What should we update? Outer probability measures and learning from ignorance

Jeremie Houssineau

June 25, 2026

Division of Mathematical Sciences
Nanyang Technological University
Singapore

Outline

Review of possibility theory

Outer probability measures

Role in IP

A non-exhaustive overview of the literature

Core topics:

- Possibility theory: Zadeh (1978), De Cooman (1997), Yager (2013), and Dubois and Prade (2015)
- Inferential models: Martin and Liu (2015) and Martin (2022)
- Dempster-Shafer theory: Dempster (2008), Shafer (1976), Yager (1987), Smets and Kennes (1994), and Denœux (2006)
- General IP: Walley (2000), Destercke et al. (2008), Augustin et al. (2014), and Troffaes and De Cooman (2014)

A non-exhaustive overview of the literature

Related topics:

- IPML: Cozman (2000), Zaffalon (2002), Caprio et al. (2023), Sale et al. (2023), and Chau et al. (2025)
- Fiducial inference: Fisher (1935) and Hannig (2009)
- $(\max, +)$-algebra: Del Moral (1997) and Del Moral and Doisy (1999)
- Large deviation theory: Puhalskii (2001) and Nguyen and Bouchon-Meunier (2003)

Review of possibility theory

Possibility function

Objective: Describe the information/uncertainty about a fixed parameter $\theta^\dagger \in \Theta$

Ingredients:

1) The *true outcome* $\omega^\dagger \in \Omega_\mathrm{u}$ $\rightsquigarrow$ the *true* state of nature
2) A *deterministic uncertain variable* (d.u.v.) $\boldsymbol{\theta} : \Omega_\mathrm{u} \to \Theta$ s.t. $\boldsymbol{\theta}(\omega^\dagger) = \theta^\dagger$
3) A *possibility function* $f_{\boldsymbol{\theta}}$ *describing* the information about $\theta^\dagger$ on Θ s.t.

$$f_{\boldsymbol{\theta}}(\theta) \geq 0 \qquad \text{and} \qquad \sup_{\theta \in \Theta} f_{\boldsymbol{\theta}}(\theta) = 1$$

As an odd ratio: $f_{\boldsymbol{\theta}}(\theta) = \frac{f_{\boldsymbol{\theta}}(\theta)}{f_{\boldsymbol{\theta}}(\theta^*)}$ with $\theta^* \in \arg\max f_{\boldsymbol{\theta}}$
Special case: no information $\rightsquigarrow$ $f_{\boldsymbol{\theta}}(\theta) = 1$ for any $\theta \in \Theta$

Introducing possibility functions

Normal possibility function:

$$\overline{\mathrm{N}}(\theta; \mu, \Sigma) = \exp\Big(-\frac{1}{2}(\theta-\mu)^\intercal \Sigma^{-1}(\theta-\mu)\Big) \propto \mathrm{N}(\theta; \mu, \Sigma)$$

Change of variable formula:

If $\boldsymbol{\theta}$ is described by $f_{\boldsymbol{\theta}}$ then $\boldsymbol{\psi} = T(\boldsymbol{\theta})$ is described by

$$f_{\boldsymbol{\psi}}(\psi) = \sup_{\theta : T(\theta)=\psi} f_{\boldsymbol{\theta}}(\theta)$$

Example: Let $\boldsymbol{\theta}$ be described by $\overline{\mathrm{N}}(\mu, \Sigma)$ and let S be an invertible matrix, then $\boldsymbol{\psi} = T(\boldsymbol{\theta}) = S\boldsymbol{\theta}$ is described by

$$f_{\boldsymbol{\psi}}(\psi) = \overline{\mathrm{N}}(S\mu, S\Sigma S^\intercal).$$

Properties of possibility functions

⊖ Cannot be sampled

⊕ Can be used for inference: if $(\boldsymbol{\theta}, \boldsymbol{\psi})$ is described by $f_{\boldsymbol{\theta},\boldsymbol{\psi}}(\theta, \psi)$

$$f_{\boldsymbol{\psi}}(\psi) = \sup_{\theta \in \Theta} f_{\boldsymbol{\theta},\boldsymbol{\psi}}(\theta, \psi) \qquad \text{and} \qquad f_{\boldsymbol{\theta}}(\theta \,|\, \psi) = \frac{f_{\boldsymbol{\theta},\boldsymbol{\psi}}(\theta, \psi)}{f(\psi)}$$

