On Decision Making under Ambiguous Prior and Sampling Information

Th. Augustin
Department of Statistics, University of Munich,
thomas@stat.uni-muenchen.de

Abstract

The standard framework of decision theory suffers from the fact that ambiguity (non-stochastic uncertainty, indeterminacy) can not be taken properly into account. By generalizing expected utility theory to interval probability and imprecise probabilities, this paper extends classical decision theory to situations where prior and/or sampling information may be ambiguous. Firstly, a general algorithm will be presented to calculate optimal actions under ambiguous prior information. Then it is shown how to incorporate ambiguous sampling information. As a by-product, the results also lead to a question concerning coherent inference with imprecise probabilities and robust Bayesian inference. It will be shown that, under ambiguity, inference based on the imprecise posterior distribution may lead to suboptimal actions; under ambiguity, the posterior distribution does no longer contain all the relevant information.

Keywords. Decision making, generalized expected utility, ambiguity, imprecise probabilities, interval probability, linear programming, Choquet expected utility, Γ-minimax principle, critic of robust Bayesian analysis

1 Introduction

Consider the basic model of decision analysis: One has to choose an *action* from of a non-empty, finite set $\mathbb{A} = \{a_1, \ldots, a_s, \ldots, a_n\}$ of possible actions. The consequences of every action depend on the true, but unknown *state* of nature $\vartheta \in \Theta = \{\vartheta_1, \ldots, \vartheta_j, \ldots, \vartheta_m\}$. The corresponding outcome is described by the *utility function*

$$\begin{array}{rccc} u & : & (\mathbb{A} \times \Theta) & \to & \mathbb{R} \\ & & (a, \vartheta) & \mapsto & u(a, \vartheta) \end{array}$$

and by the associated random variable $\mathbf{u}(\mathbf{a})$ on $(\Theta, \mathcal{P}o(\Theta))$ taking the values $u(a, \vartheta)$.

This model contains the essentials of every (formalized) decision situation under uncertainty and is applied in a huge variety of disciplines. For instance, in medicine the states may be diseases, and then the actions are different treatments. In management science, the actions can be different types of investments; the states of nature then may be certain economic constellations.

In most cases, decisions can be improved by using data (e.g., performing a medical test or administering a questionnaire to experts on the expected economic development). Therefore, later on, the case of incorporating additional sampling information will also be studied in some detail.

If the states of nature are produced by a perfect random mechanism (e.g. an ideal lottery), and the corresponding probability measure $\pi(\cdot)$ on $(\Theta, \mathcal{P}o(\Theta))$ is completely known, the Bernoulli principle is nearly unanimously favored. One chooses that action a^* which maximizes the expected utility $\mathbb{E}_p \mathbf{u}(a) = \sum_{j=1}^{m} u(a, \vartheta_j) \cdot \pi(\{\vartheta_j\})$ among all $a \in \mathbb{A}$.

As in the examples mentioned above, in most practical applications, however, the true state of nature can not be understood of arising from an ideal random mechanism. And even if so, the corresponding probability distribution will be not known exactly. There are two main directions to proceed in this situation.

Since for a classical subjectivist, or Bayesian, every situation under uncertainty can be described by a single, precise probability measure $\pi(\cdot)$, the lack of such a known random mechanism does not make any important difference to the decision maker. (S)he acts according to *subjective expected utility*. $\pi(\cdot)$ then often is called *prior probability*, and the optimal actions are called *Bayes actions* with respect to $\pi(\cdot)$.

In contrast, from the viewpoint of an 'objectivist' it does not make any sense at all to think now of a probability on $(\Theta, \mathcal{P}o(\Theta))$. Therefore, the objectivist concludes, that the decision maker is completely ignorant about which state of nature will occur; (s)he has to act according to a criterion based on complete ignorance (like the *maximin rule*).

It has often been complained that both ways to proceed, relying on subjective expected utility as well as acting according to a criterion based on complete ignorance, are inappropriate, because they both distort the *partial* nature of the knowledge on the decision maker's hand: The objectivist's criteria treat partial knowledge like complete ignorance, often leading to unsatisfactory solutions. Subjective utility theory equates identifies partial knowledge with complete probabilistic knowledge. This conflicts with Ellsberg's (1961) experiments, which made it perfectly clear that ambiguity (i.e. the deviation from ideal stochasticity)[1] plays a constitutive role in decision making — neglecting it may lead to deceptive conclusions.

Imprecise probabilities and related concepts are understood to be provide a powerful language which is able to reflect the partial nature of the knowledge suitably and to express the amount of ambiguity adequately. (See deCooman & Walley (2001) and Weichselberger (2001, Ch. 1) for recent reviews on the development in this field.) The present paper studies the question how to extend the standard framework of decision analysis to the generalized probabilistic framework. Section 2 begins with some basic definitions and technical remarks on ambiguity representation by interval probability and imprecise probabilities. The case of generalized expected utility under imprecise prior information is studied in Section 3. Then, in Section 4, these results are extended to incorporate sampling information. Section 5 briefly discusses the surprising fact that robust Bayesian solutions prove to be suboptimal in the generalized setting.

