Augmenting Bayesian inference with possibility theory

Jérémie Houssineau

22 April 2022

Outline

1. Why?
2. A dual representation of uncertainty
3. Applications

Outline

Why?

Objective

Introduce a generalised representation of uncertainty

$\hookrightarrow$ Develop a more intuitive notion of information

$\hookrightarrow$ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:

Why?

Objective

Introduce a generalised representation of uncertainty

$\hookrightarrow$ Develop a more intuitive notion of information

$\hookrightarrow$ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:

$$\Theta \quad \{ \quad \circ \quad \circ \quad \}$$

Why?

Objective

Introduce a generalised representation of uncertainty

$\hookrightarrow$ Develop a more intuitive notion of information

$\hookrightarrow$ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:

$$\Theta \quad \{ \overset{1/2}{\circ} \quad \overset{1/2}{\circ} \}$$

Why?

Objective

Introduce a generalised representation of uncertainty

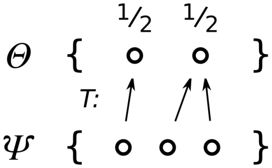
$\hookrightarrow$ Develop a more intuitive notion of information

$\hookrightarrow$ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:



Why?

Objective

Introduce a generalised representation of uncertainty

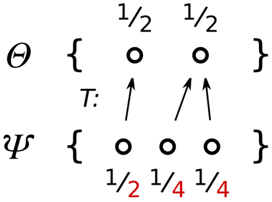
$\hookrightarrow$ Develop a more intuitive notion of information

$\hookrightarrow$ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:



Why?

Objective

Introduce a generalised representation of uncertainty

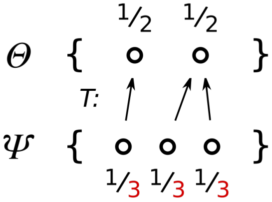
↪ Develop a more intuitive notion of information

↪ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:



Why?

Objective

Introduce a generalised representation of uncertainty

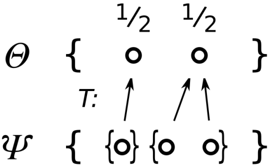
$\hookrightarrow$ Develop a more intuitive notion of information

$\hookrightarrow$ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:



Why?

Objective

Introduce a generalised representation of uncertainty

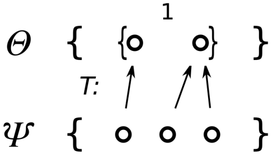
↪ Develop a more intuitive notion of information

↪ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:



Why?

Objective

Introduce a generalised representation of uncertainty

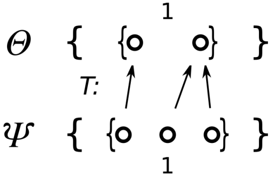
$\hookrightarrow$ Develop a more intuitive notion of information

$\hookrightarrow$ Address difficulties when prior information is lacking

Criteria: leverage power of probabilistic Statistics

Example

Laplace's principle of insufficient reason:



Outline

Uncertain variable

Ingredients:

- A sample space Ω_{u} for deterministic but uncertain phenomena with $\omega^*_{\mathrm{u}} \in \Omega_{\mathrm{u}}$ the true outcome
- A probability space $(\Omega_{\mathrm{r}}, \Sigma, \mathbb{P}(\cdot \,|\, \omega_{\mathrm{u}}))$ for random phenomena, $\omega_{\mathrm{u}} \in \Omega_{\mathrm{u}}$

Definition

An *uncertain variable* (u.v.) is a mapping $\mathcal{X}$ from $\Omega_{\mathrm{u}} \times \Omega_{\mathrm{r}}$ to a given set S such that $\mathcal{X}(\omega_{\mathrm{u}}, \cdot)$ is a random variable for any $\omega_{\mathrm{u}} \in \Omega_{\mathrm{u}}$.

The u.v. $\mathcal{X}$ is *described* by an *outer probability measure* (o.p.m.) $\bar{P}_{\mathcal{X}}$:

- Even if A and B disjoint,

$$\bar{P}_{\mathcal{X}}(A \cup B) \leq \bar{P}_{\mathcal{X}}(A) + \bar{P}_{\mathcal{X}}(B) \qquad \text{(sub-additivity)}$$

- $\bar{P}_{\mathcal{X}}(B)$ is the *credibility* of the event $\mathcal{X} \in B$

Deterministic uncertain variable

Definition

A *deterministic u.v.* (d.u.v.) in a given set Θ is a u.v. $\mathcal{X}$ in Θ for which there exists $\boldsymbol{\theta} : \Omega_{\mathrm{u}} \to \Theta$ such that $\mathcal{X}(\cdot, \omega_{\mathrm{r}}) = \boldsymbol{\theta}$.

