On the Information Value of Additional Data and Expert Knowledge in Updating Imprecise Prior Information

Pertisau, September 2002

Thomas Augustin

University of Bielefeld

`thomas@stat.uni-muenchen.de`

www.stat.uni-muenchen.de/ ~ thomas

- What is updating?
- How to learn from data? (inference)
- How to make optimal decisions?

- What's the value of additional information?

1. Updating and Bayesian Statistics

- Two paradigms

[P1] Every uncertainty can adequately be described by a classical probability distribution $\Rightarrow$ prior distribution $\pi(\cdot)$ for a parameter

[P2] After having observed the sample $\{x\}$, the posterior $\pi(\cdot|x)$ contains all the relevant information. Every inference procedure depends on $\pi(\cdot|x)$, and only on $\pi(\cdot|x)$.

- Paradigm $[P2]$ can be justified by decision theoretic arguments: Decision functions with minimal risk under prior $\pi(\cdot)$ can be constructed from considering optimal actions with respect to $\pi(\cdot|x)$ as 'updated prior'.

optimality with respect to prior risk

=

optimality with respect to posterior loss

inference	=	decision

Aim of the talk

- General: develop a comprehensive framework for decision making under partial prior information (imprecise prior instead of [P1])

- In particular, study a straightforward adoption of [P2] used in sensitivity analysis, in the robust Bayesian approach and in imprecise probability theory (Walley's generalized Bayes rule)

1. Updating and Bayesian Statistics

2. Classical Decision Theory

3. Decision Making under Interval Probability – Basic Concepts

4. The Robust Bayesian Approach/Walley's GBR

5. How to Calculate Decision Functions Minimizing Prior Risk?

6. Concluding Remarks

2. Classical Decision Theory

The Basic Decision Problem

no-data problem (on finite spaces)

- set $I\!A = \{a_1, \ldots, a_s, \ldots, a_n\}$ of *actions*,
- possibly $\Lambda(I\!A)$ set of randomized actions

$$a(\cdot) = (\lambda(a_1), \ldots, \lambda(a_s), \ldots, \lambda(a_n))$$

- set $\Theta = \{\vartheta_1, \ldots, \vartheta_j, \ldots, \vartheta_m\}$ of *states* of nature
- precise *loss function*

$$\begin{array}{rcccc} l & : & (I\!A \times \Theta) & \rightarrow & \mathbb{R} \\ & & (a, \vartheta) & \mapsto & l(a, \vartheta) \end{array} \quad ,$$

- represented in an *loss table*

	ϑ_1	ϑ_j	ϑ_m
a_1	$l(a_1, \vartheta_1) \ldots$	$l(a_1, \vartheta_j) \ldots$	$l(a_1, \vartheta_m)$
	$\vdots$	$\vdots$	$\vdots$
a_s	$l(a_s, \vartheta_1) \ldots$	$l(a_s, \vartheta_j) \ldots$	$l(a_s, \vartheta_m)$
	$\vdots$	$\vdots$	$\vdots$
a_n	$l(a_n, \vartheta_1) \ldots$	$l(a_n, \vartheta_j) \ldots$	$l(a_n, \vartheta_m)$

- associated random variable $\mathbf{l}(a)$ on $(\Theta, \mathcal{P}o(\Theta))$
- **Aim:** Choose an optimal action a^*!

Data problem

- Incorporate additional information from a sample!
- Choose an optimal strategy!
- What is the value of a certain information?