⊕ Lead to a natural notion of independence: $f_{\boldsymbol{\theta},\boldsymbol{\psi}}(\theta, \psi) = f_{\boldsymbol{\theta}}(\theta) f_{\boldsymbol{\psi}}(\psi)$
(min can be recovered)

⊕ Can be combined:

$$f_{\boldsymbol{\theta}}(\theta \,|\, \boldsymbol{\theta} = \boldsymbol{\psi}) = \frac{f_{\boldsymbol{\theta},\boldsymbol{\psi}}(\theta, \theta)}{\sup_{\theta' \in \Theta} f_{\boldsymbol{\theta},\boldsymbol{\psi}}(\theta', \theta')}$$

⊕ Are not unique

Bayesian inference in a nutshell

Bayes' rule:

$$\pi(\theta \mid y) = \frac{L(y \mid \theta)\pi(\theta)}{\pi(y)}$$

with

- $L(y \mid \theta)$ the likelihood
- $\pi(\theta)$ the *prior* and $\pi(\theta \mid y)$ the *posterior*
- the *evidence* $\pi(y)$ defined as

$$\pi(y) = \begin{cases} \displaystyle\int L(y \mid \theta)\pi(\theta)\mathrm{d}\theta & \text{if } \pi \text{ probabilistic} \\ \displaystyle\sup_{\theta\in\Theta} L(y \mid \theta)\pi(\theta) & \text{if } \pi \text{ possibilistic} \end{cases}$$

Bayesian inference in a nutshell

Bayes' rule:

$$\pi(\theta \mid y) = \frac{L(y \mid \theta)\pi(\theta)}{\pi(y)}$$

with

- $L(y \mid \theta)$ the likelihood
- $\pi(\theta)$ the *prior* and $\pi(\theta \mid y)$ the *posterior*
- the *evidence* $\pi(y)$ defined as

$$\pi(y) = \begin{cases} \int L(y \mid \theta)\pi(\theta)\mathrm{d}\theta & \text{if } \pi \text{ probabilistic} \\ \sup_{\theta \in \Theta} L(y \mid \theta)\pi(\theta) & \text{if } \pi \text{ possibilistic} \end{cases}$$

Bayesian inference in a nutshell

Bayes' rule:

$$\pi(\theta \mid y) = \frac{L(y \mid \theta)\pi(\theta)}{\pi(y)}$$

with

- $L(y \mid \theta)$ the likelihood
- $\pi(\theta)$ the *prior* and $\pi(\theta \mid y)$ the *posterior*
- the *evidence* $\pi(y)$ defined as

$$\pi(y) = \begin{cases} \displaystyle\int L(y \mid \theta)\pi(\theta)\mathrm{d}\theta & \text{if } \pi \text{ probabilistic} \\ \displaystyle\sup_{\theta \in \Theta} L(y \mid \theta)\pi(\theta) & \text{if } \pi \text{ possibilistic} \end{cases}$$

Bayesian inference in a nutshell (J. H. & B.-E. Cherief Abdellatif, to appear in UAI 2026)

Bayes' rule:

$$\pi(\theta \mid y) = \frac{L(y \mid \theta)\pi(\theta)}{\pi(y)}$$

motivated

à la Zellner (1988)

$$\Delta(\mathrm{IPR}) = \text{Input} - \text{Output}$$

$\mathcal{P}(\Theta)$	$\mathcal{F}(\Theta)$
cross-entropy	$-\log$
scalar	*functional*
no info. creation	*creation* or loss

Bayesian inference in a nutshell (J. H. & B.-E. Cherief Abdellatif, to appear in UAI 2026)

Bayes' rule:

$$\pi(\theta \mid y) = \frac{L(y \mid \theta)\pi(\theta)}{\pi(y)}$$

motivated

à la Zellner (1988)

$$\Delta(\mathrm{IPR}) = \text{Input} - \text{Output}$$

$\mathcal{P}(\Theta)$	$\mathcal{F}(\Theta)$
cross-entropy	$-\log$
scalar	*functional*
no info. creation	*creation* or loss

à la Bissiri et al. (2016):

5 axioms lead to

$$\pi(\theta \mid y) \propto \exp(-\lambda \ell(\theta, y))\pi(\theta)$$

with λ *predicted* by the framework:

$$f \in \mathcal{F}(\Theta) \Rightarrow f^{\lambda} \in \mathcal{F}(\Theta)$$

Bayesian inference in a nutshell (J. H. & B.-E. Cherief Abdellatif, to appear in UAI 2026)