- Introduce a *possibility function* $f_{\boldsymbol{\theta}}$, i.e. $f_{\boldsymbol{\theta}} \geq 0$ and $\sup f_{\boldsymbol{\theta}} = 1$

$$\bar{P}_{\boldsymbol{\theta}}(B) = \sup_{\theta \in B} f_{\boldsymbol{\theta}}(\theta), \qquad B \subseteq \Theta$$

- LLN and CLT yield the notions[1]

$$\mathbb{E}^*(\boldsymbol{\theta}) = \operatorname*{argsup}_{\theta \in \Theta} f_{\boldsymbol{\theta}}(\theta) \qquad \text{and} \qquad \mathbb{V}^*(\boldsymbol{\theta}) = \mathbb{E}^*\bigg(-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \bigg)^{-1}$$

with $\mathbb{E}^*(T(\boldsymbol{\theta})) = T(\mathbb{E}^*(\boldsymbol{\theta}))$.

[1] H., Chada & Delande (2019)

Possibility functions: pros & cons

- Cannot sample
- $f_{\boldsymbol{\theta}}$ is not unique $\rightsquigarrow$ Information can be traded for desirable properties
- Does not require a reference measure
- Can be truncated, discretized
- Standard operations apply directly: if $(\boldsymbol{\theta}, \boldsymbol{\psi})$ d.u.v. described by $f_{\boldsymbol{\theta},\boldsymbol{\psi}}$

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

- Can be combined:

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

Uncertain variable

Recall

An *uncertain variable* (u.v.) is a mapping $\mathcal{X} : \Omega_{\mathrm{u}} \times \Omega_{\mathrm{r}} \to S$.

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

Addresses:

- Limitations of Dempster-Shafer (DS) theory (Pearl 1990)
- Paradox about belief functions (Gelman 2006)

Uncertain variables of the form $\mathcal{X} = \boldsymbol{\psi}(\cdot, X)$

Ingredients:

- X a r.v. with law p
- $\boldsymbol{\psi}|X$ a d.u.v. on Ψ described by $f_{\boldsymbol{\psi}}(\cdot \,|\, X)$

Consequences:

- The u.v. $\mathcal{X}$ can be described by

$$\bar{P}_{\mathcal{X}}(\varphi) = \int \sup_{\psi \in \Psi} \big[\varphi(\psi) f_{\boldsymbol{\psi}}(\psi \,|\, x)\big] p(\mathrm{d}x)$$

$\rightsquigarrow$ fuzzy Dempster-Shafer theory

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

$$p(\mathrm{d}x \,|\, \psi) \stackrel{\int}{=} \frac{f_{\boldsymbol{\psi}}(\psi \,|\, x) p(\mathrm{d}x)}{\int f_{\boldsymbol{\psi}}(\psi \,|\, z) p(\mathrm{d}z)}$$

$\rightsquigarrow$ generalised Bayesian inference

Uncertain variables of the form $\mathcal{X} = Y(\boldsymbol{\theta}, \cdot)$

Inference

Ingredients:

- $\boldsymbol{\theta}$ a d.u.v. on Θ described by $f_{\boldsymbol{\theta}}$
- $Y|\boldsymbol{\theta}$ a r.v. with law $p(\cdot \,|\, \boldsymbol{\theta})$

Consequences:

- The u.v. $\mathcal{X}$ can be described by

$$\bar{P}_{\mathcal{X}}(\varphi) = \sup_{\theta \in \Theta} f_{\boldsymbol{\theta}}(\theta) \int \varphi(y) p(\mathrm{d}y \,|\, \theta)$$

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

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

$\rightsquigarrow$ possibilistic Bernstein-von Mises theorem

Uncertain variables of the form $\mathcal{X} = Y(\boldsymbol{\theta}, \cdot)$

Connections with the frequentist approach

With the uninformative prior $f_{\boldsymbol{\theta}} = \mathbf{1}$:

- The posterior expected value $\mathbb{E}^*(\boldsymbol{\theta} \,|\, Y = y)$ is the MLE $\theta^*(y)$
- The function $f_{\boldsymbol{\theta}}(\theta_0|\cdot)$ for a given point $\theta_0 \in \Theta$ is a likelihood ratio test
- The inverse of the posterior variance of $\boldsymbol{\theta}$ is

$$\mathbb{V}^*(\boldsymbol{\theta} \,|\, Y = y)^{-1} = \mathbb{E}^*\bigg(-\frac{\partial^2}{\partial \theta^2} \log p(y \,|\, \boldsymbol{\theta}) \,\bigg|\, Y = y \bigg) = -\frac{\partial^2}{\partial \theta^2} \log p(y \,|\, \theta^*(y)),$$

which is the observed information

$\rightsquigarrow$ Is $f_{\boldsymbol{\theta}} = \mathbf{1}$ the only interesting case?

