Data-Based Decisions under Imprecise Probability and Least Favorable Models

Robert Hable *
Department of Statistics
LMU Munich

* supported by Cusanuswerk
Robert Hable
Department of Statistics, LMU Munich (Germany)

- Education: Mathematics (in Bayreuth, Germany)
- Diploma Thesis in Robust Asymptotic Statistics (Helmut Rieder)
- Now: Ph.D. Student under the Guidance of Thomas Augustin
  → Topic: “Data-Based Decisions under Complex Uncertainty”

- Research Interests:
  - Decision Theory under Imprecise Probabilities
  - Mathematical Foundations of Imprecise Probabilities
  - Robust Statistics


**Usual Decision Theory**

- **States of nature:** $\Theta = \{\theta_1, \ldots, \theta_n\}$
- **Decisions:** $t \in \mathbb{D}$
- **Bounded loss functions:** $W_\theta : \mathbb{D} \to \mathbb{R}$, $t \mapsto W_\theta(t)$

|       | $\theta_1$ | $\cdots$ | $\theta_i$ | $\cdots$ | $\theta_n$ |
|-------|-------------|-----------|-------------|-----------|
| $t_1$ | $W_{\theta_1}(t_1)$ | $\cdots$ |             |           | $W_{\theta_n}(t_1)$ |
| $\vdots$ |             |           |             |           | $\ddots$ |
| $t_k$ |             |           |             |           | $W_{\theta_i}(t_k)$ |
| $\vdots$ |             |           |             |           | $\ddots$ |
| $t_m$ | $W_{\theta_1}(t_m)$ | $\cdots$ |             |           | $W_{\theta_n}(t_m)$ |
Usual Decision Theory

- States of nature: $\Theta = \{\theta_1, \ldots, \theta_n\}$
- Decisions: $t \in \mathcal{D}$
- Bounded loss functions: $W_\theta : \mathcal{D} \rightarrow \mathbb{R}$, $t \mapsto W_\theta(t)$
- Prior distribution over $\Theta$: $\pi = (\pi_{\theta_1}, \ldots, \pi_{\theta_n})$
- Expected loss for decision $t \in \mathcal{D}$:
  $$\sum_{\theta \in \Theta} \pi_{\theta} W_\theta(t)$$

Often: Decision making on base of observations $y \in \mathcal{Y}$

- Decision functions: $\delta : \mathcal{Y} \rightarrow \mathcal{D}$, $y \mapsto \delta(y)$
- Distribution of the observation $y$: $q_\theta$ where $\theta \in \Theta$
- Expected loss for decision function $\delta : \mathcal{Y} \rightarrow \mathcal{D}$ is
  $$\sum_{\theta \in \Theta} \pi_{\theta} \int W_\theta(\delta(y)) q_\theta(dy)$$
Decision Theory under Imprecise Probability

- Instead of a precise prior distribution $\pi$:
  Imprecise prior distribution (coherent upper prevision):
  \[
  \Pi[f] = \sup_{\pi \in \mathcal{P}} \pi[f], \quad f : \Theta \rightarrow \mathbb{R}
  \]
  \(\mathcal{P}\): a set of precise prior distributions (credal set)

- Instead of a precise distribution $q_{\theta}$ of the observation $y$:
  Imprecise distribution of the observation $y$ (coherent upper prevision):
  \[
  \overline{Q}_{\theta}[g] = \sup_{q_{\theta} \in \mathcal{M}_{\theta}} q_{\theta}[g] \quad \forall \theta \in \Theta, \quad g : \mathcal{Y} \rightarrow \mathbb{R}
  \]
  \(\mathcal{M}_{\theta}\): a set of precise distributions of the observation $y$
  (credal set)
Imprecise prior distribution (coherent upper prevision):

\[ \Pi[f] = \sup_{\pi \in \mathcal{P}} \pi[f] \]

Imprecise distribution of the observation \( y \) (coherent upper prevision):

\[ Q_\theta[g] = \sup_{q_\theta \in \mathcal{M}_\theta} q_\theta[g] \quad \forall \theta \in \Theta \]

Upper expected loss for decision function \( \delta : \mathcal{Y} \to \mathcal{D} \) is

\[ \sup_{\pi \in \mathcal{P}} \sum_{\theta \in \Theta} \pi_\theta \sup_{q_\theta \in \mathcal{M}_\theta} \int W_\theta(\delta(y)) q_\theta(dy) \]
Task

Find a decision function $\tilde{\delta}$ which minimizes the upper expected loss

$$\sup_{\pi \in \mathcal{P}} \sum_{\theta \in \Theta} \pi_\theta \sup_{q_\theta \in \mathcal{M}_\theta} \int W_\theta(\delta(y)) q_\theta(dy) = \min_{\delta}$$

Optimality criterion:

$\Gamma$-minimax criterion: worst-case consideration

Problem:

often, direct solution computationally intractable
Common Idea

Find another optimization problem which has the following properties:

- Solving this new optimization problem leads to a solution of the original problem.
- The new optimization problem should be computationally tractable!