For the whole argumentation linear programming plays an essential role. On the one hand, it is powerful for efficient calculation of optimal solutions, on the other hand, it provides valuable theoretical insights. Since two-monotone and totally monotone capacities are special cases of general interval probability, where Choquet integral and interval-valued expectation correspond with each other, the results also show, as a welcome by-product, how to deal efficiently with Choquet Expected Utility and how to perform a neat decision analysis in the case of belief functions (interpreted in a probabilistic way).

[1] It must be noted that there is a plenty of ambiguity connected with the use and the concrete meaning of the term 'ambiguity'. While in many applied as well as theoretic work the relevance of the concept is not doubted, the terminology varies. Here the term is used as in the original work by Ellsberg (1961), where the term was introduced to denote imprecision in the probabilistic statements. In contrast, Budescu, Weinberg & Wallsten (1988) give some good reason to use the term ambiguity for imprecision in the events, and to speak of 'vague (or imprecise) probabilities' – a notation which, however, conflicts with the terminology typically used in artificial intelligence (e.g., Kruse, Schwecke & Heinsohn (1991)).

2 Ambiguity Representation by Interval Probabilities and Imprecise Probabilities

2.1 Basic Terminology

A very natural idea to represent ambiguity is to describe the probability of events or the expectation of variables by intervals instead of single real numbers. The difference between upper and lower interval limits provides a means to describe the amount of ambiguity — the higher the ambiguity, the larger this difference. In the extreme case of complete ignorance (full ambiguity) the interval $[0;1]$ is appropriate as the probability of any non-trivial event, while under ideal probabilistic information without any ambiguity this interval shrinks to a single point. So, the usual formalization of probability as a real number is contained as a special border case, which is only appropriate if there is perfect information on the stochastic behaviour.

With respect to the intended application the whole consideration is restricted here to the case of finitely generated algebra $\mathcal{A}$ based on a sample space Ω. Then, without loss of generality, Ω is finite, and $\mathcal{A}$ is the power set of $\Omega = \{\omega_1, \ldots, \omega_k\}$.

To distinguish in terminology, every probability measure in the usual sense, i.e. every set function $p(\cdot)$ satisfying Kolmogorov's axioms is called a *classical probability.* The set of all classical probabilities on the measurable space $(\Omega, \mathcal{A})$ will be denoted by $\mathcal{C}(\Omega, \mathcal{A})$.

Axioms for interval-valued probabilities $P(\cdot) = [L(\cdot), U(\cdot)]$ can be obtained by looking at the relation between the non-additive set-function $L(\cdot)$ and $U(\cdot)$ and the set of classical probabilities being in accordance with them. On a finite sample space, as considered throughout this paper, several concepts of interval probability coincide. They all are concerned with set-functions

$$\begin{array}{rcl} P(\cdot): \mathcal{A} & \rightarrow & \mathcal{Z}_0 := \{[L,U] \,|\, 0 \le L \le U \le 1\} \\ A & \mapsto & P(A) = [L(A), U(A)] \end{array}$$

with

$$\mathcal{M} := \{\, p(\cdot) \in \mathcal{C}(\Omega, \mathcal{A}) \,| \\ L(A) \le p(A) \le U(A),\ \forall A \in \mathcal{A}\} \neq \emptyset\,. \qquad (1)$$

and

$$\left.\begin{array}{l}\inf\limits_{p(\cdot)\in\mathcal{M}} p(A) = L(A)\\ \sup\limits_{p(\cdot)\in\mathcal{M}} p(A) = U(A)\end{array}\right\} \quad \forall A \in \mathcal{A}\,. \qquad (2)$$

Such $P(\cdot)$, and the corresponding set functions $L(\cdot)$ and $U(\cdot)$, are called lower and upper probability (Huber & Strassen (1973)), envelopes (Walley & Fine (1982), Denneberg (1994)), coherent probability (Walley (1991)) and F-probability (Weichselberger (1995, 2000, 2001)). In game theoretic setting $\mathcal{M}$ is the *core*. Here Weichselberger's terminology is used calling $\mathcal{M}$ *structure*. Note that, by (2), there is a one-to-one correspondence between $P(\cdot)$ and the structure $\mathcal{M}$.

Two-monotone capacities (Huber & Strassen (1973), also called supermodular capacities (Denneberg (1994)) or convex capacities (Jaffray (1989)), as well as belief functions (Shafer (1976), Yager, Fedrizzi & Kacprzyk (1994)) are special cases.

