# IRRELEVANCE, INDEPENDENCE AND COHERENCE

Gert de Cooman

Ghent University, SYSTeMS

gert.decooman@UGent.be
http://users.UGent.be/~gdcooma
gertekoo.wordpress.com

5th SIPTA School
18 July 2012

## What would I like to achieve and convey?

## What would I like to achieve and convey?

| | | |
|---|---|---|
| Sets of probabilities are not necessarily the best model<br><br>lower previsions<br><br>DESIRABLE GAMBLES | | |

# On desirable gambles

## Probability measure versus expectation functional

Model for a variable $X$ assuming values in $\mathscr{X}$:

probability: $P(X \in A)$ for all events $A \subseteq \mathscr{X}$
expectation: $E(f(X))$ for all gambles $f\colon \mathscr{X} \to \mathbb{R}$

probability $P(X \in A)$ and expectation/prevision $E(f(X))$ are equally expressive

## Probability measure versus expectation functional

Model for a variable $X$ assuming values in $\mathscr{X}$:

probability: $P(X \in A)$ for all events $A \subseteq \mathscr{X}$
expectation: $E(f(X))$ for all gambles $f\colon \mathscr{X} \to \mathbb{R}$

lower probability $\underline{P}(X \in A)$ is less expressive than lower expectation/prevision $\underline{P}(f(X))$

# Probability measure versus expectation functional

Model for a variable $X$ assuming values in $\mathscr{X}$:

probability: $P(X \in A)$ for all events $A \subseteq \mathscr{X}$
expectation: $E(f(X))$ for all gambles $f\colon \mathscr{X} \to \mathbb{R}$

lower probability $\underline{P}(X \in A)$ is less expressive than
lower expectation/prevision $\underline{P}(f(X))$

WHEN WORKING WITH IMPRECISE PROBABILITIES,
USE (LOWER) EXPECTATIONS/PREVISIONS AND GAMBLES

# Set of desirable gambles as a belief model

Two types of imprecise-probability models (Walley, 1991):

lower expectation: $\underline{P}(f(X))$ for all gambles $f\colon \mathscr{X} \to \mathbb{R}$
set of desirable gambles: $\mathscr{D} \subseteq \mathscr{L}(\mathscr{X})$ is a set of gambles that a subject strictly prefers to zero

## Set of desirable gambles as a belief model

Two types of imprecise-probability models (Walley, 1991):

lower expectation: $\underline{P}(f(X))$ for all gambles $f\colon \mathscr{X} \to \mathbb{R}$

set of desirable gambles: $\mathscr{D} \subseteq \mathscr{L}(\mathscr{X})$ is a set of gambles that a subject strictly prefers to zero

Working with sets of desirable gambles $\mathscr{D}$:

- is simpler, more intuitive and more elegant
- is more general and expressive than lower previsions and even full conditional measures
- gives a geometrical flavour to probabilistic inference
- shows that probabilistic inference is 'logical' inference
- avoids problems with conditioning on sets of probability zero

The material for this part can be found in (Walley, 1991), (Moral, 2005) and (De Cooman and Miranda, 2012).

## Coherence for a set of desirable gambles

A set of desirable gambles $\mathscr{D}$ is called coherent if:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]

## Coherence for a set of desirable gambles

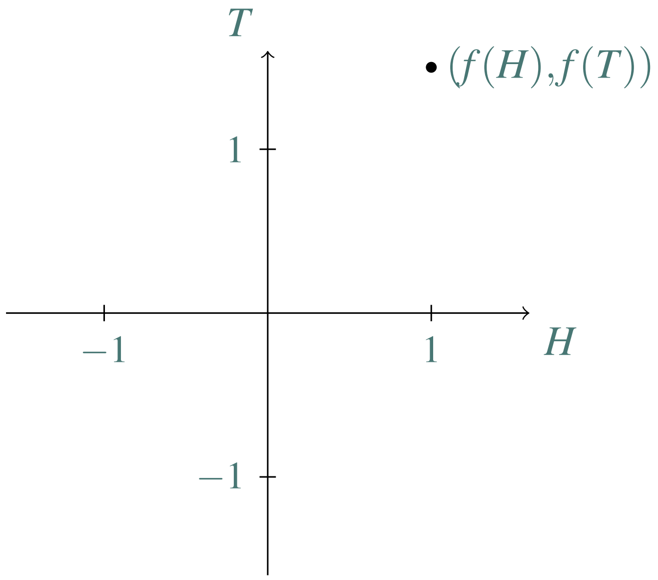
A set of desirable gambles $\mathscr{D}$ is called coherent if:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]




## Coherence for a set of desirable gambles

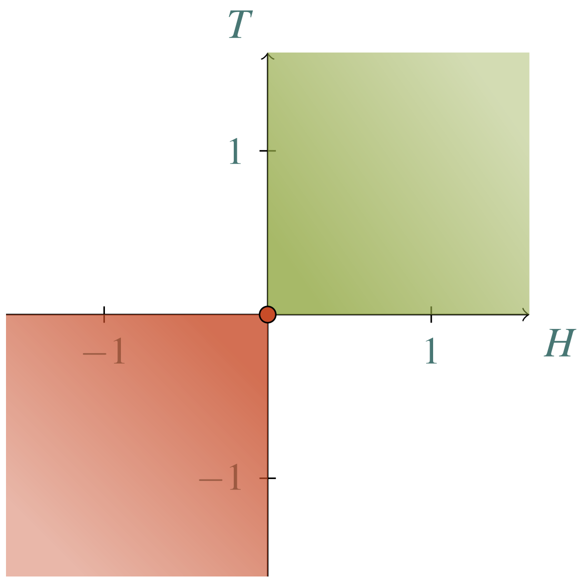
A set of desirable gambles $\mathscr{D}$ is called coherent if:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]



## Coherence for a set of desirable gambles

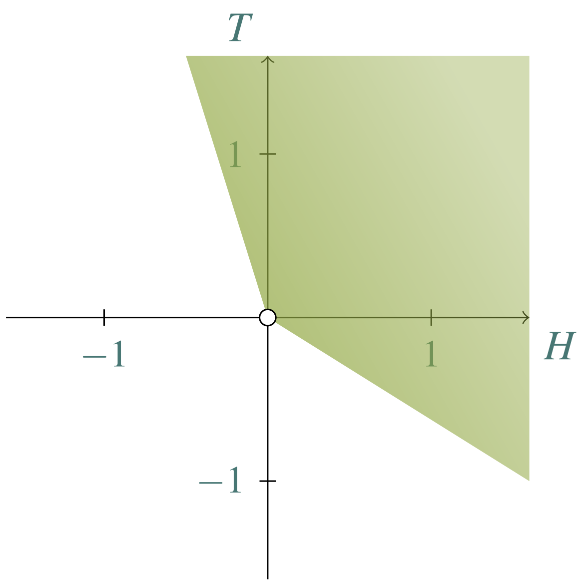
A set of desirable gambles $\mathscr{D}$ is called coherent if:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]




