International Journal of Approximate Reasoning 50 (2009) 642–654

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Data-based decisions under imprecise probability and least favorable models

R. Hable

University of Bayreuth, Department of Mathematics, Chair of Stochastics, D-95440 Bayreuth, Germany

ARTICLE INFO

Article history:
Received 17 December 2007
Accepted 25 March 2008
Available online 3 April 2008

Keywords:
Decision theory
Robust statistics
Imprecise probability
Coherent upper prevision
Huber–Strassen theory
Le Cam
Equivalence of models
Least favorable model

ABSTRACT

Data-based decision theory under imprecise probability has to deal with optimization problems where direct solutions are often computationally intractable. Using the Γ-minimax optimality criterion, the computational effort may significantly be reduced in the presence of a least favorable model. Buja [A. Buja, Simultaneously least favorable experiments. I. Upper standard functionals and sufficiency, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 65 (1984) 367–384] derived a necessary and sufficient condition for the existence of a least favorable model in a special case. The present article proves that essentially the same result is valid in case of general coherent upper previsions. This is done mainly by topological arguments in combination with some of Le Cam's decision theoretic concepts. It is shown how least favorable models could be used to deal with situations where the distribution of the data as well as the prior is allowed to be imprecise.

1. Introduction

1.1. Motivation

Decision theory provides a formal framework for determining optimal actions under uncertainty on the states of nature. It has a wide range of potential areas of application which includes also statistical problems. However, a serious problem in practical applications of decision theory is that the uncertainty often is too complex to be adequately described by a precise probability distribution. Ambiguity, i.e. the extent of deviation from ideal stochasticity, plays an important role in decision making that cannot be neglected. To take ambiguity into account properly, generalizations of the concept of probability have been developed, among others, by Walley [1] (imprecise probability) and Weichselberger [2] (interval probability, cf. also [3]). Here, the probability of an event is no longer a number $p \in [0,1]$ but an interval $[\underline{p}, \overline{p}] \subset [0,1]$. These concepts are applied in a number of recent articles in decision theory, e.g. [4–6].

Generalizations of probabilities as in [1,2] have a strong relationship with some concepts of robust statistics – a fact which is frequently disregarded. Actually, Buja [7] develops a concept of robust statistics (named "upper expectations") which lies between the concepts of [1,2]. Buja [7] considers decision making which is explicitly data-based. This can be understood as a matter of its own as has been pointed out by Augustin [4]. In the spirit of the celebrated article by Huber and Strassen [8], Buja [7] characterizes the existence of precise models which are simultaneously least favorable for a class of loss functions (or for a class of prior distributions).

Huber and Strassen [8] deals with hypothesis testing where a (rather special) upper prevision is tested against another one. This is equivalent to testing between certain sets of (precise) probabilities $\mathcal{M}_0$ and $\mathcal{M}_1$. Huber and Strassen [8] shows

E-mail address: Robert.Hable@uni-bayreuth.de

doi:10.1016/j.ijar.2008.03.009

that there is a pair $(p_0, p_1) \in \mathscr{M}_0 \times \mathscr{M}_1$ which is least favorable: Testing between p_0 and p_1 is as hard as testing between $\mathscr{M}_0$ and $\mathscr{M}_1$ and, as a consequence, there is an optimal test between p_0 and p_1 which is also an optimal test between $\mathscr{M}_0$ and $\mathscr{M}_1$. That way, testing between $\mathscr{M}_0$ and $\mathscr{M}_1$ can be done by testing only between p_0 and p_1. This reduces the computational effort substantially. In fact, it is one of the most important drawbacks of data-based decision theory (including hypothesis testing) that the computational effort of direct solutions is frequently not manageable. Therefore, least favorability has attracted enormous attention after the publication of [8]. For a review of [8] and the work following [8], confer [9]. In quite general data-based decision theory, where there are n states of nature (instead of two), an analogous question of that one solved by Huber and Strassen [8] is: Does there exist a model $(q_1, q_2, \ldots, q_n) \in \mathscr{M}_1 \times \mathscr{M}_2 \times \cdots \times \mathscr{M}_n$ which is simultaneously least favorable for a class of loss functions? This is not always the case but [7] proves a necessary and sufficient condition for the existence of such simultaneously least favorable models.

Unfortunately, Buja [7] contains an error which reduces its applicability significantly. The validity of the conclusions in [7] can only be guaranteed by adding a restrictive assumption on the involved upper previsions; cf. Section 8.1.

The present article follows the lines of [7] – but within the concept of [1] which dispenses with σ-additivity. It is shown that the same result as in [7] is possible without any additional assumption on the involved (coherent) upper previsions. Surprisingly, most of the proofs are similar to those given in [7]. This demonstrates that, in [7], insistence on σ-additivity of probabilities happens to be an unnecessary burden (cf. also Remark 2.2).

By ignoring σ-additivity, we are in line with Le Cam's decision theoretic framework (cf. [10,11]), which provides us with some effective methods. Within this framework some terms (e.g. randomization) are slightly generalized.

Sections 2 and 3 develop the decision theoretic framework. Section 4 contains a generalization of the Le Cam–Blackwell–Sherman–Stein–Theorem. This theorem plays an important role in Section 5 where the analogue to [7, Theorem 8.2] is proven which characterizes the existence of least favorable models. This is the main theorem of the present article. Section 6 explains how least favorability could be used to deal with situations where the distribution of the data as well as the prior is assumed to be imprecise.

Since the content of this article might be obscured by the mathematical details, the following subsection presents a rather detailed, but informal outline.

1.2. Outline

In order to explain the decision theoretic setup we are concerned with, the classical decision theoretic setup is recalled at first.

There is a set Θ where each element $\theta \in \Theta$ represents a possible state of nature. We know that one state of nature will occur but we do not know which one it will be. Furthermore, there is a set $\mathbb{D}$ where each element $t \in \mathbb{D}$ is a decision we can choose. Depending on what state of nature θ occurs, every decision t leads to a loss $W_\theta(t)$. The goal is to choose a "good" decision so that the loss is as small as possible.

Sometimes, we might know a precise expectation π for the states of nature $\theta \in \Theta$. Then, we can choose the decision that minimizes the expected loss

$$\int_\Theta W_\theta(t)\ \pi(\mathrm{d}\theta)$$

In addition, we often can choose our decision on the base of an observation $y \in \mathscr{Y}$. For example, the observation y may be the outcome of an experiment. The distribution of the observation y might be a precise expectation q_θ which depends on the state of nature θ. That is $(q_\theta)_{\theta\in\Theta}$ is a model which describes the distribution of the observation y.

Such "data-based decision making" can be formalized by choosing a decision function $\delta : \mathscr{Y} \to \mathbb{D}, x \mapsto \delta(y)$ which minimizes

$$\int_\Theta \int_{\mathscr{Y}} W_\theta(\delta(y)) q_\theta(\mathrm{d}y) \pi(\mathrm{d}\theta)$$

Decision theory commonly also deals with randomized decisions. Randomized decision procedures (randomizations) are defined in Section 2.1. Confer [12] for an introduction to these basic concepts of decision theory.