More general sets of classical probabilities are obtained by the theory of coherent previsions (Walley (1991)), i.e. by assigning interval-valued expectations $\mathbb{E}_{\mathcal{M}}(\cdot) := [{}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}(\cdot), {}^{\mathrm{U}}\mathbb{E}_{\mathcal{M}}(\cdot)]$ on a set $\mathcal{K}$ of random variables on $(\Omega, \mathcal{A})$. By the lower envelope theorem (Walley (1991, p.134)) and the fact that classical expectation and classical probabilities uniquely correspond with each other, the definition of coherence can be rewritten in a similar way. Denoting by $\mathbb{E}_p$ the classical expectation operator with respect to the classical probability $p(\cdot)$, the assignment $\mathbb{E}_{\mathcal{M}}(\cdot)$ is coherent iff

$$\begin{aligned}\mathcal{M} := \{\, p(\cdot) \in \mathcal{C}(\Omega, \mathcal{A}) \mid \qquad\qquad\qquad\qquad (3)\\ {}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}X \le \mathbb{E}_p X \le {}^{\mathrm{U}}\mathbb{E}_{\mathcal{M}}X,\ \forall X \in \mathcal{K}\} \neq \emptyset\,.\end{aligned}$$

and

$$\left.\begin{array}{l}\inf\limits_{p(\cdot)\in\mathcal{M}} \mathbb{E}_p X = {}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}X\\ \sup\limits_{p(\cdot)\in\mathcal{M}} \mathbb{E}_p X = {}^{\mathrm{U}}\mathbb{E}_{\mathcal{M}}X\,,\end{array}\right\} \quad \forall X \in \mathcal{K} \qquad (4)$$

Throughout the paper it will be assumed that $\mathcal{K}$ is finite. Since Walley (1991) did not coin a name for the set $\mathcal{M}$ from (3), it will be called *structure*, too.

The interval-valued functions or functionals and the structure are dual concepts, they uniquely determine each other. The results obtained in this paper will be given in terms of the structure.

Many concepts of classical probability theory can appropriately be generalized. For decision making the notion of expectation is the most important one. Looking at the structure $\mathcal{M}$, one way how to define expectation for interval probability and how to extend the functional $\mathbb{E}_{\mathcal{M}}$ to random variables $X \notin \mathcal{K}$ suggests itself (cf. also the natural extension in Walley (1991)).

Def. 1 *Given a structure* $\mathcal{M} \subseteq \mathcal{C}(\Omega, \mathcal{A})$

$$\begin{aligned}\mathbb{E}_{\mathcal{M}}X &:= \left[{}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}X, {}^{\mathrm{U}}\mathbb{E}_{\mathcal{M}}X\right]\\ &:= \left[\inf_{p(\cdot)\in\mathcal{M}} \mathbb{E}_p X\,,\ \sup_{p(\cdot)\in\mathcal{M}} \mathbb{E}_p X\right]\end{aligned} \qquad (5)$$

is the (interval-valued) expectation of X (with respect to $\mathcal{F}$).

An alternative way to define expectation for non-additive set functions is the *Choquet integral (or fuzzy integral)* (c.f., e.g., Denneberg (1994)). For the case of two-monotone and totally monotone capacities both notions are equivalent (cf., e.g., Denneberg (1994, Prop. 10.3, p. 126)). Therefore, the results developed below are also valid for the Choquet integral.

2.2 Some Basic Properties of Structures

In the case of finite spaces $\Omega = \{\omega_1, \ldots, \omega_k\}$ considered throughout this paper, the structure can be given a simple geometric interpretation: Every classical probability $p(\cdot)$ is uniquely determined by the vector $p(\{\omega_1\}), \ldots, p(\{\omega_k\})$ of probability components of the singletons $\omega_1, \ldots, \omega_k$. Therefore, every classical probability can be understood as a point in $\mathbb{R}^k$, and every structure as a certain subset of $\mathbb{R}^k$.

Prop. 1 *Every structure* $\mathcal{M}$ *is (isomorphic to) a convex polyhedron. In particular, the set* $\mathcal{E}(\mathcal{M})$ *of the extreme points (vertices) is non-empty, finite, and it uniquely determines* $\mathcal{M}$*.*

Proof: For interval probabilities compare Shapley (1970, p. 16), Weichselberger (2001, Ch. 2). More generally, note that, for every $X \in \mathcal{K}$, conditions of the form $\mathbb{E}_p X \geq {}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}X$ or $\mathbb{E}_p X \leq {}^{\mathrm{U}}\mathbb{E}_{\mathcal{M}}X$ describe half-spaces in $\mathbb{R}^k$. Then, by (1) or (3), and because $\mathcal{A}$ and $\mathcal{K}$ are assumed to be finite, $\mathcal{M}$ is just the intersection of a finite number of half spaces, i.e. a convex polyhedron. Since $\mathcal{M} \subset [0,1]^k$, it is bounded. Therefore the second statement is a general result from the theory of convex polyhedra.

This relation to convex polyhedra makes linear programming a powerful tool for working with interval probability. Weichselberger (1996) gives characterizations of R- and F-probabilities in terms of linear programs. Efficient algorithms are available to calculate $\mathcal{M}$ and $\mathcal{E}(\mathcal{M})$ from the interval-valued assignments given.

Linear programming is also important for calculating interval-valued expectations. Given a random variable X, calculation of $\mathbb{E}_{\mathcal{M}}X$ means to maximize or minimize $\sum_{\omega\in\Omega} X(\omega)\cdot p(\{\omega\})$ over all $p(\cdot)\in\mathcal{M}$, i.e. to determine extrema of a linear function on a convex polyhedron. This can easily be done by linear programming. From linear programming one also knows that it is sufficient for the calculation to concentrate on the extreme points $\mathcal{E}(\mathcal{M})$ of the structure $\mathcal{M}$.