Bayes' rule:

$$\pi(\theta \mid y) = \frac{L(y \mid \theta)\pi(\theta)}{\pi(y)}$$

motivated

à la Zellner (1988)

$$\Delta(\mathrm{IPR}) = \text{Input} - \text{Output}$$

$\mathcal{P}(\Theta)$	$\mathcal{F}(\Theta)$
cross-entropy	$-\log$
scalar	*functional*
no info. creation	*creation* or loss

à la Bissiri et al. (2016):

5 axioms lead to

$$\pi(\theta \mid y) \propto \exp(-\lambda \ell(\theta, y))\pi(\theta)$$

with λ *predicted* by the framework:

$$f \in \mathcal{F}(\Theta) \Rightarrow f^{\lambda} \in \mathcal{F}(\Theta)$$

Question: What is the point of the possibilistic formulation then?

A candidate for "objective" Bayesian inference (Hieu et al. 2025)

Criteria:

1) Proper
2) Satisfies likelihood principle
3) Invariant under re-parametrisation
4) Coherence under partitioning

Our candidate: $f_\theta = \mathbf{1}$

Consequences:

- Can be used in conjunction with informative priors on other parameters
- Avoids Dawid et al. (1973)'s marginalisation paradoxes
- The posterior characterisation of uncertainty remains subjective, but...

A candidate for “objective” Bayesian inference (Hieu et al. 2025)

Criteria:

1) Proper
2) Satisfies likelihood principle
3) Invariant under re-parametrisation
4) Coherence under partitioning

Our candidate: $f_{\boldsymbol{\theta}} = \mathbf{1}$

Consequences:

- Can be used in conjunction with informative priors on other parameters
- Avoids Dawid et al. (1973)'s marginalisation paradoxes
- The posterior characterisation of uncertainty remains subjective, but...

A candidate for "objective" Bayesian inference (Hieu et al. 2025)

Criteria:

1) Proper ✓
2) Satisfies likelihood principle
3) Invariant under re-parametrisation
4) Coherence under partitioning

Our candidate: $f_{\boldsymbol{\theta}} = \mathbf{1}$

Consequences:

- Can be used in conjunction with informative priors on other parameters
- Avoids Dawid et al. (1973)'s marginalisation paradoxes
- The posterior characterisation of uncertainty remains subjective, but...

A candidate for "objective" Bayesian inference (Hieu et al. 2025)

Criteria:

1) Proper ✓
2) Satisfies likelihood principle ✓
3) Invariant under re-parametrisation
4) Coherence under partitioning

Our candidate: $f_{\boldsymbol{\theta}} = \mathbf{1}$

Consequences:

- Can be used in conjunction with informative priors on other parameters
- Avoids Dawid et al. (1973)'s marginalisation paradoxes
- The posterior characterisation of uncertainty remains subjective, but...

A candidate for "objective" Bayesian inference (Hieu et al. 2025)

Criteria:

1) Proper ✓
2) Satisfies likelihood principle ✓
3) Invariant under re-parametrisation: $f_{T(\theta)} = \mathbf{1}$ ✓
4) Coherence under partitioning

Our candidate: $f_{\boldsymbol{\theta}} = \mathbf{1}$

Consequences:

- Can be used in conjunction with informative priors on other parameters
- Avoids Dawid et al. (1973)'s marginalisation paradoxes
- The posterior characterisation of uncertainty remains subjective, but...

A candidate for "objective" Bayesian inference (Hieu et al. 2025)

Criteria:

1) Proper ✓
2) Satisfies likelihood principle ✓
3) Invariant under re-parametrisation: $f_{T(\theta)} = \mathbf{1}$ ✓
4) Coherence under partitioning: pullback of $\mathbf{1}$ is always $\mathbf{1}$ ✓

Our candidate: $f_{\theta} = \mathbf{1}$

Consequences:

- Can be used in conjunction with informative priors on other parameters
- Avoids Dawid et al. (1973)'s marginalisation paradoxes
- The posterior characterisation of uncertainty remains subjective, but...