Information on ϑ_j from an experiment where the probability depends on ϑ_j:

For every j a classical probability $p_j(\cdot)$ is given

	$x_1 \ldots$	$x_i \ldots$	x_k
ϑ_1	$p_1(\{x_1\}) \ldots$	$p_1(\{x_i\}) \ldots$	$p_1(\{x_k\})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
ϑ_j	$p_j(\{x_1\}) \ldots$	$p_j(\{x_i\}) \ldots$	$p_j(\{x_k\})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
ϑ_m	$p_m(\{x_1\}) \ldots$	$p_m(\{x_i\}) \ldots$	$p_m(\{x_k\})$

Often $p_j(\{x_i\})$ is interpreted as $p(\{x_j\}|\{\vartheta_j\})$.

decision functions (strategies)

- describing randomized action in dependence on the observation $\{x_i\}$

$$\begin{array}{rcl} d: \; \{x_1, \ldots, x_k\} & \rightarrow & \Lambda(\mathbb{A}) \\ x_i & \mapsto & d(x_i) = a \; . \end{array}$$

- randomized decision functions $d(x_i, a_s)$; classical probability to choose a_s if $\{x_i\}$ occurs.
- $\mathcal{D}$ set of all decision functions
- associated random variable $\mathbf{l}(d, \vartheta_j)$ on $(\Omega, \mathcal{A})$
- risk of $d(\cdot)$

$$R(d, \vartheta_j) := \mathbb{E}_{p_j}\left(\mathbf{l}(d, \vartheta_j)\right).$$

- New decision problem $(\mathcal{D}, \Theta, R(\cdot, \cdot))$.

The value of the information experiment

loss of the optimal action in the no-data problem — risk of the optimal decision function in the data problem = value of information

Always nonnegative.

Optimality criteria

1) Minimax optimality

- In the no-data problem: $\max_{\vartheta \in \Theta} l(a, \vartheta) \to$ min
- In the data problem: $\max_{\vartheta \in \Theta} R(d, \vartheta) \to$ min

2) Bayes optimality with respect to prior $\pi(\cdot)$ on $(\Theta, \mathcal{P}o(\Theta))$.

- In the no-data problem: $\mathbb{E}_\pi(l(a, \vartheta)) \to$ min
- In the data problem: $\mathbb{E}_\pi(R(d, \vartheta)) \to$ min

"Main theorem of Bayesian decision analysis"

- Optimal $d^*(\cdot)$ can be obtained by solving, for every observation $\{x\}$, the no-data problem with the posterior $\pi(\cdot|x)$ as the 'updated prior'.