Lemma 1 *Let $\mathcal{M} \subseteq \mathcal{C}(\Omega,\mathcal{A})$ be a structure and $\mathcal{E}(\mathcal{M})$ its set of extreme points. Then*

$$\mathbb{E}_{\mathcal{M}}X = \left[\min_{p(\cdot)\in\mathcal{E}(\mathcal{M})} \mathbb{E}_p X \ ; \ \max_{p(\cdot)\in\mathcal{E}(\mathcal{M})} \mathbb{E}_p X\right] . \quad (6)$$

In the decision theoretic setting the following immediate conclusion will become very important.

Corollary 1 *(Vertice Reduction Lemma) In the situation of Lemma 1, for every real g,*

$$^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}X \geq g \iff \mathbb{E}_p X \geq g, \quad \forall p(\cdot)\in\mathcal{E}(\mathcal{M}) . \quad (7)$$

3 Actions Maximizing Expected Utility

3.1 The Generalized Decision Theoretic Framework

In this section the decision problem as described in in the Introduction will be analyzed in the situation where the decision maker's knowledge on the state of nature is ambiguous, expressed by a structure $\mathcal{M}$ of classical probabilities on $(\Theta,\mathcal{P}o(\Theta))$, and no sample information is available. For brevity of reference, the relevant components, the set $\mathbb{A}$ of actions, the set Θ of states of nature and the precise utility function[2] $u(\cdot)$, is collected in the triple $(\mathbb{A},\Theta,u(\cdot))$ which is called *basic decision problem.* For many applications it will prove of value to extend the problem by allowing for *randomized actions.* Formally, every randomized action can be identified with a classical probability[3] $\lambda(\cdot)$ on $(\mathbb{A},\mathcal{P}o(\mathbb{A}))$ where $\lambda(\{a\})$, $a\in\mathbb{A}$, is interpreted as the probability to choose action a. The set of all randomized actions will be denoted by $\Lambda(\mathbb{A})$. Pure actions, i.e. elements a of $\mathbb{A}$ itself, are identified with the Dirac measure in the point $\{a\}$, and therefore are also understood to be elements of $\Lambda(\mathbb{A})$.

The utility function is extended to the domain $\Lambda(\mathbb{A})\times\Theta$ by $u(\lambda,\vartheta_j) := \sum_{s=1}^{n} \lambda(a_s)\cdot u(a_s,\vartheta_j)$. Analogously to $\mathbf{u}(a)$ definied in the Introduction, $\mathbf{u}(\lambda)$ is that random variable which gives the utility of λ in dependence on the true state ϑ.

The generalization of the concept of probability now allows to consider generalized prior probabilities describing the decision maker's state of knowledge. With the notion of interval-valued expectation one immediately obtains the basic element of a generalized decision theory:

Def. 2 *Given the basic decision problem* $(\mathbb{A},\Theta,u(\cdot))$ *and a structure $\mathcal{M}\subseteq\mathcal{C}(\mathbb{A},\mathcal{P}o(\mathbb{A}))$. For every pure action $a\in\mathbb{A}$ and for every randomized action λ in $\Lambda(\mathbb{A})$, $\mathbb{E}_{\mathcal{M}}\mathbf{u}(a)$ and $\mathbb{E}_{\mathcal{M}}\mathbf{u}(\lambda)$ are the* generalized expected utility *(with respect to the prior information $\mathcal{M}$).*

Note that $\mathbb{E}_{\mathcal{M}}\mathbf{u}(a)$ and $\mathbb{E}_{\mathcal{M}}\mathbf{u}(\lambda)$ are interval-valued quantities. In most cases, comparing the generalized expected utilities of actions directly will lead only to partial orders on $\mathbb{A}$ and $\Lambda(\mathbb{A})$.

If a linear (complete) order of actions is desired, an appropriate *representation* is needed. This is a mapping from $\mathbb{R}\times\mathbb{R}$ to $\mathbb{R}$ which evaluates intervals by real numbers to use of the natural order on $\mathbb{R}$ for distinguishing optimal actions. A convenient and flexible choice may be a linear combination (compare with Ellsberg (1961, p. 664), Jaffray (1989), Weichselberger & Augustin (1998), Weichselberger (2001, Ch. 2.6)). It represents $\mathbb{E}_{\mathcal{M}}\mathbf{u}(\lambda)$ by

$$\eta\cdot {}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}\mathbf{u}(\lambda) + (1-\eta)\cdot {}^{\mathrm{U}}\mathbb{E}_{\mathcal{M}}\mathbf{u}(\lambda) . \quad (8)$$

The parameter η, called *caution* by Weichselberger (2001), reflects the degree of ambiguity aversion[4]; the more ambiguity averse the decision maker is, the higher is the influence of the lower interval limit of generalized expected utility. $\eta = 1$ corresponds to strict ambiguity aversion, $\eta = 0$ expresses maximal ambiguity seeking attitudes.

