Independence Concepts in Imprecise Probability

Fabio G. Cozman - Universidade de Sao Paulo

fgcozman@usp.br

Or, perhaps...

Structural Assessments

in the Theory of Credal Sets

Overview

1. A review of some basic definitions: credal sets, lower expectations and probabilities, decision making, and the like.
2. Structural assessments: vacuity, uniformity, exchangeability.
3. A brief review of stochastic (conditional) independence.
4. Confirmational/strict/strong/epistemic/Kuznetsov/others independence.
5. Comparison.
6. A look into the messy world of zero probabilities.

Easy warm-up

- Possibility space Ω with states ω; events are subsets of Ω.
- Random variables and indicator functions.
  - Bounded function $X : \Omega \to \Re$.
  - Special type: indicator function of event A:
    - Denoted by A as well.
    - $A(\omega) = 1$ if $\omega \in A$; 0 otherwise.

Buying/selling variables

- Buy X for α: $X - \alpha$.
- Sell X for β: $\beta - X$.
- Must satisfy: $\beta > \alpha$.

- Pay less than $\underline{E}[X]$.
- Sell for more than $\overline{E}[X]$.

Fair prices

- Suppose that $\underline{E}[X] = \overline{E}[X]$ for some X.
- Then $E[X] \doteq \underline{E}[X]$ is the fair price of X.

- What if all variables had fair prices?
- What would the resulting *expectation functional* satisfy?

Axioms for expectations

EU1 If $\alpha \leq X \leq \beta$, then $\alpha \leq E[X] \leq \beta$.

EU2 $E[X + Y] = E[X] + E[Y]$.

Axioms for expectations

EU1 If $\alpha \leq X \leq \beta$, then $\alpha \leq E[X] \leq \beta$.

EU2 $E[X + Y] = E[X] + E[Y]$.

Some consequences:

1. $X \geq Y \Rightarrow E[X] \geq E[Y]$.
2. $E[\alpha X] = \alpha X$.

Supremum buying/infimum selling prices

- If one holds a set of expectations for X: willing to pay up to $\inf E[X]$ for X.
- Likewise: willing to sell X for more than $\sup E[X]$.

So, naturally:

$$\underline{E}[X] = \inf E[X] \quad \text{(lower expectation)},$$

$$\overline{E}[X] = \sup E[X] \quad \text{(upper expectation)}.$$

Familiar properties

- $\underline{E}[X] \geq \inf X$;
- $\underline{E}[\alpha X] = \alpha \underline{E}[X]$ for $\alpha \geq 0$;
- $\underline{E}[X + Y] \geq \underline{E}[X] + \underline{E}[Y]$.

Probabilities

- Expectation $E[A]$ indicates how much we expect A to "happen."

- **Definition:** The *probability* $P(A)$ is $E[A]$.

- Properties of a probability measure:

  PU1 $P(A) \geq 0$.

  PU2 $P(\Omega) = 1$.

  PU3 If $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

Conditional expectations/probabilities

- Conditional expectation of X given B,

$$E[X|B] = \frac{E[BX]}{P(B)} \quad \text{if } P(B) > 0.$$

- Bayes rule: If $P(B) > 0$, then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Credal sets

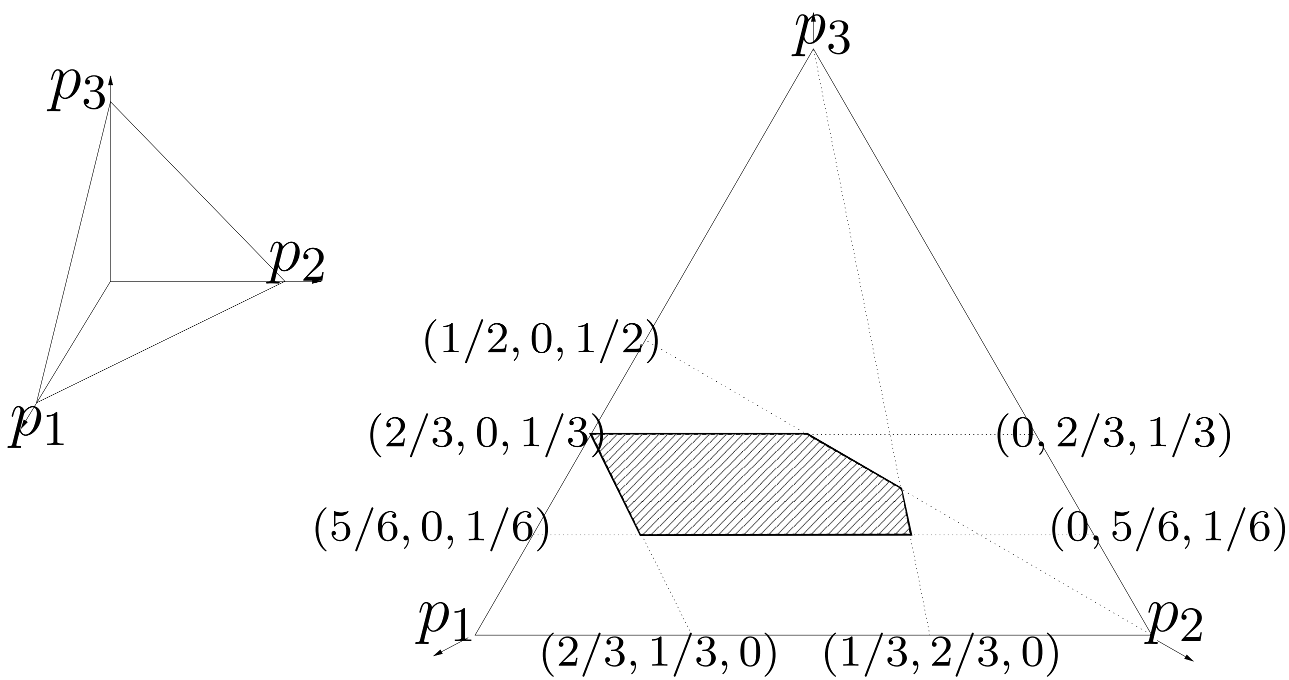
- A credal set is a set of probability measures (distributions).
- A credal set is usually defined by a set of *assessments*.

