Independence
(and Some Other Structural Assessments)
in Imprecise Probability

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

July 16, 2012
Overview

1. A review of some basic definitions.
2. Structural assessments: vacuity, uniformity, exchangeability.
3. A review of stochastic (conditional) independence.
4. Confirmational/complete/strong/epistemic/Kuznetsov/others independence.
5. Comparison.
6. A look into the messy world of zero probabilities.
Easy warm-up

- Possibility space $\Omega$ with states $\omega$.
- Events are subsets of $\Omega$.
- Random variables $(X : \Omega \rightarrow \mathbb{R})$.
- Expectations $E[X]$, probabilities $P(A)$.
- Conditional expectations $E[X|B]$, probabilities $P(A|B)$. 
A credal set is a set of probability measures (distributions).
A credal set is usually defined by a set of assessments.

Example:
1. \( \Omega = \{\omega_1, \omega_2, \omega_3\} \).
2. \( P(\omega_i) = p_i \).
3. \( p_1 \geq p_3, 2p_1 \geq p_2, p_1 \leq 2/3 \) and \( p_3 \in [1/6, 1/3] \).
4. Take points \( P = (p_1, p_2, p_3) \).
Some geometry

1. $\Omega = \{\omega_1, \omega_2, \omega_3\}$.
2. $P(\omega_i) = p_i$.
3. $p_1 \geq p_3$, $2p_1 \geq p_2$, $p_1 \leq 2/3$ and $p_3 \in [1/6, 1/3]$.
4. Take points $P = (p_1, p_2, p_3)$. 
Properties of credal sets

- Credal set with distributions for $X$ is denoted $K(X)$. 

Given credal set $K(X)$:

- $E[X] = \inf_{P \in K(X)} E_P[X]$.
- $E[X] = \sup_{P \in K(X)} E_P[X]$.

For closed convex credal sets, lower and upper expectations are attained at vertices. That is, there is only one lower expectation for a given closed convex credal set.
Properties of credal sets

- Credal set with distributions for $X$ is denoted $K(X)$.
- Given credal set $K(X)$:
  - $\underline{E}[X] = \inf_{P \in K(X)} E_P[X]$.
  - $\overline{E}[X] = \sup_{P \in K(X)} E_P[X]$.

For closed convex credal sets, lower and upper expectations are attained at vertices. A closed convex credal set is completely characterized by the associated lower expectation. That is, there is only one lower expectation for a given closed convex credal set.
Properties of credal sets

- Credal set with distributions for $X$ is denoted $K(X)$.
- Given credal set $K(X)$:
  - $\underline{E}[X] = \inf_{P \in K(X)} E_P[X]$.
  - $\overline{E}[X] = \sup_{P \in K(X)} E_P[X]$.
- For closed convex credal sets, lower and upper expectations are attained at vertices.
Credal set with distributions for $X$ is denoted $K(X)$.

Given credal set $K(X)$:

- $\underline{E}[X] = \inf_{P \in K(X)} E_P[X]$.
- $\overline{E}[X] = \sup_{P \in K(X)} E_P[X]$.

For closed convex credal sets, lower and upper expectations are attained at vertices.

A closed convex credal set is completely characterized by the associated lower expectation.

That is, there is only one lower expectation for a given closed convex credal set.
Generating credal sets I

From partial preferences:

- $X \succ Y$ means “$X$ is preferred to $Y$.”
- Axiomatize $\succ$ as partial order.
- Then:

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

- Credal sets with identical vertices produce the same $\succ$.
- Focus has been on unique *maximal* credal set that represents $\succ$.
  - Smaller credal sets have no “behavioral” significance.
From one-sided betting:

- Variables are *gambles*.
- Buy/sell gambles using $E[X]$ and $\overline{E}[X]$.
- Some constraints, such as: $\sum_{i=1}^{n} X_i - E[X_i] \geq 0$.
- Credal set is produced by the set of *dominating* expectations:

\[ \{ E : E[X] \geq \overline{E}[X] \} . \]

- Several credal sets produce the same lower expectations.
  - But only maximal closed one is given "behavioral" significance.
Overview

1. A review of some basic definitions.
2. **Structural assessments: vacuity, uniformity, exchangeability.**
3. A review of stochastic (conditional) independence.
4. Confirmational/complete/strong/epistemic/Kuznetsov/others independence.
5. Comparison.
6. A look into the messy world of zero probabilities.
A structural assessment constrains a large (possibly infinite) number of expectations.