In the following, we are concerned with a more general decision theoretic setup because we also want to deal with imprecise probabilities.

Since the prior knowledge about the states of nature will frequently not be precise, we allow for a whole set $\mathscr{P}$ of possible precise expectations π. Also the knowledge about the distribution of the observation may only be imprecise so that there are sets $\mathscr{M}_\theta$ of possible precise expectations q_θ. While minimizing the expected loss in case of precise expectations is widely accepted, there are several reasonable optimality criteria in case of imprecise expectations; confer [5] for a discussion of the most important ones. In the present article the so-called Γ-minimax criterion is used which represents a worst-case consideration.[1] That is, we choose a decision function δ (or rather a randomization later on) which minimizes the twofold upper expectation

[1] For the use of the Γ-minimax criterion in Bayesian analysis, cf. [13] and the literature cited therein.

$$\sup_{\pi\in\mathscr{P}}\int_{\Theta}\sup_{q_\theta\in\mathscr{M}_\theta}\int_{\mathscr{Y}}W_\theta(\delta(y))q_\theta(\mathrm{d}y)\pi(\mathrm{d}\theta)$$

Unfortunately, a direct solution of this problem is quite often computationally intractable. In Section 6, it is shown how the situation might become manageable: In the presence of a model $(\tilde{q}_\theta)_{\theta\in\Theta}\in(\mathscr{M}_\theta)_{\theta\in\Theta}$ which is simultaneously least favorable for $\mathscr{P}$ (or for a corresponding set of loss functions) the above minimization problem may be solved by minimizing

$$\sup_{\pi\in\mathscr{P}}\int_{\Theta}\int_{\mathscr{Y}}W_\theta(\delta(y))\tilde{q}_\theta(\mathrm{d}y)\pi(\mathrm{d}\theta)$$

However, such a least favorable model $(\tilde{q}_\theta)_{\theta\in\Theta}$ need not exist. In Section 5, a necessary and sufficient condition for existence is proven (Theorem 5.4). This condition is formulated in terms of standard models.

Indeed, standard models are our main tool. They are introduced in Section 2.3. An important fact is that every model (consisting of precise expectations) is equivalent to a standard model. In Section 2.2, we define an equivalence relation on the set of all (precise) models $(q_\theta)_{\theta\in\Theta}$ according to which two (precise) models $(p_\theta)_{\theta\in\Theta}$ and $(q_\theta)_{\theta\in\Theta}$ are equivalent if the following is true: Observations of model $(p_\theta)_{\theta\in\Theta}$ can artificially be generated (by a randomization) from observations of model $(q_\theta)_{\theta\in\Theta}$ and vice versa. Here and also as decision procedures, randomizations become important. For topological reasons, we have to rely on Le Cam's slight generalization of the term "randomization" (cf. [10]). All these tools from decision theory (namely randomizations, equivalence of models, standard models) are presented in Section 2.

In Section 3, minimal Bayes risks are defined for precise models and for imprecise models as well. It is shown that minimal Bayes risks can be expressed in terms of standard models. This is the reason why standard models are used.

Section 4 contains a generalization of the Le Cam–Blackwell–Sherman–Stein–Theorem. This theorem is important in the proof of the main theorem, Theorem 5.4, which characterizes the existence of simultaneously least favorable models.

1.3. Some notation

This subsection lists some notation which is used throughout the article.

Let $(\mathscr{Y},\mathscr{B})$ be a measurable space and $\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})$ be the Banach space of all bounded Borel-measurable real functions $g:\mathscr{Y}\to\mathbb{R}$ where $\|g\|=\sup_{y\in\mathscr{Y}}g(y)$. For a subset B of $\mathscr{Y}$, I_B denotes the characteristic function of B on $\mathscr{Y}$.

The set of all bounded, finitely additive, signed measures $\mathrm{ba}(\mathscr{Y},\mathscr{B})$ can be identified with the dual space of $\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})$, i.e. the Banach space of all linear continuous real functionals on $\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})$ where $\|\mu\|=\sup\{|\mu[g]\|\ g\in\mathscr{L}_\infty(\mathscr{Y},\mathscr{B}),\|g\|\leqslant 1\}$ for $\mu\in\mathrm{ba}(\mathscr{Y},\mathscr{B})$; cf. [14, Theorem IV.5.1]. $\mu\in\mathrm{ba}(\mathscr{Y},\mathscr{B})$ is called *positive* if $\mu[g]\geqslant 0$ for every $g\geqslant 0$. This is denoted by $\mu\geqslant 0$.

Let Θ be an index set. Throughout the article, $(\overline{Q}_\theta)_{\theta\in\Theta}$ is a family of coherent upper previsions $\overline{Q}_\theta:\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})\to\mathbb{R}$ (cf. [1]). The corresponding sets of majorized linear previsions are denoted by $\mathscr{M}_\theta:=\{q_\theta\in\mathrm{ba}(\mathscr{Y},\mathscr{B})|q_\theta[g]\leqslant\overline{Q}_\theta[g]\ \forall g\in\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})\}$. Analogously to [2], $\mathscr{M}_\theta$ is called *structure*. $(\overline{Q}_\theta)_{\theta\in\Theta}$ is called *imprecise model* on $(\mathscr{Y},\mathscr{B})$. A family $(q_\theta)_{\theta\in\Theta}$ of linear previsions $q_\theta:\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})\to\mathbb{R}$ is called *precise model* on $(\mathscr{Y},\mathscr{B})$. These terms are adapted from the notion "statistical model". Buja [7] and Le Cam [10] use the term "experiment" instead of "model".

Let $(\mathscr{X},\mathscr{A})$ be another measurable space. $\mathscr{F}=(q_\theta)_{\theta\in\Theta}$ always denotes a precise model on $(\mathscr{Y},\mathscr{B})$, $\mathscr{E}=(p_\theta)_{\theta\in\Theta}$ always denotes a precise model on $(\mathscr{X},\mathscr{A})$. If $q_\theta\in\mathscr{M}_\theta$ for every $\theta\in\Theta$, we may also write $(q_\theta)_{\theta\in\Theta}\in(\mathscr{M}_\theta)_{\theta\in\Theta}$ or $\mathscr{F}\in(\mathscr{M}_\theta)_{\theta\in\Theta}$. Expressions of the form $(a_\theta)_{\theta\in\Theta}$ are often abbreviated by $(a_\theta)_\theta$.

For some fixed $n\in\mathbb{N}$, put $\mathscr{U}:=\{u\in\mathbb{R}^n|u=(u_{\theta_1},\dots,u_{\theta_n})\prime,\ u_\theta\geqslant 0\ \forall\theta\in\Theta,\ u_{\theta_1}+\dots+u_{\theta_n}=1\}$ and $\mathscr{C}:=\mathbb{B}^{\otimes n}\cap\mathscr{U}$ where $\mathbb{B}^{\otimes n}$ is the Borel-σ-algebra of $\mathbb{R}^n$. For $\theta\in\Theta$, put $\iota_\theta:\ \mathscr{U}\to[0,1],\ u\mapsto u_\theta$ where u_θ is the θ-component of u.