## Coherence for a set of desirable gambles

A set of desirable gambles $\mathscr{D}$ is called coherent if:

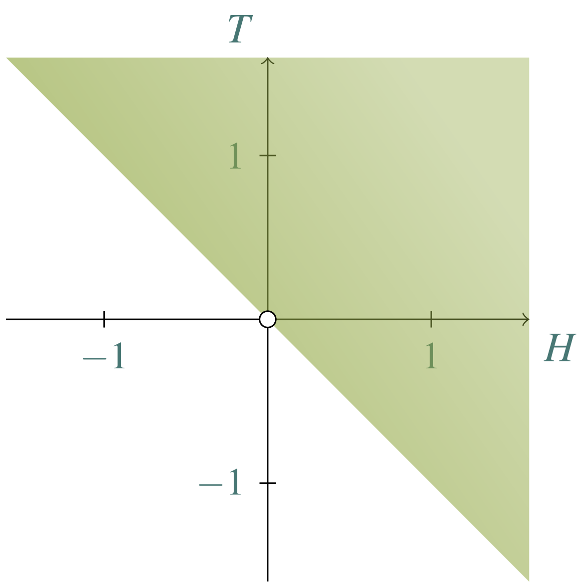
D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]

Precise models correspond to the special case that the convex cones $\mathscr{D}$ are actually halfspaces!



## Coherence for a set of desirable gambles

A set of desirable gambles $\mathscr{D}$ is called coherent if:

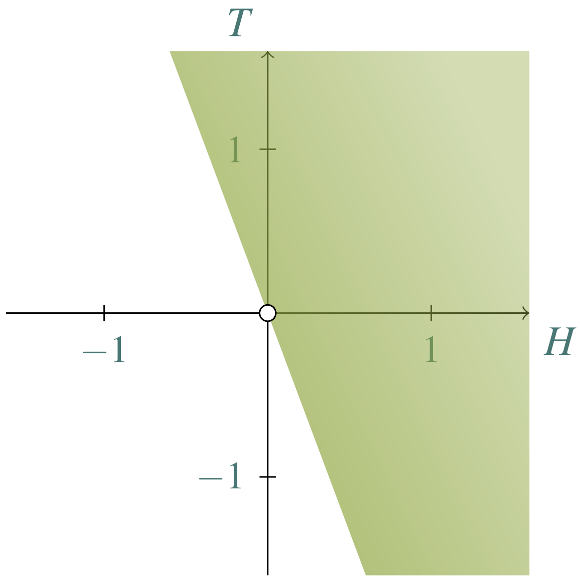
D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

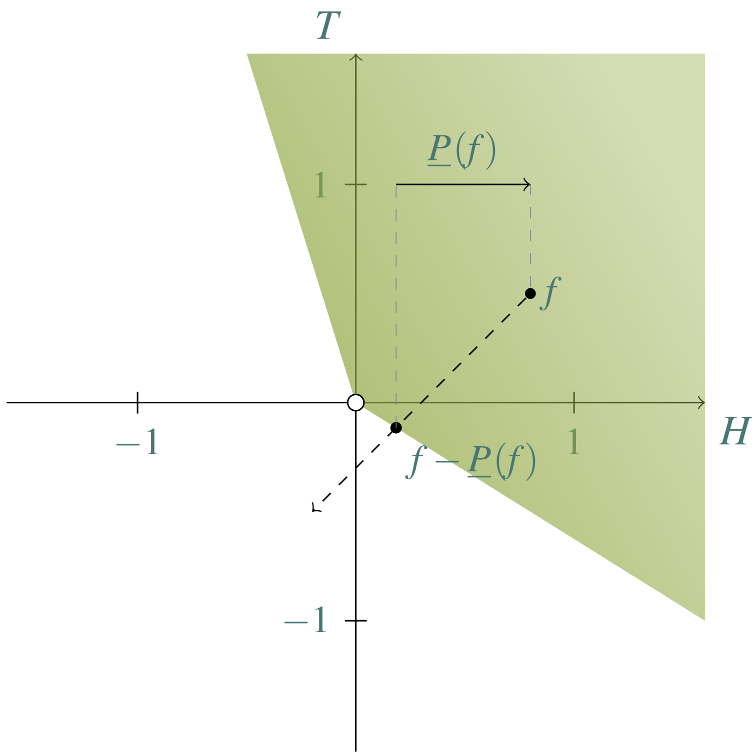
D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]

Precise models correspond to the special case that the convex cones $\mathscr{D}$ are actually halfspaces!




## Connection with lower previsions

$$\underline{P}(f) = \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathscr{D}\}$$



## Connection with lower previsions

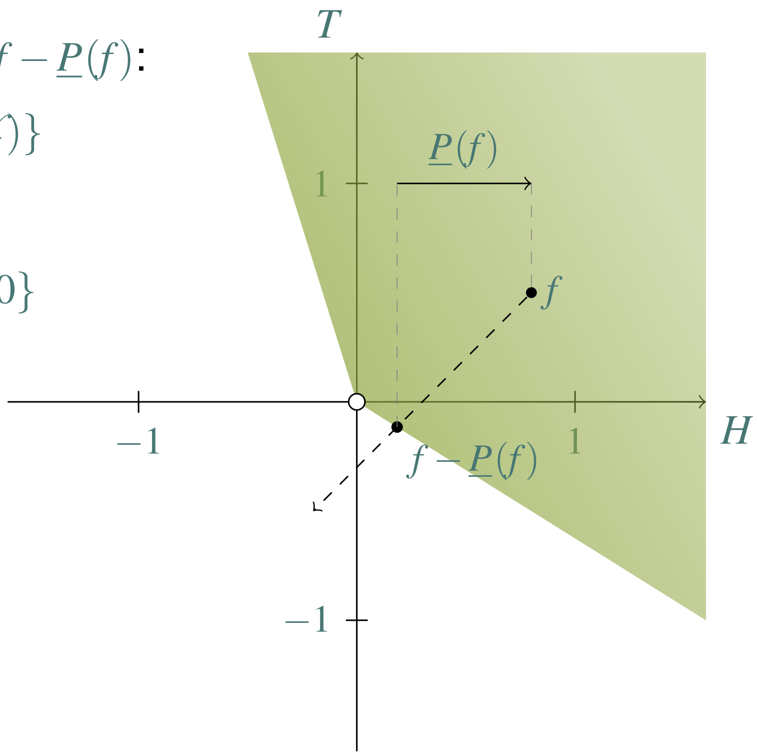
$\underline{P}(f) = \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathscr{D}\}$

marginal gambles $f - \underline{P}(f)$:

$\{f - \underline{P}(f) : f \in \mathscr{L}(\mathscr{X})\}$

$\{g : \underline{P}(g) = 0\}$

$\mathrm{cl}(\mathscr{D}) = \{f : \underline{P}(f) \geq 0\}$




## Exercise 1: sets of desirable gambles

Flipping a coin has two possible outcomes $X$: heads $H$ and tails $T$.