A simple example: vacuity.
Uniformity

- Every $\omega$ is subject to identical assessments.
- Extreme case: vacuity.
- Extreme case: uniform distribution.
- Intermediate case: $P(\omega_i) \in [1/4, 1/2]$. 
Exchangeability

- A basic structural assessment.
- To simplify, take categorical variables $\mathbf{X} = [X_1, \ldots, X_m]$.

- Denote by $\pi_m$ a permutation of integers $\{1, \ldots, m\}$, and by $\pi_m(i)$ the $i$th number in the permutation.
- Denote
  \[\{\mathbf{X} = \mathbf{x}\} \doteq \bigcap_{i=1}^{m} \{X_i = x_i\},\]
  and
  \[\{\pi_m\mathbf{X} = \mathbf{x}\} \doteq \bigcap_{i=1}^{m} \{X_{\pi_m(i)} = x_i\}.\]
Definition of exchangeability

- Variables $X_1, \ldots, X_m$ are exchangeable when

$$E[\{X = x\} - \{\pi_m X = x\}] = 0$$

for any permutation $\pi_m$.

- That is, the order of variables does not matter: trading

  $\{X = x\}$ for $\{\pi_m X = x\}$ does not seem advantageous.

- Walley’s theorem: this implies elementwise exchangeability.
Walley’s exchangeability theorem

- In words: Exchangeability implies *elementwise* exchangeability.
- Proof:

\[
0 = E\left[\{X = x\} - \{\pi_m X = x\}\right] \\
\leq E\left[\{X = x\} - \{\pi_m X = x\}\right] \\
= -E\left[\{\pi_m X = x\} - \{X = x\}\right] = 0.
\]

Hence every distribution in the credal set \( K(X_1, \ldots, X_m) \) satisfies

\[
P(X = x) = P(\pi_m X = x) \quad \text{for any permutation } \pi_m.
\]
Exercise

- Suppose we have 4 binary variables that are exchangeable.
- What are the conditions on the probabilities $P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)$?
Exercise

- Suppose we have 4 binary variables that are exchangeable.
- What are the conditions on the probabilities $P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)$?

Here they are:

Facts about exchangeability

- For $X_1, \ldots, X_m$, what matters is

$$P\left( \sum_{i=1}^{m} X_i = r \right).$$

- Any subset of exchangeable variables is exchangeable.

- Consider $m$ exchangeable variables, and take initial $n$ variables; then $P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0)$ is equal to

$$\sum_{r=k}^{m-n+k} \frac{\binom{m-n}{r-k}}{\binom{m}{r}} P\left( \sum_{i=1}^{n} X_i = r \right).$$
Denote $P(X_1 = 1)$ by $\theta$ and take $m \to \infty$:

Then $P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0)$ is equal to

$$\int_0^1 \theta^k (1 - \theta)^{n-k} dF(\theta).$$

That is, behaves like $\prod_{i=1}^n P(X_i = 1)$ with a (yet unspecified) distribution over $\theta$. 
Exercise

Draw the credal set \( K(X, Y) \) for binary variables \( X \) and \( Y \) given the structural assessments:

1. \( X \) and \( Y \) are exchangeable.
2. \( X \) and \( Y \) are the first two variables in a sequence of three exchangeable variables.
3. \( X \) and \( Y \) are the first two variables in a sequence of five exchangeable variables.
4. \( X \) and \( Y \) are the first two variables in a sequence of infinitely many exchangeable variables.
Overview

1. A review of some basic definitions.
2. Structural assessments: vacuity, uniformity, exchangeability.
3. **A review of stochastic (conditional) independence.**
4. Confirmational/complete/strong/epistemic/Kuznetsov/others independence.
5. Comparison.
6. A look into the messy world of zero probabilities.
Independence

- **Events:**
  - $P(A|B) = P(A)$.
  - $P(A \cap B) = P(A)P(B)$.

- **Variables:**
  - $E[f(X)|A(Y)] = E[f(X)]$.
  - $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$. 
A and $B$ are independent

$$P(A|B) = P(A) \quad \text{whenever } P(B) > 0;$$

or, equivalently (definition is symmetric!),

$$P(A \cap B) = P(A) P(B).$$
Independence for events

- $A$ and $B$ are independent

\[ P(A|B) = P(A) \quad \text{whenever } P(B) > 0; \]

or, equivalently (definition is symmetric!),

\[ P(A \cap B) = P(A) \cdot P(B). \]

- For all subsets of events $\{A_i\}_{i=1}^n$,

\[ P(\cap_i \{X_i \in A_i\}) = \prod_i P(\{X_i \in A_i\}). \]
Independence for variables

- $X$ is stochastically irrelevant to $Y$ when:

\[ E[f(Y)|\{X \in A\}] = E[f(Y)] \]

for any bounded function $f$, whenever $P(\{X \in A\}) > 0$. 