A candidate for "objective" Bayesian inference (Hieu et al. 2025)

Criteria:

1) Proper ✓
2) Satisfies likelihood principle ✓
3) Invariant under re-parametrisation: $f_{T(\theta)} = \mathbf{1}$ ✓
4) Coherence under partitioning: pullback of $\mathbf{1}$ is always $\mathbf{1}$ ✓

Our candidate: $f_{\theta} = \mathbf{1}$

Consequences:

- Can be used in conjunction with informative priors on other parameters
- Avoids Dawid et al. (1973)'s marginalisation paradoxes
- The posterior characterisation of uncertainty remains subjective, but...

A candidate for “objective” Bayesian inference (Hieu et al. 2025)

Criteria:

1) Proper ✓
2) Satisfies likelihood principle ✓
3) Invariant under re-parametrisation: $f_{T(\theta)} = \mathbf{1}$ ✓
4) Coherence under partitioning: pullback of $\mathbf{1}$ is always $\mathbf{1}$ ✓

Our candidate: $f_\theta = \mathbf{1}$

Consequences:

- Can be used in conjunction with informative priors on other parameters
- Avoids Dawid et al. (1973)'s marginalisation paradoxes
- The posterior characterisation of uncertainty remains subjective, but...

A candidate for objective Bayesian inference

... it can be validified (Martin 2022):

$$\tilde{\pi}_x(\theta) = \mathbb{P}_\theta(\pi(\theta \mid X) \leq \pi(\theta \mid x))$$

Remarks:

- This is *strongly valid*: for any $\delta \in [0, 1]$

$$\sup_{\theta \in \Theta} \mathbb{P}_\theta \left(\tilde{\pi}_X(\theta) \leq \delta \right) \leq \delta.$$

- It can be extended to priors different from $\mathbf{1}$ (Martin 2023)

Principle: *First* define the raw posterior $\pi(\theta|x)$, *then* validify.

Question: How to define summary statistics for a possibility function?

Moments (Hieu et al. 2025)

Proposition (LLN) If $\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ is a sequence of i.i.d. deterministic uncertain variables on $\mathbb{R}^d$ with a twice continuously diff. possibility function $f_{\boldsymbol{\theta}}$ such that $\lim_{\|x\| \to \infty} f_{\boldsymbol{\theta}}(x) = 0$, then the possibility function $f_{\boldsymbol{s}_n}$ describing $\boldsymbol{s}_n = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\theta}_i$ verifies

$$\lim_{n \to \infty} f_{\boldsymbol{s}_n} = \mathbf{1}_{\mathrm{Conv}(\mathbb{E}^*(\boldsymbol{\theta}))}.$$

Proposition (CLT) If $\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ is a sequence of i.i.d. deterministic uncertain variables on $\mathbb{R}$ with a strictly log-concave and twice-diff. possibility function $f_{\boldsymbol{\theta}}$, then $\mathbb{E}^*(\boldsymbol{\theta})$ is a singleton and the possibility function $f_{\boldsymbol{t}_n}$ describing $\boldsymbol{t}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\boldsymbol{\theta}_i - \mathbb{E}^*(\boldsymbol{\theta}))$ verifies

$$\lim_{n \to \infty} f_{\boldsymbol{t}_n} = \overline{\mathrm{N}}(0, \bar{\mathcal{I}}(\boldsymbol{\theta})^{-1}).$$

Moments (Hieu et al. 2025)

Proposition (LLN) If $\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots$ is a sequence of i.i.d. deterministic uncertain variables on $\mathbb{R}^d$ with a twice continuously diff. possibility function $f_{\boldsymbol{\theta}}$ such that $\lim_{\|x\|\to\infty} f_{\boldsymbol{\theta}}(x) = 0$, then the possibility function $f_{\boldsymbol{s}_n}$ describing $\boldsymbol{s}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\theta}_i$ verifies

$$\lim_{n\to\infty} f_{\boldsymbol{s}_n} = \mathbf{1}_{\mathrm{Conv}(\mathbb{E}^*(\boldsymbol{\theta}))}.$$

$\downarrow$ *motivates* $\downarrow$

$$\boxed{\mathbb{E}^*(\boldsymbol{\theta}) \doteq \arg\max_{\theta\in\Theta} f_{\boldsymbol{\theta}}(\theta)}$$

s.t. $\mathbb{E}^*(T(\boldsymbol{\theta})) = T(\mathbb{E}^*(\boldsymbol{\theta}))$ for any T

