ISIPTA'09 July 16th, 2009

Decision Theory under Imprecise Probabilities

Tutorial

Robert Hable
Department of Mathematics
University of Bayreuth

Imprecise Probabilities

Coherent Lower/Upper Previsions (Walley, 1991)

Lower prevision:

$$\underline{P} : \mathcal{L}_\infty(\Omega) \to \mathbb{R}, \qquad f \mapsto \underline{P}[f]$$

Upper prevision:

$$\overline{P} : \mathcal{L}_\infty(\Omega) \to \mathbb{R}, \qquad f \mapsto \overline{P}[f]$$

Credal set

$$\mathcal{M} = \Big\{ P \text{ precise prevision} \,\Big|\, \underline{P}[f] \le P[f] \le \overline{P}[f] \;\; \forall f \in \mathcal{L}_\infty(\Omega) \Big\}$$

Decision Theory

Decision theory	=	Theory of "How to make good decisions, if this is necessary"

Decision Theory

Decision theory	=	Theory of "How to make good decisions, if this is necessary"

- "if this is necessary" → decision problem
- "decisions" → What can be done?
- "good decisions" → Optimality criterion?

Decision Theory

Decision theory = Theory of "How to make good decisions, if this is necessary"

- "if this is necessary" → decision problem
- "decisions" → What can be done?
- "good decisions" → Optimality criterion?

Examples:

- Making an investment in a certain company
- Statistical decision theory:
  - Hypothesis testing: Reject the null hypothesis H_0 or not?
  - Estimation of a parameter

In the following: mathematical formalization of such things

Decisions and States of Nature

Decisions

$$\mathbb{D} \;=\; \text{set of all possible decisions (actions) } t$$

States of nature

- State of nature θ = description of the actual situation
- Whole set of possible states of nature

$$\theta \in \Theta$$

Which $\theta \in \Theta$ is true? $\longrightarrow$ This is unknown!

Example: Decisions and States of Nature

Making an Investment in a Certain Company

- States of nature θ: success of the company
- Decisions t: whether to make an investment

	investment	no investment
Company will be very successful.	excellent	slightly bad
Company will do reasonably well.	okay	okay
Company will collapse.	disastrous	slightly good

Example: Decisions and States of Nature

Making an Investment in a Certain Company

- States of nature θ: success of the company
- Decisions t: whether to make an investment

	investment	no investment
Company will be very successful.	100	-10
Company will do reasonably well.	0	0
Company will collapse.	-200	10

Utility Function

What are the consequences of my decision $t \in \mathbb{D}$?

Utility function:

$$u: \quad \Theta \times \mathbb{D} \ \rightarrow \ \mathbb{R}\,, \qquad (\theta, t) \ \mapsto \ u_\theta(t)$$

Every decision $t \in \mathbb{D}$ leads to a certain utility

$$u_\theta(t) \ \in \ \mathbb{R}$$

depending on the (unknown) state of nature $\theta \in \Theta$

Utility Function

What are the consequences of my decision $t \in \mathbb{D}$?

Utility function:

$$u : \quad \Theta \times \mathbb{D} \;\rightarrow\; \mathbb{R}\,, \quad (\theta, t) \;\mapsto\; u_\theta(t)$$

Every decision $t \in \mathbb{D}$ leads to a certain utility

$$u_\theta(t) \;\in\; \mathbb{R}$$

depending on the (unknown) state of nature $\theta \in \Theta$

- high utility: good
- small utility: bad
- negative utility = positive loss

Alternative: Loss Function

What are the consequences of my decision $t \in \mathbb{D}$?

Loss function instead of utility function

$$W : \ \Theta \times \mathbb{D} \ \rightarrow \ \mathbb{R} \, , \quad (\theta, t) \ \mapsto \ W_\theta(t)$$

Every decision $t \in \mathbb{D}$ leads to a certain loss

$$W_\theta(t) \ \in \ \mathbb{R}$$

depending on the (unknown) state of nature $\theta \in \Theta$

Alternative: Loss Function

What are the consequences of my decision $t \in \mathbb{D}$?

