Decoupling Epistemic and Aleatoric Uncertainties using Possibility Theory

Nong Minh Hieu [1, 3] Jeremie Houssineau [1] Neil K. Chada [2] Emanuel Delande [4]

[1]School of Physical and Mathematical Sciences, NTU Singapore [2]Department of Mathematics, CityU Hong Kong
[3]School of Computing and Information Systems, SMU Singapore [4]Centre National D'Etudes Spatiales, France

Overview & Contributions

Bayesian Inference: Let $x_1, \dots, x_n \sim p_{\theta^*}$ be i.i.d. observations whose distribution is parameterized by a true parameter $\theta^* \in \Theta$. In traditional Bayesian framework, θ^* is inferred probabilistically. For example, let p denote the (probabilistic) prior, the MAP estimator is defined as:

$$\theta^*_{\text{MAP}} = \arg\max_{\theta \in \Theta} p(\theta) L(x_1, \dots, x_n; \theta), \quad L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n p_\theta(x_i).$$

Decoupling EU and AU: Two types of uncertainties are Aleatoric Uncertainty (AU) and Epistemic Uncertainty (EU). Specifically:

- AU stems from the inherent randomness of the data.
- EU stems from the lack of information (non-random phenomena).

Contributions: In this work, we proposed an inference framework that models AU and EU jointly through combining probability theory and possibility theory. Under this framework, we proved:

1. A possibilistic version of the Bernstein-von Mises (BvM) theorem.
2. Possibilistic versions of the Law of Large Numbers (LLN) and Central Limit Theorem (CLT).
3. The asymptotic behavior of the possibilistic MAP estimator.

Deterministic Uncertain Variable

Definition (Uncertain Variable): Let Ω_d be the space of deterministic phenomena and $\omega^* \in \Omega_\text{d}$ be the true outcome. A *deterministic uncertain variable* $\boldsymbol{\theta}$ is the mapping $\boldsymbol{\theta} : \Omega_\text{d} \to \Theta$ such that the true outcome is associated to the true parameter θ^*, i.e., $\boldsymbol{\theta}(\omega^*) = \theta^*$.

Example (Coin Toss): Suppose that one among n unfair coins with head probabilities $\rho_1, \dots, \rho_n$ is selected once (no randomness involved) to conduct some experiment. Then, we have the space of deterministic phenomena $\Omega_\text{d} = [n]$ and the deterministic uncertain variable $\boldsymbol{\theta}$ is defined as:

$$\boldsymbol{\theta}(i) = \rho_i, \quad \forall i \in [n].$$

$\longrightarrow$ We can also write $\boldsymbol{\theta} : i \mapsto (\Omega_\text{r}, 2^{\Omega_\text{r}}, \mathbb{P}_{\rho_i})$ where $\Omega_\text{r} = \bigotimes_{i=1}^m \{\text{H}, \text{T}\}$ if the random experiment involves tossing the coin m times and $\mathbb{P}_{\rho_i}$ is the Bernoulli PMF with head probability ρ_i.

Possibility Functions & OPMs

Definition (Possibility Function): A given uncertain variable $\boldsymbol{\theta}$ on Θ is described by the set function $\bar{P} : 2^\Theta \to [0, 1]$, which is an outer probability measure over Θ:

$$\forall A \subseteq \Theta : \bar{P}(A) = \sup_{\theta \in \Theta} f_{\boldsymbol{\theta}}(\theta),$$

where $f_{\boldsymbol{\theta}} : \Theta \to [0, 1]$ is a (non-unique) possibility function that describes $\boldsymbol{\theta}$. For each $A \subseteq \Theta$, $\bar{P}(A)$ represents the *credibility* of the event $\boldsymbol{\theta} \in A$. Specifically,

- $\bar{P}(A) = 1$ implies no evidence against $\boldsymbol{\theta} \in A$.
- $\bar{P}(A) = 0$ implies no possibility that $\boldsymbol{\theta} \in A$.