Uncertain variables of the form $\mathcal{X} = Y(\boldsymbol{\theta}, \cdot)$

Conjugate prior families

- For **any** conjugate prior family $\mathcal{F} \rightsquigarrow$ possibilistic analogue $\bar{\mathcal{F}}$
- $\mathbf{1} \in \bar{\mathcal{F}}$
- for $\delta \in (0,1)$, $f \in \bar{\mathcal{F}} \implies f^\delta \in \bar{\mathcal{F}}$, e.g.,

$$\overline{\mathrm{Ga}}(\theta; \alpha, \beta)^\delta = \left[\left(\frac{\beta\theta}{\alpha} \right)^\alpha \exp(\alpha - \beta\theta) \right]^\delta = \overline{\mathrm{Ga}}(\theta; \delta\alpha, \delta\beta)$$

$\rightsquigarrow$ information can be discounted

$\rightsquigarrow$ also holds for mixtures

- There are new families, e.g.

$$\overline{\mathrm{SIG}}(\theta; \alpha, \beta, \sigma^2) = \left(\frac{\beta \vee \alpha\sigma^2}{\alpha(\sigma^2 + \theta)} \right)^\alpha \exp\left(\frac{\alpha\beta}{\beta \vee \alpha\sigma^2} - \frac{\beta}{\sigma^2 + \theta} \right)$$

with $\mathbb{E}^*(\boldsymbol{\theta}) = (\beta/\alpha - \sigma^2)^+$

Uncertain variables of the form $\mathcal{X} = Y(\boldsymbol{\theta}, \cdot)$

A candidate for "objective" Bayesian inference

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 marginalisation paradoxes (e.g. Dawid et al. 1973)
- The posterior characterisation of uncertainty remains subjective

Uncertain variables of the form $\mathcal{X} = Y(\boldsymbol{\theta}, \cdot)$

A candidate for "objective" Bayesian inference

Criteria:

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

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

Consequences:

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

Uncertain variables of the form $\mathcal{X} = Y(\boldsymbol{\theta}, \cdot)$

A candidate for "objective" Bayesian inference

Criteria:

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

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

Consequences:

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

A simple example

(and yet a challenging one)

Consider:

- A r.v. X modelling a fair die
- A d.u.v. $\boldsymbol{\theta}$ representing an unknown but fixed number in $\{1, \dots, 6\}$

Difficulty: There are two possible o.p.m.s describing the u.v. $\mathcal{X} = X + \boldsymbol{\theta}$:

$$\bar{P}_{\boldsymbol{\theta},X}(\varphi) = \frac{1}{6} \max_{\theta \in \{1,\dots,6\}} \sum_{n=1}^{6} \varphi(\theta + n) \quad \text{and} \quad \bar{P}_{X,\boldsymbol{\theta}}(\varphi) = \frac{1}{6} \sum_{n=1}^{6} \max_{\theta \in \{1,\dots,6\}} \varphi(\theta + n).$$

sum (s)	2	3	...	6	7	8	...	11	12
$\bar{P}_{\boldsymbol{\theta},X}(s)$	1/6	1/6	...	1/6	1/6	1/6	...	1/6	1/6
$\bar{P}_{X,\boldsymbol{\theta}}(s)$	1/6	2/6	...	5/6	1	5/6	...	2/6	1/6

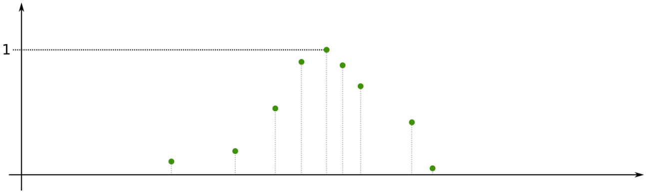
Outline

Data assimilation

Ongoing work with C. Kimchaiwong & A. Johansen

Forecast:

1. Sample $N-1$ particles/support points $\{x_{0,i}\}_{i=1}^{N-1}$ according to $\overline{\mathrm{N}}_{\mathrm{p}}(\mu_0, P_0)$
2. Set $x_{0,N} = \mu_0$
3. Define weights $\{w_i\}_{i=1}^{N}$ via $w_i = \overline{\mathrm{N}}_{\mathrm{p}}(x_{0,i}; \mu_0, P_0)$

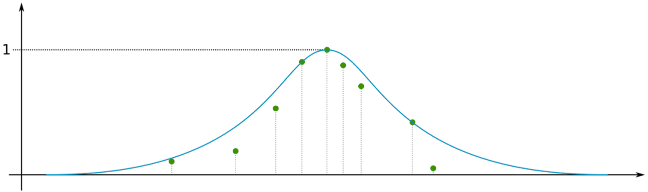


Data assimilation

Ongoing work with C. Kimchaiwong & A. Johansen

Forecast:

1. Sample $N-1$ particles/support points $\{x_{0,i}\}_{i=1}^{N-1}$ according to $\overline{\mathrm{N}}_{\mathrm{p}}(\mu_0, P_0)$
2. Set $x_{0,N} = \mu_0$
3. Define weights $\{w_i\}_{i=1}^{N}$ via $w_i = \overline{\mathrm{N}}_{\mathrm{p}}(x_{0,i}; \mu_0, P_0)$



Data assimilation

Ongoing work with C. Kimchaiwong & A. Johansen

Forecast:

1. Via a fixed but partially-unknown function

$$\boldsymbol{x}_k = F(\boldsymbol{x}_{k-1}) + \boldsymbol{u}_k$$

→ It holds that $\mathbb{E}^*(\boldsymbol{x}_k) = F(\mathbb{E}^*(\boldsymbol{x}_{k-1}))$

2. Set $x_{k,i} = F(\hat{x}_{k-1,i})$ for all $i \in \{1, \dots, N\}$

3. Localised forecast precision can be recovered by optimisation:

$$\sum_{i=1}^{N} w_i \mathbf{1}_{x_{k,i}}(x) \leq \overline{\mathrm{N}}_{\mathrm{p}}(x; x_{k,N}, P_k),$$

with P_k enforcing conditional independence $\rightsquigarrow$ *localisation & inflation*

Data assimilation

Ongoing work with C. Kimchaiwong & A. Johansen

Forecast:

4. Uncertainty in the dynamics by transport $\rightsquigarrow$ *inflation*
   Map between $\overline{\mathrm{N}}_{\mathrm{p}}(\mu, P)$ and $\overline{\mathrm{N}}_{\mathrm{p}}(\mu', Q)$ is

$$T(x) = P^{1/2}(P^{-1/2}Q^{-1}P^{-1/2})^{1/2}P^{1/2}(x-\mu) + \mu'$$

Assimilation:

$\rightarrow$ Only deterministic methods such as square-root KF make sense:

$$K = \Sigma_k H^{\intercal}(S_k^{-1/2})^{\intercal}(S_k^{1/2} + R^{1/2})^{-1}$$

with $\Sigma_k = P_k^{-1}$ and $S_k = H\Sigma_k H^{\intercal} + R$

J. S. Whitaker et al. (2002). "Ensemble data assimilation without perturbed observations". In: *Monthly weather review* 130.7, pp. 1913–1924

Robust inference

Ongoing work with D. Nott

Observation: The possibilistic marginal likelihood indicates *coherence*

Principle: *Discount* the likelihood based on the *coherence* between the information in the likelihood and the one in the prior.

Advantages:

- Parameter free in its simplest form
- Same computational complexity as non-robust inference in many cases

Weaknesses:

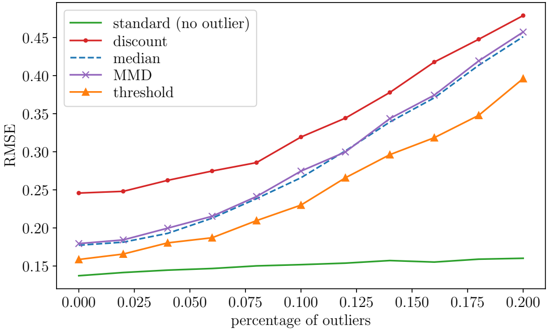
- All observations are discounted to some extent
  - ↪ A threshold can be introduced to control for this
- Ordering matters

J. Houssineau et al. (2022). "Robust Bayesian inference in complex models with possibility theory". In: *arXiv preprint arXiv:2204.06911*