2. Some tools from decision theory

2.1. Randomizations

2.1.1. Introduction

Let $\mathscr{X}$ be a set of possible outcomes of an experiment and $\mathbb{D}$ be a set of possible decisions t. Then, a decision function may be a map $\delta:\mathscr{X}\to\mathbb{D}$ where $\delta(x)=t$ means: If x appears, choose action t. In addition, decision theory commonly deals with randomized decisions $\delta:\ \mathscr{X}\to\mathrm{ba}(\mathbb{D},\mathscr{D}),\ x\mapsto\tau_x$. Here, it is supposed that each τ_x is a linear prevision and that $\tau_\cdot[h]:x\mapsto\tau_x[h]$ lies in $\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ for every $h\in\mathscr{L}_\infty(\mathbb{D},\mathscr{D})$. Then, $\delta(x)=\tau_x$ means: After observing x, start an auxiliary random experiment according to the distribution τ_x and choose that action d which is the outcome of the auxiliary random experiment.

For our purposes, we will need a slight generalization. Note that every randomized decision function $x\mapsto\tau_x$ defines a map

$$\sigma:\ \mathrm{ba}(\mathscr{X},\mathscr{A})\to\mathrm{ba}(\mathbb{D},\mathscr{D}),\qquad \mu\mapsto\sigma(\mu) \tag{1}$$

via $\sigma(\mu):h\mapsto\sigma(\mu)[h]=\mu[\tau_\cdot[h]]=\int\tau_x[h]\ \mu(\mathrm{d}x)$.

It is easy to see that σ is linear, positive ($\sigma(\mu)\geqslant 0$ for every $\mu\geqslant 0$) and normalized ($\|\sigma(\mu)\|=\|\mu\|$ for every $\mu\geqslant 0$).

2.1.2. Definition

Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be measurable spaces. According to [10], a *randomization* from $\mathcal{X}$ to $\mathcal{Y}$ is a linear, positive and normalized map

$$T: \ \mathrm{ba}(\mathcal{X}, \mathcal{A}) \to \mathrm{ba}(\mathcal{Y}, \mathcal{B})$$

where "positive" means $T(\mu) \geqslant 0$ for every $\mu \geqslant 0$ and "normalized" means $\|T(\mu)\| = \|\mu\|$ for every $\mu \geqslant 0$. Let $\mathcal{T}(\mathcal{X}, \mathcal{Y})$ denote the set of all randomizations from $\mathcal{X}$ to $\mathcal{Y}$.

We also mark a class of randomizations of a very simple form: To this end, let κ be a map

$$\kappa : \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B}) \to \mathcal{L}_\infty(\mathcal{X}, \mathcal{A}), \qquad g \mapsto \kappa(g)$$

so that there is some finite set $S \subset \mathcal{Y}$ and

$$\kappa(g) = \sum_{y \in S} g(y) \alpha_y \qquad \forall \ g \in \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B})$$

where $\alpha_y \in \mathcal{L}_\infty(\mathcal{X}, \mathcal{A}) \ \forall \ y \in S$, $\alpha_y \geqslant 0 \ \forall \ y \in S$ and $\sum_{y \in S} \alpha_y \equiv 1$. Then,

$$\kappa^* : \ \mathrm{ba}(\mathcal{X}, \mathcal{A}) \to \mathrm{ba}(\mathcal{Y}, \mathcal{B}), \qquad \mu \mapsto \kappa^*(\mu)$$

where $\kappa^*(\mu)[g] = \mu[\kappa(g)] \ \forall \ g \in \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B})$, is called *restricted randomization*. It is easy to see that every restricted randomization is generated by a (very simple) randomized decision function as in (1). Every restricted randomization is in fact a randomization, i.e. $\mathcal{T}_r(\mathcal{X}, \mathcal{Y}) \subset \mathcal{T}(\mathcal{X}, \mathcal{Y})$ where $\mathcal{T}_r(\mathcal{X}, \mathcal{Y})$ denotes the set of all restricted randomizations.

2.1.3. Topological issues

Models which consist of imprecise probabilities are so extensive that sequential limit arguments are no longer adequate. So, we have to resort to topological arguments. In addition to the norm-topology, $\mathrm{ba}(\mathcal{Y}, \mathcal{B})$ can also be provided with the $\sigma(\mathrm{ba}, \mathcal{L}_\infty)$-topology. This is the smallest topology so that

$$\mathrm{ba}(\mathcal{Y}, \mathcal{B}) \to \mathbb{R}, \quad \mu \mapsto \mu[g]$$

is continuous for every $g \in \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B})$.

Let $\overline{\mathrm{Q}} : \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B}) \to \mathbb{R}$ be a coherent upper prevision with structure $\mathcal{M} := \{q \in \mathrm{ba}(\mathcal{Y}, \mathcal{B}) | q[g] \ \leqslant \ \overline{\mathrm{Q}}[g] \ \forall g \in \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B})\}$.

Theorem 2.1. *$\mathcal{M}$ is $\sigma(\mathrm{ba}, \mathcal{L}_\infty)$-compact* (cf. [1, Section 3.6.1]).

Remark 2.2. According to Theorem 2.1, compactness of $\mathcal{M}$ comes for free. If we restricted $\mathcal{M}$ to σ-additive measures, we would have to impose additional assumptions to ensure compactness in reasonable topologies (cf. Section 8.1). So, insistence on σ-additivity appears to be a burden.

$\mathcal{T}(\mathcal{X}, \mathcal{Y})$ can be provided with the topology of pointwise convergence. This is the smallest topology so that

$$\mathcal{T}(\mathcal{X}, \mathcal{Y}) \to \mathbb{R}, \qquad T \mapsto T(\mu)[g]$$

is continuous for every $\mu \in \mathrm{ba}(\mathcal{X}, \mathcal{A})$ and every $g \in \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B})$. The following theorem is the reason why we use the generalization of randomized procedures:

Theorem 2.3. *$\mathcal{T}(\mathcal{X}, \mathcal{Y})$ is a compact Hausdorff space* (cf. [11, Theorem 1.4.2]).

The following theorem indicates that the term "randomization" has only been slightly generalized:

Theorem 2.4. *$\mathcal{T}_r(\mathcal{X}, \mathcal{Y})$ is dense in $\mathcal{T}(\mathcal{X}, \mathcal{Y})$.*

Proof. This is a consequence of [10, Theorem 1]. □

Especially, Theorem 2.4 implies that the randomized procedures defined as in (1) are dense in $\mathcal{T}(\mathcal{X}, \mathcal{Y})$.