Remark 1 *It should be noted that the criterion considered here contains two classical decision criteria as border cases: If there is perfect probabilistic information and therefore no ambiguity, then $\mathcal{M}$ consists of one single classical prior probability $\pi(\cdot)$ only, and (8) coincides with Bayes optimality with respect to $\pi(\cdot)$. On the other hand, in the case of*

[2]Throughout the paper it is assumed that a (precise) utility function is given. On the *construction* of utility functions in the presence of ambiguity, generalizing the Neumann Morgenstern approach, see Jaffray (1989) and Schmeidler (1989).

[3]Since the concept of randomization is based on an ideal random experiment, it is coherent with the theory not to allow for interval-valued randomization probabilities.

[4]Here 'attitude towards ambiguity' is used in an intuitive sense — as an analog to risk aversion. This reflects the observation, repeatedly having been made since Ellsberg (1961), that people vary with respect to their reactions to ambiguity. Clearly this concept needs much deeper investigation, especially from empirical psychology. Weichselberger (2001, Ch. 2.6) proposes some measurement procedures for the caution η.

completely lacking information, the prior information consist of all classical probabilities on $(\Theta, \mathcal{P}o(\Theta))$ *('*non-selective*' or '*vacuous*' prior). Then one can show that* ${}^{\mathrm{L}}\mathbb{E}\mathbf{u}(\lambda) = \min_{j\in\{1,\ldots,m\}} u(\lambda, \vartheta_j)$, ${}^{\mathrm{U}}\mathbb{E}\mathbf{u}(\lambda) = \max_{j\in\{1,\ldots,m\}} u(\lambda, \vartheta_j)$*, and (8) leads to the Hurwicz criterion with optimism parameter* $1 - \eta$*. In particular, additionally taking* $\eta = 1$ *yields the maximin criterion.*

Expressing the probabilistic knowledge by a structure means that inside the structure there is complete ignorance: none of the elements of the structure is 'more likely' than another one. Therefore several authors (see the literature cited below) suggested to apply 'the maximin criterion to the structure', i.e. taking the very conservative choice $\eta = 1$. Then the interval-valued expectations $[{}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}\mathbf{u}(\lambda); {}^{\mathrm{U}}\mathbb{E}_{\mathcal{M}}\mathbf{u}(\lambda)]$ are represented by the lower interval limit alone, and an action λ^* is optimal iff

$$ {}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}(\mathbf{u}(\lambda^*)) \geq {}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}(\mathbf{u}(\lambda)) \,, \quad \forall \lambda \in \Lambda(\mathrm{IA}) \,. \qquad (9) $$

This criterion corresponds to the Gamma-Minimax principle (e.g. Berger (1984, Section 4.7.6), Vidakovic (2000)), to the Maxmin expected utility model (Gilboa & Schmeidler (1989)), and to the MaxEMin criterion considered by Kofler & Menges (1976; cf. also Kofler (1989) and the references therein.))

In what follows the criterion (9) will be used. This is done, however, without claiming that this is the only appropriate choice. Indeed, already in the original paper by Ellsberg (1961) there are strong arguments for additionally taking into account other criteria. So, the use of (9) is mainly technically motivated. It allows to derive results which are understood as a first step towards a comprehensive framework for decision making under ambiguity (see also the remarks in Section 6).

3.2 Optimal Actions – Calculation and Basic Properties

Searching for an optimal action with respect to the criterion (9) means to maximize

$$ {}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}\mathbf{u}(\lambda) = \min_{\pi(\cdot)\in\mathcal{M}} \sum_{j=1}^{m} \left(u(\lambda; \vartheta_j) \cdot \pi(\{\vartheta_j\})\right) \qquad (10) $$

among all $\lambda(\cdot) \in \Lambda(\mathrm{IA})$, i.e. to solve

$$ \min_{\pi(\cdot)\in\mathcal{M}} \sum_{j=1}^{m} \left(\sum_{s=1}^{n} \left(u(a_s; \vartheta_j)\lambda(a_s)\right) \cdot \pi(\{\vartheta_j\}) \right) \to \max \qquad (11) $$

under the restrictions

$$ \sum_{s=1}^{n} \lambda(a_s) = 1, \qquad \lambda(a_s) \geq 0 \,, \quad s = 1, \ldots, n \,. \qquad (12) $$

This originally non-linear problem can be made linear by introducing an auxiliary variable g for ${}^{\mathrm{L}}\mathbb{E}\mathbf{u}(\lambda)$: The maximization in (11) is equivalent to '$g \to \max$' under the additional restrictions

$$ \sum_{j=1}^{m} \left(\sum_{s=1}^{n} \left(u(a_s, \vartheta_j) \cdot \lambda(a_s)\right) \cdot \pi(\{\vartheta_j\}) \right) \geq g, \qquad (13) $$
$$ \forall \pi(\cdot) \in \mathcal{M} \,. $$

