Introduction to the theory of imprecise probability

Erik Quaeghebeur

TU Delft, the Netherlands

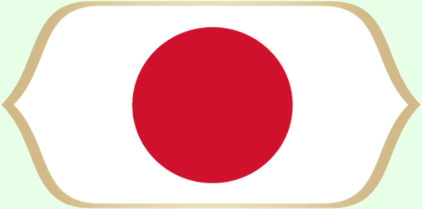
UTOPIAE Training School 2018, Durham, England

Why would you want your probability to be imprecise?

versus

Uncertainty about outcome of...

versus



Agents (Gamblers)

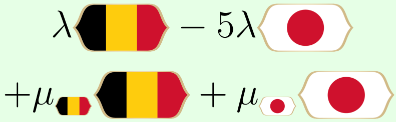
Wiske

Yoko Tsuno

Assessment (gambles accepted)	**Agents (Gamblers)**
1 if  $-\,5$ if 	Wiske
-4 if  $+\,1$ if 	Yoko Tsuno

Natural extension ($\lambda, \mu \geq 0$)

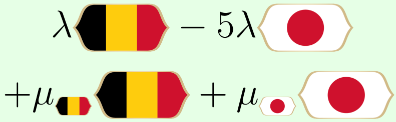
Rational agents



Wiske

Yoko Tsuno

Natural extension $(\lambda, \mu \geq 0)$



COHERENCE

Rational agents

Wiske

Yoko Tsuno

Assessment (gambles accepted) **Agents (Gamblers)**

 $- 5$ ,

-4  $+$ 

Heroine pool

Natural extension ($\lambda, \mu \geq 0$)

Irrational agents

$(\lambda_{\mathsf{W}} - 4\lambda_{\mathsf{Y}} + \mu_{$  $})$ 

$+ (-5\lambda_{\mathsf{W}} + \lambda_{\mathsf{Y}} + \mu_{$  $})$ 

Heroine pool

Natural extension ($\lambda, \mu \geq 0$)

$(\lambda_{\mathsf{W}} - 4\lambda_{\mathsf{Y}} + \mu$$)$

$+ (-5\lambda_{\mathsf{W}} + \lambda_{\mathsf{Y}} + \mu$$)$

SURE LOSS!

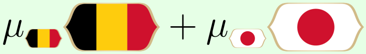
Irrational agents

Heroine pool

Assessment (gambles accepted)	**Agents (Gamblers)**
$\emptyset$	Heroine pool

Natural extension ($\lambda, \mu \geq 0$)

Rational agents



Heroine pool

Natural extension ($\lambda, \mu \geq 0$)

Rational agents

μ   $+ \mu$  

Heroine pool

VACUOUS

Basic concepts

- *Agent* reasoning about experiment with uncertain outcome
- *Possibility space* $\mathcal{X}$ of outcomes
- *Gambles* are real-valued functions of the outcomes;

$$\mathcal{L} = \mathcal{X} \to \mathbb{R}$$

  ($\mathcal{L}$ is assumed to be a linear space)

- *Assessment* is a description of a set of acceptable gambles
- *Natural extension* of an assessment is
  the set of all acceptable gambles implied by
  the agent's rationality criteria (and other assumptions)

Coherence, the classical rationality criteria

Constructive

Positive scaling

If f is acceptable and $\lambda > 0$,
then λf is acceptable.

Addition

If f and g are acceptable,
then $f + g$ is acceptable.

Background

Accepting gain

If f is nonnegative for all outcomes,
then f is acceptable.

Avoiding sure loss

If g is negative for all outcomes,
then g is not acceptable.

Where are the
imprecise probabilities
I came here for!!!