2.2. Sufficiency and equivalence of models

Let $\mathcal{E} = (p_\theta)_{\theta \in \Theta}$ be a precise model on $(\mathcal{X}, \mathcal{A})$ and $\mathcal{F} = (q_\theta)_{\theta \in \Theta}$ a precise model on $(\mathcal{Y}, \mathcal{B})$. Analogously to [7], $(p_\theta)_{\theta \in \Theta}$ is called *sufficient* for $(q_\theta)_{\theta \in \Theta}$ if there is a randomization $T \in \mathcal{T}(\mathcal{X}, \mathcal{Y})$ so that $T(p_\theta) = q_\theta \forall \theta \in \Theta$.

This definition of "sufficiency" essentially goes back to [15]. It does not strictly coincide with the more common definition in terms of conditional expectations but, under suitable assumptions of regularity, the definitions do coincide (cf. [16]). At least, if the randomization T is generated by a randomized function $x \mapsto \tau_x$ as in (1), the above definition has a very descriptive interpretation.

Let x be an observation distributed according to p_θ. After observing x, start an auxiliary random experiment according to τ_x. Then, the outcome y of the auxiliary random experiment is distributed according to q_θ. That is, if we have observations of the model $(p_\theta)_\theta$, we can artificially generate observations of the model $(q_\theta)_\theta$ "by coin tossing".

$(p_\theta)_{\theta\in\Theta}$ and $(q_\theta)_{\theta\in\Theta}$ are called *equivalent* if they are mutually sufficient, i.e. there are some $T_1 \in \mathcal{T}(\mathcal{X},\mathcal{Y})$, $T_2 \in \mathcal{T}(\mathcal{Y},\mathcal{X})$ so that $T_1(p_\theta) = q_\theta \forall\theta \in \Theta$ and $T_2(q_\theta) = p_\theta \forall\theta \in \Theta$. This definition of equivalence is in accordance with Le Cam's definition (cf. Lemma 8.1). The descriptive interpretation of sufficiency already indicates that equivalent models essentially coincide from a decision theoretic point of view.

Let $(\overline{Q}_\theta)_{\theta\in\Theta}$ be an imprecise model with corresponding structures $\mathcal{M}_\theta$, $\theta \in \Theta$. Analogously to [7], $(p_\theta)_{\theta\in\Theta}$ is called *worst-case-sufficient* for $(\overline{Q}_\theta)_{\theta\in\Theta}$ if $(p_\theta)_{\theta\in\Theta}$ is sufficient for some $(q_\theta)_{\theta\in\Theta} \in (\mathcal{M}_\theta)_{\theta\in\Theta}$. So, $(p_\theta)_{\theta\in\Theta}$ is worst-case-sufficient for $(\overline{Q}_\theta)_{\theta\in\Theta}$ if and only if there is some $T \in \mathcal{T}(\mathcal{X},\mathcal{Y})$ so that $\forall\theta \in \Theta$

$$T(p_\theta)[g] \leqslant \overline{Q}_\theta[g], \qquad \forall g \in \mathcal{L}_\infty(\mathcal{Y},\mathcal{B})$$

2.3. Standard models

Let the index set Θ be finite with cardinality n. In Section 2.2, we have defined an equivalence relation on the precise models with a fixed index set Θ. Each equivalence class contains a uniquely defined representative (called standard model later on) which has some nice properties.[2] This is the content of the following theorem.

Theorem 2.5. *Every precise model $\mathcal{F} = (q_\theta)_{\theta\in\Theta}$ on $(\mathcal{Y},\mathcal{B})$ admits a uniquely defined (σ-additive) probability measure $s^{\mathcal{F}}$ on $(\mathcal{U},\mathcal{C})$ so that $\mathrm{d}s_\theta^{\mathcal{F}} = n\iota_\theta\,\mathrm{d}s^{\mathcal{F}}$ defines a precise model $(s_\theta^{\mathcal{F}})_{\theta\in\Theta}$ on $(\mathcal{U},\mathcal{C})$ which is equivalent to $\mathcal{F}$* (cf. [17, Theorem 6.5]).

Sketch of proof. For $\mathcal{F} = (q_\theta)_{\theta\in\Theta}$, there is a uniquely defined localization $\tilde{\eta}_{\mathcal{F}}$ on $(\mathcal{U},\mathcal{C})$ (cf. [11, p. 33f]). Let ι_θ be defined as in Section 1.3. Then, $\mathrm{d}s_\theta^{\mathcal{F}} = \iota_\theta\,\mathrm{d}\tilde{\eta}_{\mathcal{F}}$ uniquely defines a probability measure $s_\theta^{\mathcal{F}}$. Put $s^{\mathcal{F}} = \frac{1}{n}\sum_\theta s_\theta^{\mathcal{F}}$.

Since $(s_\theta^{\mathcal{F}})_{\theta\in\Theta}$ and $\mathcal{F}$ have the same conical measure, they are equivalent in the sense of [11, p. 19] according to [11, p. 28 and p. 32f]; cf. also [18, Remark 4]. A twofold application of Lemma 8.1 implies that $(s_\theta^{\mathcal{F}})_{\theta\in\Theta}$ and $\mathcal{F}$ are also equivalent in the sense of Section 2.2 of the present article.

A detailed proof is contained in [17,19]. □

Analogously to [7], $s^{\mathcal{F}}$ is called *standard measure* and $(s_\theta^{\mathcal{F}})_{\theta\in\Theta}$ is called *standard (precise) model* of $\mathcal{F}$.

Standard models share two important properties: Firstly, they are defined on the very nice measurable space $(\mathcal{U},\mathcal{C})$ (cf. Section 1.3). Secondly, they consist of linear previsions s_θ which are σ-additive probability measures.

For the imprecise model $(\overline{Q}_\theta)_{\theta\in\Theta}$ with corresponding structures $\mathcal{M}_\theta$, we can uniquely define

$$\overline{S}[h] = \sup\{s^{\mathcal{F}}[h] | \mathcal{F} \in (\mathcal{M}_\theta)_{\theta\in\Theta}\} \quad \forall h \in \mathcal{L}_\infty(\mathcal{U},\mathcal{C})$$
$$\overline{S}_\theta[h] = \sup\left\{s_\theta^{\mathcal{F}}[h] | \mathcal{F} \in (\mathcal{M}_\theta)_{\theta\in\Theta}\right\} \ \forall h \in \mathcal{L}_\infty(\mathcal{U},\mathcal{C})$$

$\overline{S}$ is called *standard upper prevision*, $(\overline{S}_\theta)_{\theta\in\Theta}$ is called *standard imprecise model* of $(\overline{Q}_\theta)_{\theta\in\Theta}$. Note that $\overline{S}$ is a coherent upper prevision on $\mathcal{L}_\infty(\mathcal{U},\mathcal{C})$ and $(\overline{S}_\theta)_{\theta\in\Theta}$ is an imprecise model on $(\mathcal{U},\mathcal{C})$.

3. Minimal Bayes risks

Let the index set $\Theta = \{\theta_1,\ldots,\theta_n\}$ be finite with cardinality n and let π be a prior distribution on $(\Theta, 2^\Theta)$, i.e. π is a linear prevision on $\mathcal{L}_\infty(\Theta, 2^\Theta)$. Put $\pi_\theta := \pi[I_{\{\theta\}}]$.

