!
  - Some concepts satisfy semi-graphoid properties, but are inadequate or hard to justify; other concepts seem intuitive but fail semi-graphoid properties...