The general form of any coherent lower prevision for $X$ is given by

$$\underline{P}(f) = (1-\varepsilon)\,[p(H)f(H) + p(T)f(T)] + \varepsilon \min\{f(H), f(T)\},$$

where $p$ is any probability mass function on $\{H, T\}$ and $0 \leq \varepsilon \leq 1$.

If we have no reason to prefer heads over tails, we use a symmetrical model, with $\underline{P}(\{H\}) = \underline{P}(\{T\})$.

### Questions:

1. Find out which model corresponds to this symmetry requirement.
2. Draw a corresponding coherent set of desirable gambles, and find its extreme rays.

# Exercise 1: solution

From the symmetry requirement $\underline{P}(\{H\}) = \underline{P}(\{T\})$ we get, with

$$\underline{P}(\{H\}) = (1-\varepsilon)p(H) \text{ and } \underline{P}(\{T\}) = (1-\varepsilon)p(T),$$

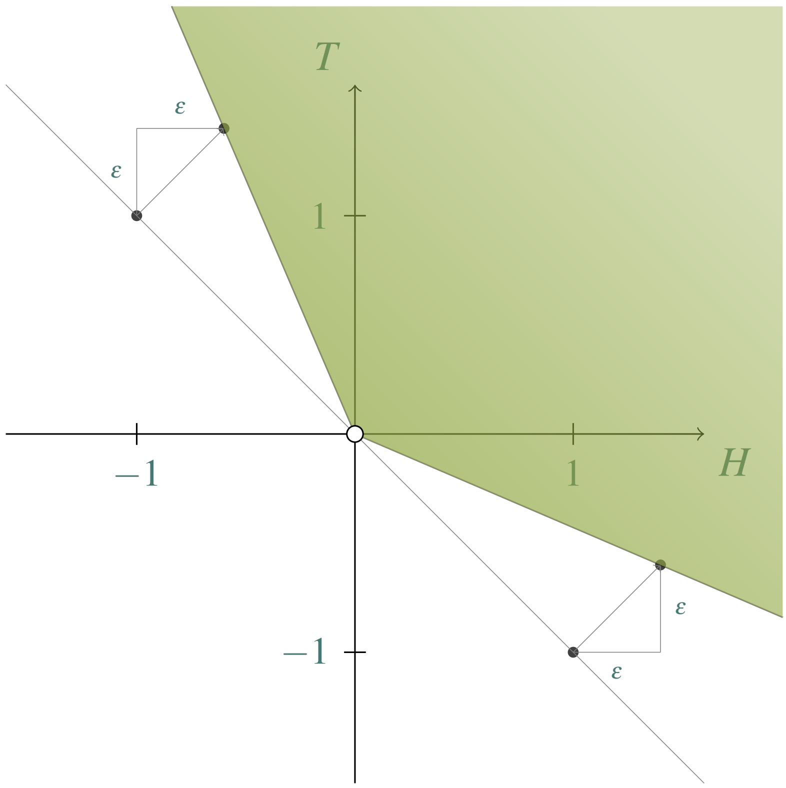
that:

$$p(H) = p(T) = \frac{1}{2}.$$

Hence

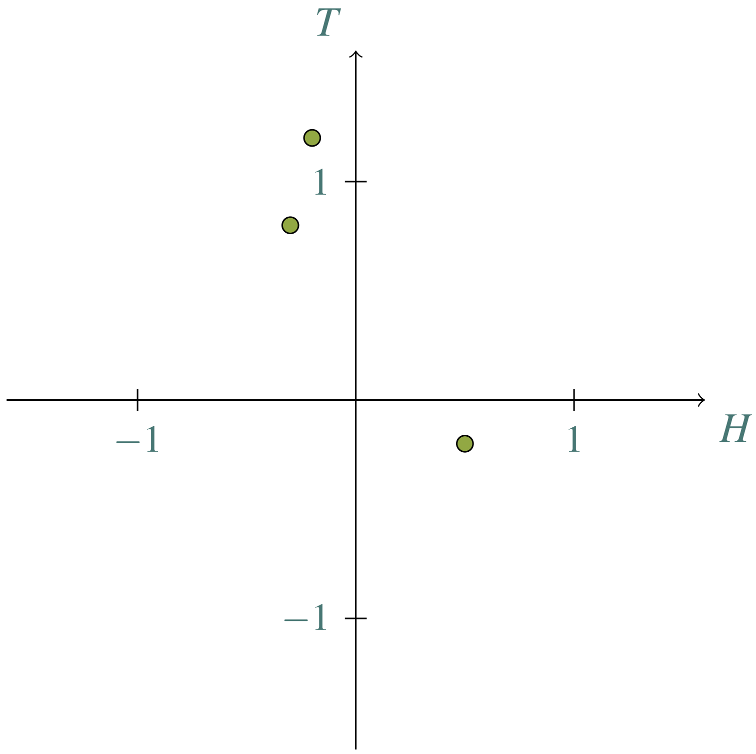
$$\underline{P}(f) = (1-\varepsilon)\frac{f(H)+f(T)}{2} + \varepsilon \min\{f(H), f(T)\}$$
$$\overline{P}(f) = (1-\varepsilon)\frac{f(H)+f(T)}{2} + \varepsilon \max\{f(H), f(T)\}$$