A *decision space* is a measurable space $(\mathbb{D},\mathcal{D})$ where $\mathbb{D}$ is the set of possible decisions. A loss function is a family $(W_\theta)_{\theta\in\Theta} \subset \mathcal{L}_\infty(\mathbb{D},\mathcal{D})$.

The measurable space $(\mathcal{Y},\mathcal{B})$ may represent the results of an experiment. According to [10], a *decision procedure* is a randomization

$$\sigma : \mathrm{ba}(\mathcal{Y},\mathcal{B}) \to \mathrm{ba}(\mathbb{D},\mathcal{D}), \qquad \sigma \in \mathcal{T}(\mathcal{Y},\mathcal{D})$$

Now, Bayes risks can be defined for precise models (Section 3.1) and for imprecise models (Section 3.2). The main goal of the present section is to express minimal Bayes risks in terms of standard measures (Theorem 3.2) and standard upper previsions (Theorem 3.4).

3.1. Precise models

Let $(q_\theta)_{\theta\in\Theta}$ be a precise model on $(\mathcal{Y},\mathcal{B})$. For a decision procedure $\sigma \in \mathcal{T}(\mathcal{Y},\mathbb{D})$ and a loss function $(W_\theta)_{\theta\in\Theta} \subset \mathcal{L}_\infty(\mathbb{D},\mathcal{D})$, the *risk function* of $(q_\theta)_{\theta\in\Theta}$ is

$$\sigma(q)[W_\cdot] : \theta \mapsto \sigma(q_\theta)[W_\theta]$$

The *Bayes risk* is

[2] As stated in Section 2.2, equivalent models coincide from a decision theoretic point of view. Therefore, every decision problem coincides with a "standard decision problem" where a standard model is involved; properties of the original decision problem can be deduced from the corresponding "standard decision problem".

$$R((q_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta}) = \pi[\sigma(q)[W_\cdot]] = \sum_{\theta\in\Theta}\pi_\theta\sigma(q_\theta)[W_\theta]$$

Note that this definition coincides with the usual one if σ is defined by a randomized decision function as in (1).

The minimal Bayes risk is the same if we let σ vary among the randomizations or the restricted randomizations:

Proposition 3.1

$$\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})} R((q_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta}) = \inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})} R((q_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta})$$

Proof. The definition of the topology of pointwise convergence implies continuity of the map

$$\sigma\mapsto(\sigma(q_{\theta_1})[W_{\theta_1}],\ldots,\sigma(q_{\theta_n})[W_{\theta_n}])$$

and, therefore, continuity of $\sigma\mapsto R((q_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta})$.

Since $\mathscr{T}_r(\mathscr{Y},\mathbb{D})$ is dense in $\mathscr{T}(\mathscr{Y},\mathbb{D})$ (Theorem 2.4), the statement follows. $\square$

For $(W_\theta)_{\theta\in\Theta}\subset\mathscr{L}_\infty(\mathbb{D},\mathscr{D})$, put

$$K((W_\theta)_\theta): u\mapsto \inf_{\tau\in\mathbb{D}}\sum_{\theta\in\Theta} n\pi_\theta W_\theta(\tau)\imath_\theta(u) \tag{2}$$

on $\mathbb{R}^n$ where $\imath_\theta(u)=u_\theta$ is the θ-component of $u\in\mathbb{R}^\Theta\cong\mathbb{R}^n$. Note that $K((W_\theta)_\theta)$ is concave and, therefore, continuous on $\mathbb{R}^n$. Hence, the restriction of $K((W_\theta)_\theta)$ on $\mathscr{U}$ is Borel-measurable and $s^{(q_\theta)_\theta}[K((W_\theta)_\theta)]$ is well defined where $s^{(q_\theta)_\theta}$ is the standard measure of $(q_\theta)_{\theta\in\Theta}$.

Theorem 3.2

$$\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})} R((q_\theta)_\theta,\sigma,(W_\theta)_\theta) = s^{(q_\theta)_\theta}\left[K((W_\theta)_\theta)\right]$$

Proof. According to Theorem 2.5, the standard model $(s_\theta^{\mathscr{F}})_{\theta\in\Theta}$ is equivalent to $\mathscr{F}:=(q_\theta)_{\theta\in\Theta}$. That is $(s_\theta^{\mathscr{F}})_{\theta\in\Theta}$ and $\mathscr{F}$ are mutual sufficient. So, a twofold application of Lemma 8.3 yields

$$\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})} R(\mathscr{F},\sigma,(W_\theta)_\theta) = \inf_{\rho\in\mathscr{T}(\mathscr{U},\mathbb{D})} R((s_\theta^{\mathscr{F}})_\theta,\rho,(W_\theta)_\theta)$$

and an application of Lemma 8.2 closes the proof. $\square$

3.2. Imprecise models

Let $(\overline{Q}_\theta)_{\theta\in\Theta}$ be an imprecise model on $(\mathscr{Y},\mathscr{B})$ with corresponding structures $\mathscr{M}_\theta$, $\theta\in\Theta$, and standard upper prevision $\overline{S}$. For a decision procedure $\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})$ and a loss function $(W_\theta)_{\theta\in\Theta}\subset\mathscr{L}_\infty(\mathbb{D},\mathscr{D})$, the *risk function* of $(\overline{Q}_\theta)_{\theta\in\Theta}$ is

$$\theta\mapsto\sup_{q_\theta\in\mathscr{M}_\theta}\sigma(q_\theta)[W_\theta]$$

and the Bayes risk is

$$R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta) = \sum_{\theta\in\Theta}\pi_\theta\sup_{q_\theta\in\mathscr{M}_\theta}\sigma(q_\theta)[W_\theta]$$

Hence,

$$R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta) = \sup_{(q_\theta)_\theta\in(\mathscr{M}_\theta)_\theta} R((q_\theta)_\theta,\sigma,(W_\theta)_\theta)$$

These definitions assume that we have chosen the Γ-minimax optimality criterion which represents a worst-case consideration (cf. Section 1.2) – as done in [8,7].

Now, we can derive the analogues of Proposition 3.1 and Theorem 3.2 in case of imprecise models:

Proposition 3.3

$$\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})} R((\overline{Q}_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta}) = \inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})} R((\overline{Q}_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta})$$

Proof. This is a direct consequence of Lemma 8.4 (a), Proposition 3.1 and Lemma 8.4 (b). $\square$

Theorem 3.4

$$\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})} R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta) = \overline{S}[K((W_\theta)_\theta)]$$

Proof. This is a direct consequence of Lemma 8.4, Theorem 3.2 and the definition of the standard upper prevision. $\square$

4. The general Le Cam–Blackwell–Sherman–Stein–theorem

This section contains a generalization of the Le Cam–Blackwell–Sherman–Stein–Theorem. It is needed in the proof of the main theorem, Theorem 5.4.