Definition is symmetric! We might as well use the symmetric condition:

\[ E[f(X)g(Y)] = E[f(X)]E[g(Y)] \]

for every bounded $f$ and $g$. 

So, take these conditions to mean the symmetric concept of stochastic independence of $X$ and $Y$. 
Independence for variables

- $X$ is stochastically irrelevant to $Y$ when:

$$E[f(Y)|\{X \in A\}] = E[f(Y)]$$

for any bounded function $f$, whenever $P(\{X \in A\}) > 0$.

- Definition is symmetric! We might as well use the symmetric condition:

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

for every bounded $f$ and $g$.

- So, take these conditions to mean the symmetric concept of *stochastic independence* of $X$ and $Y$. 
Other definitions...

1. $X$ is stochastically irrelevant to $Y$ when

$$P(\{Y \in B\} \mid \{X \in A\}) = P(\{Y \in B\})$$

whenever $P(\{X \in A\}) > 0$.

2. $X$ is stochastically irrelevant to $Y$ when

$$P(\{Y \in B\} \cap \{X \in A\}) = P(\{Y \in B\}) \cdot P(\{X \in A\}).$$
For many variables...

**Variables** \( \{X_i\}_{i=1}^n \) are *independent* if

\[
E[f_i(X_i) \mid \cap_{j \neq i} \{X_j \in A_j\}] = E[f_i(X_i)],
\]

for

- all functions \( f_i(X_i) \),
- all events \( \cap_{j \neq i} \{X_j \in A_j\} \) with positive probability.

For all functions \( f_i(X_i) \),

\[
E \left[ \prod_{i=1}^n f_i(X_i) \right] = \prod_{i=1}^n E[f_i(X_i)].
\]
Conditional stochastic independence

- All these definitions can be extended to *conditional independence* by conditioning on some $Z$, for every value of $Z$.

- *For instance*, $X$ and $Y$ are conditionally stochastically independent given $Z$ iff

\[
P(\{Y \in B\} \cap \{X \in A\}|Z = z) = P(\{Y \in B\}|Z = z) \\
\times P(\{X \in A\}|Z = z)
\]

whenever $P(Z = z) > 0$. 
What do we do with independence?

- We consider repetitions of experiments.
- We prove convergence under quite broad conditions.
- We construct large models by combining pieces (see exciting talk tomorrow on credal networks!!).
- We discard parts of a model when focusing on particular inferences.
Laws of large numbers...

- Weak law of large numbers:
  \[ \forall \epsilon > 0, \lim_{n \to \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0 \]

- Strong law of large numbers:
  \[ P \left( \lim_{n \to \infty} \frac{\sum_{i=1}^{n} X_i}{n} = \mu \right) = 1. \]

- (Finitely-additive) strong law of large numbers: for all \( \epsilon > 0 \), there is integer \( N \) such that for every positive integer \( k \),
  \[ P \left( \exists n \in [N, N + k] : \left| \frac{\sum_{i=1}^{n} X_i}{n} - \mu \right| > \epsilon \right) < \epsilon. \]
The graphoid properties

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

**Symmetry:** \((X \independent Y \mid Z) \Rightarrow (Y \independent X \mid Z)\)

**Decomposition:** \((X \independent (W, Y) \mid Z) \Rightarrow (X \independent Y \mid Z)\)

**Weak union:** \((X \independent (W, Y) \mid Z) \Rightarrow (X \independent W \mid (Y, Z))\)

**Contraction:** \((X \independent Y \mid Z) \& (X \independent W \mid (Y, Z)) \Rightarrow (X \independent (W, Y) \mid Z)\)

**Redundancy:** \((X \independent Y \mid X)\)

Often added (true when probabilities are positive):

**Intersection** \((X \independent W \mid (Y, Z)) \& (X \independent Y \mid (W, Z)) \Rightarrow (X \independent (W, Y) \mid Z)\)

Satisfied by many structures (graphs, lattices, etc).
Exercise

Prove contraction for stochastic independence.