Example:

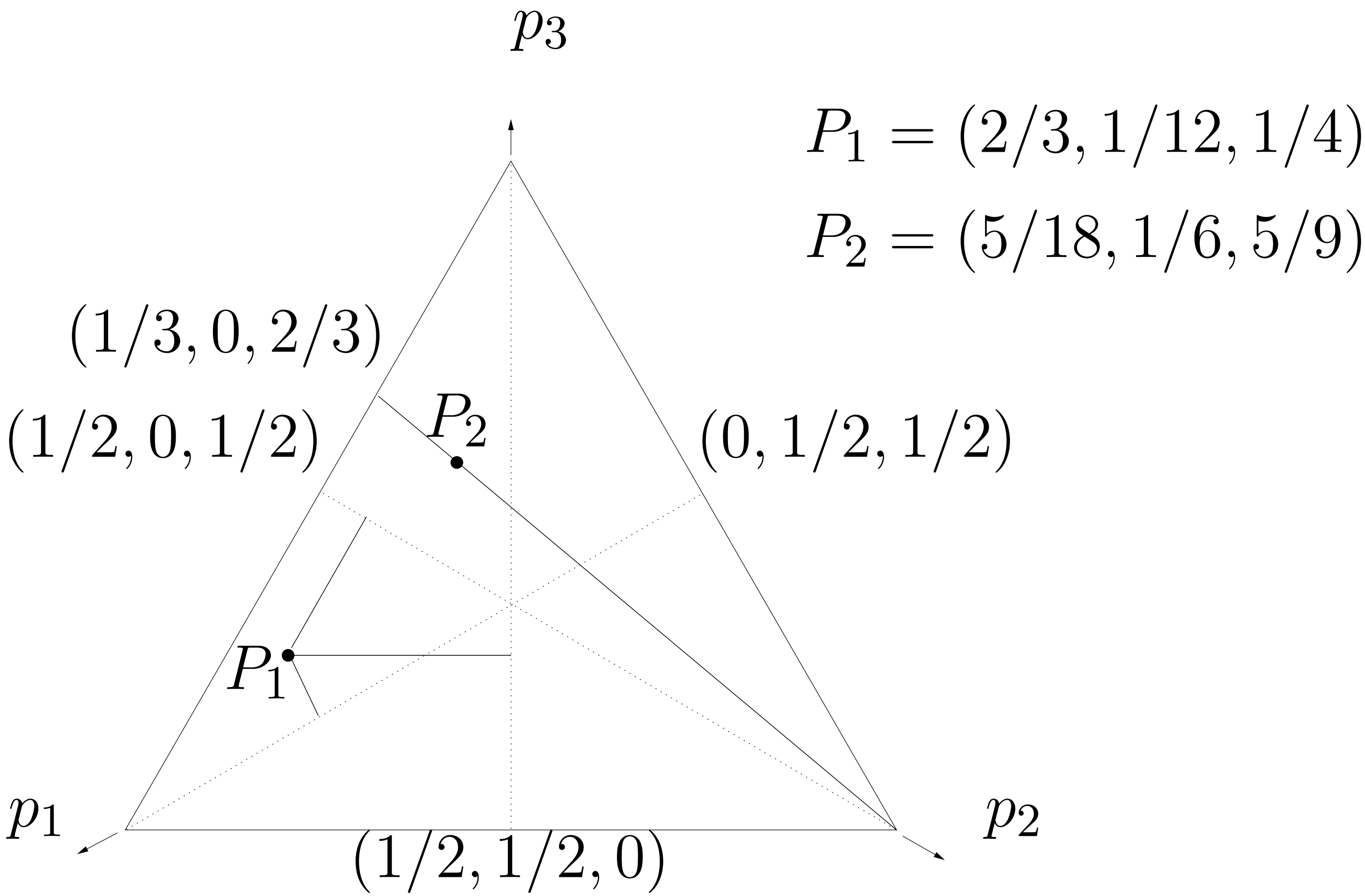
1. $\Omega = \{\omega_1, \omega_2, \omega_3\}$.
2. $P(\omega_i) = p_i$.
3. $p_1 > p_3$, $2p_1 \geq p_2$, $p_1 \leq 2/3$ and $p_3 \in [1/6, 1/3]$.
4. Take points $P = (p_1, p_2, p_3)$.

Some geometry

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



Baricentric coordinates



The coordinates of a distribution are read on the lines bissecting the angles of the triangle.

Exercise

Consider a variable X with 3 possible values x_1, x_2 and x_3. Suppose the following assessments are given:

$$p(x_1) \leq p(x_2) \leq p(x_3)\,;$$

$$p(x_i) \geq 1/20 \quad \text{for } i \in \{1, 2, 3\};$$

$$p(x_3 | x_2 \cup x_3) \leq 3/4.$$

Show the credal set determined by these assessments in baricentric coordinates.

Back to 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.

Exercise

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

- But given a lower expectation, many credal sets generate it.
- Usually only the maximal closed convex set is chosen.

- **Exercise:** Given the assessments in the previous exercise, find two credal sets that yield the same lower expectation.

Common ways to generate credal sets I

From partial preferences:

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

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

- 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.

Common ways to generate credal sets II

From one-sided betting:

- Variables are *gambles*.
- Buy/sell gambles using $\underline{E}[X]$ and $\overline{E}[X]$.
- Some constraints, such as $\sum_{i=1}^{n} \alpha_i (X_i - \underline{E}[X_i]) \geq 0$ for $\alpha_i \geq 0$.
- Credal set is produced by the set of *dominating* expectations:
$$\{E : E[X] \geq \underline{E}[X]\}.$$
- Several credal sets produce the same lower expectations.
  - But only maximal closed one is given "behavioral" significance.

Decision making with credal sets

- Set of acts $\mathcal{A}$, need to choose one.
  - There are several criteria!
- Γ-*minimax*:

$$\arg\max_{X \in \mathcal{A}} \underline{E}[X] \,.$$

- *Maximality*: maximal elements of the partial order $\succ$. That is, X is *maximal* if

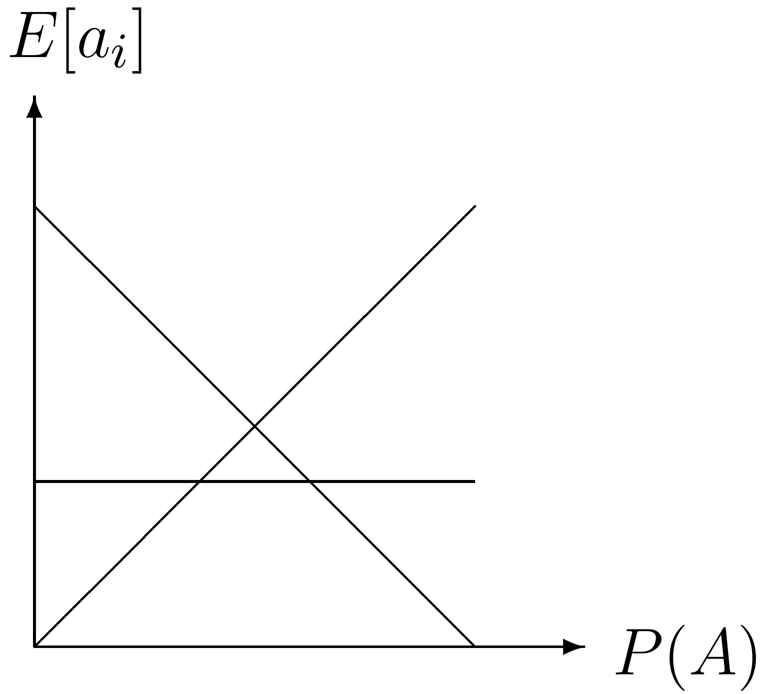
  there is no $Y \in \mathcal{A}$ such that $E_P[Y - X] > 0$ for all $P \in K$.