Let Θ be a finite index set. Let π be a prior distribution on $(\Theta, 2^\Theta)$ so that $\pi_\theta := \pi[I_{\{\theta\}}] > 0\ \forall \theta \in \Theta$. Let $(p_\theta)_{\theta\in\Theta}$ be a precise model on $(\mathscr{X}, \mathscr{A})$ and $(\overline{Q}_\theta)_{\theta\in\Theta}$ an imprecise model on $(\mathscr{Y}, \mathscr{B})$ where $(\mathscr{M}_\theta)_{\theta\in\Theta}$ is the corresponding family of structures. Let $s^{(p_\theta)_\theta}$ be the standard measure of $(p_\theta)_{\theta\in\Theta}$ and $\overline{S}$ the standard upper prevision of $(\overline{Q}_\theta)_{\theta\in\Theta}$ on $(\mathscr{U}, \mathscr{C})$.

Let Ψ be the set of all functions $k \in \mathscr{L}_\infty(\mathscr{U}, \mathscr{C})$ such that there is some decision space $(\mathbb{D}, \mathscr{D})$ and a loss function $(W_\theta)_{\theta\in\Theta} \subset \mathscr{L}_\infty(\mathbb{D}, \mathscr{D})$ where $k(u) = \inf_{\tau\in\mathbb{D}} \sum_{\theta\in\Theta} n\pi_\theta W_\theta(\tau)\iota_\theta(u)\ \forall u \in \mathscr{U}$.

Theorem 4.1 is the analogue to [7, Theorem 7.1], the proof is similar.

Theorem 4.1. *The following statements are equivalent:*

(a) $(p_\theta)_{\theta\in\Theta}$ *is worst-case-sufficient for* $(\overline{Q}_\theta)_{\theta\in\Theta}$.
(b) $s^{(p_\theta)_\theta}[k] \leqslant \overline{S}[k] \qquad \forall k \in \Psi$
(c) For every finite decision space $(\mathbb{D}, \mathscr{D})$ *and every loss function* $(W_\theta)_{\theta\in\Theta} \subset \mathscr{L}_\infty(\mathbb{D}, \mathscr{D})$,

$$\inf_{\rho\in\mathscr{T}(\mathscr{X},\mathbb{D})} R((p_\theta)_\theta, \rho, (W_\theta)_\theta) \leqslant \inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})} R((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$$

(d) For every decision space $(\mathbb{D}, \mathscr{D})$ *and every loss function* $(W_\theta)_{\theta\in\Theta} \subset \mathscr{L}_\infty(\mathbb{D}, \mathscr{D})$,

$$\inf_{\rho\in\mathscr{T}(\mathscr{X},\mathbb{D})} R((p_\theta)_\theta, \rho, (W_\theta)_\theta) \leqslant \inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})} R((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$$

Proof. The proof has the following structure: (a) $\Longleftrightarrow$ (d), (d) $\Longleftrightarrow$ (c), (d) $\Longleftrightarrow$ (b)

$(a) \Rightarrow (d)$: This is a direct consequence of Lemma 8.3.

$(a) \Leftarrow (d)$: Put $\mathbb{D} = \mathscr{Y}$ and $\sigma_0(\mu) = \mu\ \forall \mu \in \mathrm{ba}(\mathscr{Y}, \mathscr{B})$. Then (d) implies that for all $(g_\theta)_{\theta\in\Theta} \subset \mathscr{L}_\infty(\mathscr{Y}, \mathscr{B})$,

$$\inf_{T\in\mathscr{T}(\mathscr{X},\mathscr{Y})} R((p_\theta)_\theta, T, (g_\theta)_\theta) \leqslant R((\overline{Q}_\theta)_\theta, \sigma_0, (g_\theta)_\theta)$$

which may be rewritten as $\inf_{T\in\mathscr{T}(\mathscr{X},\mathscr{Y})} \sum_{\theta\in\Theta} \pi_\theta (T(p_\theta)[g_\theta] - \overline{Q}_\theta[g_\theta]) \leqslant 0$.

Put $\Gamma(T, (g_\theta)_\theta) := \sum_{\theta\in\Theta} \pi_\theta (T(p_\theta)[g_\theta] - \overline{Q}_\theta[g_\theta])$. Then,

$$sup_{(g_\theta)_\theta\subset\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})} \inf_{T\in\mathscr{T}(\mathscr{X},\mathscr{Y})} \Gamma(T, (g_\theta)_\theta) \leqslant 0 \tag{3}$$

$\mathscr{T}(\mathscr{X}, \mathscr{Y})$ is compact, $T \mapsto \Gamma(T, (g_\theta)_\theta)$ is continuous and convex, $(g_\theta)_\theta \mapsto \Gamma(T, (g_\theta)_\theta)$ is concave. So, the minimax theorem [20, Theorem 2] and (3) yield

$$\inf_{T\in\mathscr{T}(\mathscr{X},\mathscr{Y})} \sup_{(g_\theta)_\theta\subset\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})} \Gamma(T, (g_\theta)_\theta) \leqslant 0$$

Compactness of $\mathscr{T}(\mathscr{X}, \mathscr{Y})$ and lower semicontinuity of

$$T \mapsto \sup_{(g_\theta)_\theta\subset\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})} \Gamma(T, (g_\theta)_\theta)$$

imply the existence of some $T_0 \in \mathscr{T}(\mathscr{X}, \mathscr{Y})$ so that

$$\sup_{(g_\theta)_\theta\subset\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})} \Gamma(T_0, (g_\theta)_\theta) \leqslant 0 \tag{4}$$

(cf. [21, Theorem 3.7]). Since $\pi_\theta > 0\ \forall \theta \in \Theta$, it follows from (4) that

$$T_0(p_\theta)[g_\theta] \leqslant \overline{Q}[g_\theta] \qquad \forall g_\theta \in \mathscr{L}_\infty(\mathscr{Y}, \mathscr{B}) \qquad \forall \theta \in \Theta$$

$(d) \Rightarrow (c)$: This is obvious.

$(d) \Leftarrow (c)$: Let $\sigma : \mu \mapsto \kappa^*(\mu)$ be a restricted randomization from $\mathscr{Y}$ to $\mathbb{D}$ where

$$\kappa^*(\mu)[g] = \mu\left[\sum_{t\in D} g(t)\alpha_t\right]$$

and D is a finite subset of $\mathbb{D}$. $(D, 2^D)$ may be regarded as a finite decision space and σ may be regarded as an element of $\mathscr{T}(\mathscr{Y}, D)$. Then, (c) implies

$$\inf_{\hat{\rho}\in\mathscr{T}(\mathscr{X},D)} R((p_\theta)_\theta, \hat{\rho}, (W_\theta)_\theta) \leqslant R((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) \tag{5}$$

Since every element of $\mathscr{T}_r(\mathscr{X}, D)$ may be regarded as an element of $\mathscr{T}_r(\mathscr{X}, \mathbb{D})$, Proposition 3.1 implies

$$\inf_{\rho \in \mathscr{T}(\mathscr{X},\mathbb{D})} R((p_\theta)_\theta, \rho, (W_\theta)_\theta) \leqslant \inf_{\hat{\rho} \in \mathscr{T}(\mathscr{X},D)} R((p_\theta)_\theta, \hat{\rho}, (W_\theta)_\theta) \tag{6}$$