Previsions/Expectations are prices for gambles

- '*Prevision*' and '*Expectation*' are synomyms
- *Prices* are real values interpreted as constant gambles
- Lower prevision $\underline{P}(f)$ is the
  supremum acceptable buying price of f:

  $$\underline{P}(f) = \sup\{\nu \in \mathbb{R} : f - \nu \text{ is acceptable}\}$$

  Upper prevision $\overline{P}(f)$ is the
  infimum acceptable selling price of f
- Conjugacy of coherent lower and upper previsions:

  $$\overline{P}(f) = -\underline{P}(-f)$$

- If $\underline{P}(f) = \overline{P}(f)$, then $P(f) = \underline{P}(f)$ is the *prevision* of f

Probabilities are previsions of indicator gambles

- *Event* A is a subset of $\mathcal{X}$
- *Indicator gamble*

$$1_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A \end{cases}$$

- Lower probability $\underline{P}(A) = \underline{P}(1_A)$

  Upper probability $\overline{P}(A) = \overline{P}(1_A)$
- Conjugacy of coherent lower and upper probabilities ($A^\mathsf{c} = \mathcal{X} \setminus A$):

$$\overline{P}(A) = 1 - \underline{P}(A^\mathsf{c})$$

- If $\underline{P}(A) = \overline{P}(A)$, then $P(A) = \underline{P}(A)$ is the *probability* of A

Natural extension ($\lambda, \mu \geq 0$)	**Agents**
$(\lambda + \mu_{\text{🇧🇪}})$ 🇧🇪 $+(-5\lambda + \mu_{\text{🇯🇵}})$ 🇯🇵	Wiske
$(-4\lambda + \mu_{\text{🇧🇪}})$ 🇧🇪 $(\lambda + \mu_{\text{🇯🇵}})$ 🇯🇵	Yoko Tsuno
$(\lambda_{\mathsf{W}} - 4\lambda_{\mathsf{Y}} + \mu_{\text{🇧🇪}})$ 🇧🇪 $+ (-5\lambda_{\mathsf{W}} + \lambda_{\mathsf{Y}} + \mu_{\text{🇯🇵}})$ 🇯🇵	Irrational pool
$\mu_{\text{🇧🇪}}$ 🇧🇪 $+ \mu_{\text{🇯🇵}}$ 🇯🇵	Rational pool

Wiske's lower probability that Belgium will win

$$\underline{P}(\text{🇧🇪}) = \sup\left\{\nu : \begin{bmatrix}1-\nu\\ -\nu\end{bmatrix} = \begin{bmatrix}\lambda + \mu_{\text{🇧🇪}}\\ -5\lambda + \mu_{\text{🇯🇵}}\end{bmatrix}, \mu \geq 0, \lambda \geq 0\right\}$$

Wiske's lower probability that Belgium will win

$$\underline{P}(\text{🇧🇪}) = \sup\left\{\nu : \begin{bmatrix} 1-\nu \\ -\nu \end{bmatrix} = \begin{bmatrix} \lambda + \mu_{\text{🇧🇪}} \\ -5\lambda + \mu_{\text{🇯🇵}} \end{bmatrix}, \mu \geq 0, \lambda \geq 0\right\}$$

$$= \sup\{5\lambda + \mu_{\text{🇯🇵}} : 1 - 5\lambda + \mu_{\text{🇯🇵}} = \lambda + \mu_{\text{🇧🇪}}, \mu \geq 0, \lambda \geq 0\}$$

Wiske's lower probability that Belgium will win

$$\underline{P}(\text{🇧🇪}) = \sup\left\{\nu : \begin{bmatrix} 1-\nu \\ -\nu \end{bmatrix} = \begin{bmatrix} \lambda + \mu_{\text{🇧🇪}} \\ -5\lambda + \mu_{\text{🇯🇵}} \end{bmatrix}, \mu \geq 0, \lambda \geq 0\right\}$$

$$= \sup\{5\lambda + \mu_{\text{🇯🇵}} : 1 - 5\lambda + \mu_{\text{🇯🇵}} = \lambda + \mu_{\text{🇧🇪}}, \mu \geq 0, \lambda \geq 0\}$$

$$= \sup\left\{5\lambda + \mu_{\text{🇯🇵}} : \lambda = \frac{1}{6}(1 + \mu_{\text{🇯🇵}} - \mu_{\text{🇧🇪}}), \mu \geq 0, \lambda \geq 0\right\}$$