- *E-admissibility*: maximality for at least a distribution. That is, X is *E-admissible* if

  there is $P \in K$ such that $E_P[X - Y] \geq 0$ for all $Y \in \mathcal{A}$.

Comparing criteria

Three acts: $a_1 = 0.4$; $a_2 = 0/1$ if A/A^c; $a_3 = 1/0$ if A/A^c.



$P(A) \in [0.3, 0.7]$.

Γ-minimax: a_1; Maximal: all of them; E-admissible: $\{a_2, a_3\}$.

Exercise

Credal set $\{P_1, P_2\}$:

$$P_1(s_1) = 1/8, \quad P_1(s_2) = 3/4, \quad P_1(s_3) = 1/8,$$

$$P_2(s_1) = 3/4, \quad P_2(s_2) = 1/8, \quad P_2(s_3) = 1/8,$$

Acts $\{a_1, a_2, a_3\}$:

	s_1	s_2	s_3
a_1	3	3	4
a_2	2.5	3.5	5
a_3	1	5	4.

Which one to select?
And if we take convex hull of credal set?

Overview

1. Some basic (mostly known) definitions: credal sets, lower expectations and probabilities, decision making, and the like.
2. **Structural assessments: vacuity, uniformity, exchangeability.**
3. A brief review of stochastic (conditional) independence.
4. Confirmational/strict/strong/epistemic/Kuznetsov/others independence.
5. Comparison.
6. A look into the messy world of zero probabilities.

Structural assessments

- What is it?
- An assessment that alone constrains a large (possibly infinite) number of expectations.
- A simple example: vacuity.
- A credal set $K(X)$ is vacuous when it contains every possible distribution for X.

Vacuity

- Suppose $K(X)$ is vacuous.
- Then:

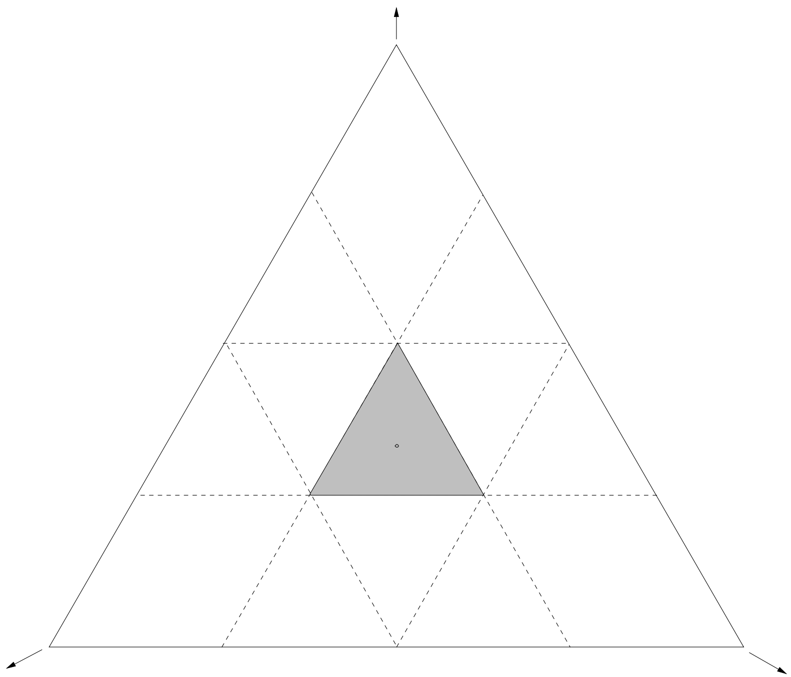
$$\underline{E}[f(X)] = \min_{\omega \in \Omega} f(X(\omega)), \qquad \overline{E}[f(X)] = \max_{\omega \in \Omega} f(X(\omega)).$$

- An ϵ-contaminated credal set is a “mixture” of a precise distribution and a vacuous credal set:

$$(1-\epsilon)P_0 + \epsilon Q, \quad \text{any } Q.$$

Uniformity

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



Exercise

- Urn with $m > 0$ balls, numbered from 1 to m
- r balls are red and $m - r$ balls are black.
- n samples with replacement.
- ω is a numbered sequence produced this way.
- m^n possible numbered sequences.
- Assume uniformity: $P(\omega) \geq (1 - \epsilon)m^{-n}$.
- What is the lower probability that k balls are red?

Exchangeability

- A basic structural assessment.
- To simplify, take categorical variables $\mathbf{X} = [X_1, \ldots, X_m]$.
- Denote by π_m a permutation of integers $\{1, \ldots, m\}$, and by $\pi_m(i)$ the ith number in the permutation.
- Denote

$$\{\mathbf{X} = \mathbf{x}\} \doteq \cap_{i=1}^{m} \{X_i = x_i\},$$

and

$$\{\pi_m \mathbf{X} = \mathbf{x}\} \doteq \cap_{i=1}^{m} \{X_{\pi_m(i)} = x_i\}.$$

Definition of exchangeability

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

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

for any permutation π_m.

- That is, the order of variables does not matter: trading $\{\mathbf{X} = \mathbf{x}\}$ for $\{\pi_m \mathbf{X} = \mathbf{x}\}$ does not seem advantageous.

Walley's exchangeability theorem

- We have

$$\begin{aligned} 0 &= \underline{E}[\{\mathbf{X} = \mathbf{x}\} - \{\pi_m \mathbf{X} = \mathbf{x}\}] \\ &\leq \overline{E}[\{\mathbf{X} = \mathbf{x}\} - \{\pi_m \mathbf{X} = \mathbf{x}\}] \\ &= -\underline{E}[\{\pi_m \mathbf{X} = \mathbf{x}\} - \{\mathbf{X} = \mathbf{x}\}] = 0. \end{aligned}$$

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

$$P(\mathbf{X} = \mathbf{x}) = P(\pi_m \mathbf{X} = \mathbf{x}) \qquad \text{for any permutation } \pi_m.$$

- In words: Exchangeability implies *elementwise* exchangeability.