Loss function instead of utility function

$$W: \ \Theta \times \mathbb{D} \ \to \ \mathbb{R}, \quad (\theta, t) \ \mapsto \ W_\theta(t)$$

Every decision $t \in \mathbb{D}$ leads to a certain loss

$$W_\theta(t) \ \in \ \mathbb{R}$$

depending on the (unknown) state of nature $\theta \in \Theta$

Equivalence of loss function and utility function:

$$W_\theta(t) \ = \ -u_\theta(t) \qquad \forall\, t \in \mathbb{D} \quad \forall\, \theta \in \Theta$$

Loss function is used in the following!

What is a good decision?

	investment	no investment
Company will be very successful.	excellent	slightly bad
Company will do reasonably well.	okay	okay
Company will collapse.	disastrous	slightly good

$\longrightarrow$ **Rating depends on the unknown state of nature θ .**

What is a good decision?

	investment	no investment
Company will be very successful.	excellent	slightly bad
Company will do reasonably well.	okay	okay
Company will collapse.	disastrous	slightly good

$\longrightarrow$ **Rating depends on the unknown state of nature θ .**

What can be done?

- If we have a prior distribution (prevision) π for θ :
  $\rightarrow$ Deciding according to the expected loss / Bayes risk

What is a good decision?

Precise prior distribution

- Find a decision $\tilde{t}$ which minimizes the "Bayes risk"

$$\pi[W_{\cdot}(t)] \;=\; \int_{\Theta} W_{\theta}(t)\, \pi(d\theta) \;\in\; \mathbb{R}$$

What is a good decision?

Precise prior distribution

- Find a decision $\tilde{t}$ which minimizes the “Bayes risk”

$$\pi[W_{\cdot}(t)] \;=\; \int_{\Theta} W_{\theta}(t)\,\pi(d\theta) \;\in\; \mathbb{R}$$

Imprecise prior distribution

- leads to a whole interval of “Bayes risks”

$$\big[\underline{\pi}[W_{\cdot}(t)]\,,\,\overline{\pi}[W_{\cdot}(t)]\big] \;\subset\; \mathbb{R}$$

$\longrightarrow$ How to compare overlapping intervals?

$\longrightarrow$ Different possibilities (i.e. different optimality criteria)

Γ-minimax, Γ-minimin and a mixture

Γ-minimax (Γ-maximin)

- Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W.(t)] \;=\; \text{upper Bayes risk}$$

$\rightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

Γ-minimax, Γ-minimin and a mixture

Γ-minimax (Γ-maximin)

- Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W.(t)] \;=\; \text{upper Bayes risk}$$

$\rightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

Popular in robust Bayesian statistics.

Γ-minimax, Γ-minimin and a mixture

Γ-minimax (Γ-maximin)

- Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W.(t)] \;=\; \text{upper Bayes risk}$$

$\rightarrow \tilde{t}_1$ is optimal in the worst case (pessimistic)!

Γ-minimax, Γ-minimin and a mixture

Γ-minimax (Γ-maximin)

- Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W_\cdot(t)] \;=\; \text{upper Bayes risk}$$

$\rightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

Γ-minimin (Γ-maximax)

- Find a decision $\tilde{t}_2$ which minimizes the lower bound

$$\underline{\pi}[W_\cdot(t)] \;=\; \text{lower Bayes risk}$$

$\rightarrow$ $\tilde{t}_2$ is optimal in the best case (optimistic)!

Γ-minimax, Γ-minimin and a mixture

Γ-minimax (Γ-maximin)

- Find a decision $\tilde{t}_1$ which minimizes the upper bound

$$\overline{\pi}[W_{\cdot}(t)] \;=\; \text{upper Bayes risk}$$

$\rightarrow$ $\tilde{t}_1$ is optimal in the worst case (pessimistic)!