Now both, the objective function as well as the restrictions, are linear in a finite number of variables, $\lambda(a_1), \ldots, \lambda(a_s), \ldots \lambda(a_n)$ and g. Nevertheless this is no linear programming problem in the usual sense. Except in the case of a classical prior probability, the structure $\mathcal{M}$ consists of non-countable many elements. So, in the way as it is written, (13) contains an infinite number of restrictions. Now the vertex reduction lemma (Corollary 1) becomes very valuable. It states that (13) is equivalent to requiring $\sum_{j=1}^{m} \left(\sum_{s=1}^{n} \left(u(a_s, \vartheta_j) \cdot \lambda(a_s)\right) \cdot \pi(\{\vartheta_j\})\right) \geq g$ only for all $\pi(\cdot)$ in the set $\mathcal{E}(\mathcal{M})$ of extreme points of $\mathcal{M}$. Since, according to Proposition 1, the set $\mathcal{E}(\mathcal{M})$ is finite, the original decision problem was successfully transformed into a single linear programming problem. This is of double interest. Firstly, this means that the calculation of optimal actions can be done by standard routines. Secondly, it allows to apply general results from the theory of linear programming to the decision problem under consideration. Summing up one obtains

Theorem 1 *Consider a basic decision problem* $(\mathrm{IA}, \Theta, u(\cdot))$*, a structure* $\mathcal{M} \subseteq \mathcal{C}(\mathrm{IA}, \mathcal{P}o(\mathrm{IA}))$ *and the linear programming problem*

$$ g \to \max $$

under the constraints (12) and

$$ \left(\sum_{j=1}^{m} \left(\sum_{s=1}^{n} \left(u(a_s, \vartheta_j) \cdot \lambda(a_s)\right) \right) \pi(\{\vartheta_j\}) \right) \geq g \,, $$
$$ \forall \pi(\cdot) \in \mathcal{E}(\mathcal{M}) \,. $$

- *a) Let* $(\lambda^*(\{a_1\}), \ldots, \lambda^*(\{a_n\}), g^*)$ *be an optimal solution, then the randomized action*[5] $\lambda^*(\cdot) = (\lambda^*(\{a_1\}), \ldots, \lambda^*(\{a_n\})$ *is maximizing generalized expected utility with respect to the prior information* $\mathcal{M}$ *and the criterion (9). Furthermore,* ${}^{\mathrm{L}}\mathbb{E}_{\mathcal{M}}u(\lambda^*) = g^*$*.*
- *b) Every expected utility maximal action* $\lambda^*(\cdot)$ *uniquely corresponds with an optimal solution of this optimization problem.*

[5] Here, and later on, every element $\lambda(\cdot)$ of $\Lambda(\mathrm{IA})$ is identified with the corresponding vector with the components $\lambda(\{a_1\}), \ldots, \lambda(\{a_n\})$.

c) There always exists an action maximizing general expected utility.

d) The set Λ^* of actions maximizing general expected utility is convex and compact.

The proof mainly follows from the considerations preceding the theorem and from some general results provided by the theory of linear programming. It is given in Augustin (2001), where also further properties of optimal actions are discussed. Special attention is paid there to the case where the decision maker's prior knowledge can be expressed by two-monotone or totally monotone capacities (belief functions). Then the set of extreme points can be given in a closed form, and the algorithm from Theorem 1 provides an efficient and elegant tool to work with Choquet expected utility.

4 Extension to the Data Problem

In many decision problems it is possible to obtain some information on the states of nature by collecting additional data. Formally, this can be described by an additional 'experiment' where the probability of the outcomes depends on the true state of nature. Let $\Omega = \{\omega_1 \ldots, \omega_k\}$ be the sample space of this experiment, $\mathcal{A}$ the power set on Ω. Let further structures $\mathcal{M}_j \subseteq \mathcal{C}(\Omega, \mathcal{A})$ be given which describe the distribution on Ω in dependence on the true ϑ_j. The triple $(\Omega, \mathcal{A}, (\mathcal{M}_j)_{j=1,\ldots,m})$ is called *sample information*, the basic decision problem together with the sample information *data problem*.

Now the decision problem consists of the choice between *decision functions (strategies)*

$$\begin{array}{rccl} d: & \{\omega_1, \ldots, \omega_k\} & \to & \Lambda(\mathrm{I\!A}) \\ & \omega_i & \mapsto & d(\omega_i) = \lambda\ , \end{array}$$

i.e. functions which state the action to be chosen in dependence on the observation $\{\omega_i\}$. Denote the set of all decision functions by $\mathcal{D}$.

It is standard to define the utility produced by a decision function $d(\cdot)$ via the random variable $\mathbf{u}(d, \vartheta_j)$ on $(\Omega, \mathcal{A})$ which takes the values $u(d(\omega), \vartheta_j)$. Given ϑ_j, one compares decision functions by looking at the expectation of $\mathbf{u}(d, \vartheta_j)$ with respect to the structure $\mathcal{M}_j$.

Since only the case of strict ambiguity aversion is considered here,[6] one can represent this interval-valued quantity by the random variable

$$\mathbf{u}(d) := {}^{\mathrm{L}}\mathrm{I\!E}_{\mathcal{M}_j}\mathbf{u}(d, \vartheta_j)\,. \qquad (14)$$

[6]Concerning this choice, see the remarks in Section 3.1 and in Section 6.