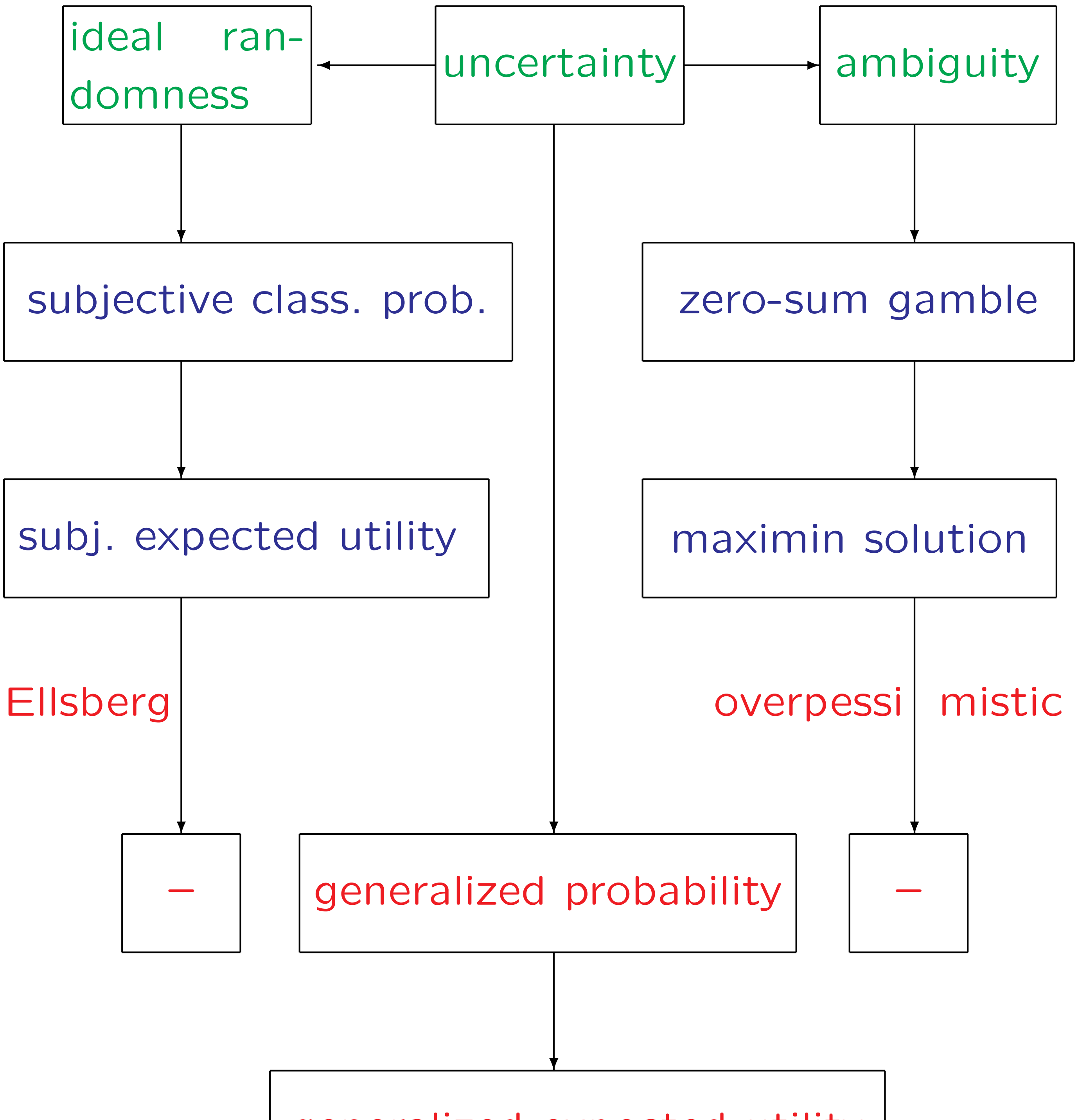
  optimality with respect to prior risk

  =

  optimality with respect to posterior loss

- For maximin solutions NO reduction of the data problem to no-data problems possible.

3. Decision Making under Interval Probability – Basic Concepts



Ellsberg's Experiments

- Ellsberg's (1961, Quart. J. Econ.)

- Ellsberg (2002, Series of most influential Harvard theses)

- Does the difference between an ideal lottery situation and the general decision situation under uncertainty matter?

- Urn with balls of three different colours: one with known proportion, two with partially unknown proportions

- participants express preferences which can not be modelled by any classical probability measure

- deliberate (not only empirical!) violations of the axioms of (classical) probability!

- Conclusion: (Classical) probability is insufficient to adequately model ambiguous uncertainty.

Ambiguity

- Ellsberg (1961, Quart. J. Econ.)
  Ellsberg (2002, Series of most influential Harvard theses)
- in psychology, management science and economics
  * bibliography: Smithson (1999, tech. report)
  * collection of important papers: Hamouda & Rowley (1997, Edward Elgar)
  * in principle even in Knight (1921) & Keynes (1921)
- in statistics
  * Walley (1991, Chap. & Hall, Ch. 5)
  * Weichselberger (2001, Physica, Ch. 1, 2.6)
  * ISIPTA Proceedings (1999, 2001)
  * Special volumes Statistical Papers (2002), J.Stat.Plan.Inf. (2002)
- in artificial intelligence
  * Uncertainty in Artificial Intelligence Proceedings (Annual)

Basic decision theoretic framework

$$+$$

Generalized concept of probability to model ambiguous uncertainty

$$=$$

General framework for decision making under ambiguous uncertainty

`www.stat.uni-muenchen.de/ ~ thomas`

Generalized Probabilities to Handle Ambiguity

Probability and uncertainty as a two-dimensional phenomenon: (ideal type separation of two overlapping phenomena)

uncertainty

=

ideal randomness + ambiguity

Mainly two approaches

- sets of classical probabilities
- for every event A interval $[L(A); U(A)]$
  $\Rightarrow$ non-additive set functions $L(\cdot)$ and $U(\cdot)$