**Contraction:** \((X \perp \!\!\!\!\!\!\perp Y \mid Z) \& (X \perp \!\!\!\!\!\!\perp W \mid (Y, Z)) \Rightarrow (X \perp \!\!\!\!\!\!\perp (W, Y) \mid Z)\)
Overview

1. A review of some basic definitions.
2. Structural assessments: vacuity, uniformity, exchangeability.
3. A review of stochastic (conditional) independence.
4. Confirmational/complete/strong/epistemic/Kuznetsov/others independence.
5. Comparison.
6. A look into the messy world of zero probabilities.
X and Y are completely independent if for all $P \in K(X, Y)$,

$$P(X \in A | Y \in B) = P(X \in A) \quad \text{whenever } P(Y \in B) > 0.$$ 

That is, elementwise stochastic independence.

This concept violates convexity (presumably has no “behavioral” justification).
Example of Jeffrey’s:

- Binary variables $X$ and $Y$, completely independent.
- $K(X, Y)$: convex hull of $P_1$ and $P_2$,

\[ P_1(X = 0) = P_1(Y = 0) = 1/3, \ P_2(X = 0) = P_2(Y = 0) = 2/3. \]

- Take $P_{1/2} = P_1/2 + P_2/2$ (by convexity, $P_{1/2} \in K(X, Y)$).
- However,

\[
P_{1/2}(X = 0, Y = 0) = P_1(X = 0)P_1(Y = 0)/2 + P_2(X = 0)P_1(Y = 0)/2 \]
\[
= 5/18 \neq 1/4 \]
\[
= P_{1/2}(X = 0)P_{1/2}(Y = 0).
\]
Independence surface for two events

\[ P(A \cap B) \]

\[ P(A^c \cap B) \]

\[ P(A \cap B^c) \]

\[ P(A^c \cap B^c) \]
I. Levi, the pioneer on convex credal sets, detected this problem with complete independence.

His proposal: \( Y \) is *confirmationally irrelevant* to \( X \) if

\[
K(X|Y \in B) = K(X) \quad \text{for nonempty} \quad \{ Y \in B \},
\]

His position: use complete independence if needed, but take convex hull (does not affect partial preferences...).
Strong independence

- $X$ and $Y$ are *strongly independent* when $K(X, Y)$ is the convex hull of a set of distributions satisfying complete independence.
Walley and Fine (1982) called this expression an *independent product* when restricted to indicators:

$$E[A(X, Y)] = \min \left( E_P[A(X, Y)] : P = P_X P_Y \right).$$

This is Weichselberger’s definition of *mutual independence*.

In his book, Walley (1991) called the general expression a *type-1 product*.

...and *type-2 products* refer to the case of identical marginals.
Walley also proposes a different concept: $Y$ is epistemically irrelevant to $X$ if for any bounded function $f(X)$,

$$E[f(X) | Y \in B] = E[f(X)]$$

for nonempty $\{Y \in B\}$.

If credal sets are closed and convex, then epistemic irrelevance is identical to Levi’s confirmational irrelevance.
Consider three urns (with red and white balls):

<table>
<thead>
<tr>
<th>Urn</th>
<th>Red</th>
<th>White</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Take ball $X$ from A, then
- if red, take ball $Y$ from B,
- otherwise take ball $Y$ from C.

How to construct set of joint distributions for $(X, Y)$?

What are $P(Y = R|X = R)$, $P(Y = R|X = W)$, $P(Y = R)$?

Epistemic independence

- Walley’s clever idea: “symmetrize” irrelevance (this is actually a strategy by Keynes).

- $X$ and $Y$ are **epistemically independent** if $Y$ is epistemically irrelevant to $X$ and $X$ is epistemically irrelevant to $Y$.

- Quite an intuitive concept that “generates convexity” automatically.
Two binary variables $X$ and $Y$.

$P(X = 0) \in [2/5, 1/2]$ and $P(Y = 0) \in [2/5, 1/2]$.

Epistemic independence of $X$ and $Y$: $K(X, Y)$ is convex hull of

$$[1/4, 1/4, 1/4, 1/4], [4/25, 6/25, 6/25, 9/25],$$

$$[1/5, 1/5, 3/10, 3/10], [1/5, 3/10, 1/5, 3/10],$$

$$[2/9, 2/9, 2/9, 1/3], [2/11, 3/11, 3/11, 3/11],$$
Kuznetsov (1991) proposed yet another concept.