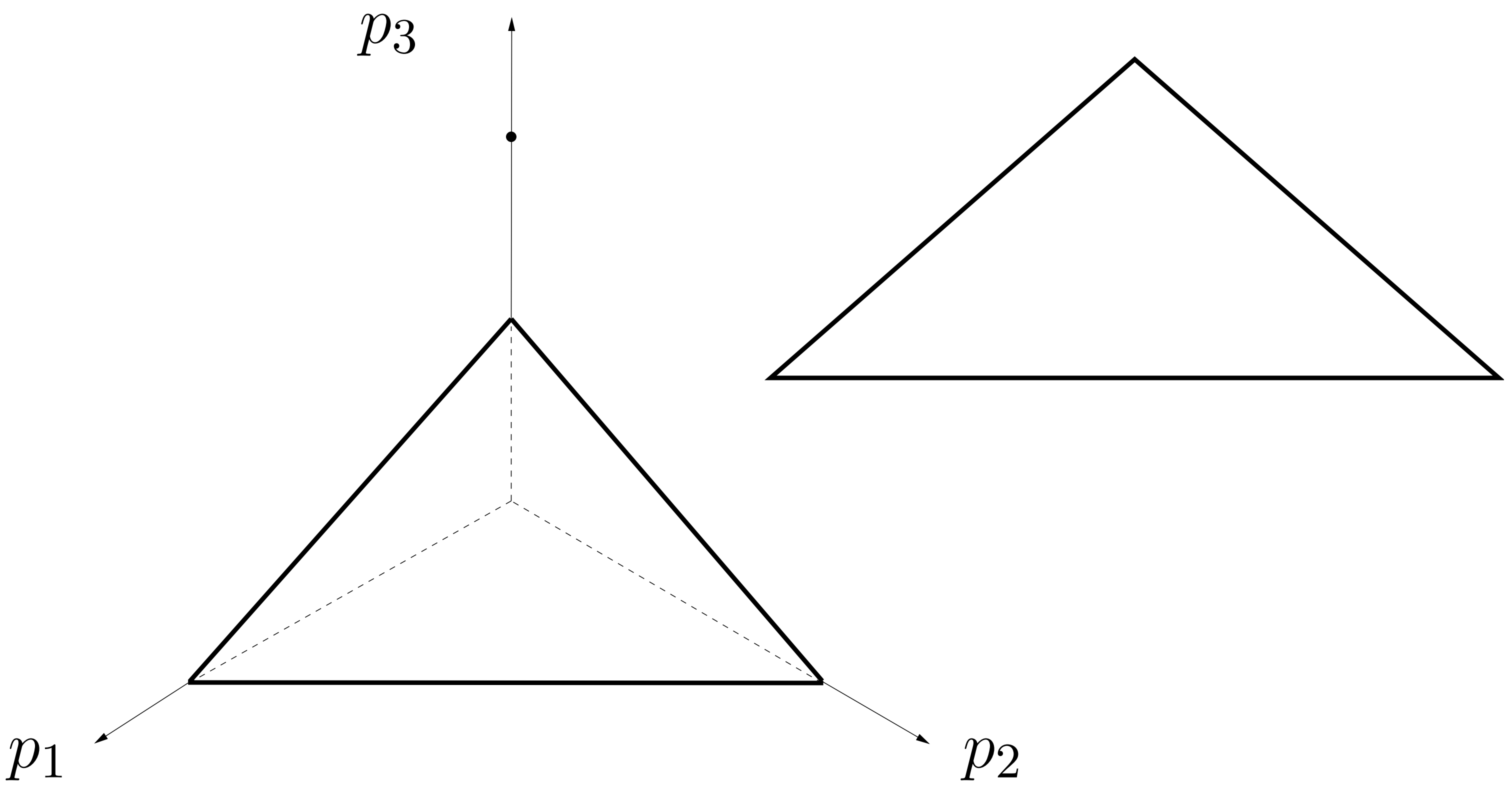
Exercise

What is the largest credal set that satisfies exchangeability of two binary variables?

Exercise

What is the largest credal set that satisfies exchangeability of two binary variables?

$p_1 = P(X = 0, Y = 0)$, $p_2 = P(X = 1, Y = 1)$,
$p_3 = P(X = 1, Y = 0) = P(X = 0, Y = 1)$.



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:

- One success: $P(0001) = P(0010) = P(0100) = P(1000)$.
- Two successes: $P(1001) = P(1010) = P(1100) = P(0101) = P(0110) = P(0011)$.
- Three successes: $P(1110) = P(1101) = P(1011) = P(0111)$.

Exercise

- Suppose we have 4 binary variables that are exchangeable.
- Suppose $P(0000) = 1/10$ and $P(1111) = 1/2$.
- Draw the credal set.

Exercise

- Suppose we have 4 binary variables that are exchangeable.
- Suppose $P(0000) = 1/10$ and $P(1111) = 1/2$.
- Draw the credal set.

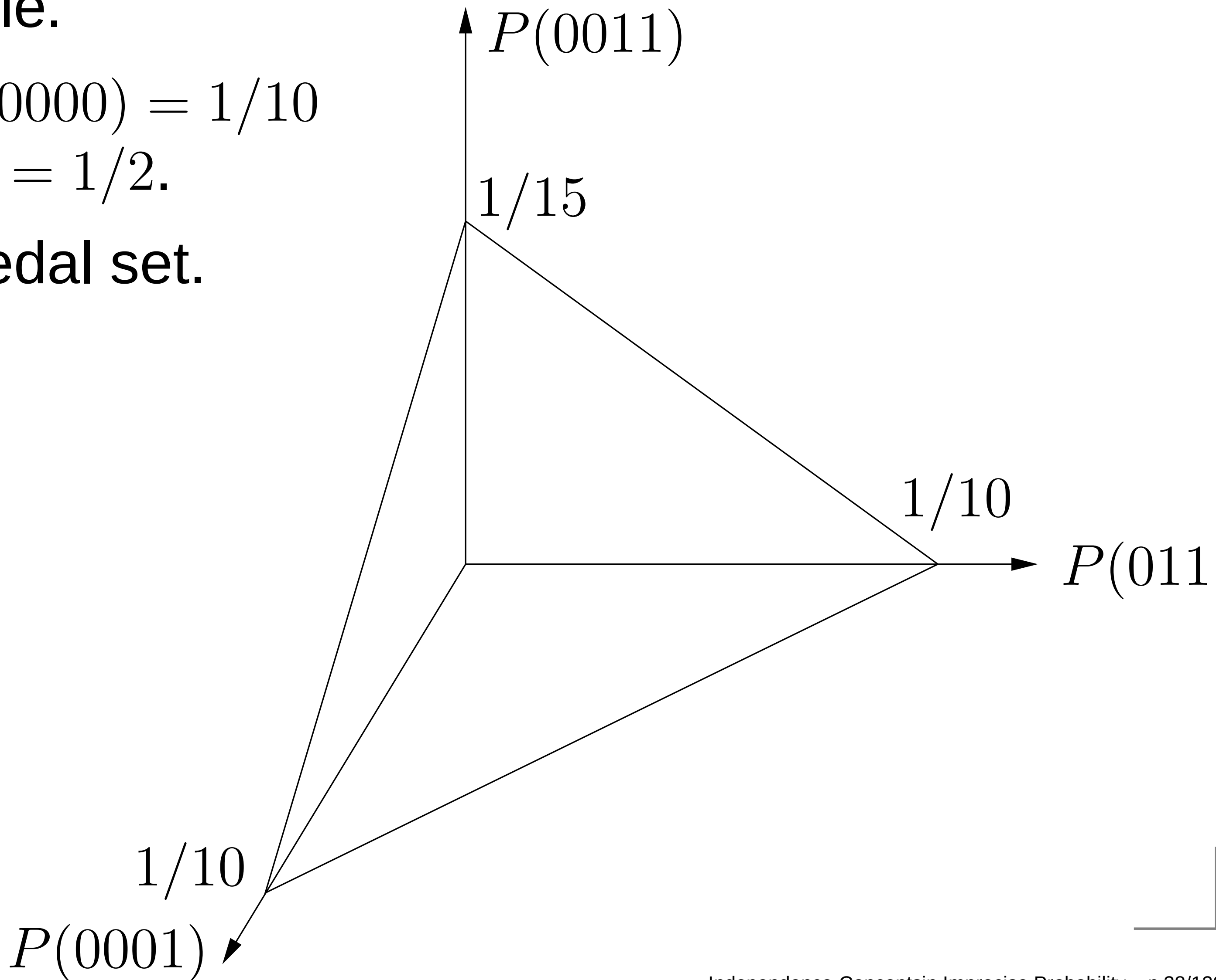
Set of triplets $[P(0001)\,, P(0011)\,, P(0111)]$ satisfying