Γ-minimin (Γ-maximax)

- Find a decision $\tilde{t}_2$ which minimizes the lower bound

$$\underline{\pi}[W_{\cdot}(t)] \;=\; \text{lower Bayes risk}$$

$\rightarrow$ $\tilde{t}_2$ is optimal in the best case (optimistic)!

and a mixture

- Find a decision $\tilde{t}_3$ which minimizes a convex combination of the lower and the upper bound

$$\alpha\overline{\pi}[W_{\cdot}(t)] + (1-\alpha)\underline{\pi}[W_{\cdot}(t)]$$

$\rightarrow$ compromise between pessimism and optimism

E-admissibility and Maximality

Let

$$\mathcal{P} \;=\; \Big\{\pi \text{ precise prior} \;\Big|\; \underline{\pi}[h] \leq \pi[h] \leq \overline{\pi}[h] \;\; \forall\, h \in \mathcal{L}_\infty(\Theta)\Big\}$$

be the credal set of the imprecise prevision.

E-admissibility

- Find a decision $\tilde{t}_4$ which minimizes

$$\pi[W_\cdot(t)] \;=\; \int_\Theta W_\theta(t)\,\pi(d\theta)$$

for at least one $\pi \in \mathcal{P}$.

E-admissibility and Maximality

Let

$$\mathcal{P} \;=\; \left\{\pi \text{ precise prior} \;\middle|\; \underline{\pi}[h] \leq \pi[h] \leq \overline{\pi}[h] \;\; \forall\, h \in \mathcal{L}_\infty(\Theta)\right\}$$

be the credal set of the imprecise prevision.

E-admissibility

- Find a decision $\tilde{t}_4$ which minimizes

$$\pi[W_\cdot(t)] \;=\; \int_\Theta W_\theta(t)\,\pi(d\theta)$$

for at least one $\pi \in \mathcal{P}$.

Minimality (Maximality)

- Find a decision $\tilde{t}_5$ such that: For every $t \in \mathbb{D}$, there is at least one $\pi_t \in \mathcal{P}$ such that

$$\pi_t[W_\cdot(\tilde{t}_5)] \;\leq\; \pi_t[W_\cdot(t)]$$

Interval dominance

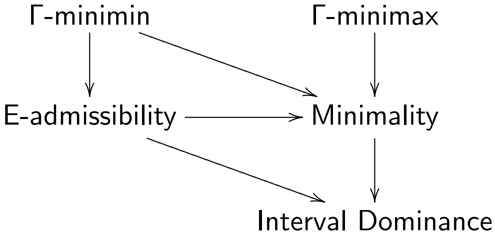
- A decision $\tilde{t}_6$ is optimal with respect to **interval dominance** if

  there is no decision $t \in \mathbb{D}$ such that

$$\overline{\pi}\big[W_\cdot(t)\big] \;\leq\; \underline{\pi}\big[W_\cdot(\tilde{t}_6)\big]$$

Which is the right one?

The following implications hold:



Confer Troffaes (2007).

How to calculate an optimal decision?

- In case of a finite $\Theta = \{\theta_1, \ldots, \theta_m\}$:
  - Calculations by linear programming
  - See Utkin & Augustin (2005) and Kikuti et al. (2005).
- In case of an infinite Θ:
  - Θ can be discretized
  - Decision problems can be approximately solved by linear programming
  - See Troffaes (2008).