Important Notations

Given an uncertain variable $\boldsymbol{\theta}$ over $\Theta \subset \mathbb{R}^d$ described by the possibility function $f_{\boldsymbol{\theta}}$, we have the following important definitions:

Notation	Formula
Expected Value	$\mathbb{E}^*_{f_{\boldsymbol{\theta}}}[\boldsymbol{\theta}] = \arg\max_{\theta \in \Theta} f_{\boldsymbol{\theta}}(\theta)$
Variance	$\mathbb{V}^*_{f_{\boldsymbol{\theta}}}(\boldsymbol{\theta}) = \bar{\mathcal{I}}_{f_{\boldsymbol{\theta}}}(\boldsymbol{\theta})^{-1}$
Fisher Information	$\bar{\mathcal{I}}_{f_{\boldsymbol{\theta}}}(\boldsymbol{\theta}) = \mathbb{E}^*_{f_{\boldsymbol{\theta}}}\Big[- H\Big(\log f_{\boldsymbol{\theta}}(\theta) \Big)\Big]$
Gaussian Possibility Function	$\overline{\text{N}}_d\Big(\theta; \mu, \Lambda\Big) = \exp\Big(- \frac{1}{2}(\theta - \mu)^\top \Lambda (\theta - \mu)\Big).$

Table 1. For Fisher Information - $H(\cdot)$ denotes the Hessian matrix. For Gaussian Possibility Function - $\mu \in \mathbb{R}^d$ denotes the mean vector and $\Lambda \in \mathbb{R}^{d \times d}$ is a positive semi-definite *precision* matrix.

Inference with OPMs

Let $\boldsymbol{\theta}$ be the uncertain variable over Θ described by the (prior) possibility function $f_{\boldsymbol{\theta}}$ and X be a random variable whose density comes from the family $\{p_X(\cdot|\theta) : \theta \in \Theta\}$. By analogy to the Bayes rule, the possibility function describing $\boldsymbol{\theta}$ given $X = x$ is defined as:

$$f_{\boldsymbol{\theta}|X}(\theta|x) = \frac{p_X(x|\theta) f_{\boldsymbol{\theta}}(\theta)}{\sup_{\varphi \in \Theta} p_X(x|\varphi) f_{\boldsymbol{\theta}}(\varphi)}.$$

For each $\theta \in \Theta$, we assume that the mapping $\theta \mapsto f_{\boldsymbol{\theta}}(\theta) p_X(x|\theta)$ is bounded. As a result, the equivalent definition of the MAP estimator is:

$$\widehat{\theta}_{\text{MAP}} = \arg\max_{\theta \in \Theta} f_{\boldsymbol{\theta}|X}(\theta|x).$$

Asymptotic Behavior of the Possibilistic Posterior

Theorem (Bernstein-von Mises): Let $x_1, \dots, x_n$ be observations drawn i.i.d. from $p_X(\cdot|\theta^*)$ and let $p_n(\cdot|\theta)$ denotes the distribution of the random sample $x_1, \dots, x_n$ given any $\theta \in \Theta$. Assume that the following conditions are met:

1. $\Theta \subset \mathbb{R}^d$ is compact and convex.
2. $\theta^* \in \Theta$ and the MLE is consistent.
3. The possibilistic prior $f_{\boldsymbol{\theta}}$ is continuous and positive in a neighborhood of θ^*.

Then, for large values of n, it holds that:

$$f_{\boldsymbol{\theta}}(\theta|x_1, \dots, x_n) \approx \overline{\text{N}}\left(\theta; \theta^* + \frac{\Delta_n}{\sqrt{n}}, \mathcal{J}_n\right),$$

where $\mathcal{J}_n = -H_\theta\Big(\log p_n\big(x_1, \dots, x_n \big| \hat{\theta}_{\text{MLE}}(x_1, \dots, x_n)\big)\Big)$ is the observed information at the MLE based on $x_1, \dots, x_n$ and Δ_n is defined as:

$$\Delta_n = \sqrt{n} \mathcal{J}_n^{-1} \nabla_\theta \log p_n(x_1, \dots, x_n | \theta^*).$$

$\longrightarrow$ Intuitively, as $n \to \infty$, the possibilistic MAP tends to $\hat{\theta}_{\text{MLE}}(x_1, \dots, x_n)$, making the information in the possibilistic prior forgotten.