Proposition (CLT) If $\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots$ is a sequence of i.i.d. deterministic uncertain variables on $\mathbb{R}$ with a strictly log-concave and twice-diff. possibility function $f_{\boldsymbol{\theta}}$, then $\mathbb{E}^*(\boldsymbol{\theta})$ is a singleton and the possibility function $f_{\boldsymbol{t}_n}$ describing $\boldsymbol{t}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\boldsymbol{\theta}_i - \mathbb{E}^*(\boldsymbol{\theta}))$ verifies

$$\lim_{n\to\infty} f_{\boldsymbol{t}_n} = \overline{\mathrm{N}}(0, \bar{\mathcal{I}}(\boldsymbol{\theta})^{-1}).$$

Moments (Hieu et al. 2025)

Proposition (LLN) If $\theta, \theta_1, \theta_2, \ldots$ is a sequence of i.i.d. deterministic uncertain variables on $\mathbb{R}^d$ with a twice continuously diff. possibility function $f_{\boldsymbol{\theta}}$ such that $\lim_{\|x\|\to\infty} f_{\boldsymbol{\theta}}(x) = 0$, then the possibility function $f_{\boldsymbol{s}_n}$ describing $\boldsymbol{s}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\theta}_i$ verifies

$$\lim_{n\to\infty} f_{\boldsymbol{s}_n} = \mathbf{1}_{\mathrm{Conv}(\mathbb{E}^*(\boldsymbol{\theta}))}.$$

$\downarrow$ *motivates* $\downarrow$

$$\boxed{\mathbb{E}^*(\boldsymbol{\theta}) \doteq \arg\max_{\theta\in\Theta} f_{\boldsymbol{\theta}}(\theta)}$$

s.t. $\mathbb{E}^*(T(\boldsymbol{\theta})) = T(\mathbb{E}^*(\boldsymbol{\theta}))$ for any T

Proposition (CLT) If $\theta, \theta_1, \theta_2, \ldots$ is a sequence of i.i.d. deterministic uncertain variables on $\mathbb{R}$ with a strictly log-concave and twice-diff. possibility function $f_{\boldsymbol{\theta}}$, then $\mathbb{E}^*(\boldsymbol{\theta})$ is a singleton and the possibility function $f_{\boldsymbol{t}_n}$ describing $\boldsymbol{t}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\boldsymbol{\theta}_i - \mathbb{E}^*(\boldsymbol{\theta}))$ verifies

$$\lim_{n\to\infty} f_{\boldsymbol{t}_n} = \overline{\mathrm{N}}(0, \bar{\mathcal{I}}(\boldsymbol{\theta})^{-1}).$$

$\downarrow$ *motivates* $\downarrow$

$$\boxed{\bar{\mathcal{I}}(\boldsymbol{\theta}) \doteq \mathbb{E}^*\left(-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right)}$$

Applications of possibility theory

- Possibilistic instrumental variables
  $\hookrightarrow$ Steiner et al. (2026) UAI
- Distributed inference / Federated learning
  $\hookrightarrow$ H., Cai, et al. (2026) Signal Processing
- Prior specification
  $\hookrightarrow$ Ongoing work with G. Steiner, Y. McLatchie & E. Fong
- Robust inference
  $\hookrightarrow$ H. and Nott (2022)
- Possibilistic Variational inference
  $\hookrightarrow$ Theory: Singh et al. (2025)
  $\hookrightarrow$ Applications: Ni et al. (2026) ICML

Outer probability measures

Uncertain variable (H. 2018)

Definition:

An *uncertain variable* (u.v.) is a mapping $\mathcal{X} : \Omega_{\mathrm{u}} \times \Omega_{\mathrm{r}} \to S$, with $(\Omega_{\mathrm{r}}, \mathcal{E}, \mathbb{P}(\cdot \mid \omega_{\mathrm{u}}))$

- Standard approach: make the sample space disappear
- Difficulty: The law of $\mathcal{X}(\omega_{\mathrm{u}}, \cdot)$ depends on ω_{u}
- Solution: Assume more structure, e.g. $\mathcal{X}(\omega_{\mathrm{u}}, \omega_{\mathrm{r}}) = Y(\boldsymbol{\theta}(\omega_{\mathrm{u}}), \omega_{\mathrm{r}})$

$$\mathcal{X}(\omega_{\mathrm{u}}, \omega_{\mathrm{r}}) = \boldsymbol{\psi}(\omega_{\mathrm{u}}, X(\omega_{\mathrm{r}}))$$

$$\boldsymbol{\theta} \longrightarrow Y \qquad\qquad X \longrightarrow \boldsymbol{\psi}$$