  The interval width reflects the extent of ambiguity

  * $P(A) = [a; a]$: classical probability, situation of ideal randomness

    $\vdots$ increasing ambiguity

  * $P(A) = [0; 1]$: complete ignorance

Axiomizing interval probability

- Classical probability $p(\cdot)$: set-function satisfying Kolmogorov's axioms

- interval probability $P(\cdot) = [L(\cdot), U(\cdot)]$

- look at the *relation* between the dual pair of non-additive set functions $L(\cdot)$ and $U(\cdot)$ and the structure $\mathcal{M}$, i.e, the set of all classical probabilities $p(\cdot)$ compatible with $[L(\cdot), U(\cdot)]$

  $$\mathcal{M} := \{p(\cdot) \mid L(A) \leq p(A) \leq U(A), \quad \forall A \in \mathcal{A}\}.$$

- Here: consider only assignments with:

  $$\mathcal{M} \neq \emptyset \qquad \text{and}$$

  $$\inf_{p(\cdot) \in \mathcal{M}} p(A) = L(A) \wedge \sup_{p(\cdot) \in \mathcal{M}} p(A) = U(A), \forall A \in \mathcal{A}.$$

- * lower and upper probability (Huber & Strassen (1973, Ann. Stat.))

  * envelopes (Walley & Fine (1982, Ann. Stat.), Denneberg (1994, Kluwer))

  * coherent probability (Walley (1991, Chap. & Hall))

  * F-probability (Weichselberger (2000, Int. J. Approx. Reas.; 2001, Physika))

- $\mathcal{M}$: *structure* (Weichselberger), (in co-operative

- slightly more general: Walley (1991, Chapm. & Hall): imprecise previsions obtained from interval-valued expectations = linear partial information (Kofler & Menges (1976, SpringerLN Econ); Huschens (1985, R.G. Fischer); Kofler (1989, Campus))

- **Special Case: Capacities of Higher Order**

  * Belief-functions (totally monotone probabilities), corresponding to a basic probability assignment (Shafer (1976, Princeton UP), Yager, Fedrizzi und Kacprzyk (1994, Wiley))

  * Neighborhood models in robust statistics (pseudo capacities, Choquet-capacities)(Huber (1981, Wiley), for a survey (and extensions): Augustin (2001, J. Stat. Plan. Inf.))

  * Probability intervals (PRI) (Weichselberger & Pöhlmann (1990; Springer LN AI))

  * Other common names 'supermodular' (Denneberg (1994; Kluwer)) or 'convex' (Jaffray (1989, OR Letters))

Expectation

Classical expectation for $X(\cdot) \geq 0$:

$$\mathbb{E}X = \int X dp = \int p(\{\omega \mid X(\omega) > t\})\, dt$$

Two possible ways to generalize this for F-probability $P(\cdot) = [L(\cdot), U(\cdot)]$ with structure $\mathcal{M}$

- "outer method": substitute $p(\cdot)$ by $L(\cdot)$ and $U(\cdot)$ (*Choquet integral, fuzzy integral*)

$$\mathbb{E}_L X := \int_0^\infty L(\{\omega \mid X(\omega) > t\})\, dt\,.$$

- "inner method": refers to the structure; considers $\inf_{p(\cdot)\in\mathcal{M}}$ and $\sup_{p(\cdot)\in\mathcal{M}}$ (here in what follows)

$$\begin{aligned}\mathbb{E}_{\mathcal{M}} X &:= \left[{}^{\mathsf{L}}\mathbb{E}_{\mathcal{M}} X, {}^{\mathsf{U}}\mathbb{E}_{\mathcal{M}} X\right] \\ &:= [\inf_{p(\cdot)\in\mathcal{M}} \mathbb{E}_p X\,, \sup_{p(\cdot)\in\mathcal{M}} \mathbb{E}_p X]\end{aligned}$$

Theorem (e.g, Denneberg (1994, Kluwer, Prop. 10.3)):
In the case of two-monotone probability both definitions coincide.