Given the prior information $\mathcal{M}$ on $(\Theta, \mathcal{P}o(\Theta))$, the overall expected utility $\mathrm{I\!E}_{\mathcal{M}}\mathbf{u}(d)$ can be calculated. In the case of strict ambiguity aversion with respect to the prior information $\mathcal{M}$, a strategy $d^*(\cdot)$ is *optimal* iff

$${}^{\mathrm{L}}\mathrm{I\!E}_{\mathcal{M}}\mathbf{u}(d^*) \geq {}^{\mathrm{L}}\mathrm{I\!E}_{\mathcal{M}}\mathbf{u}(d)\,, \qquad \forall d \in \mathcal{D}\,. \qquad (15)$$

Writing this as an optimization problem one describes every decision function $d(\cdot)$ by the matrix

$$(d(\omega_i, a_s)_{i=1,\ldots,k;s=1,\ldots,n})$$

where $d(\omega_i, a_s)$ is the classical probability to choose action a_s if $\{\omega_i\}$ was observed.

One then has to maximize

$$\min_{\pi\in\mathcal{M}} \sum_{j=1}^{m} \min_{p_j(\cdot)\in\mathcal{M}_j} XY$$

with $XY :=$

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

under the constraints

$$\sum_{s=1}^{n} d(\omega_i,\, a_s) = 1\,, \quad \forall i = 1, \ldots, k, \qquad (16)$$

$$d(\omega_i,\, a_s) \geq 0\,, \quad \forall i = 1, \ldots, k,\; s = 1, \ldots, n\,. \qquad (17)$$

Similar techniques to those described in Section 3.2 can be used to extend Theorem 1 to the data problem. In particular, one obtains

Theorem 2 *Consider a basic decision problem $(\mathrm{I\!A}, \Theta, u(\cdot))$, a structure $\mathcal{M} \subseteq \mathcal{C}(\mathrm{I\!A}, \mathcal{P}o(\mathrm{I\!A}))$ and the sampling information $((\Omega, \mathcal{A}), (\mathcal{M}_j)_{j=1,\ldots,m})$. A decision function $d^*(\cdot)$ is optimal iff it corresponds to an optimal solution $(d^*(\omega_i, a_s)_{i=1,\ldots,k;s=1,\ldots,n})$ of the linear problem*

$$g \longrightarrow \max$$

under the constraints (16), (17) and

$$\begin{aligned} \sum_{i=1}^{k}\left(\sum_{s=1}^{n} u(a_s,\, \vartheta_j)\cdot d(\omega_i,\, a_s)\right)\cdot p_j(\{\omega_i\}) &\geq g_j\,, \\ \forall p_j(\cdot) \in \mathcal{E}(\mathcal{M}_j);\quad \forall j = 1, \ldots, m\,, & \\ \sum_{j=1}^{m} g_j \cdot \pi(\{\vartheta_j\}) &\geq g\,, \\ \forall \pi(\cdot) \in \mathcal{E}(\mathcal{M})\,. & \end{aligned}$$

5 On the Suboptimality of the Robust Bayesian Approach

This section presents a question for discussion:

The main paradigm of Bayesian inference states that, once the data have been observed, the posterior probability contains all the relevant information for inference. In the area of classical probability this can be justified from the decision theoretic point of view: Let $d^*(\cdot)$ be the optimal decision function in a decision problem $(\mathrm{IA}, \Theta, u(\cdot))$ with sampling information $(\Omega, \mathcal{A}, (\{p_j(\cdot)\})_{j=1,\ldots,m})$. The action $d^*(\omega)$, chosen after having observed $\{\omega\}$, is also the optimal action in the basic decision problem without data if one replaces the original prior $\pi(\cdot)$ by the posterior $\pi(\cdot\,|\{\omega\})$. In other words: one just uses the posterior as the 'updated prior'. Optimality with respect to posterior loss and optimality with respect to prior risk coincide.

This paradigm has also been accepted in generalized Bayesian inference based on imprecise prior probabilities expressed by a structure $\mathcal{M}$. In robust Bayesian inference (e.g., Wasserman (1997)) the 'posterior structure' $\mathcal{M}_{|\{\omega\}}$ based on the observation $\{\omega\}$ is calculated as the set of all posteriors of elements of $\mathcal{M}$. Walley (1991, Ch. 7f.) justifies such a procedure as being coherent.

A look at the framework considered here shows, however, that such a procedure lacks its decision theoretic justification. It may lead to *suboptimal* solutions. A very brief counterexample can be obtained from a border case: Consider a precise sample information $(\Omega, \mathcal{A}, (\{p_j(\cdot)\})_{j=1,\ldots,m})$, and the vacuous prior information $\mathcal{C}(\Theta, \mathcal{P}o(\Theta))$. Then, independent of ω, also the posterior is vacuous. Using it as the 'updated prior' yields, according to Remark 1, the maximin solution λ^{mm} of the basic decision problem as the optimal randomized action. In contrast, the optimal decision function can be shown to coincide with the maximin decision function $d^{mm}(\cdot)$ of the data problem. Typically, at least for some ω, the actions $d^{mm}(\omega)$ and λ^{mm} differ.

It should be noted that the vacuous prior was chosen only to keep the example as short as possible. Other examples[7] can be constructed supporting the conclusion that the robust Bayesian solution may be suboptimal; the imprecise posterior does not contain all the relevant information for inference.