Data-Based Decision Theory

Often: Decision making on base of observations is possible

$$\text{Observation / data } x \qquad \longrightarrow \qquad \text{decision } t = \delta(x)$$

Example: Statistical decision theory is always data-based

Data-Based Decision Theory

Often: Decision making on base of observations is possible

$$\text{Observation / data } x \quad \longrightarrow \quad \text{decision } t = \delta(x)$$

Example: Statistical decision theory is always data-based

So, we have

- a **sample space** $(\mathcal{X}, \mathcal{A})$
- an **observation** $x \in \mathcal{X}$
- a **distribution of the observation**

$$x \sim P_\theta \qquad \text{(depending on the state of nature } \theta\text{)}$$

We have to **chose** not only one decision t but **a decision function**

$$\delta : \; \mathcal{X} \to \mathbb{D}\,, \qquad x \mapsto \delta(x)$$

Data-Based Decision Theory

The Bayes risk of a decision function $\delta : x \mapsto \delta(x)$

- **Precise previsions**:

$$\int_{\Theta} \int_{\mathcal{X}} W_\theta(\delta(x)) \, P_\theta(dx) \, \pi(d\theta)$$

Data-Based Decision Theory

The Bayes risk of a decision function $\delta : x \mapsto \delta(x)$

- **Precise previsions**:

$$\int_{\Theta} \int_{\mathcal{X}} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta)$$

- **Imprecise previsions**:

$$\left\{ \int_{\Theta} \int_{\mathcal{X}} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \;\middle|\; \pi \in \mathcal{P}\,,\ \ P_\theta \in \mathcal{M}_\theta \ \ \forall\, \theta \in \Theta \right\}$$

  - $\mathcal{P}$ credal set of an imprecise prevision of θ
  - $\mathcal{M}_\theta$ credal set of an imprecise prevision of the observation x

  $\rightarrow$ imprecise model $\big(\mathcal{X}, \mathcal{A}, (\mathcal{M}_\theta)_{\theta\in\Theta}\big)$

Example: Estimating (Statistical Decision Theory)

Data $x_1, \ldots, x_n$ are modeled by a vector of random variables

$$X = (X_1, \ldots, X_n) \quad \sim \quad P_0$$

P_0 is the true (precise) distribution.

Example: Estimating (Statistical Decision Theory)

Data $x_1, \ldots, x_n$ are modeled by a vector of random variables

$$X = (X_1, \ldots, X_n) \sim P_0$$

P_0 is the true (precise) distribution.

Assumptions: precise case

- $\{ P_\theta \mid \theta \in \Theta \}$ a known parametric set of precise distributions
- There is a true parameter $\theta_0 \in \Theta$ such that $P_0 = P_{\theta_0}$
- There might be a precise prior π on Θ.

Example: Estimating (Statistical Decision Theory)

Data $x_1, \ldots, x_n$ are modeled by a vector of random variables

$$X = (X_1, \ldots, X_n) \sim P_0$$

P_0 is the true (precise) distribution.

Assumptions: precise case

- $\left\{ P_\theta \,\middle|\, \theta \in \Theta \right\}$ a known parametric set of precise distributions
- There is a true parameter $\theta_0 \in \Theta$ such that $P_0 = P_{\theta_0}$
- There might be a precise prior π on Θ .

Decision theoretic formalization:

- States of nature: the possible parameters $\theta \in \Theta$
- Decision function: an estimator S

$$\delta = S \;:\; \mathcal{X} \to \mathbb{D} = \Theta \,, \quad x \mapsto S(x)$$

- Loss function: e.g. the least squares loss

$$W_\theta\big(S(x)\big) = \big(\theta - S(x)\big)^2 \qquad \forall\, \theta \in \Theta$$

Example: Estimating (Statistical Decision Theory)

Data $x_1, \ldots, x_n$ are modeled by a vector of random variables

$$X \;=\; (X_1, \ldots, X_n) \quad \sim \quad P_0$$

P_0 is the true (precise) distribution.

Assumptions: imprecise case

- $\left\{\mathcal{M}_\theta \middle| \, \theta \in \Theta\right\}$ a known parametric set of imprecise distributions
- There is a true parameter $\theta_0 \in \Theta$ such that $P_0 \in \mathcal{M}_{\theta_0}$
- There is a imprecise prior $\mathcal{P}$ on Θ .