$$P(0001) \geq 0, \quad P(0011) \geq 0, \quad P(0111) \geq 0,$$

$$4P(0001) + 6P(0011) + 4P(0111) = 1 - (1/2 + 1/10) = 2/5.$$

Exercise

- Suppose we have 4 binary variables that are exchangeable.
- Suppose $P(0000) = 1/10$ and $P(1111) = 1/2$.
- Draw the credal set.



Facts about exchangeability

- Any subset of exchangeable variables is exchangeable.
- Exchangeability is a “convex” concept.
- For $X_1, \ldots, X_m$, what matters is

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

- For each r, $\begin{pmatrix} m \\ r \end{pmatrix}$ probabilities with identical value

$$\frac{P\left(\sum_{i=1}^{n} X_i = r\right)}{\begin{pmatrix} m \\ r \end{pmatrix}}.$$

Representation for binary variables

- 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

$$\sum_{r=k}^{m-n+k} \frac{\binom{m-n}{r-k}}{\binom{m}{r}} P\left(\sum_{i=1}^{n} X_i = r\right) .$$

de Finetti's theorem (binary variables)

- Take $m \to \infty$:
  Then $P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0)$ is equal to
$$\int_0^1 \theta^k (1-\theta)^{n-k} dF(\theta).$$
- Here θ is the probability of $\{X_1 = 1\}$, and the distribution function $F(\theta)$ acts as a "prior" over θ.
- So: we have a credal set $K(\theta)$.
- Moreover: this credal set is convex!

Exercise

Draw the credal set $K(X,Y)$ given the structural assessments:

- X and Y are exchangeable.
- X and Y are the first two variables in a sequence of three exchangeable variables.
- X and Y are the first two variables in a sequence of five exchangeable variables.
- X and Y are the first two variables in a sequence of infinitely many exchangeable variables.

Overview

1. Some basic (mostly known) definitions: credal sets, lower expectations and probabilities, decision making, and the like.
2. Structural assessments: vacuity, uniformity, exchangeability.
3. **A brief review of stochastic (conditional) independence.**
4. Confirmational/strict/strong/epistemic/Kuznetsov/others independence.
5. Comparison.
6. A look into the messy world of zero probabilities.

Now, stochastic independence

1. X is stochatically irrelevant to Y when:

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

   for any bounded function f, whenever $P(\{X \in A\}) > 0$.
2. Definition is symmetric!
3. So, take it to mean
   stochastic independence of X and Y.

Symmetry

1. X is irrelevant to Y iff

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

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

2. X is irrelevant to Y iff

$$P(\{Y \in B\} \cap \{X \in A\}) = P(\{Y \in B\})\, P(\{X \in A\})\,.$$

Complete definition

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

$$E\left[f_i(X_i) | \cap_{j \neq i} \{X_j \in A_j\}\right] = E[f_i(X_i)],$$

for

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

Other forms

Independence of variables $\{X_i\}_{i=1}^n$ is equivalent to:

- For all functions $f_i(X_i)$,

$$E\left[\prod_{i=1}^{n} f_i(X_i)\right] = \prod_{i=1}^{n} E[f_i(X_i)] \, .$$

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

$$P(\cap_{i=1}^n \{X_i \in A_i\}) = \prod_{i=1}^{n} P(\{X_i \in A_i\}) \, .$$

Independence for events

1. A and B are independent

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

or, equivalently,

$$P(A \cap B) = P(A)\, P(B)\,.$$

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

$$P(\cap_i \{X_i \in A_i\}) = \prod_i P(\{X_i \in A_i\})\,.$$

Weak law of large numbers

1. Remember Chebyshev inequality:

$$P(|X - E[X]\,| \geq t) \leq \frac{V[X]}{t^2},$$

2. Apply inequality to $\bar{X} = \sum_i X_i/n$:

$$P\big(|\bar{X} - \mu| \geq \epsilon\big) \leq \frac{\sigma^2}{n\epsilon^2},$$

3. The larger the n, the smaller this probability!

$$\forall \epsilon > 0, \quad \lim_{n \to \infty} P\big(|\bar{X} - \mu| \geq \epsilon\big) = 0$$

4. There are other versions with different assumptions.

(Finite) strong law of large numbers

- Finitistic version:
  - for all $\epsilon > 0$,
  - there is integer N
  - such that for every positive integer k,

$$P\left(\exists n \in [N, N+k] : \left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \epsilon\right) < \epsilon.$$

Strong law of large numbers

In a sequence of variables $X_1, \ldots, X_n$, the mean converges to the expectation with probability one:

$$P\left(\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = \mu\right) = 1.$$

1. It requires countable additivity; that is,

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{\infty} P(A_i)\,.$$

2. It is really a *strong* result.

The graphoid properties

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

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:

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

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

Other graphoid properties

Often added:

Redundancy: $(X \perp\!\!\!\perp Y \,|\, X)$

Often added (true when probabilities are positive):

Intersection

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

Not discussed further in this talk.

Exercise

Prove decomposition, weak union and contraction for stochastic independence.

Overview

1. Some basic (mostly known) definitions: credal sets, lower expectations and probabilities, decision making, and the like.
2. Structural assessments: vacuity, uniformity, exchangeability.
3. A brief review of stochastic (conditional) independence.
4. **Confirmational/strict/strong/epistemic/Kuznetsov/others independence.**
5. Comparison.
6. A look into the messy world of zero probabilities.

Strict independence

- X and Y are *strictly independent* if for all $P \in K(X, Y)$,

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

- That is, elementwise stochastic independence.

- This concept violates convexity (presumably has no "behavioral" justification).