Addresses:

- Some limitations of Dempster-Shafer theory (Pearl 1990)
- Some "paradoxes" about belief functions (Gelman 2006)

Belief functions

Ingredients:

- X a random variable with law p
- $\boldsymbol{\psi}$ a deterministic u.v. on Θ described by $f(\cdot \mid X)$

Consequences:

- The u.v. $\mathcal{X}$ can be described by an *outer (probability) measure*

$$\bar{P}_{X,\boldsymbol{\psi}}(\varphi) = \int p(x)\Big(\sup_{\psi \in \Theta} \varphi(x,\psi) f(\psi \mid x)\Big)\mathrm{d}x, \qquad \text{with e.g. } \varphi = \mathbf{1}_A$$

- The posterior distribution of X given $\boldsymbol{\psi} = \psi$ is

$$p(x \mid \psi) = \frac{f(\psi \mid x) p(x)}{\int f(\psi \mid x) p(x) \mathrm{d}x}$$

- Fusion of two u.v.s like $\mathcal{X}$ when $f(\cdot \mid X)$ is an indicator function almost surely

$\iff$ Dempster's rule of combination

Inference (Hieu et al. 2025)

Ingredients:

- $\boldsymbol{\theta}$ a deterministic u.v. on Θ described by f
- $Y|\boldsymbol{\theta}$ a random variable with law $p(\cdot \mid \boldsymbol{\theta})$

Consequences:

- The u.v. $\mathcal{X}$ can be described by an *outer (probability) measure*

$$\bar{P}_{\boldsymbol{\theta},Y}(\varphi) = \sup_{\theta \in \Theta} f(\theta) \int \varphi(\theta, y) p(y \mid \theta) \mathrm{d}y$$

- The posterior possibility function of $\boldsymbol{\theta}$ given $Y = y$ is

$$f(\theta \mid y) = \frac{p(y \mid \theta) f(\theta)}{\sup_{\theta \in \Theta} p(y \mid \theta) f(\theta)}$$

$\rightsquigarrow$ possibilistic Bernstein-von Mises theorem

OPMs are closed under Bayesian updating

Example

Bayes' rule with $\bar{P} \doteq \bar{P}_{\boldsymbol{\theta},Y}$ as prior and $L(\theta, y; D)$ as likelihood :

$$\bar{P}(\varphi|D) = \sup_{\theta\in\Theta} \underbrace{\frac{\bar{L}(\theta; D)f(\theta)}{\sup_{\vartheta\in\Theta} \bar{L}(\vartheta; D)f(\vartheta)}}_{\doteq\, f(\theta\,|\,D)} \int \varphi(\theta, y) \underbrace{\frac{L(\theta, y; D)p(y\,|\,\theta)}{\int L(\theta, y'; D)p(y'\,|\,\theta)\mathrm{d}y'}}_{\doteq\, p(y\,|\,\theta, D)}$$

with $\bar{L}(\theta; D) = \mathbb{E}_Y[L(\theta, Y; D)\,|\,\theta]$ the marginal likelihood.

But, this is not the case for prediction: if $\bar{Q}(\cdot\,|\,\theta, y)$ is a conditional OPM then

$$\bar{P}_+(\varphi) = \bar{P}(\bar{Q}(\varphi\,|\,\cdot))$$

is a more complex OPM than $\bar{P}$ in general.

Role in IP

Overview

Probability $\rightarrow$ random phenomenon

Credal set $\mathcal{C}$ $\rightarrow$ lack of information about *which* random phenomenon in $\mathcal{C}$

OPM $\rightarrow$ *arbitrary combination* of random and deterministic phenomena

Q.1 Why not a random set formulation?

$\hookrightarrow$ OPMs and membership functions disagree on test functions in general

Q.2 Why not a lower probability?

$\hookrightarrow$ Bayes' rule for OPMs has better properties (IPR, Dempter's rule, etc.)