Denote by $EI[X]$ the interval $[\underline{E}[X], \overline{E}[X]]$.

$X$ and $Y$ are Kuznetsov independent if, for any bounded functions $f(X)$ and $g(Y)$,

$$EI[f(X)g(Y)] = EI[f(X)] \times EI[g(Y)].$$
Exercise

Prove:

- Kuznetsov independence implies epistemic independence (assume all probabilities are nonzero!).

- Epistemic independence does not imply Kuznetsov independence.
It would be nice if Kuznetsov and strong independence were equivalent.

But they are not!

(Actually, they are equivalent if one of the variables is binary.)
Example

- Ternary variables $X$ and $Y$, credal sets $K(X)$ and $K(Y)$:

- Largest set that satisfies strong independence (strong extension) has 16 vertices and 24 facets; for instance, a facet with normal

  $$[-434, 301, 21, 2836, -1154, -1734, -1164, 96, 1116].$$

- This facet cannot be written as $f(X)g(Y) + \alpha$.

- Intuitively, a Kuznetsov “extension” wraps the strong extension using only functions $f(X)g(Y)$. 
Some history

Several variants between 1990/2000... inspired by intense activity in Dempster-Shafer and possibility theory.

For each possible definition of conditioning or product-measure, a concept of independence...

Quick example: Dempster conditioning defines

$$\overline{P}(X|D Y) = \overline{P}(X, Y)/\overline{P}(Y)$$

then we can impose

$$\overline{P}(X|D Y) = \overline{P}(X, Y)/\overline{P}(Y) = \overline{P}(X).$$

Related (mathematically at least) to Shafer’s concept of cognitive independence
Attempt to organize the field.

Their *type-2* independence is strong independence.

Their *type-3* independence obtains when $K(X, Y)$ is the convex hull of all product distributions $P_X P_Y$, where $P_X \in K(X)$ and $P_Y \in K(Y)$.

That is, type-3 independence is simply strong extension.

Their *type-5* independence is a variant on confirmational irrelevance.
Y is type-5 irrelevant to X if

$$R(X|Y \in B) = K(X) \quad \text{whenever} \quad \overline{P}(Y \in B) > 0,$$

where $R(X|Y \in B)$ denotes the set

$$\{ P(\cdot|Y \in B) : P \in K(X, Y); P(Y \in B) > 0 \}.$$

Then take type-5 independence to be the “symmetrized” concept.

The set $R$ is often used to define conditioning (related to what Walley calls regular extension).
Due to de Campos and Moral (1995).

- X and Y are binary.
- K(X, Y) is the convex hull of two distributions P₁ and P₂ such that P₁(X = 0, Y = 0) = P₂(X = 1, Y = 1) = 1.

Show:

- X and Y are strongly independent.
- Neither Y is type-5 irrelevant to X, nor X is type-5 irrelevant to Y.
In 1999 Couso et al presented an influential review.

- Their *independence in the selection* is strong independence.
- Their *strong independence* is strong extension.
- Their *repetition independence* refers to Walley’s *type-2 product*.

- They also discuss *non-interactivity* and *random set independence* (called *belief function product* by Walley and Fine, 1982).
The zoo, so far

- Complete independence.
- Confirmational and epistemic irrelevance/independence.
- Strong independence.
- Kuznetsov independence.
- Type-5 independence.
The zoo, so far

- Complete independence.
- Confirmational and epistemic irrelevance/independence.
- Strong independence.
- Kuznetsov independence.
- Type-5 independence.

Comments:

- Epistemic independence is most intuitive (under convexity).
- Complete independence is closer to stochastic independence (without convexity).
- How to justify strong independence?
Any concept of independence can be modified to express *conditional independence*.

For example, *conditional* epistemic irrelevance of $Y$ to $X$ given $Z$:

$$E[f(X)|Y \in B, Z = z] = E[f(X)|Z = z]$$

for all bounded functions $f(X)$ and all nonempty $\{Z = z\}$.

Likewise for conditional Kuznetsov/complete/strong independence of $X$ and $Y$ given $Z$.

Complications with probability zero, but let’s postpone that...
Some basic (mostly known) definitions: credal sets, lower expectations and probabilities, decision making, and the like.

Structural assessments: vacuity, uniformity, exchangeability.

A brief review of stochastic (conditional) independence.

Confirmational/complete/strong/epistemic/Kuznetsov/others independence.

Comparison.

A look into the messy world of zero probabilities.
Comparing concepts

There are perhaps too many concepts around.