# Exercise 1: solution



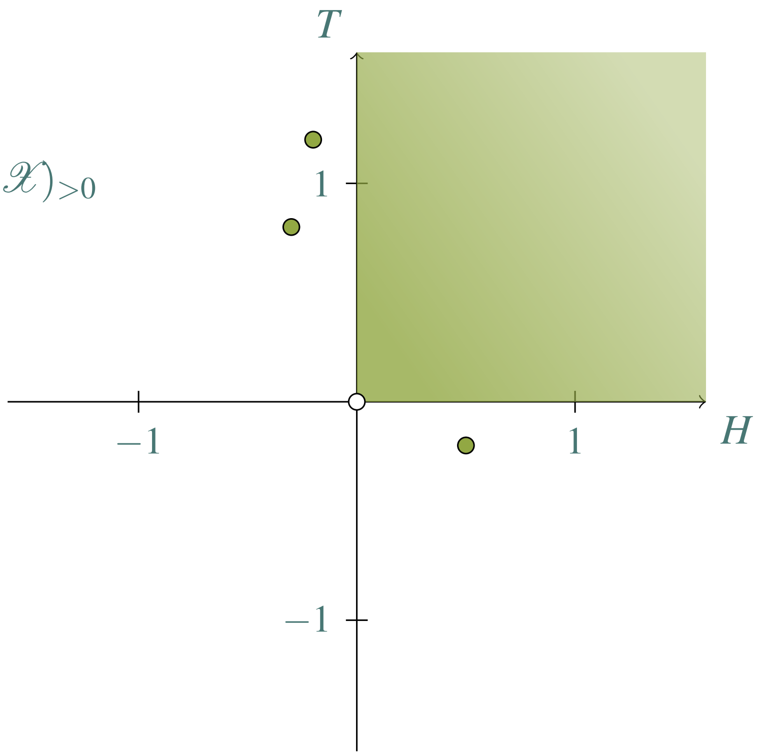
# Inference: natural extension

$\mathscr{A}$



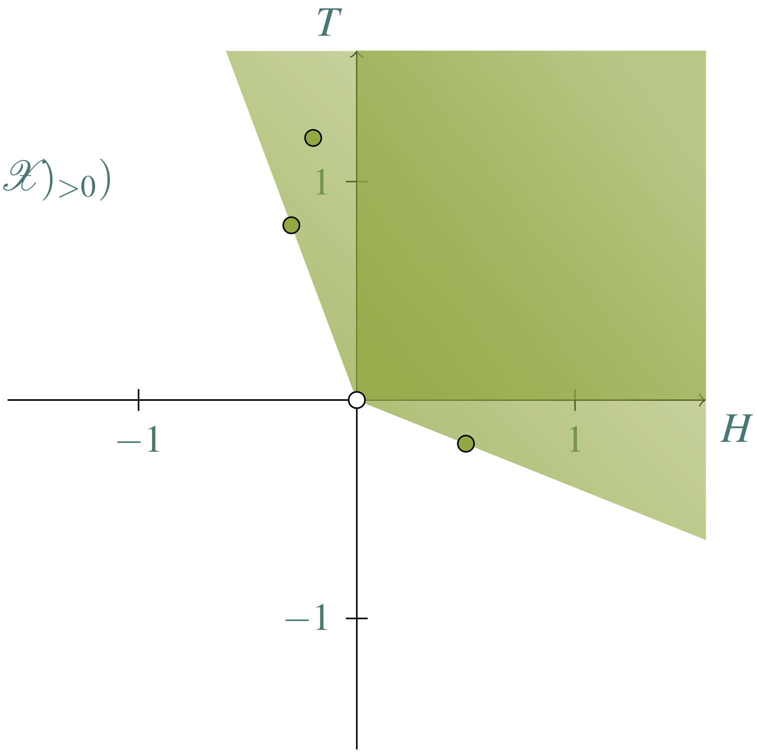

# Inference: natural extension

$\mathscr{A} \cup \mathscr{L}(\mathscr{X})_{>0}$



# Inference: natural extension

$$\mathcal{E}_{\mathscr{A}} := \operatorname{posi}(\mathscr{A} \cup \mathscr{L}(\mathscr{X})_{>0})$$




$$\operatorname{posi}(\mathscr{K}) := \left\{ \sum_{k=1}^{n} \lambda_k f_k : f_k \in \mathscr{K}, \lambda_k > 0, n > 0 \right\}$$

# Several variables: notation

Consider variables $X_n$ assuming values in the finite set $\mathscr{X}_n$, with $n \in N$.

For any subset $I \subseteq N$:

- the joint variable $X_I$ is the $I$-tuple of $X_i$ for $i \in I$, where

  $X_I(i) = X_i$ is the $i$-th component of $X_I$

- it assumes values in the product set $\mathscr{X}_I := \times_{i \in I} \mathscr{X}_i$
- we denote generic values of $X_I$ in $\mathscr{X}_I$ by $x_I$ and $z_I$, where

  $x_I(i) = x_i$ is the $i$-th component of $x_I$

- $\mathscr{L}(\mathscr{X}_I)$ denotes the set of all gambles $f : \mathscr{X}_I \to \mathbb{R}$ on $X_I$, also $f(X_I)$

## Marginalisation

Let $\mathscr{D}_N$ be a coherent set of desirable gambles on $\mathscr{X}_N$, with lower prevision $\underline{P}_N$.

For any subset $I \subseteq N$, we have the $\mathscr{X}_I$-marginals:

$$\mathscr{D}_I = \mathrm{marg}_I(\mathscr{D}_N) := \mathscr{D}_N \cap \mathscr{L}(\mathscr{X}_I),$$

so

$$f(X_I) \in \mathscr{D}_I \Leftrightarrow f(X_I) \in \mathscr{D}_N.$$

$$\underline{P}_I = \mathrm{marg}_I(\underline{P}_N) := \underline{P}_N|_{\mathscr{L}(\mathscr{X}_I)},$$

so

$$\underline{P}_I(f) = \underline{P}_N(f) \text{ for all gambles } f(X_I) \text{ on } \mathscr{X}_I.$$

## The conditioning rule

Consider any subset $I$ of $N$.

How to condition a coherent set $\mathscr{D}_N$ on the observation that $X_I = x_I$?

- This leads to an updated set of desirable gambles $\mathscr{D}_N|x_I \subseteq \mathscr{L}(\mathscr{X}_N)$ on $\mathscr{X}_N$:

$$\begin{aligned} f \in \mathscr{D}_N|x_I &\Leftrightarrow f > 0 \text{ or } \mathbb{I}_{\{x_I\}}f \in \mathscr{D}_N \\ &\Leftrightarrow f > 0 \text{ or } \mathbb{I}_{\{x_I\}}f(x_I, \cdot) \in \mathscr{D}_N \end{aligned}$$

## The conditioning rule

Consider any subset $I$ of $N$.

How to condition a coherent set $\mathscr{D}_N$ on the observation that $X_I = x_I$?

- This leads to an updated set of desirable gambles $\mathscr{D}_N|x_I \subseteq \mathscr{L}(\mathscr{X}_N)$ on $\mathscr{X}_N$:

$$f \in \mathscr{D}_N|x_I \Leftrightarrow f > 0 \text{ or } \mathbb{I}_{\{x_I\}}f \in \mathscr{D}_N$$
$$\Leftrightarrow f > 0 \text{ or } \mathbb{I}_{\{x_I\}}f(x_I, \cdot) \in \mathscr{D}_N$$

- Equivalently, this leads to an updated set of desirable gambles $\mathscr{D}_N \rfloor x_I \subseteq \mathscr{L}(\mathscr{X}_{N\setminus I})$ on $\mathscr{X}_N$:

$$g \in \mathscr{D}_N \rfloor x_I \Leftrightarrow \mathbb{I}_{\{x_I\}}g \in \mathscr{D}_N$$

Observe that $g \in \mathscr{D}_N \rfloor x_I \Leftrightarrow \mathbb{I}_{\{x_I\}}g \in \mathscr{D}_N|x_I$.

## Conditional lower previsions

Just like in the unconditional case, we can use a coherent set of desirable gambles $\mathscr{D}_N$ to derive conditional lower previsions.