→ Least Favorable Models
**Least favorable model**

- $M_\theta$: credal set for the distribution of the observation $y$
- $\mathcal{P}$: credal set for the prior distribution

Find some $\tilde{q}_\theta \in M_\theta$ for every $\theta \in \Theta$ so that

$$\inf_\delta \sum_{\theta \in \Theta} \pi_\theta \sup_{q_\theta \in M_\theta} \int W_\theta(\delta(y)) q_\theta(dy) = \inf_\delta \sum_{\theta \in \Theta} \pi_\theta \int W_\theta(\delta(y)) \tilde{q}_\theta(dy) \quad \forall \pi \in \mathcal{P}$$

$(\tilde{q}_\theta)_{\theta \in \Theta} \in (M_\theta)_{\theta \in \Theta}$ is called least favorable model.

($\rightarrow$ Huber-Strassen (1973))
Then, we have:

The new optimization problem

\[
\sup_{\pi \in \mathcal{P}} \sum_{\theta \in \Theta} \pi_{\theta} \int W_{\theta}(\delta(y)) \tilde{q}_{\theta}(dy) = \min_{\delta}!
\]

is computationally easier than the original optimization problem

\[
\sup_{\pi \in \mathcal{P}} \sum_{\theta \in \Theta} \pi_{\theta} \sup_{q_{\theta} \in \mathcal{M}_{\theta}} \int W_{\theta}(\delta(y)) q_{\theta}(dy) = \min_{\delta}!
\]
Data-Based Decisions under Imprecise Probability and Least Favorable Models

... and we have:

There is a solution $\tilde{\delta}$ of the new optimization problem

$$\sup_{\pi \in \mathcal{P}} \sum_{\theta \in \Theta} \pi_{\theta} \int W_{\theta}(\delta(y)) \tilde{q}_{\theta}(dy) = \min_{\delta}$$

which also solves the original optimisation problem is

$$\sup_{\pi \in \mathcal{P}} \sum_{\theta \in \Theta} \pi_{\theta} \sup_{q_{\theta} \in \mathcal{M}_{\theta}} \int W_{\theta}(\delta(y)) q_{\theta}(dy) = \min_{\delta}$$

Robert Hable

LMU Munich
A least favorable model \((\tilde{q}_\theta)_{\theta \in \Theta} \in (\mathcal{M}_\theta)_{\theta \in \Theta}\) does not always exist!

That is: The presented procedure does not always work!

**Question:** When does it work?
Main Result

The **Main Theorem** provides:

A necessary and sufficient condition for the existence of a least favorable model \((\tilde{q}_\theta)_{\theta \in \Theta} \in (\mathcal{M}_\theta)_{\theta \in \Theta}\)

**Remarks:**

- exact condition is rather involved;
  uses some of Le Cam’s concepts such as
  - equivalence of models
  - conical measures (or standard measures)
- generalizes Buja (1984) and Huber-Strassen (1973)
Comparison with Buja 1984 – Some Technicalities

Buja 1984

- Credal sets $\mathcal{M}_\theta$ only contain $\sigma$-additive probability measures
- Condition: Compactness of credal sets $\mathcal{M}_\theta$
  This is restrictive in Buja’s setup! (cf. Hable (2007B, E-print))

Now

- Credal sets $\mathcal{M}_\theta$ may contain finitely-additive probability measures (which are not $\sigma$-additive).
- Compactness of credal sets $\mathcal{M}_\theta$ comes for free in Walley’s setup.

A first conclusion:

- $\sigma$-additivity is not necessary here.
- Getting around $\sigma$-additivity is possible by Le Cam’s setup
Le Cam's setup

- Le Cam: strictly functional analytic approach to probability theory (cf. e.g. Hable (2007C, E-print))
- Rather involved and abstract (uses advanced functional analytic methods)
- "Traditional concepts" ($\sigma$-additivity, Markov-kernels, . . . ): appropriate for small models (dominated by a $\sigma$-finite measure)

**Le Cam's concepts**: also appropriate for large models

**A second conclusion:**

Imprecise probabilities lead to large models

$\rightarrow$ Le Cam’s concepts are appropriate for the theory of imprecise probabilities.

$\rightarrow$ Maybe, they could/should be used further on.
References


  
  www.statistik.lmu.de/~hable/publications.html

  
  www.statistik.lmu.de/~hable/publications.html