- Idea: verify which concepts satisfy laws of large numbers.
  - Not really discriminating: all satisfy forms of laws of large numbers (results by de Cooman and Miranda).
- Other idea: check graphoid properties.
Reminder: graphoid properties

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

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

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

Contraction: \((X \perp Y \mid Z) \& (X \perp W \mid (Y, Z)) \Rightarrow (X \perp (W, Y) \mid Z)\)
Exercise

Show that complete and strong independence satisfy all graphoid properties.
Failure of contraction

- Epistemic independence satisfies symmetry, redundancy, decomposition, weak union, but fails contraction even when all probabilities are positive.
  - Thus type-5 independence also fails contraction.

- Kuznetsov independence fails contraction even when all probabilities are positive.

Note: there are different results when probabilities can be equal to zero!
Failure of contraction: epistemic indep.

- Binary variables $W$, $X$ and $Y$.
- $K(W, X, Y)$ is convex hull of three distributions:

<table>
<thead>
<tr>
<th>$W$</th>
<th>$X$</th>
<th>$Y$</th>
<th>$p_1(X, Y, W)$</th>
<th>$p_2(X, Y, W)$</th>
<th>$p_3(X, Y, W)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_0$</td>
<td>$X_0$</td>
<td>$Y_0$</td>
<td>0.008</td>
<td>0.018</td>
<td>0.0093</td>
</tr>
<tr>
<td>$W_1$</td>
<td>$X_0$</td>
<td>$Y_0$</td>
<td>0.072</td>
<td>0.072</td>
<td>0.0757</td>
</tr>
<tr>
<td>$W_0$</td>
<td>$X_1$</td>
<td>$Y_0$</td>
<td>0.032</td>
<td>0.042</td>
<td>0.037</td>
</tr>
<tr>
<td>$W_1$</td>
<td>$X_1$</td>
<td>$Y_0$</td>
<td>0.288</td>
<td>0.168</td>
<td>0.228</td>
</tr>
<tr>
<td>$W_0$</td>
<td>$X_0$</td>
<td>$Y_1$</td>
<td>0.096</td>
<td>0.084</td>
<td>0.09</td>
</tr>
<tr>
<td>$W_1$</td>
<td>$X_0$</td>
<td>$Y_1$</td>
<td>0.024</td>
<td>0.126</td>
<td>0.075</td>
</tr>
<tr>
<td>$W_0$</td>
<td>$X_1$</td>
<td>$Y_1$</td>
<td>0.384</td>
<td>0.196</td>
<td>0.290</td>
</tr>
<tr>
<td>$W_1$</td>
<td>$X_1$</td>
<td>$Y_1$</td>
<td>0.096</td>
<td>0.294</td>
<td>0.195</td>
</tr>
</tbody>
</table>

- $X$ and $Y$ are epistemically independent; $X$ and $W$ are conditionally epistemically independent given $Y$.
- But $X$ and $(W, Y)$ are not not epistemically independent.
Binary variables \( W, X, \) and \( Y \)

\( K(W, X, Y) \) with four vertices (each is the product of \( p(W|Y)p(Y)p(X) \)):

| Vertex | \( p_i(w_0|y_0) \) | \( p_i(w_0|y_1) \) | \( p_i(x_0) \) | \( p_i(y_0) \) |
|--------|-----------------|-----------------|--------------|--------------|
| \( p_1 \) | 0.7 | 0.4 | 0.2 | 0.2 |
| \( p_2 \) | 0.7 | 0.4 | 0.3 | 0.3 |
| \( p_3 \) | 0.8 | 0.5 | 0.2 | 0.3 |
| \( p_4 \) | 0.8 | 0.5 | 0.3 | 0.2 |

\( X \) and \( Y \) are Kuznetsov independent; \( X \) and \( W \) are conditionally Kuznetsov independent given \( Y \).

But \( X \) and \( (W, Y) \) are not Kuznetsov independent.
Exercise

Show that epistemic irrelevance satisfies: if $Y$ is epistemically irrelevant to $X$ and $W$ is epistemically irrelevant to $X$ given $Y$ then $(W, Y)$ are epistemically irrelevant to $X$. 
Little is known about the computational complexity of various concepts.

Complete/strong independence have been addressed in the context of credal networks.

Some algorithms are known for epistemic independence.

It seems that complete/strong independence are “more tractable” in an informal way.
The zoo, so far...