Consider disjoint subsets $I$ and $O$ of $N$:

$$\underline{P}_O(g|x_I) := \sup\left\{\mu \in \mathbb{R} \colon \mathbb{I}_{\{x_I\}}[g-\mu] \in \mathscr{D}_N\right\} \text{ for all } g \in \mathscr{L}(\mathscr{X}_O)$$

is the lower prevision of $g$, conditional on $X_I = x_I$.

$\underline{P}_O(g|X_I)$ is the gamble on $\mathscr{X}_I$ that assumes the value $\underline{P}_O(g|x_I)$ in $x_I \in \mathscr{X}_I$.

# Coherent conditional lower previsions

Consider $m$ couples of disjoint subsets $I_s$ and $O_s$ of $N$, and corresponding conditional lower previsions $\underline{P}_{O_s}(\cdot|X_{I_s})$ for $s = 1, \ldots, m$.

### Theorem

*These conditional lower previsions are (jointly) coherent if and only if there is some coherent set of desirable gambles $\mathscr{D}_N$ that produces them, in the sense that for all $s = 1, \ldots, m$:*

$$\underline{P}_{O_s}(g|x_{I_s}) := \sup\left\{\mu \in \mathbb{R} \colon \mathbb{I}_{\{x_{I_s}\}}[g-\mu] \in \mathscr{D}_N\right\}$$

*for all $g \in \mathscr{L}(\mathscr{X}_{O_s})$ and all $x_{I_s} \in \mathscr{X}_{I_s}$.*

# All you know about probability theory . . .

All you know about propositional logic and about probability theory can be inferred from:

- the coherence rules D1–D4
- the conditioning rule
- (and perhaps some extra continuity requirements)

for sets of desirable gambles.

1. Bayes's Rule and Theorem
2. laws of large numbers
3. other limit laws
4. . . .

But they provide a solid foundation for imprecise probabilities too!

## What would I like to achieve and convey?

| Sets of probabilities are not necessarily the best model | EPISTEMIC IRRELEVANCE: what is it and how can it be used? | |
|---|---|---|
| lower previsions | coherence | |
| DESIRABLE GAMBLES | convenient properties | |

# Epistemic irrelevance

## What is an epistemic irrelevance statement?

Consider disjoint subsets $I$ and $O$ and $C$ of $N$:

### Definition (Epistemic irrelevance)

I say that $X_I$ is (epistemically) irrelevant to $X_O$ when I assess that learning the value $x_I$ that $X_I$ assumes will not change my beliefs about $X_O$.

Notation: $X_I \nrightarrow X_O$

More generally:

### Definition (Conditional epistemic irrelevance)

I say that $X_I$ is (epistemically) irrelevant to $X_O$ conditional on $X_C$ when I assess that, knowing the value $x_C$ of $X_C$, learning in addition the value $x_I$ that $X_I$ assumes will not change my beliefs about $X_O$.

Notation: $X_I \nrightarrow X_O | X_C$

## How can it be modelled: desirable gambles

1. Suppose I have a joint model $\mathscr{D}_N$, then the restrictive effect is to impose the following condition on it:

$$X_I \nrightarrow X_O \Leftrightarrow (\forall x_I \in \mathscr{X}_I)\, \mathrm{marg}_O(\mathscr{D}_N \rfloor x_I) = \mathrm{marg}_O(\mathscr{D}_N).$$

# How can it be modelled: desirable gambles

1. Suppose I have a joint model $\mathscr{D}_N$, then the restrictive effect is to impose the following condition on it:

$$X_I \nrightarrow X_O \Leftrightarrow (\forall x_I \in \mathscr{X}_I)\, \mathrm{marg}_O(\mathscr{D}_N \rfloor x_I) = \mathrm{marg}_O(\mathscr{D}_N).$$

2. Suppose I have a marginal model $\mathscr{D}_O$ for $X_O$, then the effect of an assessment $X_I \nrightarrow X_O$ is that

$$f(X_O) \in \mathscr{D}_O \Rightarrow f(X_O) \in \mathscr{D}_N \rfloor x_I \Rightarrow \mathbb{I}_{\{x_I\}} f(X_O) \in \mathscr{D}_N$$

so the constructive effect is to state that all gambles in the set

$$\mathscr{A}^{\mathrm{irr}}_{I \to O} := \left\{\mathbb{I}_{\{x_I\}} f \colon x_I \in \mathscr{X}_I \text{ and } f \in \mathscr{D}_O\right\},$$

and therefore also all gambles in $\mathrm{posi}\, \mathscr{A}^{\mathrm{irr}}_{I \to O}$, should be desirable.

# How can it be modelled: lower previsions

1. Suppose I have a marginal model $\underline{P}_O$ for $X_O$, then the constructive effect of an assessment $X_I \nrightarrow X_O$ is the definition of a conditional lower prevision $\underline{P}_O(\cdot|X_I)$ by:

$$\underline{P}_O(g|x_I) := \underline{P}_O(g) \text{ for all } g \in \mathscr{L}(\mathscr{X}_O) \text{ and all } x_I \in \mathscr{X}_I.$$

## How can it be modelled: lower previsions

1. Suppose I have a marginal model $\underline{P}_O$ for $X_O$, then the constructive effect of an assessment $X_I \nrightarrow X_O$ is the definition of a conditional lower prevision $\underline{P}_O(\cdot|X_I)$ by:

$$\underline{P}_O(g|x_I) := \underline{P}_O(g) \text{ for all } g \in \mathscr{L}(\mathscr{X}_O) \text{ and all } x_I \in \mathscr{X}_I.$$

2. Suppose I have a joint model $\underline{P}_N$ for $X_N$, then the restrictive effect of an assessment $X_I \nrightarrow X_O$ is to impose the coherence of $\underline{P}_N$ with the conditional lower prevision $\underline{P}_O(\cdot|X_I)$ defined by:

$$\underline{P}_O(g|x_I) := \underline{P}_N(g) \text{ for all } g \in \mathscr{L}(\mathscr{X}_O) \text{ and all } x_I \in \mathscr{X}_I.$$

# Irrelevant joints and products

# A family of irrelevance statements . . .

Consider $m$ pairs of disjoint subsets $I_s$ and $O_s$ of $N$, and the $m$ corresponding epistemic irrelevance assessments:

$X_{I_s}$ is epistemically irrelevant to $X_{O_s}$, for $s = 1, \dots, m$

or equivalently

$$\mathscr{I} = \{X_{I_s} \nrightarrow X_{O_s} : s \in \{1, \dots, m\}\}$$

HOW TO COMBINE THIS WITH COHERENCE?

# . . . and a coherent joint model

- Any coherent set of desirable gambles $\mathscr{D}_N$ on $\mathscr{X}_N$ that is compatible with the irrelevance assessments in the family $\mathscr{I}$, in the sense that:

  $$\mathrm{marg}_{O_s}(\mathscr{D}_N \rfloor x_{I_s}) = \mathrm{marg}_{O_s}(\mathscr{D}_N) \text{ for all } x_{I_s} \in \mathscr{X}_{I_s} \text{ and all } s = 1, \dots, m$$

  is called an $\mathscr{I}$-irrelevant joint.