Decision theoretic formalization:

- States of nature: the possible parameters $\theta \in \Theta$
- Decision function: an estimator S

$$\delta = S \;:\; \mathcal{X} \;\rightarrow\; \mathbb{D} = \Theta\,, \quad x \;\mapsto\; S(x)$$

- Loss function: e.g. the least squares loss

$$W_\theta\big(S(x)\big) \;=\; \big(\theta - S(x)\big)^2 \qquad \forall\, \theta \in \Theta$$

Example: Estimating (Statistical Decision Theory)

The Bayes risk of an estimator $S : x \mapsto S(x)$

- **Precise previsions**:

$$\int_{\Theta} \int_{\mathcal{X}} \big(\theta - S(x)\big)^2 \, P_\theta(dx) \, \pi(d\theta)$$

- **Imprecise previsions**:

$$\left\{ \int_{\Theta} \int_{\mathcal{X}} \big(\theta - S(x)\big)^2 \, P_\theta(dx) \, \pi(d\theta) \;\middle|\; \pi \in \mathcal{P}, \; P_\theta \in \mathcal{M}_\theta \;\; \forall\, \theta \in \Theta \right\}$$

  - $\mathcal{P}$ credal set of an imprecise prevision of θ
  - $\mathcal{M}_\theta$ credal set of an imprecise prevision of the observation x
    $\rightarrow$ imprecise model $\big(\mathcal{X}, \mathcal{A}, (\mathcal{M}_\theta)_{\theta \in \Theta}\big)$

How to solve data-based decision problems?

In case of precise previsions:

$$\int_\Theta \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta)$$

- **By updating:** Find a decision $\delta(x) \in \mathbb{D}$ which minimizes

$$\int_\Theta W_\theta(\delta(x))\, \pi(d\theta|x) \quad = \quad \text{posterior Bayes risk}$$

- **Without updating:** Find a decision function $\delta : \mathcal{X} \to \mathbb{D}$ which minimizes

$$\int_\Theta \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \quad = \quad \text{Bayes risk}$$

How to solve data-based decision problems?

In case of precise previsions:

$$\int_\Theta \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta)$$

- **By updating:** Find a decision $\delta(x) \in \mathbb{D}$ which minimizes

$$\int_\Theta W_\theta(\delta(x))\, \pi(d\theta|x) \quad = \quad \text{posterior Bayes risk}$$

- **Without updating:** Find a decision function $\delta : \mathcal{X} \to \mathbb{D}$ which minimizes

$$\int_\Theta \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \quad = \quad \text{Bayes risk}$$

Both ways are equivalent!

How to solve data-based decision problems?

In case of imprecise previsions:

$$\left\{ \int_{\Theta} \int_{\mathcal{X}} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \;\middle|\; \pi \in \mathcal{P},\; P_\theta \in \mathcal{M}_\theta \;\; \forall\, \theta \in \Theta \right\}$$

$\longrightarrow$ Again, we have a whole set of (possible) Bayes risks.

$\longrightarrow$ Again, different optimality criteria can be used.

How to solve data-based decision problems?

In case of imprecise previsions:

$$\left\{ \int_{\Theta} \int_{\mathcal{X}} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \;\middle|\; \pi \in \mathcal{P},\; P_\theta \in \mathcal{M}_\theta \;\; \forall\, \theta \in \Theta \right\}$$

$\longrightarrow$ Again, we have a whole set of (possible) Bayes risks.

$\longrightarrow$ Again, different optimality criteria can be used.

The Γ-**minimax** criterion is used in the following.

How to solve data-based decision problems?

In case of imprecise previsions:

- **By updating:** Find a decision $\delta(x) \in \mathbb{D}$ which minimizes

$$\sup_{\pi \in \mathcal{P}_x} \int_{\Theta} W_\theta(\delta(x))\, \pi(d\theta|x) \quad = \quad \text{upper posterior Bayes risk}$$

where $\mathcal{P}_x = \left\{ \pi(\cdot|x) \,\middle|\, \pi \in \mathcal{P},\ P_\theta \in \mathcal{M}_\theta \ \forall\, \theta \in \Theta \right\}$

- **Without updating:** Find a decision function $\delta : \mathcal{X} \to \mathbb{D}$ which minimizes

$$\sup_{\pi \in \mathcal{P}} \int_{\Theta} \sup_{P_\theta \in \mathcal{M}_\theta} \int_{\mathcal{X}} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \ = \ \text{upper Bayes risk}$$

How to solve data-based decision problems?