Robust inference

Example: normal distribution with outliers

Objective: Learn the mean of a normal distribution with outliers from a Cauchy distribution

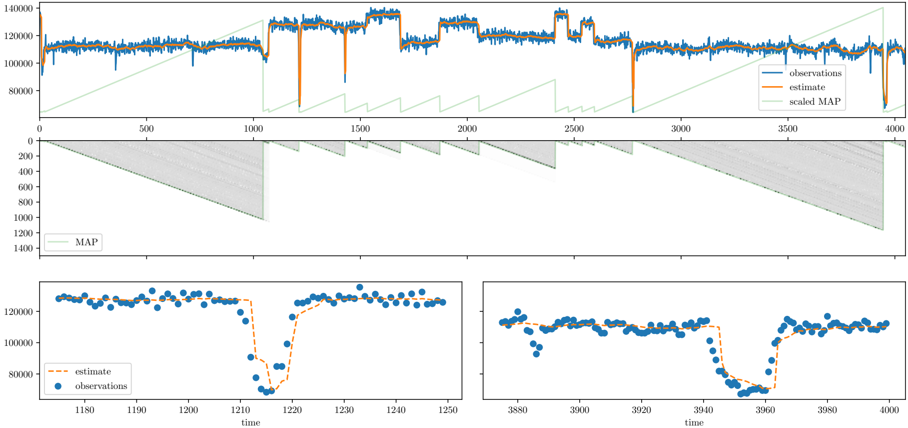


Computational time:

- MMD (Cherief-Abdellatif et al. 2020): $\sim 10s$
- Proposed approach: $\sim 10ms$

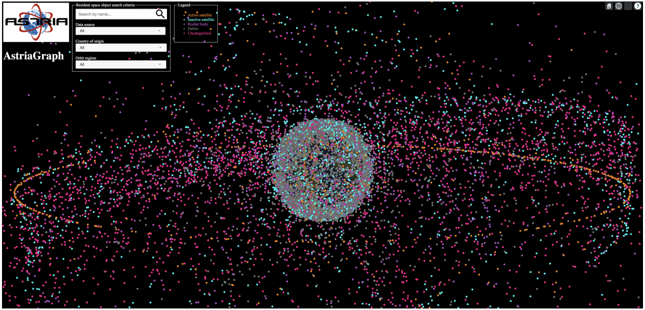
Robust inference

Example: change-point detection



→ See also, e.g., Knoblauch et al. 2018

Space situational awareness

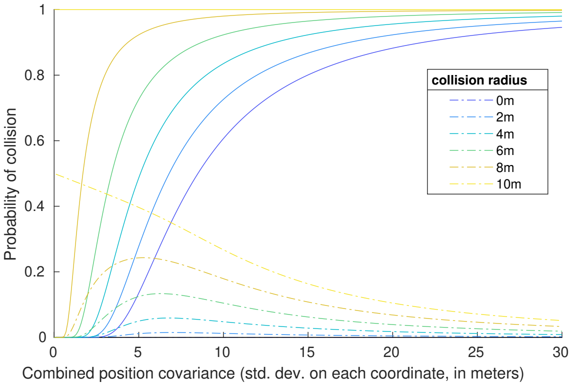


http://astria.tacc.utexas.edu/AstriaGraph/

Challenges:

1. Credibility of collision between space assets
2. Faithful representation of the uncertainty

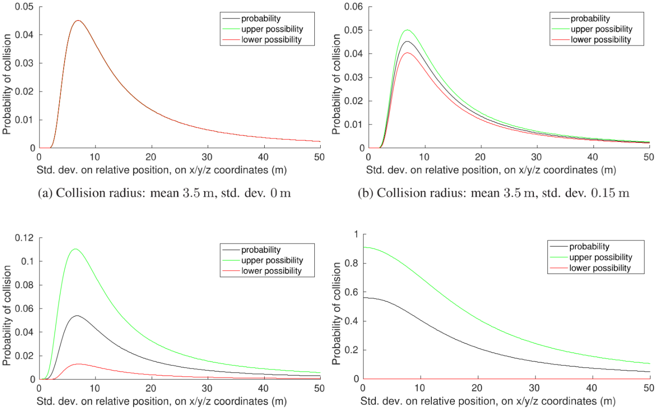
Space situational awareness



Credibility of collision in probabilistic and possibilistic contexts

E. Delande et al. (2019). "A new representation of uncertainty for collision assessment". In: *AAS/AIAA Space Flight Mechanics Meeting*

Space situational awareness



(a) Collision radius: mean 3.5 m, std. dev. 0 m
(b) Collision radius: mean 3.5 m, std. dev. 0.15 m
(c) Collision radius: mean 3.5 m, std. dev. 1.5 m
(d) Collision radius: mean 3.5 m, std. dev. 15 m

Probability of collision with uncertain radius

E. Delande et al. (2019). "A new representation of uncertainty for collision assessment". In: *AAS/AIAA Space Flight Mechanics Meeting*

Space situational awareness

Credibility of a collision event E

- Recent work[0] suggests minimising/maximising

$$\mathbb{P}(E \,|\, y_{1:k}; \theta)$$

with θ the epistemic parameter.