Wiske's lower probability that Belgium will win

$$\underline{P}(\text{🇧🇪}) = \sup\left\{\nu : \begin{bmatrix} 1-\nu \\ -\nu \end{bmatrix} = \begin{bmatrix} \lambda + \mu_{\text{🇧🇪}} \\ -5\lambda + \mu_{\text{🇯🇵}} \end{bmatrix}, \mu \geq 0, \lambda \geq 0\right\}$$

$$= \sup\{5\lambda + \mu_{\text{🇯🇵}} : 1 - 5\lambda + \mu_{\text{🇯🇵}} = \lambda + \mu_{\text{🇧🇪}}, \mu \geq 0, \lambda \geq 0\}$$

$$= \sup\left\{5\lambda + \mu_{\text{🇯🇵}} : \lambda = \frac{1}{6}(1 + \mu_{\text{🇯🇵}} - \mu_{\text{🇧🇪}}), \mu \geq 0, \lambda \geq 0\right\}$$

$$= \sup\left\{\frac{5}{6} - \frac{1}{6}\mu_{\text{🇯🇵}} - \frac{5}{6}\mu_{\text{🇧🇪}} : \mu \geq 0\right\}$$

$$= \frac{5}{6}$$

$\underline{P}($🇧🇪$), \overline{P}($🇧🇪$)$	$\underline{P}($🇯🇵$), \overline{P}($🇯🇵$)$	**Agents**
$\frac{5}{6}, 1$	$0, \frac{1}{6}$	Wiske
$0, \frac{1}{5}$	$\frac{4}{5}, 1$	Yoko Tsuno
$+\infty, -\infty$	$+\infty, -\infty$	Irrational pool
$0, 1$	$0, 1$	Rational pool

Assessments of lower previsions

- Assume a lower prevision $\underline{P}$ with values assessed for a set of gambles $\mathcal{K}$
- How can we apply the theory we have seen?
- Translate the lower prevision $\underline{P}(f)$ for a gamble $f \in \mathcal{K}$ into a set

$$\{f - \underline{P}(f) + \varepsilon : \varepsilon > 0\}$$

  of acceptable gambles
- The gambles $f - \underline{P}(f)$ are called *marginal gambles*

Expressions for assessments of lower previsions

Avoiding sure loss

$$\sup_{x \in \mathcal{X}} \sum_{k=1}^{n} (f_k(x) - \underline{P}(f_k)) \geq 0 \quad \text{for all } n \geq 0 \text{ and } f_k \in \mathcal{K}$$

Expressions for assessments of lower previsions

Avoiding sure loss

$$\sup_{x\in\mathcal{X}} \sum_{k=1}^{n} (f_k(x) - \underline{P}(f_k)) \geq 0 \quad \text{for all } n \geq 0 \text{ and } f_k \in \mathcal{K}$$

Coherence

$$\sup_{x\in\mathcal{X}} \left(\sum_{k=1}^{n} (f_k(x) - \underline{P}(f_k)) - m(f_0 - \underline{P}(f_0)) \right) \geq 0$$

for all $n, m \geq 0$ and $f_k \in \mathcal{K}$

Expressions for assessments of lower previsions

Avoiding sure loss

$$\sup_{x\in\mathcal{X}} \sum_{k=1}^{n} \left(f_k(x) - \underline{P}(f_k)\right) \geq 0 \quad \text{for all } n \geq 0 \text{ and } f_k \in \mathcal{K}$$

Coherence

$$\sup_{x\in\mathcal{X}} \left(\sum_{k=1}^{n} \left(f_k(x) - \underline{P}(f_k)\right) - m(f_0 - \underline{P}(f_0)) \right) \geq 0$$

for all $n, m \geq 0$ and $f_k \in \mathcal{K}$

Natural extension

$$\underline{E}(f) = \sup\left\{ \inf_{x\in\mathcal{X}} \left\{ f(x) - \sum_{k=1}^{n} \lambda_k \left(f_k(x) - \underline{P}(f_k)\right) \right\} : n \geq 0, f_k \in \mathcal{K}, \lambda_k > 0 \right\}$$

Does it really have to be so involved?