- Complete independence.
- Confirmational and epistemic irrelevance/independence.
- Strong independence.
- Kuznetsov independence (not very promising).
- Type-5 independence (only relevant with zero probabilities).

Comments:

- Epistemic independence is more intuitive (under convexity).
- Complete independence is closer to stochastic independence (without convexity).
- How to justify strong independence?
Sensitivity analysis interpretation: several experts agree on stochastic independence.

- This is an argument for complete independence.

Is there a justification that uses partial preferences, lower expectations, credal sets, etc?

A possible idea: changes in assessments (Cozman (2000), Moral and Cano (2002)).
Moral and Cano (2002):
Variables $X$ and $Y$ are [fully] strongly independent iff they are epistemically independent after $K(X, Y)$ is combined with an arbitrary collection of compatible assessments on $X$ and on $Y$.

Compatibility requires some maneuvers “similar to” Jeffrey’s rule: we change the marginal, then see what happens to the other marginal.
Consider a vector of $m$ exchangeable binary variables $\mathbf{X} = [X_1, \ldots, X_m]$. If we look at the first $n$ variables and let $m \to \infty$, then $P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0)$ is

$$\int_0^1 \theta^k (1 - \theta)^{n-k} dF(\theta).$$

Remember: $\theta$ is the probability of $\{X_1 = 1\}$.

We have a convex credal set $K(\theta)$.

Strong independence obtains if each vertex of $K(\theta)$ assigns probability 1 to a particular value of $\theta$.

We have in fact obtained a type-2 product.

Similar argument works for general variables.
Complete independence is very attractive.
But it violates convexity.
So, it does not have a “behavioral” interpretation...
Is it true?
NO!
Seidenfeld cuts

Three acts: $a_1 = 0.6$; $a_2 = 0/1$ if $A/A^c$; $a_3 = 1/0$ if $A/A^c$.

We can “cut” pieces of the probability interval!
That is, *There is a difference between a set of distributions and its convex hull when one chooses among several acts.*
Can we axiomatize preferences amongst sets of acts, so as to obtain general credal sets?

Yes. It has been done by Seidenfeld et al (2007) [it seems first idea by Kyburg and Pittarelli (1992)].
Are events $A$ and $B$ are completely independent?

Construct five acts $a_0, \ldots, a_4$:

<table>
<thead>
<tr>
<th></th>
<th>$AB$</th>
<th>$A^c B$</th>
<th>$AB^c$</th>
<th>$A^c B^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$1 - \alpha$</td>
<td>$-\alpha$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$-(1 - \alpha)$</td>
<td>$\alpha$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0</td>
<td>0</td>
<td>$1 - \beta$</td>
<td>$-\beta$</td>
</tr>
<tr>
<td>$a_4$</td>
<td>0</td>
<td>0</td>
<td>$-(1 - \beta)$</td>
<td>$\beta$</td>
</tr>
</tbody>
</table>

Test: if we observe that for every $\alpha, \beta \in (0, 1)$ such that $\alpha \neq \beta$ we have some act rejected, we can conclude that $A$ and $B$ are completely independent.
Some basic (mostly known) definitions: credal sets, lower expectations and probabilities, decision making, and the like.

Structural assessments: vacuity, uniformity, exchangeability.

A brief review of stochastic (conditional) independence.

Confirmational/complete/strong/epistemic/Kuznetsov/others independence.

Comparison.

A look into the messy world of zero probabilities.
Events may have zero *lower* probability but nonzero *upper* probability (cannot ignore those).

Example of difficulties one may face:

- Suppose we refuse to define $K(X|Y = y)$ when $P(Y = y) = 0$.
- Now consider the following concept: $Y$ is “irrelevant” to $X$ if

$$K(X|Y \in B) = K(X) \quad \text{whenever } P(Y \in B) > 0.$$ 

But this is quite weak:

- We may have $P(Y \in B) = 0$ for every $B \neq \Omega$; 
- then $Y$ is irrelevant to any other variable!
The most elegant solution is to consider *full probability measures*.

A full probability measure is a function $P(\cdot|\cdot)$ on $\mathcal{E} \times \mathcal{E}\setminus\emptyset$ where $\mathcal{E}$ is an 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 $B \cap C \neq \emptyset$.

Full probability measures allow $P(A|C)$ to be defined even if $P(C) = 0$!
The Krauss-Dubins representation

We can partition a $\Omega$ into events $L_0, \ldots, L_K$, where $K \leq N$, such that the full conditional measure is represented as a sequence of completely positive probability measures $P_0, \ldots, P_K$, where the support of $P_i$ is restricted to $L_i$.