- Can we go further by considering *outer probability measures*, e.g.

$$\bar{\mathbb{P}}(E \,|\, y_{1:k}) = \sup_{\theta \in \Theta} f_{\boldsymbol{\theta}}(\theta \,|\, y_{1:k}) \mathbb{P}(E \,|\, y_{1:k}; \theta)$$

The *subjective* probability of collision, say $\pi(E)$, verifies

$$\min_{\theta \in \Theta} \mathbb{P}(E \,|\, y_{1:k}; \theta) \leq 1 - \bar{\mathbb{P}}(E^{\mathrm{c}} \,|\, y_{1:k}) \leq \pi(E) \leq \bar{\mathbb{P}}(E \,|\, y_{1:k}) \leq \max_{\theta \in \Theta} \mathbb{P}(E \,|\, y_{1:k}; \theta)$$

[0] C. Greco et al. (2021). "Robust Bayesian particle filter for space object tracking under severe uncertainty". In: *Journal of Guidance, Control, and Dynamics*, pp. 1–18

Space situational awareness

Challenges

1. Careful consideration of the "nature" of different sources of uncertainty
    - Initial orbit determination
    - Unknown manoeuvre
    - Number and orbit of debris after a collision
    - Solar radiation pressure $\leftarrow$ Space weather?
    - Drag?

2. Methodology for general outer probability measures
    - Ordering is important
    - Problem becomes increasingly difficult when mixing "max" and "$\int \cdot \, \mathrm{d}x$"

3. Derive the corresponding multi-target tracking solutions?

Conclusion

The proposed approach:

- connects the frequentist and Bayesian approaches (and others!)
  - $\hookrightarrow$ MLE, LRT, empirical Bayes
- models the intuitive notion of information
  - $\hookrightarrow$ information fusion, $\mathbf{1}$, observed information, discounting
- generalises to complex inference problems
  - $\hookrightarrow$ data assimilation, robust inference
  - $\hookrightarrow$ but also, e.g., reinforcement learning and multi-target tracking

Thank you!

jeremie.houssineau@warwick.ac.uk
jeremiehoussineau.com

Cherief-Abdellatif, B.-E. and Alquier, P. (2020). "MMD-Bayes: Robust Bayesian Estimation via Maximum Mean Discrepancy". In: *Proceedings of The 2nd Symposium on Advances in Approximate Bayesian Inference*. Vol. 118. Proceedings of Machine Learning Research. PMLR, pp. 1–21.

Delande, E., Jah, M., and Jones, B. (2019). "A new representation of uncertainty for collision assessment". In: *AAS/AIAA Space Flight Mechanics Meeting*.

Dubois, D. and Prade, H. (2015). "Possibility theory and its applications: Where do we stand?" In: *Springer Handbook of Computational Intelligence*. Springer, pp. 31–60.

Greco, C. and Vasile, M. (2021). "Robust Bayesian particle filter for space object tracking under severe uncertainty". In: *Journal of Guidance, Control, and Dynamics*, pp. 1–18.

H., J. (2018). "Parameter estimation with a class of outer probability measures". In: *arXiv preprint arXiv:1801.00569*.

H., J., Chada, N. K., and Delande, E. (2019). "Elements of asymptotic theory with outer probability measures". In: *arXiv preprint arXiv:1908.04331*.

Houssineau, J. and Nott, D. J. (2022). "Robust Bayesian inference in complex models with possibility theory". In: *arXiv preprint arXiv:2204.06911*.

Knoblauch, J., Jewson, J. E., and Damoulas, T. (2018). "Doubly Robust Bayesian Inference for Non-Stationary Streaming Data with β-Divergences". In: *Advances in Neural Information Processing Systems*. Vol. 31.

Whitaker, J. S. and Hamill, T. M. (2002). "Ensemble data assimilation without perturbed observations". In: *Monthly weather review* 130.7, pp. 1913–1924.