[7]They will be available at `www.stat.uni-muenchen.de/~thomas`.

6 Concluding Remarks

In this paper a general framework for decision making under ambiguity was proposed. Linear optimization problems were derived to calculate optimal solutions efficiently, and to describe basic properties of the optimal solutions. Further insights may be expected from looking at the dual optimization problems. In the very special decision problem of hypotheses testing, duality theory for linear programming led to an extension of the generalized Neyman Pearson lemma (Augustin (1998, Chapter 5.1f)).

As already discussed in Section 3.1, it is highly desirable to extend the consideration from strict ambiguity aversion to more complex criteria like (8). First results for the basic problem without sampling information have been obtained in Augustin (2001). In principle, they promise to be extensible to the data problem of Section 4.

The question raised in Section 6 requires deeper investigation. If the conclusion that the imprecise posterior does not contain all the relevant information for inference is correct, then the topic of 'updating' in the area of imprecise probability has to be reconsidered in detail.

Acknowledgements

I am very grateful to the two referees, an editor, David Budescu, and Kurt Weichselberger for their very valuable remarks.

References

[1] T. Augustin (1998): *Optimale Testing under Interval Probability.* Vandenhoeck & Ruprecht, Göttingen. In German with an English summary on pages 247–249.

[2] T. Augustin (2001): Expected utility within a generalized concept of probability — a comprehensive framework for decision making under ambiguity. *submitted.*

[3] J.O. Berger (1984): *Statistical Decision Theory and Bayesian Analysis.* (2nd edition). Springer. New York.

[4] Budescu, D.V.; Weinberg, S.; Wallsten , T.S. (1988): Decisions on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance.* 14, 281-294.

[5] deCooman, G.; Walley, P. (Eds.) (2001): *The Imprecise Probability Project.* `http://ippserv.rug.ac.be/`.

[6] Denneberg, D. (1994): *Non-Additive Measure and Integral.* Kluwer. Dordrecht.

[7] Ellsberg, D. (1961): Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics.* **75**, 643-669.

[8] Gilboa, I.; Schmeidler, D. (1989): Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics.* 18, 141–153.

[9] Huber, P.J.; Strassen, V. (1973): Minimax tests and the Neyman-Pearson lemma for capacities. *Annals of Statistics.* 1, 251-263; Correction: 2, 223-224.

[10] Jaffray, J.Y. (1989): Linear utility theory and belief functions. *Operations Research Letters.* 8, 107–122.

[11] Kofler, E. (1989): *Prognosen und Stabilität bei unvollständiger Information.* Campus. Frankfurt/Main.

[12] Kofler, E.; Menges G. (1976): Entscheidungen bei unvollstndiger Information. Springer. Berlin.

[13] Kruse, R.; Schwecke, E.; Heinsohn, J. (1991): *Uncertainty and Vagueness in Knowledge Based Systems. Numerical Methods.* Springer, Berlin.

[14] Schmeidler, D. (1989): Subjective probability and expected utility without additivity. *Econometrica.* 57, 571–587.

[15] Shafer, G. (1976): *A Mathematical Theory of Evidence.* Princeton University Press. Princeton.

[16] Shapley, S. (1971): Cores of convex games. *International Journal of Game Theory.* **1**, 12-26.

[17] Vidakovic, B. (2000): Γ-minimax: A paradigm for conservative robust Bayesians. In: Insua, D.R., and Ruggeri, F. (Eds.) *Robust Bayesian Analysis.* Springer.

[18] Walley, P. (1991): *Statistical Reasoning with Imprecise Probabilities.* Chapman & Hall. London.

[19] Walley, P.; Fine, T.L. (1982): Towards a frequentist theory of upper and lower probability. *The Annals of Statistics.* 10, 741-761.

[20] Wasserman, L. (1997): Bayesian robustness. In: Kotz, S., Read, C.B., Banks, D.L. (Eds.): *Encyclopedia of Statistical Sciences. Update Volume 1.* Wiley, New York, pp. 45-51.

[21] Weichselberger, K. (1995): Axiomatic foundations of the theory of interval-probability. In: Mammitzsch, V., Schneeweiß, H. (Eds.) *Symposia Gaussiana Conference B.* de Gruyter, Berlin, pp. 47-64.

[22] Weichselberger, K. (1996): Interval-probability on finite sample-spaces. In: Rieder, H. (Ed.) *Robust Statistics, Data Analysis and Computer Intensive Methods*, Springer, New York, pp. 391-409.

[23] Weichselberger, K. (2000): The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning.* 24, 149-170.

[24] Weichselberger, K. (2001): *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung.* Physika. Heidelberg, to appear.

[25] Weichselberger, K.; Augustin, T. (1998): Analysing Ellsberg's Paradox by means of interval-probability. In: Galata, R..; Küchenhof, H. (Eds.): *Econometrics in Theory and Practice. (Festschrift for Hans Schneeweiß.)* Physika. Heidelberg; 291–304.

[26] Yager, R.R.; Fedrizzi, M.; Kacprzyk, J. (Eds.) (1994): *Advances in the Dempster-Shafer Theory of Evidence.* Wiley. New York.