Failure of convexity

Example of Jeffrey's:

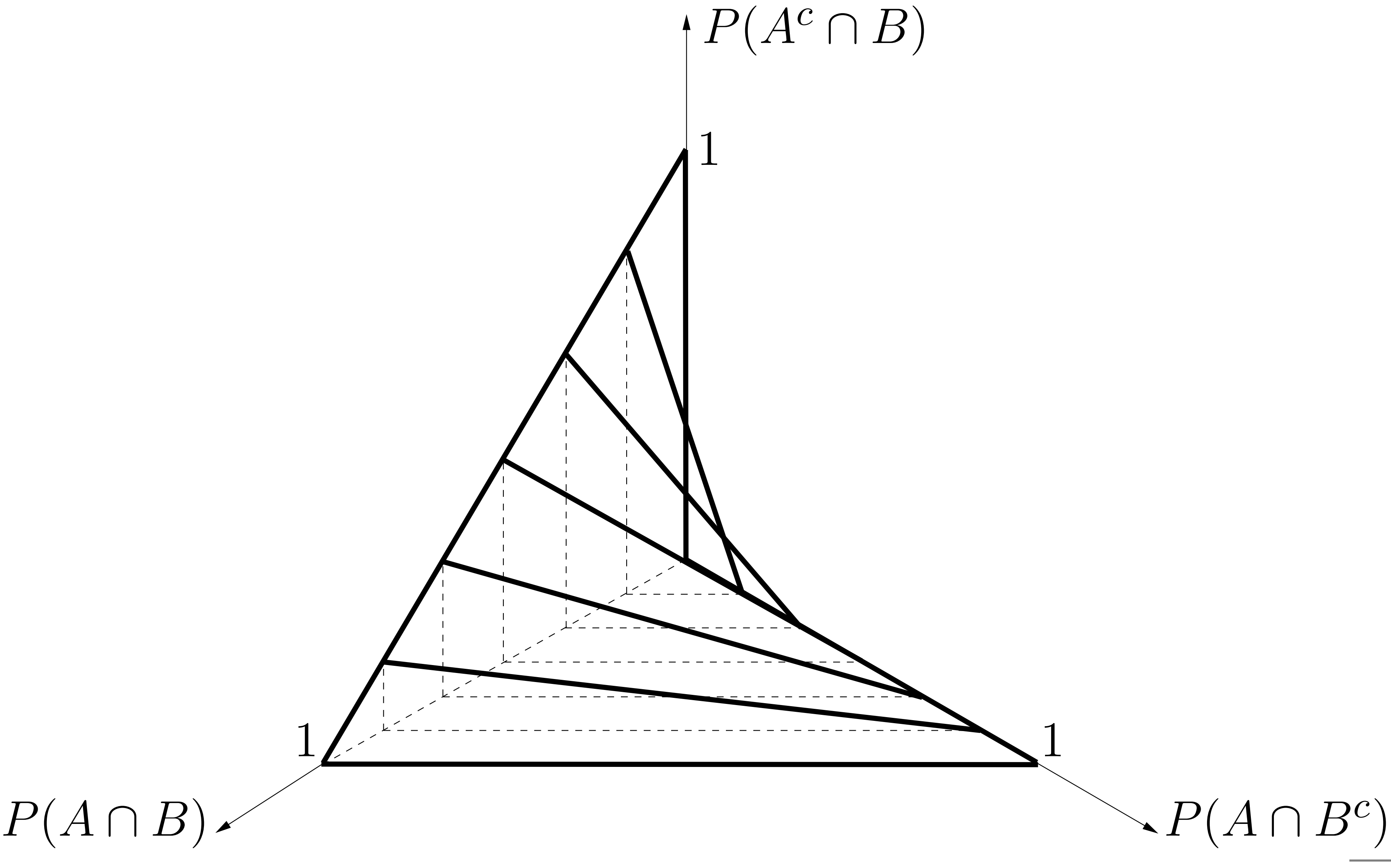
- Binary variables X and Y, strictly independent.
- $K(X, Y)$: convex hull of P_1 and P_2,

$$P_1(X = 0) = P_1(Y = 0) = 1/3, \quad P_2(X = 0) = P_2(Y = 0) = 2/$$

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

$$\begin{aligned} P_{1/2}(X = 0, Y = 0) &= P_1(X = 0)P_1(Y = 0)/2 + \\ &\quad P_2(X = 0)P_1(Y = 0)/2 \\ &= 5/18 \neq 1/4 \\ &= P_{1/2}(X = 0)P_{1/2}(Y = 0). \end{aligned}$$

Independence surface for two events



Confirmational independence

- I. Levi, the pioneer on convex credal sets, detected this problem with strict independence.

- His proposal: Y is *confirmationally irrelevant* to X if

$$K(X|Y \in B) = K(X) \qquad \text{for nonempty } \{Y \in B\},$$

- His position: use strict 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 strict independence.

- Equivalently (for closed credal sets):
  X and Y are strongly independent iff for any bounded function $f(X, Y)$,

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

Type-1/2 products and others

- Walley and Fine (1982) called this expression an *independent product* when restricted to indicators:

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

- 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.

Epistemic irrelevance

- Walley also proposes a different concept: Y is epistemically irrelevant to X if for any bounded function $f(X)$,

$$\underline{E}[f(X)|Y \in B] = \underline{E}[f(X)] \qquad \text{for nonempty } \{Y \in B\}.$$

- Definition is what Smith refers to as independence in his pioneering work on medial odds.

- *If* credal sets are closed and convex, then epistemic irrelevance is identical to Levi's confirmational irrelevance.

Exercise

- Consider a finite possibility space.
- Suppose $K(Y)$ is a singleton.
- Suppose $P(X)$, $K(X|Y \in B)$ are "almost" vacuous in that $P(X \in A|\cdot) > 0$ is the only constraint.
- Show that Y is epistemically irrelevant to X, but X is not epistemically irrelevant to Y.
- This is an extreme case of *dilation*!
- Construct an example that is not so extreme but that stills fails symmetry.

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.

Kuznetsov: some interval arithmetic

- Kuznetsov (1991) proposed yet another concept.
- Actually, he uses strong independence, but proposes a new concept as a secondary idea.
- His concept is based on interval arithmetic.
- Denote by $EI[X]$ the interval $\left[\underline{E}[X]\,,\overline{E}[X]\right]$.
- Overload the symbol $\times$ to understand $a \times b$ as the product of two intervals when a and b are intervals:

$$a = [\underline{a}, \overline{a}], b = [\underline{b}, \overline{b}] \quad \Rightarrow \quad a \times b = [\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{ab}].$$

Kuznetsov independence

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

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

- Equivalent formulation is: for any bounded functions $f(X)$ and $g(Y)$,

$$\begin{aligned} \underline{E}[f(X)g(Y)] &= \inf(E_{P_X \times P_Y}[f(X)g(Y)] : \\ &\qquad P_X \in K(X), P_Y \in K(Y)). \end{aligned}$$

Exercise

Prove:

- Kuznetsov independence implies epistemic independence.
- Epistemic independence does not imply Kuznetsov independence.

Strong $\neq$ Epistemic

- 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

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

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

$$[2/9, 2/9, 2/9, 1/3], [2/11, 3/11, 3/11, 3/11],$$

Exercise

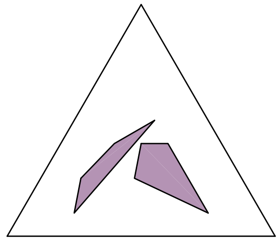
Write down the linear constraints that must be satisfied by $K(X, Y)$ in the previous example.

Strong $\neq$ Kuznetsov

- 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 $f(X)g(Y)$.

A possible variant

- X and Y are "independent" if

$$\underline{E}\big[f(X)|Y \in B'\big] = \underline{E}\big[f(X)|Y = B''\big]$$

for any bounded function $f(X)$ and any nonempt $\{Y \in B'\}$, $\{Y \in B''\}$.