On coherence

Considering the OPM $\bar{P}$:

- $\bar{P}(\cdot \mid D)$ is not *coherent* but the following OPM is:

$$\bar{P}_\epsilon(\varphi|D) = \sup_{\theta : f(\theta|D) > \epsilon} \int \varphi(\theta, y) p(y \mid \theta, D) \mathrm{d}y$$

- We can start from $\Theta = \mathcal{P}(\mathsf{X})$ and $f = \mathbf{1}$ and still learn something (beyond the incompatible parameters)

Principle: *First* define the raw posterior $\bar{P}(\cdot \mid D)$, *then* make it coherent.

On dilation

- A possibility function f is an *odd/likelihood ratio* $f(\theta)/f(\theta^*)$

  $\implies$ Change in θ^* can increase f $\rightsquigarrow$ dilation!

- Dilation occurs under prior-data conflicts (related to Walter and Coolen (2016))

  $\hookrightarrow$ No dilation when starting from $\mathbf{1}$

  $\hookrightarrow$ No dilation at the price of an open world

- But Bernstein-von Mises gives

  $\hookrightarrow$ The curvature is additive at any fixed point

  $\hookrightarrow$ The possibilistic precision accumulates exactly like Fisher information

Ongoing/future work

- Causal inference
  $\hookrightarrow$ Exploit ordering between variables to break equivalence class

- Information geometry
  $\hookrightarrow$ Leverage special properties of possibilistic exponential families

- Extreme events
  $\hookrightarrow$ Distinguish between unlikely events and uncertain ones

And, finally,

Thank you!

jeremie.houssineau@ntu.edu.sg
jeremiehoussineau.com

References

- Augustin, Thomas, Frank PA Coolen, Gert De Cooman, and Matthias CM Troffaes (2014). ***Introduction to imprecise probabilities.*** John Wiley & Sons.
- Bissiri, Pier Giovanni, Chris C Holmes, and Stephen G Walker (2016). **"A general framework for updating belief distributions".** In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 78.5, pp. 1103–1130.
- Caprio, Michele, Souradeep Dutta, Kuk Jin Jang, Vivian Lin, Radoslav Ivanov, Oleg Sokolsky, and Insup Lee (2023). **"Credal Bayesian deep learning".** In: *arXiv preprint arXiv:2302.09656*.
- Chau, Siu Lun, Michele Caprio, and Krikamol Muandet (2025). **"Integral Imprecise Probability Metrics".** In: *arXiv preprint arXiv:2505.16156*.
- Cozman, Fabio G (2000). **"Credal networks".** In: *Artificial intelligence* 120.2, pp. 199–233.

- Dawid, A Philip, Mervyn Stone, and James V Zidek (1973). **“Marginalization paradoxes in Bayesian and structural inference”.** In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 35.2, pp. 189–213.
- De Cooman, Gert (1997). **“Possibility theory I: the measure-and integral-theoretic groundwork”.** In: *International journal of general systems* 25.4, pp. 291–323.
- Del Moral, P (1997). ***A survey of Maslov optimization theory.| In: VN Kolokoltsov and VP Maslov, Idempotent Analysis and Applications*.**
- Del Moral, Pierre and Michel Doisy (1999). **“Maslov idempotent probability calculus, I”.** In: *Theory of Probability & Its Applications* 43.4, pp. 562–576.
- Dempster, Arthur P (2008). **“Upper and lower probabilities induced by a multivalued mapping”.** In: *Classic works of the Dempster-Shafer theory of belief functions*. Springer, pp. 57–72.
- Denœux, Thierry (2006). **“Constructing belief functions from sample data using multinomial confidence regions”.** In: *International Journal of Approximate Reasoning* 42.3, pp. 228–252.

- Destercke, Sébastien, Didier Dubois, and Eric Chojnacki (2008). **"Unifying practical uncertainty representations–I: Generalized p-boxes".** In: *International Journal of Approximate Reasoning* 49.3, pp. 649–663.
- Dubois, Didier and Henry Prade (2015). **"Possibility theory and its applications: Where do we stand?"** In: *Springer Handbook of Computational Intelligence*. Springer, pp. 31–60.
- Fisher, Ronald A (1935). **"The fiducial argument in statistical inference".** In: *Annals of eugenics* 6.4, pp. 391–398.
- Gelman, Andrew (2006). **"The boxer, the wrestler, and the coin flip: a paradox of robust Bayesian inference and belief functions".** In: *The American Statistician* 60.2, pp. 146–150.
- H., J. (2018). **"Parameter estimation with a class of outer probability measures".** In: *arXiv preprint arXiv:1801.00569*.
- H., J., Han Cai, Murat Uney, and Emmanuel Delande (2026). **"A possibilistic framework for multi-target multi-sensor fusion".** In: *Signal Processing*.
- H., J. and David J Nott (2022). **"Robust Bayesian inference in complex models with possibility theory".** In: *arXiv preprint arXiv:2204.06911*.