Hence, (according to Proposition 3.3)

$$\inf_{\rho \in \mathscr{T}(\mathscr{X},\mathbb{D})} R((p_\theta)_\theta, \rho, (W_\theta)_\theta) \overset{(6,5)}{\leqslant} \inf_{\sigma \in \mathscr{T}_r(\mathscr{Y},\mathbb{D})} R((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) = \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} R((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$$

$(d) \iff (b)$: This is a direct consequence of Theorems 3.2 and 3.4. □

5. Least favorable models

Let again the index set Θ be finite with cardinality n. Let π be a prior distribution on $(\Theta, 2^\Theta)$ so that $\pi_\theta := \pi[I_{\{\theta\}}] > 0 \ \forall \theta \in \Theta$. Let $(\overline{Q}_\theta)_{\theta\in\Theta}$ be an imprecise model on $(\mathscr{Y}, \mathscr{B})$ where $(\mathscr{M}_\theta)_{\theta\in\Theta}$ is the corresponding family of structures. Let $(\mathbb{D}, \mathscr{D})$ be a fixed decision space and let $\mathscr{W}$ be a set of loss functions $(W_\theta)_{\theta\in\Theta} \subset \mathscr{L}_\infty(\mathbb{D}, \mathscr{D})$.

Definition 5.1. $(q_\theta)_{\theta\in\Theta} \in (\mathscr{M}_\theta)_{\theta\in\Theta}$ is called *least favorable (precise) model of* $(\mathscr{M}_\theta)_{\theta\in\Theta}$ *for* $\mathscr{W}$ if

$$\inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} R((q_\theta)_\theta, \sigma, (W_\theta)_\theta) = \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} R((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) \quad \forall (W_\theta)_\theta \in \mathscr{W}$$

That is the minimal Bayes risk of the imprecise model is attained in the least favorable model which represents the worst-case. (This justifies the term "least favorable".) Remember that our definition of the Bayes risk corresponds to a worst-case consideration.

We are not primarily interested in a set of loss functions but in a set of prior distributions. However, a set of prior distributions can always be transformed into a set of loss functions (cf. Section 6).

For $\mathscr{F} \in (\mathscr{M}_\theta)_{\theta\in\Theta}$, put

$$\Phi_{\mathscr{F}} := \{h \in \mathscr{L}_\infty(\mathscr{U}, \mathscr{C}) | s^{\mathscr{F}}[h] = \overline{S}[h]\}$$

where $s^{\mathscr{F}}$ is the standard measure of $\mathscr{F}$ and $\overline{S}$ is the standard upper prevision of $(\overline{Q}_\theta)_{\theta\in\Theta}$ on $(\mathscr{U}, \mathscr{C})$.

Lemma 5.2. *$\Phi_{\mathscr{F}}$ is a norm-closed convex cone in $\mathscr{L}_\infty(\mathscr{U}, \mathscr{C})$.*

Proof. For $h \in \Phi_{\mathscr{F}}$ and $c \in [0, \infty)$, $\overline{S}[ch] = c\overline{S}[h] = cs^{\mathscr{F}}[h] = s^{\mathscr{F}}[ch]$.

For $h_1, h_2 \in \Phi_{\mathscr{F}}$,

$$\overline{S}[h_1 + h_2] \leqslant \overline{S}[h_1] + \overline{S}[h_2] = s^{\mathscr{F}}[h_1] + s^{\mathscr{F}}[h_2] = s^{\mathscr{F}}[h_1 + h_2] \leqslant \overline{S}[h_1 + h_2]$$

For $(h_m)_{m\in\mathbb{N}} \subset \Phi_{\mathscr{F}}$, $\lim_m \|h_m - h\| = 0$ and $h \in \mathscr{L}_\infty(\mathscr{U}, \mathscr{C})$,

$$\overline{S}[h] \leqslant \limsup_m (\overline{S}[h_m] + \overline{S}[h - h_m]) = \limsup_m s^{\mathscr{F}}[h_m] = s^{\mathscr{F}}[h]$$

i.e. $s^{\mathscr{F}}[h] = \overline{S}[h]$. □

For every $(W_\theta)_{\theta\in\Theta} \subset \mathscr{L}_\infty(\mathbb{D}, \mathscr{D})$, define $K((W_\theta)_\theta)$ as in (2).

$$\Psi_{\mathscr{W}} := \{K((W_\theta)_\theta) | (W_\theta)_\theta \in \mathscr{W}\} \subset \mathscr{L}_\infty(\mathscr{U}, \mathscr{C})$$

$\widetilde{\Psi}_{\mathscr{W}}$ denotes the smallest norm-closed convex cone in $\mathscr{L}_\infty(\mathscr{U}, \mathscr{C})$ which contains $\Psi_{\mathscr{W}}$. Lemma 5.3 is a direct consequence of Theorems 3.2 and 3.4.

Lemma 5.3. *$\mathscr{F} \in (\mathscr{M}_\theta)_{\theta\in\Theta}$ is least favorable for $\mathscr{W}$ if and only if*

$$s^{\mathscr{F}}[k] = \overline{S}[k] \qquad \forall\, k \in \Psi_{\mathscr{W}}$$

Theorem 5.4 is the analogue to [7, Theorem 8.2]. It characterizes the existence of least favorable models in full generality.

Theorem 5.4. *The following statements are equivalent:*

(a) There is some $\mathscr{F} := (q_\theta)_{\theta\in\Theta} \in (\mathscr{M}_\theta)_{\theta\in\Theta}$ which is least favorable for $\mathscr{W}$.
(b) $\overline{S}[k_1 + k_2] = \overline{S}[k_1] + \overline{S}[k_2] \qquad \forall\, k_1, k_2 \in \widetilde{\Psi}_{\mathscr{W}}$

Proof

$(a) \Rightarrow (b)$: Statement (a) and Lemma 5.3 imply $\Psi_{\mathscr{W}} \subset \Phi_{\mathscr{F}}$. According to Lemma 5.2, $\widetilde{\Psi}_{\mathscr{W}} \subset \Phi_{\mathscr{F}}$ and $k_1 + k_2 \in \Phi_{\mathscr{F}}\ \forall\, k_1, k_2 \in \widetilde{\Psi}_{\mathscr{W}}$. Hence, for every $k_1, k_2 \in \widetilde{\Psi}_{\mathscr{W}}$

$$\overline{S}[k_1+k_2] = s^{\mathscr{F}}[k_1+k_2] = s^{\mathscr{F}}[k_1] + s^{\mathscr{F}}[k_2] = \overline{S}[k_1] + \overline{S}[k_2]$$