## . . . and a coherent joint model

- Any coherent set of desirable gambles $\mathscr{D}_N$ on $\mathscr{X}_N$ that is compatible with the irrelevance assessments in the family $\mathscr{I}$, in the sense that:

  $$\mathrm{marg}_{O_s}(\mathscr{D}_N \rfloor x_{I_s}) = \mathrm{marg}_{O_s}(\mathscr{D}_N) \text{ for all } x_{I_s} \in \mathscr{X}_{I_s} \text{ and all } s = 1, \dots, m$$

  is called an $\mathscr{I}$-irrelevant joint.

- Any coherent lower prevision $\underline{P}_N$ on $\mathscr{L}(\mathscr{X}_N)$ that is compatible with the irrelevance assessments in the family $\mathscr{I}$, in the sense that it is *jointly coherent with* all conditional lower previsions $\underline{P}_{O_s}(\cdot|X_{I_s})$ defined by:

  $$\underline{P}_{O_s}(g|x_{I_s}) = \underline{P}_N(g) \text{ for all } x_{I_s} \in \mathscr{X}_{I_s}, \text{ all } g \in \mathscr{L}(\mathscr{X}_{O_s}) \text{ and all } s = 1, \dots, m$$

  is called an $\mathscr{I}$-irrelevant joint.

# Lower envelopes

### Lower envelope result

- the intersection $\bigcap_{i \in I} \mathscr{D}_i$ of any non-empty collection $\mathscr{D}_i$, $i \in I$ of $\mathscr{I}$-irrelevant joints is still $\mathscr{I}$-irrelevant;
- the lower envelope $\inf_{i \in I} \underline{P}_i$ of any non-empty collection $\underline{P}_i$, $i \in I$ of $\mathscr{I}$-irrelevant joints is still $\mathscr{I}$-irrelevant.

# Forward irrelevance

## A forward irrelevance assessment

Consider the following ordered chain of variables

$$X_1 \longrightarrow X_2 \longrightarrow \cdots \longrightarrow X_{n-1} \longrightarrow X_n$$

and the forward irrelevant assessment that you don't learn about the future from past observations:

$$X_1 \not\rightarrow X_{\{2,\dots,n\}}$$
$$X_2 \not\rightarrow X_{\{3,\dots,n\}}$$
$$\vdots$$
$$X_{n-2} \not\rightarrow X_{\{n-1,,n\}}$$
$$X_{n-1} \not\rightarrow X_n$$

$$\mathscr{I}_{\text{fir}} := \{X_k \not\rightarrow X_{\{k+1,\dots,n\}} : k \in \{1,\dots,n-1\}\}$$

# Forward irrelevant natural extension

Suppose we have marginal lower previsions $\underline{P}_k$ on $\mathscr{L}(\mathscr{X}_k)$ for each $X_k$.

We call a $\mathscr{I}_{\text{fir}}$-irrelevant joint a forward irrelevant joint.

Any forward irrelevant joint with these marginals is called a forward irrelevant product of these marginals.

There is a smallest forward irrelevant product, the forward irrelevant natural extension, given by (De Cooman and Miranda, 2009):

$$(\underline{P}_1 \circledcirc \underline{P}_2 \circledcirc \dots \circledcirc \underline{P}_n)(f) \coloneqq \underline{P}_1(\underline{P}_2(\dots(\underline{P}_n(f)\dots)))$$

# Exercise 2: forward irrelevant natural extension

We flip two coins successively, with respective outcomes $X_1$ and $X_2$.

The models for $X_1$ and $X_2$ are given by the respective lower previsions:

$$\underline{P}_1(g(X_1)) = (1-\varepsilon)\frac{g(H)+g(T)}{2} + \varepsilon \min\{g(H), g(T)\}$$
$$\underline{P}_2(h(X_2)) = (1-\delta)\frac{h(H)+h(T)}{2} + \delta \min\{h(H), h(T)\}.$$

Consider the event that we have two equal outcomes, with indicator $f \coloneqq I_{\{(H,H)\}} + I_{\{(T,T)\}}$.

Question:
Find its lower probability with the forward irrelevant natural extension:

$$(\underline{P}_1 \circledcirc \underline{P}_2)(f).$$

## Exercise 2: solution

We have that

$$\begin{aligned}\underline{P}_2(f(X_1,X_2)) &= \underline{P}_2\left(I_{\{(H,H)\}}(X_1,X_2)+I_{\{(T,T)\}}(X_1,X_2)\right)\\ &= (1-\delta)\frac{I_{\{H\}}(X_1)+I_{\{T\}}(X_1)}{2}+\delta\min\left\{I_{\{H\}}(X_1),I_{\{T\}}(X_1)\right\}\\ &= (1-\delta)\frac{1}{2}+\delta\cdot 0\\ &= \frac{1-\delta}{2},\end{aligned}$$

and therefore

$$(\underline{P}_1\otimes\underline{P}_2)(f)=\underline{P}_1\left(\underline{P}_2(f(X_1,X_2))\right)=\underline{P}_1\left(\frac{1-\delta}{2}\right)=\frac{1-\delta}{2}.$$

# Independence

# An epistemic independence assessment

Suppose we have variables $X_n$, $n \in N$ and the independence assessment that you don't learn about any variables by observing any other variables:

$X_I \nrightarrow X_O$ for any disjoint subsets $I$ and $O$ of $N$,

so we consider the collection of irrelevance assessments:

$$\mathscr{I}_{\text{ind}} := \{X_I \nrightarrow X_O \colon I, O \subseteq N \text{ and } I \cap O = \emptyset\}$$

We call a $\mathscr{I}_{\text{ind}}$-irrelevant joint an independent joint.

# Independent natural extension

Suppose we have marginal sets of desirable gambles $\mathscr{D}_n$, with corresponding marginal lower previsions $\underline{P}_n$ on $\mathscr{L}(\mathscr{X}_k)$ for each $X_n$.

Any independent joint with these marginals is called an independent product of these marginals.

There is a smallest independent product, the independent natural extension, given by (De Cooman et al., 2009, 2012):

$$\otimes_{n \in N} \mathscr{D}_n := \operatorname{posi}\left(\mathscr{L}(\mathscr{X}_N)_{>0} \cup \bigcup_{n \in N} \mathscr{A}^{\text{irr}}_{N \setminus \{n\} \to \{n\}}\right)$$

with $\mathscr{A}^{\text{irr}}_{N \setminus \{n\} \to \{n\}} = \left\{\mathbb{I}_{\{x_{N \setminus \{n\}}\}} f \colon x_{N \setminus \{n\}} \in \mathscr{X}_{N \setminus \{n\}} \text{ and } f \in \mathscr{D}_n\right\}$, and

$$\underline{E}_N(f) = \otimes_{n \in N} \underline{P}_n(f) := \sup\left\{\alpha \colon f - \alpha \geq \sum_{n \in N} \mathbb{I}_{\{x_{N \setminus \{n\}}\}} [f_n - \underline{P}_n(f_n)]\right\}$$

## Special case of linear previsions

Consider marginal linear previsions $P_n$ for each $X_n$.