In case of imprecise previsions:

- **By updating:** Find a decision $\delta(x) \in \mathbb{D}$ which minimizes

$$\sup_{\pi \in \mathcal{P}_x} \int_\Theta W_\theta(\delta(x))\, \pi(d\theta|x) \quad = \quad \text{upper posterior Bayes risk}$$

where $\mathcal{P}_x = \left\{ \pi(\cdot|x) \,\middle|\, \pi \in \mathcal{P},\; P_\theta \in \mathcal{M}_\theta \;\forall\, \theta \in \Theta \right\}$

- **Without updating:** Find a decision function $\delta : \mathcal{X} \to \mathbb{D}$ which minimizes

$$\sup_{\pi \in \mathcal{P}} \int_\Theta \sup_{P_\theta \in \mathcal{M}_\theta} \int_\mathcal{X} W_\theta(\delta(x))\, P_\theta(dx)\, \pi(d\theta) \;=\; \text{upper Bayes risk}$$

Both ways are not equivalent! (See e.g. Augustin (2003).)

Calculations in data-based decision problems?

In case of imprecise previsions:

- **Solution by updating:**
  - In general, it is hard to compute the set of all (possible) posteriors

$$\mathcal{P}_x = \left\{ \pi(\cdot|x) \,\middle|\, \pi \in \mathcal{P},\ P_\theta \in \mathcal{M}_\theta \ \forall\, \theta \in \Theta \right\}$$

  - Once $\mathcal{P}_x = \left\{ \pi(\cdot|x) \,\middle|\, \pi \in \mathcal{P},\ P_\theta \in \mathcal{M}_\theta \ \forall\, \theta \in \Theta \right\}$ is calculated, problem can be treated as a data-free problem.

Calculations in data-based decision problems?

In case of imprecise previsions:

- **Solution by updating:**
  - In general, it is hard to compute the set of all (possible) posteriors

$$\mathcal{P}_x = \left\{ \pi(\cdot|x) \,\middle|\, \pi \in \mathcal{P},\; P_\theta \in \mathcal{M}_\theta \;\forall \theta \in \Theta \right\}$$

  - Once $\mathcal{P}_x = \left\{ \pi(\cdot|x) \,\middle|\, \pi \in \mathcal{P},\; P_\theta \in \mathcal{M}_\theta \;\forall \theta \in \Theta \right\}$ is calculated, problem can be treated as a data-free problem.

- **Solution without updating:**
  - If Θ and $\mathcal{X}$ are finite, calculations by linear programming
  - Computationally costly and, quite often, intractable so far

Other/Further Generalizations of Decision Theory

- **Imprecise loss functions**
- **Fuzzy sets**
- **Dempster-Shafer theory**
- ...

Some References

- J.O. Berger. *Statistical decision theory and Bayesian analysis.* Springer-Verlag, New York, second edition, 1985.
- H. Strasser. *Mathematical theory of statistics. Statistical experiments and asymptotic decision theory.* Walter de Gruyter & Co. Berlin, 1985.
- M.C.M. Troffaes. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45:17-29, 2007.
- T. Augustin. On the suboptimality of the generalized Bayes rule and robust Bayesian procedures from the decision theoretic point of view: a cautionary note on updating imprecise priors. In J.M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *Proceedings of ISIPTA'03*, pages 31-45. Carleton Scientifc, Waterloo, 2003.
- R. Hable. *Data-based decisions under complex uncertainty.* Ph.D. Thesis, LMU Munich, 2009. `http://edoc.ub.uni-muenchen.de/9874/`