- This is *not* epistemic irrelevance!

- It is quite weak. For instance we can have vacuous credal sets $K(X|Y = y)$ for every y . It seems bizarre to say that Y is then irrelevant to X.

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*

de Campos and Moral, 1995

- 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.

Type-5 independence

- Y is *type-5* irrelevant to X if

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

  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 defined conditioning (related to what Walley calls *regular extension*).

Exercise

Due to de Campos and Moral (1995).

- X and Y are binary.
- $K(X, Y)$ is the convex hull of two distributions P_1 and P_2 such that $P_1(X = 0, Y = 0) = P_2(X = 1, Y = 1) = 1$.

Show:

- X and Y are strongly independent.
- Neither Y is type-5 irrelevant to X, nor X is type-5 irrelevant to Y.

Couso et al, 1999

- 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

- Strict independence.
- Confirmational, epistemic irrelevance/independence.
- Strong independence.
- Kuznetsov independence.
- Type-5 independence.

The zoo, so far

- Strict independence.
- Confirmational, epistemic irrelevance/independence.
- Strong independence.
- Kuznetsov independence.
- Type-5 independence.

Consider:

- Epistemic independence is most intuitive (under convexity).
- Strict independence is closer to stochastic independence (without convexity).
- How to justify strong independence?

Conditional independence

- Any concept of independence can be modified to express *conditional independence*.
- For example, *conditional* epistemic irrelevance of Y to X given Z:

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

  for all bounded functions $f(X)$ and all nonempty $\{Z = z\}$.
- Likewise for conditional Kuznetsov/strict/strong independence of X and Y given Z.
- Aside: Moral and Cano (2002) consider three related forms of conditional strict independence (closer to extensions...).

Overview

1. Some basic (mostly known) definitions: credal sets, lower expectations and probabilities, decision making, and the like.
2. Structural assessments: vacuity, uniformity, exchangeability.
3. A brief review of stochastic (conditional) independence.
4. Confirmational/strict/strong/epistemic/Kuznetsov/others independence.
5. **Comparison.**
6. 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 (recent results by de Cooman and Miranda).
- Other idea: check graphoid properties.

Reminder: graphoid properties

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

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

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

Contraction:

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

Exercise

Show that strict and strong independence satisfy all graphoid properties.

Failure of contraction

- Epistemic independence 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.
- The other graphoid properties are satisfied by these concepts.

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:

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

- 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.

Failure of contraction: Kuznetsov indep.

- 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:

- Epistemic independence satisfies decomposition and weak union in finite spaces.
- 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.
- Kuznetsov independence satisfies decomposition.

An application: Markov chains

- Take chain $W \to X \to Y \to Z$.
- With stochastic independence, W and Z are conditionally stochastically given X (among other relations).
- But a Markov condition with epistemic independence does not guarantee such a relation.

(That is, a variable is epistemically independent of its predecessors given its parent.)

Comparing complexity

- Little is known about the computational complexity of various concepts.
- Strict/strong independence have been addressed in the context of credal networks.
- Some algorithms are known for epistemic independence.
- It seems that strict/strong independence are “more tractable” in an informal way.

The zoo, so far...

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

Consider:

- Epistemic independence is more intuitive (under convexity).
- Strict independence is closer to stochastic independence (without convexity).
- How to justify strong independence?

Justifying strong independence

- Sensitivity analysis interpretation: several experts agree on stochastic independence.
  - This is an argument for strict 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)).

Example

- 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: $K(X,Y)$ is convex hull of

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

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

$$[2/9, 2/9, 2/9, 1/3], [2/11, 3/11, 3/11, 3/11],$$

- Suppose we learn that

$$P(Y=0) = 4/9.$$

Changing assessments

- So, we have $K(X,Y)$ and we learn

$$P(Y = 0) = 4/9.$$

- If we simply generate

$$K'(X,Y) = K(X,Y) \cap \{P : P(Y = 0) = 4/9\}.$$

then X and Y are not epistemically independent anymore.

Producing strong independence

- This is "like" Jeffrey's rule: we change the marginal, then see what happens to the other marginal.
- 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.

- A bit strange: after learning new assessments, shouldn't we change $K(X, Y)$ so as to preserve the epistemic independence?

Another justification: exchangeability

- Consider a vector of m exchangeable binary variables $\mathbf{X} = [X_1, \dots, X_m]$.
- If we look at the first n variables and let $m \to \infty$, then $P(X_1 = 1, \dots, X_k = 1, X_{k+1} = 0, \dots, X_n = 0)$ is

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

- Remember: θ is the probability of $\{X_1 = 1\}$.
- We have a convex credal set $K(\theta)$.

Strong indep. from exchangeability

- So, n variables amongst infinitely many exchangeable variables.
- Represented by a convex credal set $K(\theta)$ as

$$P\big(X_{1,\ldots,k} = 1, X_{k+1,\ldots,n} = 0\big) = \int_0^1 \theta^k (1-\theta)^{n-k} dF(\theta).$$

- Strong independence obtains if each vertex of $K(\theta)$ assigns probability 1 to a particular value of θ.
  - We have in fact obtained a type-2 product.
  - Similar argument works for general variables.
  - It is possible to extend the argument to general strong independence (but a bit artificial).

Back to strict independence

- Strict independence is very attractive.
- But it violates convexity.
- It does not have a “behavioral” interpretation...
- Is it true?
- NO!
- Let’s think about E-admissibility.

Example

Credal set $\{P_1, P_2\}$:

$$P_1(s_1) = 1/8, \quad P_1(s_2) = 3/4, \quad P_1(s_3) = 1/8,$$

$$P_2(s_1) = 3/4, \quad P_2(s_2) = 1/8, \quad P_2(s_3) = 1/8,$$

Acts $\{a_1, a_2, a_3\}$:

	s_1	s_2	s_3
a_1	3	3	4
a_2	2.5	3.5	5
a_3	1	5	4.

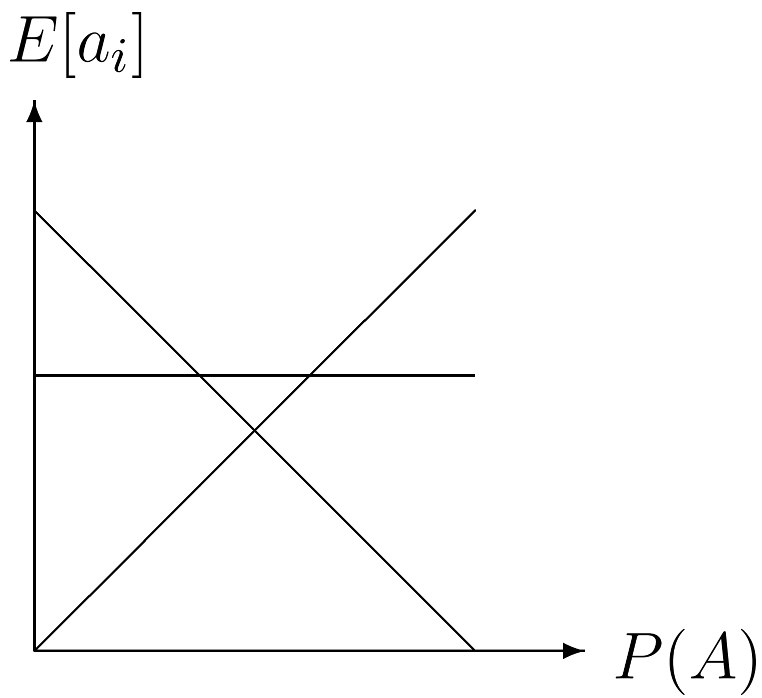
With respect to P_1 and P_2, a_1 and a_3 are E-admissible but a_2 is not; with respect to the convex hull of $\{P_1, P_2\}$, all acts are E-admissible.