Therefore: **In the case of two-monotonicity everything said here is also valid for the Choquet integral.**

Basic decision theoretic framework

$$+$$

Generalized concept of probability to model ambiguous uncertainty

$$=$$

General framework for decision making under ambiguous uncertainty

- classical decision theory can only deal with the two – both unrealistic – extreme cases
  - complete probabilistic knowledge (i.e knowledge of a single classical probability) or
  - complete ignorance

+ Now modeling of arbitrary intermediate steps becomes possible. The true level of knowledge can be adequately represented.

Generalized Expected Utility/Loss

- **Def.: Generalized expected loss**
  - * basic decision problem $(\mathbb{A}, \Theta, l(\cdot))$
  - * F-probability $\Pi(\cdot)$ on $(\Theta, \mathcal{P}_o(\Theta))$ with structure $\mathcal{M}$.

  Then, for every pure action $a \in \mathbb{A}$ and for every randomized action $a \in \Lambda(\mathbb{A})$, resp.,

  $$\mathbb{E}_{\mathcal{M}}\mathbf{l}(a)$$

  is the *generalized expected loss* (with respect to the prior $\Pi(\cdot)$).

- Notice: $\mathbb{E}_{\mathcal{M}}\mathbf{l}(a)$ is an interval-valued quantity. If a linear ordering is desired $\longrightarrow$ *representation*

- easiest choice: $\mathbb{E}_{\mathcal{M}}\mathbf{l}(a) \mapsto {}^{\mathsf{U}}\mathbb{E}_{\mathcal{M}}\mathbf{l}(a)$
  * a^* is optimal iff

$$ {}^{\mathsf{U}}\mathbb{E}_{\mathcal{M}}\mathbf{l}(a^*) \leq {}^{\mathsf{U}}\mathbb{E}_{\mathcal{M}}\mathbf{l}(a)\,, \quad \forall a \in \Lambda(\mathbb{A})\,. $$

  * strict ambiguity aversion
  * Gamma-Minimax criterion (e.g., Berger (1984, Springer, Section 4.7.6), Vidakovic (2000, in Rios-Insua & Ruggeri (eds.)),
  * Maxmin expected utility model (Gilboa & Schmeidler (1989, J. Math. Econ.))
  * MaxEMin criterion (Kofler & Menges (1976, SpringerLN Econ); Kofler (1989, Campus))
  * For two-monotone capacities: Choquet expected utility (e.g., Chateauneuf, Cohen & Meilijson (1997, Finance))

The two classical decision criteria are contained as border cases:

  * perfect probabilistic information, no ambiguity: $\mathcal{M} = \{\pi(\cdot)\} \longrightarrow$ Bayes optimality with respect to $\pi(\cdot)$.
  * Completely lacking information, $\Pi(B) = [0;1]$, for every $B \in \mathcal{P}o(\Omega)\backslash\{\emptyset, \Theta\}$, ('*non-selective* or *vacuous* prior'); leads to the

Data problem under interval probability:

- Now, for every j, an F-probability field
$$P_j(\cdot) = [L_j(\cdot), U_j(\cdot)]$$
with structure $\mathcal{M}_j$ is given.

- risk of the decision function $d(\cdot)$
  * Given ϑ_j : $\mathbb{E}_{\mathcal{M}_j}\Big(\mathbf{l}(d, \vartheta_j)\Big)$ represented by
$$\mathbf{R}(d) := {}^{\mathsf{U}}\mathbb{E}_{\mathcal{M}_j}\Big(\mathbf{l}(d, \vartheta_j)\Big)$$
  * with prior structure $\mathcal{M}$ look at
$$\mathbb{E}_{\mathcal{M}}\left(\mathbf{R}(d)\right)$$
represented by
$${}^{\mathsf{U}}\mathbb{E}_{\mathcal{M}}\left(\mathbf{R}(d)\right)$$