Hannig, Jan (2009). **"On generalized fiducial inference".** In: *Statistica Sinica*, pp. 491–544.

Hieu, Nong Minh, J. H., Neil K Chada, and Emmanuel Delande (2025). **"Decoupling epistemic and aleatoric uncertainties with possibility theory".** In: *The 28th International Conference on Artificial Intelligence and Statistics*.

Martin, Ryan (2022). **"Valid and efficient imprecise-probabilistic inference with partial priors, I. First results".** In: *arXiv preprint arXiv:2203.06703*.

— (Sept. 2023). ***Valid and efficient imprecise-probabilistic inference with partial priors, II. General framework.*** arXiv:2211.14567 [stat]. DOI: 10.48550/arXiv.2211.14567. URL: http://arxiv.org/abs/2211.14567 (visited on 10/14/2025).

Martin, Ryan and Chuanhai Liu (2015). ***Inferential models: reasoning with uncertainty.*** CRC Press.

Nguyen, Hung T and Bernadette Bouchon-Meunier (2003). **"Random sets and large deviations principle as a foundation for possibility measures".** In: *Soft Computing* 8.1, pp. 61–70.

Ni, Yao, J. H., Yew Soon Ong, and Piotr Koniusz (2026). **"Possibilistic Predictive Uncertainty for Deep Learning".** In: *ICML, arXiv preprint arXiv:2605.00600*.

Pearl, Judea (1990). **"Reasoning with belief functions: An analysis of compatibility".** In: *International Journal of Approximate Reasoning* 4.5-6, pp. 363–389.

Puhalskii, Anatolii (2001). ***Large deviations and idempotent probability.*** Chapman and Hall/CRC.

Sale, Yusuf, Michele Caprio, and Eyke Hüllermeier (31 Jul–04 Aug 2023). **"Is the volume of a credal set a good measure for epistemic uncertainty?"** In: *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*. Ed. by Robin J. Evans and Ilya Shpitser. Vol. 216. Proceedings of Machine Learning Research. PMLR, pp. 1795–1804. URL: https://proceedings.mlr.press/v216/sale23a.html.

Shafer, Glenn (1976). ***A mathematical theory of evidence.*** Vol. 42. Princeton university press.

Singh, Jasraj, Shelvia Wongso, J. H., and Badr-Eddine Chérief-Abdellatif (2025). **"Maxitive Donsker-Varadhan Formulation for Possibilistic Variational Inference".** In: *arXiv preprint arXiv:2511.21223*.

Smets, Philippe and Robert Kennes (1994). **"The transferable belief model".** In: *Artificial intelligence* 66.2, pp. 191–234.

Steiner, Gregor, Jeremie H., and Mark F.J. Steel (2026). **"Possibilistic Instrumental Variable Regression".** In: *UAI, arXiv preprint arXiv:2511.16029*.

Troffaes, Matthias CM and Gert De Cooman (2014). ***Lower previsions.*** John Wiley & Sons.

Walley, Peter (2000). **"Towards a unified theory of imprecise probability".** In: *International Journal of Approximate Reasoning* 24.2-3, pp. 125–148.

Walter, Gero and Frank PA Coolen (2016). **"Sets of priors reflecting prior-data conflict and agreement".** In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, pp. 153–164.

Yager, Ronald R (1987). **"On the Dempster-Shafer framework and new combination rules".** In: *Information sciences* 41.2, pp. 93–137.

— (2013). **"Nonmonotonic reasoning via possibility theory".** In: *arXiv preprint arXiv:1304.2381*.

Zadeh, Lotfi Asker (1978). **"Fuzzy sets as a basis for a theory of possibility".** In: *Fuzzy sets and systems* 1.1, pp. 3–28.

Zaffalon, Marco (2002). **“The naive credal classifier”.** In: *Journal of statistical planning and inference* 105.1, pp. 5–21.

Zellner, Arnold (1988). **“Optimal information processing and Bayes’s theorem”.** In: *The American Statistician* 42.4, pp. 278–280.