Law of Large Numbers

Theorem (Law of Large Numbers): If $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots$ is a sequence of independent deterministic uncertain variables on $\mathbb{R}^d$ with possibility function $f_{\boldsymbol{x}}$ such that

1. $f_{\boldsymbol{x}}$ is continuous on $\mathbb{R}^d$.
2. $f_{\boldsymbol{x}}$ is twice continuously differentiable on an open neighborhood of each point in $\mathbb{E}^*[\boldsymbol{x}]$.
3. $\lim_{\|x\| \to \infty} f_{\boldsymbol{x}}(x) = 0$.

Then, the possibility function $f_{\boldsymbol{s}_n}$ describing $\boldsymbol{s}_n = \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i$ verifies:

$$\lim_{n \to \infty} f_{\boldsymbol{s}_n}(t) = \mathbf{1}_{\text{Conv}(\mathbb{E}^*[\boldsymbol{x}])}(t),$$

where the convergence is point-wise and $\text{Conv}(S)$ is the convex hull of a set $S \subseteq \mathbb{R}^d$.

Central Limit Theorem

Theorem (Central Limit Theorem): If $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots$ is a sequence of deterministic uncertain variables on $\mathbb{R}$ independently and identically described by a possibility function $f_{\boldsymbol{x}}$ verifying:

1. $f_{\boldsymbol{x}}$ is strictly log-concave.
2. $f_{\boldsymbol{x}}$ is twice-differentiable.

Then $\mathbb{E}^*[\boldsymbol{x}]$ is a singleton and the possibility function $f_{\boldsymbol{t}_n}$ describing $\boldsymbol{t}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\boldsymbol{x}_i - \mathbb{E}^*[\boldsymbol{x}])$ verifies:

$$\lim_{n \to \infty} f_{\boldsymbol{t}_n} = \overline{\text{N}}\Big(0, \bar{\mathcal{I}}(\boldsymbol{x})\Big),$$

where the convergence is uniform.

Asymtotic Behaviour of Possibilistic MAP

Corollary: Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be i.i.d. observations described by a possibility function $f_{\boldsymbol{x}}(\cdot|\theta^*)$. Let $f_n(\cdot|\theta)$ be the possibility function describing $\boldsymbol{x}_{1:n} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ for any $\theta \in \Theta$. Define the corresponding MLE as $\arg\max_{\theta \in \Theta} f_n(\boldsymbol{x}_{1:n}|\theta)$. Under the same conditions as the **BvM** theorem with $d = 1$ and assume additionally:

1. $f_{\boldsymbol{x}}(\cdot|\theta)$ is strictly log-concave for any $\theta \in \Theta$.
2. The map $(\theta, x) \mapsto f_{\boldsymbol{x}}(x|\theta)$ is twice differentiable in x as well as thrice-differentiable in θ.
3. $\partial^3_\theta \log f_n(\boldsymbol{x}_{1:n}|\theta) = \bar{O}_p(n)$.

Then, the possibility function $f_{\boldsymbol{t}_n}$ describing the deterministic uncertain variable $\boldsymbol{t}_n = \sqrt{n}(\widehat{\theta}_{\text{MAP}}(\boldsymbol{x}_{1:n}) - \theta^*)$ verifies:

$$\lim_{n \to \infty} f_{\boldsymbol{t}_n} = \overline{\text{N}}(0, \bar{\mathcal{I}}(\theta^*)^2 \tau_s),$$

where $\tau_s = \bar{\mathcal{I}}(\partial_\theta \log f_{\boldsymbol{x}}(\boldsymbol{x}|\theta^*))$ is the precision of the score.