Example:

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$A^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>0</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>$B^c$</td>
<td>0</td>
<td>$1 - \alpha$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$A^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td></td>
<td>$\beta$</td>
</tr>
<tr>
<td>$B^c$</td>
<td></td>
<td>$1 - \beta$</td>
</tr>
</tbody>
</table>

Left: $P$. Right: $P(\cdot | A)$.
Here: $P(A) = 0$, but $P(B | A) = \beta$. 
Unsurprisingly, Levi and Walley both adopt full conditional measures.

A challenge is that full conditional measures seem to call for finite additivity.

- Again, this is the path taken by Levi and Walley.
A problem with stochastic independence

- The usual product definition is now too weak!
- Consider: we may have

\[ P(X, Y = y | Z = z) = P(X | Z = z) P(Y = y | Z = z) \]

and yet

\[ P(X | Y = y, Z = z) \neq P(X | Z = z). \]

- (Failure may happen when \( P(Y = y, Z = z) = 0 \).)
Failure of symmetry

- Take epistemic irrelevance:

\[ P(X|Y = y, Z = z) = P(X|Z = z). \]

- But: this is not symmetric!!

Example:

<table>
<thead>
<tr>
<th></th>
<th>( A )</th>
<th>( A^c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B )</td>
<td>([\beta]) _1</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>( B^c )</td>
<td>([1 - \beta]) _1</td>
<td>( 1 - \alpha )</td>
</tr>
</tbody>
</table>

Note: \( P(A|B) = P(A), \) but \( P(B|A) \neq P(B) \)!
As before: symmetrize!

Definition of *epistemic* independence:

Require

\[ P(X|Y = y, Z = z) = P(X|Z = z) \]

and

\[ P(Y|X = x, Z = z) = P(Y|Z = z). \]

This is symmetric for sure.

How does it fare with respect to graphoid properties?
Reminder: graphoid properties

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

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: \((X \perp\!
\perp Y \mid Z) \& (X \perp\!
\perp W \mid (Y, Z)) \Rightarrow
(X \perp\!
\perp (W, Y) \mid Z)\)
Problem with epistemic independence

- It fails weak union!

<table>
<thead>
<tr>
<th></th>
<th>( w_0y_0 )</th>
<th>( w_1y_0 )</th>
<th>( w_0y_1 )</th>
<th>( w_1y_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_0 )</td>
<td>( \alpha )</td>
<td>( \beta )</td>
<td>( 1 - \alpha )</td>
<td>( 1 - \beta )</td>
</tr>
<tr>
<td>( x_1 )</td>
<td>( \alpha )</td>
<td>( \gamma )</td>
<td>( 1 - \alpha )</td>
<td>( 1 - \gamma )</td>
</tr>
</tbody>
</table>

Remember:

**Weak union:** \( (X \perp (W, Y) \mid Z) \Rightarrow (X \perp Y \mid (W, Z)) \)
Hammond’s independence

- Here is a proposal for independence:

\[ P(B(Y)|A(X) \cap D(Y)) = P(B(Y)|D(Y)) \] and

\[ P(A(X)|B(Y) \cap C(X)) = P(A(X)|C(X)). \]

- This is symmetric.

- It satisfies weak union! But if fails contraction...

Remember:

**Contraction:** \((X \indep Y | Z) \& (X \indep W | (Y, Z)) \Rightarrow (X \indep (W, Y) | Z)\)
There are many different structural assessments for credal sets.

Vacuity/uniformity/exchangeability are quite useful.

Independence is the most important one.

There are many different concepts of independence for credal sets.

A study of (conditional) independence touches on
- convexity and decision-making;
- conditioning and full conditional measures.
Epistemic irrelevance/independence is quite intuitive and simple to state for convex credal sets.
- Difficult to handle computationally.
- Fails the contraction property (perhaps ok?).
- Requires full conditional measures and associated challenges (perhaps then use type-5/regular independence?).
Complete independence is simple to state and inherits all the familiar properties of stochastic independence:
- Fails convexity, but this has behavioral meaning.
- Nonlinear, but this is unavoidable in the end.
- Can be adapted to full conditional measures (but need extra work).
Strong independence: popular because people want at once convexity and stochastic independence, no matter what.

- It can be justified in some cases (exchangeability).
- But hard to justify in general.