Lower previsions on linear spaces

If the lower prevision $\underline{P}$ is defined for all gambles in a linear space $\mathcal{L}$, the coherence criteria simplify:

Accepting sure gains

$$\underline{P}(f) \geq \inf f \quad \text{for all } f \in \mathcal{L}$$

Super-linearity

$$\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g) \quad \text{for all } f, g \in \mathcal{L}$$

Positive homogeneity

$$\underline{P}(\lambda f) = \lambda \underline{P}(f) \quad \text{for all } f \in \mathcal{L} \text{ and } \lambda > 0$$

Upper previsions on linear spaces

If the upper prevision $\overline{P}$ is defined for all gambles in a linear space $\mathcal{L}$, the coherence criteria simplify:

Accepting sure gains

$$\overline{P}(f) \leq \sup f \quad \text{for all } f \in \mathcal{L}$$

Sub-linearity

$$\overline{P}(f+g) \leq \overline{P}(f) + \overline{P}(g) \quad \text{for all } f, g \in \mathcal{L}$$

Positive homogeneity

$$\overline{P}(\lambda f) = \lambda \overline{P}(f) \quad \text{for all } f \in \mathcal{L} \text{ and } \lambda > 0$$

Coherent lower & upper previsions

For a coherent lower prevision $\underline{P}$ and its conjugate upper prevision $\overline{P}$ many useful properties can be derived; we present a few:

Upper dominates lower $\overline{P}(f) \geq \underline{P}(f)$ for all $f \in \mathcal{L}$

Constants $\overline{\underline{P}}(\mu) = \mu$ for all $\mu \in \mathbb{R}$

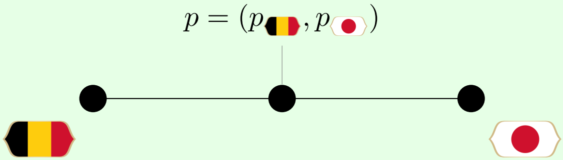
Constant additivity $\overline{\underline{P}}(f + \mu) = \overline{\underline{P}}(f) + \mu$ for all $f \in \mathcal{L}$ and $\mu \in \mathbb{R}$

Gamble dominance
if $f \geq g + \mu$ then $\overline{\underline{P}}(f) \geq \overline{\underline{P}}(g) + \mu$ for all $f, g \in \mathcal{L}$ and $\mu \in \mathbb{R}$

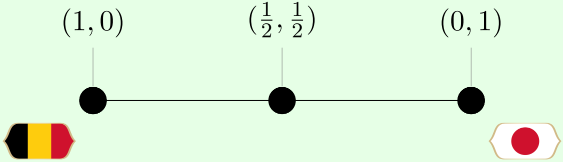
Mixed sub/super-additivity
$$\underline{P}(f + g) \leq \underline{P}f + \overline{P}(g) \leq \overline{P}(f + g) \quad \text{for all } f, g \in \mathcal{L}$$

I heard that imprecise probabilities are just *sets of probabilities*?