The value of the information experiment

loss of the optimal action
in the no-data problem
—
risk of the optimal decision function
in the data problem
=
value of information

Still always nonnegative

4. Robust Bayesian Procedures
Generalized Bayes Rule

- classical statistics:

  data problem with prior $\pi(\cdot)$

  $$\equiv$$

  no-data problem with updated prior $\pi(\cdot|x)$

  $\Rightarrow$ posterior contains full information

- Generalization: Robust Bayesian Inference (Survey: Wasserman (1997, Enc. Stat. Sc., Update 1))

  $$\begin{array}{rl} & \text{prior structure } \mathcal{M} \\ + & \text{observation } x \\ \hline & \text{posteriori structures } \mathcal{M}_{|x} \end{array}$$

  with

  $$\mathcal{M}_{|x} = \{\pi(\cdot|x) | \pi(\cdot) \in \mathcal{M}\}$$

  and $\Pi(\cdot|x) = [\underline{\pi}(\cdot|x), \overline{\pi}(\cdot|x)]$ derived from it.

- Used in Kofler & Menges' (1976) theory of partial information
- Strong justification by coherence axioms (Walley (1991): Generalized Bayes Rule)
- intuitively very plausible
- elegant modelling of prior-data conflict (Walley (1991, Ch. 1))
- successive updating: use $\Pi(\cdot|x)$ as a new prior in handling new observations

BUT

- Decision theoretic justification is lost.

- Decision functions constructed via the posterior structure may have higher risk.

- optimality with respect to imprecise prior risk

$$\neq$$

optimality with respect to imprecise posterior loss

- $\stackrel{?}{\Longrightarrow}$ The imprecise posterior does not contain all the relevant information !?!?

WHY?

Decision functions constructed via the posterior structure may have higher risk

- *First (counter)example:* Vacuous prior("$\Pi(\cdot) = [0, 1]$")
  - $*$ Minimax decision function $d^*(\cdot)$ minimizes prior risk.
  - $*$ Vacuous posterior for every observation (we do not learn from the data!); minimax action a^* minimizes posterior loss for every observation
  - $*$ Usually $d^* > (a^*, \dots, a^*)$

- Representation theorem: Optimal decision functions with respect to an imprecise prior $\pi(\cdot)$ are always minimax solutions (in a different decision problem)

- Imprecise posteriors may be dilated (Seidenfeld & Wasserman (1993, Ann.Statist.)) This leads often to a negative value of information.

$$[\underline{\pi}(\cdot|x), \overline{\pi}(\cdot|x)] \supset [\underline{\pi}(\cdot), \overline{\pi}(\cdot)]\,, \ \forall x$$

Representation Theorem: Optimal decision functions with respect to an imprecise prior $\pi(\cdot)$ are always minimax solutions (in a different decision problem):

Consider

- a basic decision problem $(\mathbb{A}, \Theta, l(\cdot,\cdot))$ with
- prior structure $\mathcal{M}$ and
- (precise) sampling information $(p_\vartheta(\cdot))_{\vartheta \in \Theta}$

i) An action a^* is optimal optimal with respect to the prior structure $\mathcal{M}$

iff

it is maximin action in the decision problem $(\mathbb{A}, \mathcal{M}, \tilde{l}(\cdot,\cdot))$ with

$$\begin{array}{rccc} \tilde{l} : & (\mathbb{A} \times \mathcal{M}) & \rightarrow & \mathbb{R} \\ & (a,\, \pi) & \mapsto & \tilde{l}(a,\, \pi) := \mathbb{E}_\pi(l(a, \vartheta) \end{array}$$

ii) A decision function $d^*(\cdot)$ is optimal

iff

$d^*(\cdot)$ is maximin decision function in the decision problem $(\mathcal{D}, \mathcal{M}, \tilde{R}(\cdot,\cdot))$ with

$$\begin{array}{rccc} \tilde{R} : & (\mathcal{D} \times \mathcal{M}) & \rightarrow & \mathbb{R} \\ & (d,\, \pi) & \mapsto & \tilde{R}(d,\, \pi) := \mathbb{E}_\pi(R(d, \vartheta))\,. \end{array}$$

Proof:

$$\underbrace{\max_{\pi(\cdot) \in \mathcal{M}}}_{\uparrow} \underbrace{\mathbb{E}_\pi(l(a, \vartheta))}_{\uparrow}$$

Remarks

- Optimal decision functions have all the ((un)pleasant) properties of minimax solutions.