That is,

There is a difference between a set of distributions and its convex hull when one chooses among several acts.

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!

Axiomatizing partial preferences

- 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)].
- Axioms on *rejection functions*: for a given set D of acts, $R(D)$ indicates the acts that are *not* admissible.
  - Example: An inadmissible act cannot become admissible when (a) new acts are added to the set of acts; (b) inadmissible acts are deleted from the set of acts.
  - And so on.

Producing strict independence

- Are events A and B are strictly independent?
- Construct five acts $a_0, \ldots, a_4$:

	AB	AB^c	A^cB	A^cB^c
a_0	0	0	0	0
a_1	$1-\alpha$	$-\alpha$	0	0
a_2	$-(1-\alpha)$	α	0	0
a_3	0	0	$1-\beta$	$-\beta$
a_4	0	0	$-(1-\beta)$	β

- 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 strictly independent.

Just to close

- How about confirmational independence for general credal sets?
- Very weak: fails decomposition/weak union/contraction!

	1	2	3	4
$P(X=0\|W=0,Y=0)$, $P(W=0,Y=0)$	α, $1/4$	α, $1/4$	α, $1/4$	β, $\frac{\beta/2}{\alpha+\beta}$
$P(X=0\|W=0,Y=1)$, $P(W=0,Y=0)$	α, $1/4$	α, $1/4$	α, $1/4$	β, $\frac{\alpha/2}{\alpha+\beta}$
$P(X=0\|W=1,Y=0)$, $P(W=0,Y=0)$	α, $\frac{\alpha/2}{\alpha+\beta}$	α, $\frac{(1-\alpha)/2}{2-(\alpha+\beta)}$	α, $1/4$	β, $1/4$
$P(X=0\|W=1,Y=1)$, $P(W=0,Y=0)$	α, $\frac{\beta/2}{\alpha+\beta}$	α, $\frac{(1-\beta)/2}{2-(\alpha+\beta)}$	α, $1/4$	β, $1/4$

	5	6	7
$P(X=0\|W=0,Y=0)$, $P(W=0,Y=0)$	β, $\frac{(1-\beta)/2}{2-(\alpha+\beta)}$	$\frac{\alpha+\beta}{2}$, $1/4$	β, $1/4$
$P(X=0\|W=0,Y=1)$, $P(W=0,Y=0)$	β, $\frac{(1-\alpha)/2}{2-(\alpha+\beta)}$	$\frac{\alpha+\beta}{2}$, $1/4$	α, $1/4$
$P(X=0\|W=1,Y=0)$, $P(W=0,Y=0)$	β, $1/4$	α, $1/4$	$\frac{\alpha+\beta}{2}$, $1/4$
$P(X=0\|W=1,Y=1)$, $P(W=0,Y=0)$	β, $1/4$	β, $1/4$	$\frac{\alpha+\beta}{2}$, $1/4$

Failure of decomposition and weak union; $\alpha, \beta \in (0,1)$.

Overview

1. Some basic (mostly known) definitions: credal sets, lower expectations and probabilities, decision making, and the like.
2. Structural assessments: vacuity, uniformity, exchangeability.
3. A brief review of stochastic (conditional) independence.
4. Confirmational/strict/strong/epistemic/Kuznetsov/others independence.
5. Comparison.
6. **A look into the messy world of zero probabilities.**

Potentially null events

- Events may have zero *lower* probability but nonzero *upper* probability (cannot ignore those).
- Example of difficulties one may face:
  - Suppose we refuse to define a conditional credal set $K(X|Y = y)$ whenever $\underline{P}(Y = y) = 0$.
  - Consider: Y is "irrelevant" to X if

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

  - But Y may have finitely many values, and for each value y of Y there is a distribution P in $K(Y)$ such that $P(Y = y) = 0$.
    - Then Y is irrelevant to any other variable!

Full conditional measures

- The most elegant solution is to consider *full probability measures*.
- A full probability measure is a function $P(\cdot|\cdot)$ on $\mathcal{E} \times \mathcal{E} \backslash \emptyset$ where $\mathcal{E}$ is an algebra of events, such that
  - $P(A|C) = 1$;
  - $P(A|C) \geq 0$ 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 Ω into events $L_0, \ldots, L_K$, where $K \leq N$,
- such that the full conditional measure is represented as a sequence of strictly positive probability measures $P_0, \ldots, P_K$, where the support of P_i is restricted to L_i.

Example:

	A	A^c
B	$\lfloor \beta \rfloor_1$	α
B^c	$\lfloor 1 - \beta \rfloor_1$	$1 - \alpha$

Here: $P(A) = 0$, but $P(B|A) = \beta$.

Using full conditional measures

- 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:

	A	A^c
B	$\lfloor \beta \rfloor_1$	α
B^c	$\lfloor 1-\beta \rfloor_1$	$1-\alpha$

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 the theory of graph-theoretical models?

Reminder: graphoid properties

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

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

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

Contraction:

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

Problem with epistemic independence

- It fails weak union!

	$w_0 y_0$	$w_1 y_0$	$w_0 y_1$	$w_1 y_1$
x_0	α	$\lfloor \beta \rfloor_2$	$1 - \alpha$	$\lfloor 1 - \beta \rfloor_2$
x_1	$\lfloor \alpha \rfloor_1$	$\lfloor \gamma \rfloor_3$	$\lfloor 1 - \alpha \rfloor_1$	$\lfloor 1 - \gamma \rfloor_3$

Remember:

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

Hammond's independence

- Here is a proposal for independence:

$$P(B(Y)|A(X) \cap D(Y)) = P(B(Y)|D(Y)) \text{ and}$$

$$P(A(X)|B(Y) \cap C(X)) = P(A(X)|C(X)) \, .$$

- This is symmetric.
- It satisfies weak union! But if fails contraction...

Remember:

Contraction:

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

Conclusion

- 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.
- My humble suggestion:

  We need to move to general credal sets, so that strict independence comes naturally

  (and many other things come naturally then...).

Conclusion

- 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.
- My humble suggestion:
  We need to move to general credal sets, so that strict independence comes naturally (and many other things come naturally then...).

Final words on independence I

- 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?).

Final words on independence II

- Strict 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).

Final words on independence III

- 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.