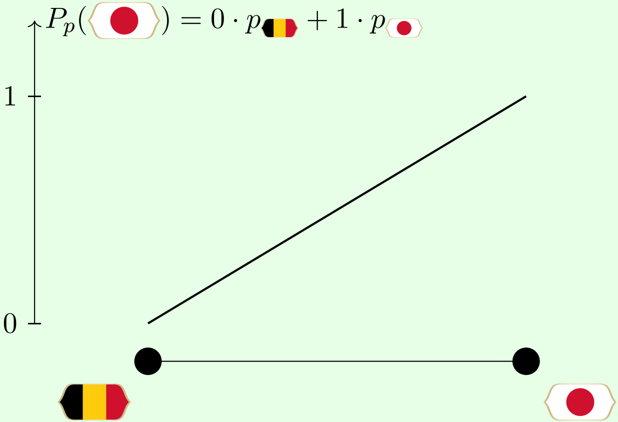
The probability simplex



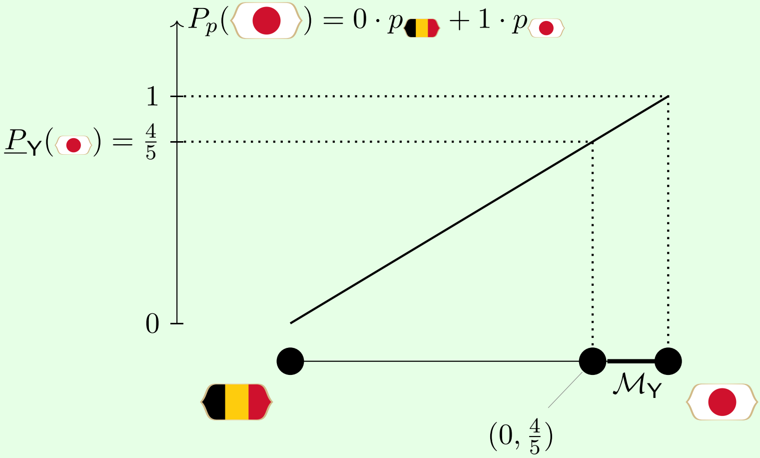
The probability simplex



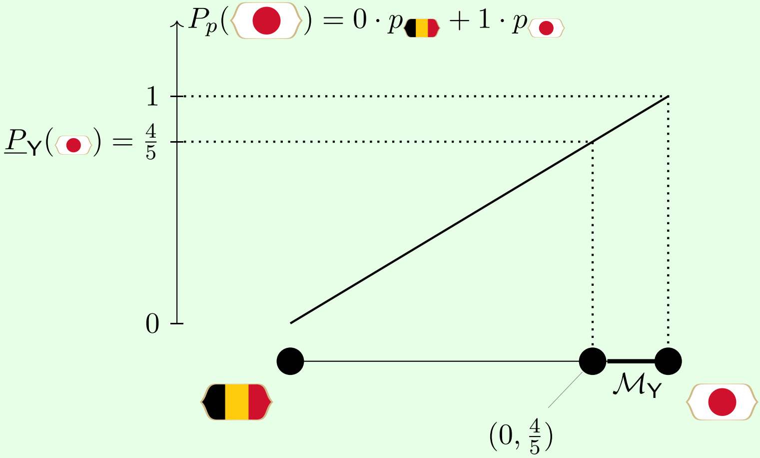
The probability simplex



The probability simplex



The probability simplex



CREDAL SET

From lower previsions to credal sets

- The prevision of a gamble
  is a linear function over the probability simplex

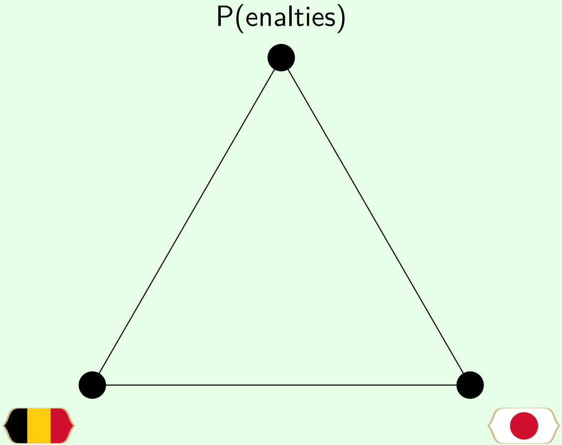
- The lower prevision of a gamble
  can be seen as bounding the prevision of that gamble
  so constraining the possible probability mass functions