For linear previsions, the independent natural extension $\otimes_{n\in N} P_n$ coincides with the classical independent product.

Recall the definition of the strong product of lower previsions $\underline{P}_n$ as a lower envelope of classical independent products:

$$\boxtimes_{n\in N}\underline{P}_n(f) = \inf\{\otimes_{n\in N} P_n(f) \colon P_n \in \mathscr{M}(\underline{P}_n), n \in N\}$$

The lower envelope result now tells us that the strong product is also an independent product, but not necessarily the smallest one.

## Exercise 3: independent natural extension

We flip two coins successively, with respective outcomes $X_1$ and $X_2$.

The models for $X_1$ and $X_2$ are given by the respective lower previsions:

$$\underline{P}_1(g(X_1)) = (1-\varepsilon)\frac{g(H)+g(T)}{2} + \varepsilon \min\{g(H), g(T)\}$$
$$\underline{P}_2(h(X_2)) = (1-\delta)\frac{h(H)+h(T)}{2} + \delta \min\{h(H), h(T)\}.$$

Consider the event that we have two equal outcomes, with indicator $f := I_{\{(H,H)\}} + I_{\{(T,T)\}}$.

Question:
Find its lower probability with the independent natural extension:

$$(\underline{P}_1 \otimes \underline{P}_2)(f).$$

## Exercise 3: solution

The extreme points for the relevant set of desirable gambles are given by:

| | $HH$ | $HT$ | $TH$ | $TT$ |
|---|---|---|---|---|
| $e_1$ | $1+\delta$ | $-1+\delta$ | $0$ | $0$ |
| $e_2$ | $-1+\delta$ | $1+\delta$ | $0$ | $0$ |
| $e_3$ | $0$ | $0$ | $1+\delta$ | $-1+\delta$ |
| $e_4$ | $0$ | $0$ | $-1+\delta$ | $1+\delta$ |
| $e_5$ | $1+\varepsilon$ | $0$ | $-1+\varepsilon$ | $0$ |
| $e_6$ | $-1+\varepsilon$ | $0$ | $1+\varepsilon$ | $0$ |
| $e_7$ | $0$ | $1+\varepsilon$ | $0$ | $-1+\varepsilon$ |
| $e_8$ | $0$ | $-1+\varepsilon$ | $0$ | $1+\varepsilon$ |

## Exercise 3: solution

The linear programme to be solved is then:

Maximise $\alpha$ subject to:

$$f-\alpha \geq \sum_{k=1}^{8} \lambda_k e_k \text{ and } \lambda_k \geq 0$$

or in other words, subject to:

$$\begin{aligned}
1 = f(H,H) &\geq \alpha + \lambda_1(\delta+1) + \lambda_2(\delta-1) + \lambda_5(\varepsilon+1) + \lambda_6(\varepsilon-1) \\
0 = f(H,T) &\geq \alpha + \lambda_1(\delta-1) + \lambda_2(\delta+1) + \lambda_7(\varepsilon+1) + \lambda_8(\varepsilon-1) \\
0 = f(T,H) &\geq \alpha + \lambda_3(\delta+1) + \lambda_4(\delta-1) + \lambda_5(\varepsilon-1) + \lambda_6(\varepsilon+1) \\
1 = f(T,T) &\geq \alpha + \lambda_3(\delta-1) + \lambda_4(\delta+1) + \lambda_7(\varepsilon-1) + \lambda_8(\varepsilon+1)
\end{aligned}$$

This yields:

$$(\underline{P}_1 \otimes \underline{P}_2)(f) = \frac{1-\min\{\delta,\varepsilon\}}{2}.$$

# Convenient properties

## Associativity

Consider a partition $N_1, N_2$ of $N$.

The independent natural extensions $\otimes_{n \in N} \mathscr{D}_n$ and $\otimes_{n \in N} \underline{P}_n$ have the following important associativity properties:

$$\otimes_{n \in N} \mathscr{D}_n = \left(\otimes_{n_1 \in N_1} \mathscr{D}_{n_1}\right) \otimes \left(\otimes_{n_2 \in N_2} \mathscr{D}_{n_2}\right)$$

and

$$\otimes_{n \in N} \underline{P}_n = \left(\otimes_{n_1 \in N_1} \underline{P}_{n_1}\right) \otimes \left(\otimes_{n_2 \in N_2} \underline{P}_{n_2}\right)$$

The strong product $\underline{S}_N := \boxtimes_{n \in N} \underline{P}_n$ has the same associativity property.

## Marginalisation

Consider a subset $O$ of $N$.

The independent natural extensions $\otimes_{n\in N}\mathscr{D}_n$ and $\otimes_{n\in N}\underline{P}_n$ have the following important marginalisation properties:

$$\otimes_{o\in O}\mathscr{D}_o = \operatorname{marg}_O(\otimes_{n\in N}\mathscr{D}_n)$$

and

$$\otimes_{o\in O}\underline{P}_o = \operatorname{marg}_O(\otimes_{n\in N}\underline{P}_n)$$

The strong product $\underline{S}_N := \boxtimes_{n\in N}\underline{P}_n$ has the same marginalisation property.

## Factorisation

Consider two disjoint subsets $I$ and $O$ of $N$.

The independent natural extension $\underline{E}_N := \otimes_{n\in N}\underline{P}_n$ has the following important strong factorisation property:

For all gambles $h(X_O)$ on $\mathscr{X}_O$ and all non-negative gambles $g(X_I) \geq 0$ on $\mathscr{X}_I$:

$$\begin{aligned}\underline{E}_N(g(X_I)h(X_O)) &= \underline{E}_N(g(X_I)\underline{E}_N(h(X_O))) \\ &= \begin{cases}\underline{E}_N(h(X_O))\underline{E}_N(g(X_I)) & \text{if } \underline{E}_N(h(X_O)) \geq 0 \\ \underline{E}_N(h(X_O))\overline{E}_N(g(X_I)) & \text{if } \underline{E}_N(h(X_O)) \leq 0\end{cases} \\ &=: \overline{\underline{E}}_N(g(X_I)) \odot \underline{E}_N(h(X_O))\end{aligned}$$

The strong product $\underline{S}_N := \boxtimes_{n\in N}\underline{P}_n$ has the same strong factorisation property.

## Exercise 4: factorisation

Consider the gamble

$$f := I_{\{(H,H)\}} - I_{\{(H,T)\}} + 2I_{\{(T,H)\}} - 2I_{\{(T,T)\}}.$$

### Question:

Find its lower prevision

1. with the forward irrelevant natural extension: $(\underline{P}_1 \circledcirc \underline{P}_2)(f)$.
2. with the strong product: $(\underline{P}_1 \boxtimes \underline{P}_2)(f)$.
3. with the independent natural extension: $(\underline{P}_1 \otimes \underline{P}_2)(f)$.