- Neither

  * equivalence of posterior loss and prior risk

  nor

  * essentially completeness of unrandomized actions (also for robust Bayesian solutions!)

  can be expected.

- Representation similar to Schneeweiß's (1964) representation of a no-data problem.

- Extensions to interval-valued sampling model and Hurwicz-like criterion.

- Framework for decision making with second order probabilities.

5. How to Calculate Decision Functions Minimizing Prior Risk?

- Vidakovic (2000, in Rios-Insua & Ruggeri (eds.))

- Noubiap & Seidel (2001, Comp. Stat.& Data Anal.), (2001, Ann. Stat)

- On finite parameter spaces solution via a single linear programming problem available (Augustin (2001, ISIPTA-Cornell))

Consider finite sample spaces. Minimize

$$^{\mathsf{U}}\mathbb{E}_{\mathcal{M}}\left(^{\mathsf{U}}\mathbb{E}_{\mathcal{M}_j}(l(d,\,\vartheta_j))\right) = \max_{\pi\in\mathcal{M}}\left(\sum_{j=1}^{m}\max_{p_j(\cdot)\in\mathcal{M}_j}\right.$$

$$\left(\sum_{i=1}^{k}\left(\sum_{s=1}^{n}\underbrace{l(a_s;\,\vartheta_j)}_{\text{given}}\cdot\underbrace{d(x_i;\,a_s)}_{\text{unknown}}\right)\cdot\underbrace{p_j(\{x_i\})}_{\substack{\in\mathcal{M}_j\\ \text{constr.}}}\right)\cdot\underbrace{\pi(\{\vartheta_j\})}_{\substack{\in\mathcal{M}_\pi\\ \text{constr.}}}\Bigg)$$

w.r.t. $\sum_{j=1}^{n} d(x_i,\,a_s) = 1\,, d(x_i,\,a_s) \geq 0\ \forall i,\,s.$

- Make this problem linear: auxiliary variables g for $^{\mathsf{U}}\mathbb{E}_{\mathcal{M}}$, as well as g_j for $^{\mathsf{U}}\mathbb{E}_{\mathcal{M}_j}$.

$$g \longrightarrow \min \qquad \text{with respect to}$$

$$\sum_{i=1}^{k}\left(\sum_{s=1}^{n} l(a_s,\,\vartheta_j)\cdot d(x_i,\,a_s)\right)\cdot p_j(\{x_i\}) \;\leq\; g_j$$

$$\forall p_j(\cdot)\in\mathcal{M}_j;\ \ \forall j\in\{1,\,\ldots,\,m\}$$

$$\sum_{j=1}^{m} g_j\cdot\pi(\{\vartheta_j\}) \;\leq\; g$$

$$\forall\pi(\cdot)\in\mathcal{M}$$

$$\sum_{j=1}^{n} d(x_i,\,a_s) = 1\,,\ \ d(x_i,\,a_s) \;\geq\; 0\ \forall i,\,s\,.$$

- objective function and constraints are linear in a finite number of variables, but still NO linear programming problem. $\mathcal{M}$ and $\mathcal{M}_j$ are uncountable!

Some Properties of Structures on Finite Sample Spaces:

- $\mathcal{M}$ is a convex polyhedron.
  * $\mathcal{M}$ is closed.
  * The set $\mathcal{E}(\mathcal{M})$ of the **extreme points (vertices)** is non-empty, finite, and it uniquely determines $\mathcal{M}$.

- Treatment of typical problems of interval probability with linear programming: Weichselberger (1996, Huber-Festschrift).