- A lower prevision corresponds to a set of constraints,
  defining a credal set (closed convex set)

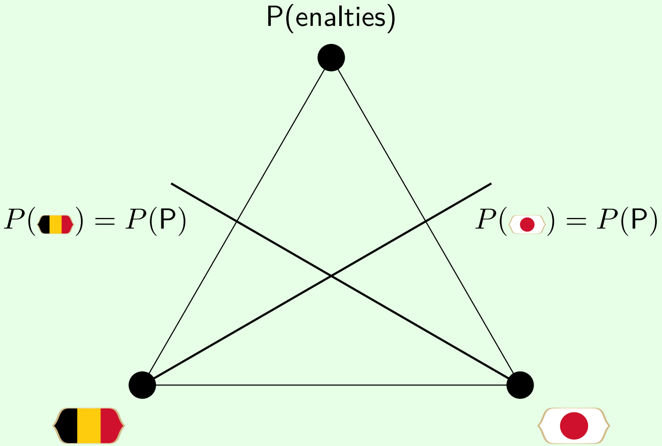
$$\mathcal{M} = \{p : P_p(f) \geq \underline{P}(f) \text{ for all } f \in \mathcal{K}\}$$

- All this generalizes to infinite $\mathcal{X}$

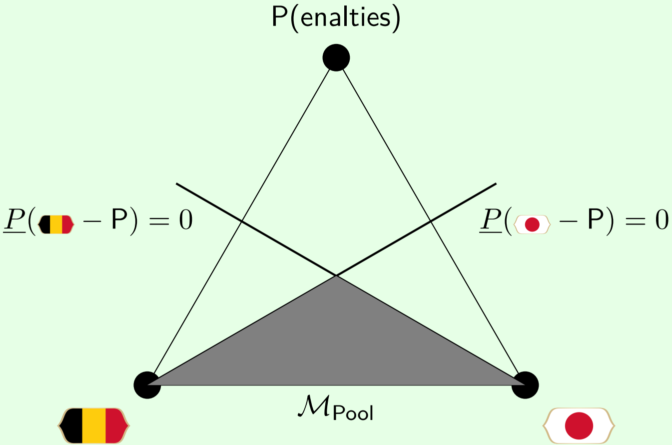
A larger probability simplex



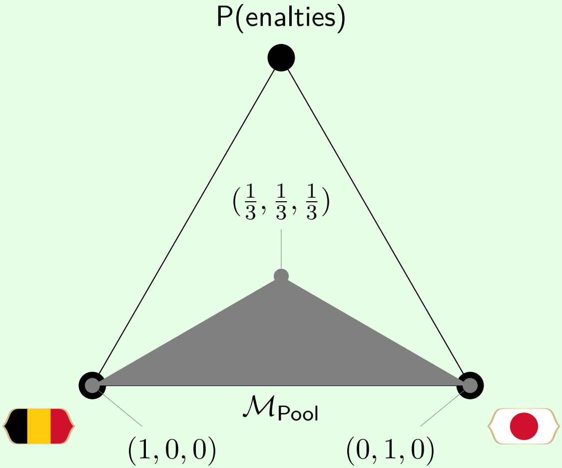
A larger probability simplex



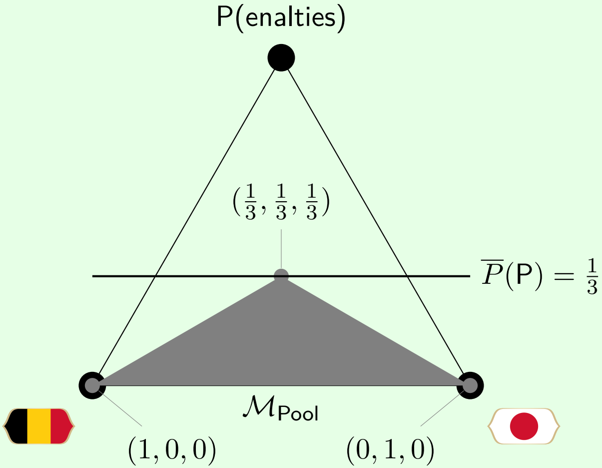
A larger probability simplex



A larger probability simplex



A larger probability simplex



From credal sets to lower previsions

- A *credal set* is a closed convex set of probability mass functions
  (or more generally, previsions)