## Exercise 4: solution

Observe that:

$$\begin{aligned} f(X_1,X_2) &= I_{\{(H,H)\}}(X_1,X_2) - I_{\{(H,T)\}}(X_1,X_2) \\ &\quad + 2I_{\{(T,H)\}}(X_1,X_2) - 2I_{\{(T,T)\}}(X_1,X_2) \\ &= \underbrace{\left[I_{\{H\}}(X_1) + 2I_{\{T\}}(X_1)\right]}_{g(X_1)\geq 0}\underbrace{\left[I_{\{H\}}(X_2) - I_{\{T\}}(X_2)\right]}_{h(X_2)} \end{aligned}$$

For the forward irrelevant natural extension:

$$\begin{aligned} (\underline{P}_1 \circledcirc \underline{P}_2)(f) &= \underline{P}_1\left(\underline{P}_2(f(X_1,X_2))\right) = \underline{P}_1\left(\underline{P}_2(g(X_1)h(X_2))\right) \\ &= \underline{P}_1\left(g(X_1)\underline{P}_2(h(X_2))\right) = \underline{P}_1\left(-\delta g(X_1)\right) = -\delta \overline{P}_1(g(X_1)) \\ &= -\delta\left[(1-\varepsilon)\frac{1+2}{2} + \varepsilon \max\{1,2\}\right] = -\frac{\delta(3+\varepsilon)}{2}. \end{aligned}$$

# Exercise 4: solution

For the other two, due to factorisation, we find similarly

$$\begin{aligned}(\underline{P}_1 \otimes \underline{P}_2)(f) &= (\underline{P}_1 \boxtimes \underline{P}_2)(f) \\ &= \underline{P}_1\left(\underline{P}_2(f(X_1,X_2))\right) = \underline{P}_1\left(\underline{P}_2(g(X_1)h(X_2))\right) \\ &= \underline{P}_1\left(g(X_1)\underline{P}_2(h(X_2))\right) = \underline{P}_1\left(-\delta g(X_1)\right) = -\delta \overline{P}_1(g(X_1)) \\ &= -\delta\left[(1-\varepsilon)\frac{1+2}{2} + \varepsilon \max\{1,2\}\right] = -\frac{\delta(3+\varepsilon)}{2}.\end{aligned}$$

# External additivity

Consider two disjoint subsets $I$ and $O$ of $N$.

The independent natural extension $\underline{E}_N := \otimes_{n\in N}\underline{P}_n$ has the following important strong external additivity property:

For all gambles $h(X_O)$ on $\mathscr{X}_O$ and all gambles $g(X_I)$ on $\mathscr{X}_I$:

$$\underline{E}_N(g(X_I) + h(X_O)) = \underline{E}_N(g(X_I)) + \underline{E}_N(h(X_O))$$

The strong product $\underline{S}_N := \boxtimes_{n\in N}\underline{P}_n$ has the same strong external additivity property.

# Exercise 5: external additivity

Consider the gamble that gives the number of heads:

$$f(X_1, X_2) := I_{\{H\}}(X_1) + I_{\{H\}}(X_2).$$

## Question:

Find its lower prevision

1. with the forward irrelevant natural extension: $(\underline{P}_1 \circledcirc \underline{P}_2)(f)$.
2. with the strong product: $(\underline{P}_1 \boxtimes \underline{P}_2)(f)$.
3. with the independent natural extension: $(\underline{P}_1 \otimes \underline{P}_2)(f)$.

# Exercise 5: solution

For the forward irrelevant natural extension:

$$\begin{aligned}
\underline{P}_2(f(X_1, X_2)) &= \underline{P}_2\left(I_{\{H\}}(X_1) + I_{\{H\}}(X_2)\right) \\
&= (1-\delta)\frac{I_{\{H\}}(X_1) + 1 + I_{\{H\}}(X_1) + 0}{2} \\
&\qquad + \delta \min\left\{I_{\{H\}}(X_1) + 1, I_{\{H\}}(X_1) + 0\right\} \\
&= (1-\delta) I_{\{H\}}(X_1) + (1-\delta)\frac{1}{2} + \delta I_{\{H\}}(X_1) \\
&= I_{\{H\}}(X_1) + \frac{1-\delta}{2},
\end{aligned}$$

and therefore

$$\begin{aligned}
(\underline{P}_1 \circledcirc \underline{P}_2)(f) = \underline{P}_1\left(\underline{P}_2(f(X_1, X_2))\right) &= \underline{P}_1\left(I_{\{H\}}(X_1) + \frac{1-\delta}{2}\right) \\
&= \frac{1-\varepsilon}{2} + \frac{1-\delta}{2} = 1 - \frac{\varepsilon+\delta}{2}.
\end{aligned}$$

# Exercise 5: solution

For the other two, due to external additivity:

$$
\begin{aligned}
(\underline{P}_1 \otimes \underline{P}_2)(f) = (\underline{P}_1 \boxtimes \underline{P}_2)(f) &= \underline{P}_1\left(I_{\{H\}}(X_1)\right) + \underline{P}_2\left(I_{\{H\}}(X_2)\right) \\
&= \frac{1-\varepsilon}{2} + \frac{1-\delta}{2} \\
&= 1 - \frac{\varepsilon+\delta}{2}.
\end{aligned}
$$

## What would I like to achieve and convey?

| | | |
|---|---|---|
| Sets of probabilities are not necessarily the best model | EPISTEMIC IRRELEVANCE: what is it and how can it be used? | ASYMMETRY in irrelevance is not a bad thing |
| lower previsions | coherence | natural requirement |
| DESIRABLE GAMBLES | convenient properties | leads to lower complexity |

# To conclude

## Conclusion

Elegance and conceptual simplicity
of working with sets of desirable gambles

Asymmetry is not necessarily a deficiency!

Complexity for epistemic independence: $n\,|\mathscr{X}|^{n-1}\,N_{ER}$ versus $N_{EP}^{n}$

Convenient properties that may help avoid combinatorial explosion:
factorisation and external additivity

# References

## Useful references


- Peter Walley.
  *Statistical Reasoning with Imprecise Probabilities*.
  Chapman and Hall, London, 1991.
- Serafín Moral.
  Epistemic irrelevance on sets of desirable gambles.
  *Ann. of Mathematics and Artificial Intelligence*, 45:197–214, 2005.
- Gert de Cooman and Enrique Miranda.
  Forward irrelevance.
  *Journal of Statistical Planning and Inference*, 139:256–276, 2009.
- Gert de Cooman, Enrique Miranda, and Marco Zaffalon.
  Independent natural extension.
  *Artificial Intelligence*, 175:1911–1950, 2011.
- Gert de Cooman and Enrique Miranda.
  Irrelevant and independent natural extension for sets of desirable gambles.
  Submitted, 2012.