- **Calculation of $\mathcal{E}(\mathcal{M})$:**
  * Algorithm from the theory of convex polyhedra. (Intersection of k hyperplains)
  * For two monotone and totally monotone probability closed form available:

$$\mathcal{E}(\mathcal{M}) = \{p_\varsigma(\cdot) \mid \varsigma \in \Upsilon\}$$

with

$$p_\varsigma(\{\omega_{\}}) = L\left(\bigcup_{j=1}^{i} \omega_{\varsigma(j)}\right) - L\left(\bigcup_{j=1}^{i-1} \omega_{\varsigma(j)}\right),$$

for all $i = 1, \ldots, k$ and Υ as the set of all permutations of $\{1, \ldots, k\}$.

Lemma $P(\cdot) = [L(\cdot), U(\cdot)]$ F-probability with structure $\mathcal{M}$ and extreme points $\mathcal{E}(\mathcal{M})$.

$$\mathbb{E}_{\mathcal{M}} X = \left[\min_{p(\cdot)\in\mathcal{E}(\mathcal{M})} \mathbb{E}_p X \; ; \; \max_{p(\cdot)\in\mathcal{E}(\mathcal{M})} \mathbb{E}_p X \right] .$$

Corollary (Vertice reduction lemma)
For every real g,

$$^{\mathsf{U}}\mathbb{E}_{\mathcal{M}} X \leq g \Longleftrightarrow \mathbb{E}_p X \leq g, \quad \forall p(\cdot) \in \mathcal{E}(\mathcal{M}) .$$

Use the vertice reduction lemma to reformulate the task as a linear optimization problem:

$$g \longrightarrow \min$$

under the constraints

$$\sum_{i=1}^{k} \left(\sum_{s=1}^{n} l(a_s, \vartheta_j) \cdot d(x_i, a_s) \right) \cdot p_j(\{x_i\}) \leq g_j$$

$$\forall p_j(\cdot) \in \mathcal{E}(\mathcal{M}_j); \ \forall j \in \{1, \dots, m\}$$

$$\sum_{j=1}^{m} g_j \cdot \pi(\{\vartheta_j\}) \leq g$$

$$\forall \pi(\cdot) \in \mathcal{E}(\mathcal{M})$$

$$\sum_{j=1}^{n} d(x_i, a_s) = 1 \, , \ d(x_i, a_s) \geq 0 \ \forall i, s \, .$$

- Single linear programming problem
- Easy calculation of optimal decision functions
- Easy proof of existence of solutions and of the convexity of the set of optimal decision functions
- further insights by dualization!
- also for Choquet Expected Utility in case of two-monotone capacities or belief functions
- optimal unrandomized actions by integer

6. Further Work

- Detailed understanding of the problems of Robust Bayesianism.
  How to 'update' in decision making?
- Extension to infinite sample spaces using results by Rüger, Utkin
- Apply dualization:
  * $\rightarrow$ least favorable constellations
  * for hypothesis testing:
    Generalization of the Generalized Neyman-Pearson Lemma (Augustin (1998, Vandenh. & R., Ch. 5))
- Use more sophisticated interval ordering to model general ambiguity attitudes (for the no-data problem: Augustin (2002, Stat. Papers))
- sequential decision making, but be careful!
  * backward induction: Hermanez (1999, ISIPTA99-Ghent), de Cooman (2002, Workshop Munich)
  * 'sophisticated versus step by step optimal': Jaffray (2002, J. Stat. Plan. Inf.)
  * for sequential testing: Augustin & Pöhlmann

- What is updating?

- How to learn from data? (inference)

- How to make optimal decisions?

- Does $[\underline{\pi}(\cdot|x), \overline{\pi}(\cdot|x)]$ deserve to be called posterior, since

  - it does not contain the full information from a sample and

  - it leads to suboptimal decisions ?

- Implicit definition of posterior ??

or

- Separate updating/inference and decision in an uncompromising way !

- But check for potential paradoxes (statistical estimating and testing problem can be formulated as inference as well as decision problems.)