- A credal set is determined completely by its set of *extreme points* $\mathcal{M}^*$

- A nonempty credal set is equivalent to a coherent lower prevision

Lower envelope theorem

$$
\begin{aligned}
\underline{P}(f) &= \min\{P_p(f) : p \in \mathcal{M}\} \\
&= \min\{P_p(f) : p \in \mathcal{M}^*\}
\end{aligned}
$$

Where are the conditional models?
We need them to learn!

Basics of conditioning and updating

- Conditioning acceptable gambles is done by restriction to the subspace of gambles that are zero outside the conditioning event
- Conditioning lower previsions is a form of natural extension

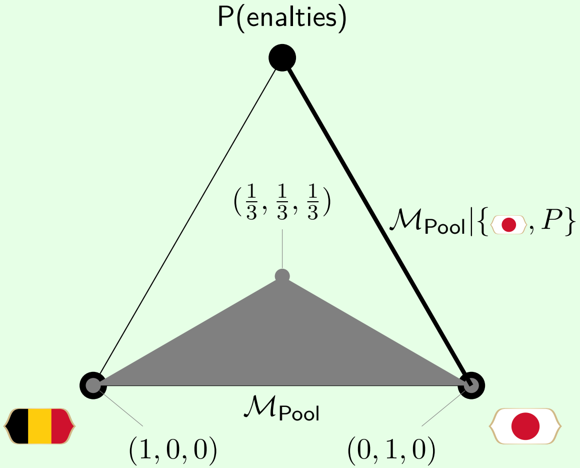
$$\underline{E}(f \mid A) = \begin{cases} \inf_{x \in A} f(x) & \text{if } \underline{P}(A) = 0, \\ \max\{\mu \in \mathbb{R} : \underline{P}(1_A(f - \mu))\} & \text{if } \underline{P}(A) > 0 \end{cases}$$

- Conditional credal set =
  credal set of conditional probability mass functions

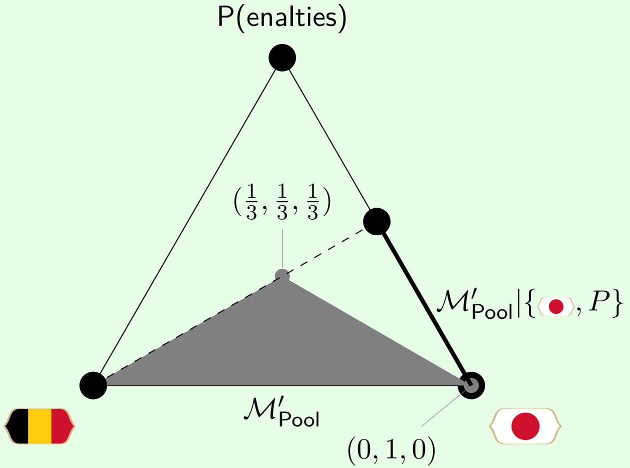
$$\mathcal{M}|A = \begin{cases} \text{whole simplex} & \text{if } \exists p \in \mathcal{M} : P_p(A) = 0, \\ \{p(\cdot \mid A) : p \in \mathcal{M}\} & \text{if } \forall p \in \mathcal{M} : P_p(A) > 0 \end{cases}$$

- Natural extension often gives vacuous conditionals;
  regular extension is a less imprecise updating rule:
  it removes those p such that $P_p(A) = 0$ from $\mathcal{M}$

Conditioning using natural extension



Conditioning using regular extension

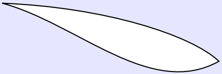


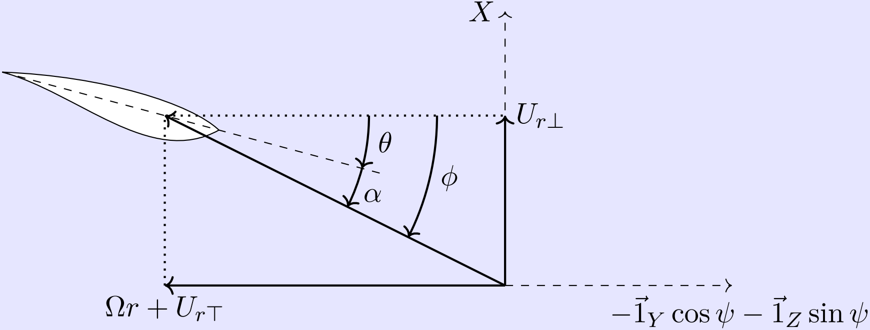
What about imprecise probabilities on continuous spaces?

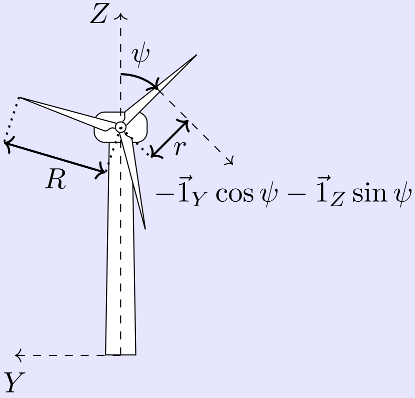
Too few remarks about imprecise probabilities on continuous spaces

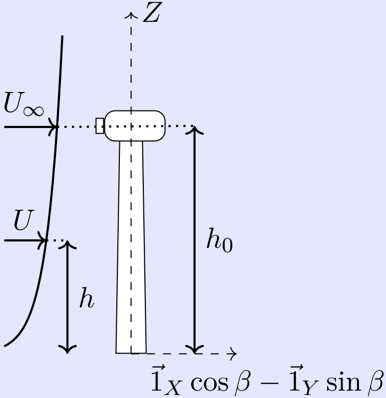
- Mostly defined using credal sets
  of parametric distributions
  where the parameters vary in a set
- Also common:
  defined using probability mass assignments
  to subsets of the space
- Examples:
  - Imprecise Dirichlet model
  - P-boxes
  - lower density functions
- Calculating lower and upper previsions (natural extension)
  can easily become difficult optimization problems

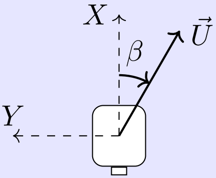
Nothing in this lecture had anything to do with aerospace engineering!



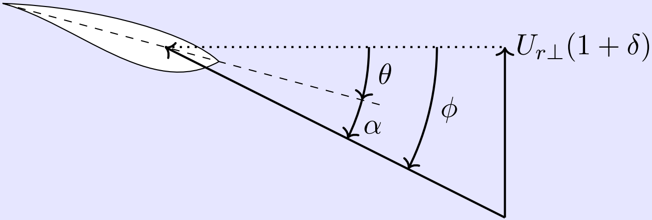

X
$U_{r\perp}$
θ
α
ϕ
$\Omega r + U_{r\top}$
$-\vec{1}_Y \cos\psi - \vec{1}_Z \sin\psi$


Z
ψ
r
R
$-\vec{1}_Y \cos\psi - \vec{1}_Z \sin\psi$
Y


Z
U_∞
U
h
h_0
$\vec{1}_X \cos\beta - \vec{1}_Y \sin\beta$


X
$\vec{U}$
β
Y

Relationship between turbulence and angle of attack



$$\alpha_\delta = \alpha - \alpha_0 = \operatorname{atan}\left((1+\delta)\tan\phi_0\right) - \phi_0$$

$$\delta = \frac{\tan(\phi_0 + \alpha_\delta) - \tan\phi_0}{\tan\phi_0}$$

Is this all you have?