$(a) \Leftarrow (b)$: Put $s[k] := \overline{S}[k]\ \forall\ k \in \widetilde{\Psi}_{\mathscr{W}}$ and

$$s[k_1 - k_2] := s[k_1] - s[k_2] = \overline{S}[k_1] - \overline{S}[k_2]$$

for all $k_1, k_2 \in \widetilde{\Psi}_{\mathscr{W}}$. Statement (b) implies that this is defined well. Hence, s is a linear functional on the vector space $\operatorname{lin}(\widetilde{\Psi}_{\mathscr{W}}) = \widetilde{\Psi}_{\mathscr{W}} - \widetilde{\Psi}_{\mathscr{W}}$. For every $k = k_1 - k_2 \in \widetilde{\Psi}_{\mathscr{W}} - \widetilde{\Psi}_{\mathscr{W}} = \operatorname{lin}(\widetilde{\Psi}_{\mathscr{W}})$,

$$s[k] = \overline{S}[k_2 + k_1 - k_2] - \overline{S}[k_2] \ \leqslant\ \overline{S}[k_2] + \overline{S}[k_1 - k_2] - \overline{S}[k_2] = \overline{S}[k]$$

According to the Hahn–Banach-Theorem [14, Theorem II.3.10], s can be extended to a linear functional on $\mathscr{L}_\infty(\mathscr{U}, \mathscr{C})$ (again denoted by s) so that

$$s[h] \ \leqslant\ \overline{S}[h] \qquad \forall\ h \in \mathscr{L}_\infty(\mathscr{U}, \mathscr{C}) \tag{7}$$

(7) implies, that $s[I_{\mathscr{U}}] = 1$ and $s[\iota_\theta] = \frac{1}{n} \forall\ \theta \in \Theta$ (cf. Theorem 2.5). Then, $s_\theta : h \mapsto s[n\iota_\theta h]$ defines a precise model $(s_\theta)_{\theta\in\Theta}$ on $(\mathscr{U}, \mathscr{C})$. For every decision space $(\widehat{\mathbb{D}}, \widehat{\mathscr{D}})$ and every $(\widehat{W}_\theta)_\theta \subset \mathscr{L}_\infty(\widehat{\mathbb{D}}, \widehat{\mathscr{D}})$,

$$\inf_{\rho \in \mathscr{T}(\mathscr{U}, \widehat{\mathbb{D}})} R((s_\theta)_\theta, \rho, (\widehat{W}_\theta)_\theta) = s[K((\widehat{W}_\theta)_\theta)] \tag{8}$$

according to Lemma 8.2 and

$$\inf_{\rho \in \mathscr{T}(\mathscr{U}, \widehat{\mathbb{D}})} R((s_\theta)_\theta, \rho, (\widehat{W}_\theta)_\theta) \overset{(8)}{=} s[K((\widehat{W}_\theta)_\theta)] \overset{(7)}{\leqslant} \overline{S}[K((\widehat{W}_\theta)_\theta)] = \inf_{\sigma \in \mathscr{T}(\mathscr{Y}, \widehat{\mathbb{D}})} R((\overline{Q}_\theta)_\theta, \sigma, (\widehat{W}_\theta)_\theta)$$

according to Theorem 3.4. Hence, Theorem 4.1 implies that $(s_\theta)_{\theta\in\Theta}$ is worst-case-sufficient for $(\overline{Q}_\theta)_{\theta\in\Theta}$, i.e. there is some $T \in \mathscr{T}(\mathscr{U}, \mathscr{Y})$ so that $q_\theta := T(s_\theta) \in \mathscr{M}_\theta \forall \theta \in \Theta$. Finally for all $(W_\theta)_{\theta\in\Theta} \in \mathscr{W}$,

$$\inf_{\sigma \in \mathscr{T}(\mathscr{Y}, \mathbb{D})} R((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) = \overline{S}[K((W_\theta)_\theta)] = s[K((W_\theta)_\theta)] \overset{(8)}{=} \inf_{\rho \in \mathscr{T}(\mathscr{U}, \mathbb{D})} R((s_\theta)_\theta, \rho, (W_\theta)_\theta) \ \leqslant\ \inf_{\sigma \in \mathscr{T}(\mathscr{Y}, \mathbb{D})} R((q_\theta)_\theta, \sigma, (W_\theta)_\theta)$$

where the last inequality follows from Lemma 8.3. □

6. Application of least favorable models

Situations where we are faced with one precise prior distribution and a set of loss functions seem to be of secondary interest. More frequently, we are interested in situations where we are faced with an *imprecise* prior and one fixed loss function. However, the second issue can be treated as a special case of the first one.

Let Θ be a finite index set with cardinality n and $(W_\theta)_{\theta\in\Theta} \subset \mathscr{L}_\infty(\mathbb{D}, \mathscr{D})$ be a loss function. Let $(\overline{Q}_\theta)_{\theta\in\Theta}$ be an imprecise model on $(\mathscr{Y}, \mathscr{B})$ where $(\mathscr{M}_\theta)_{\theta\in\Theta}$ is the corresponding family of structures. Let $\overline{\Pi}$ be a coherent upper prevision on $\mathscr{L}_\infty(\Theta, 2^\Theta)$ i.e. $\overline{\Pi}$ corresponds to a set of prior distributions $\mathscr{P} := \{\pi \in \operatorname{ba}(\Theta, 2^\Theta) |\ \pi[a] \ \leqslant\ \overline{\Pi}[a]\ \forall a \in \mathscr{L}_\infty(\Theta, 2^\Theta)\}$.

For some $\pi \in \mathscr{P}$, put $\pi_\theta := \pi[I_{\{\theta\}}]\ \forall \theta \in \Theta$. Let σ be a randomization. For the prior π, the Bayes risk is

$$R_\pi((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) = \sum_{\theta\in\Theta} \pi_\theta \sigma(\overline{Q}_\theta)[W_\theta] = \frac{1}{n} \sum_{\theta\in\Theta} \sigma(\overline{Q}_\theta)[n\pi_\theta W_\theta] = R_0((\overline{Q}_\theta)_\theta, \sigma, (n\pi_\theta W_\theta)_\theta)$$

where $R_0((\overline{Q}_\theta)_\theta, \sigma, (n\pi_\theta W_\theta)_\theta)$ denotes the Bayes risk for the uniform prior π_0 defined by $\pi_0[I_\theta] = \frac{1}{n}$.

That is every prior can be absorbed in the loss function. So, we can transform the set $\mathscr{P}$ of priors π into a set $\mathscr{W}$ of loss functions $(n\pi_\theta W_\theta)_{\theta\in\Theta}$. Next, Theorem 5.4 yields a necessary and sufficient condition for the existence of a precise model which is simultaneously least favorable for the set of loss functions $\mathscr{W}$. We may also say that such a precise model is *simultaneously least favorable for the set of priors* $\mathscr{P}$.

The next theorem shows how least favorable models can be used to deal with situations where the distribution of the data as well as the prior is assumed to be imprecise. A decision procedure is optimal if it minimizes the upper Bayes risk

$$R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) = \sup_{\pi\in\mathscr{P}} R_\pi((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$$

Theorem 6.1. *If $(\tilde{q}_\theta)_{\theta\in\Theta}$ is a simultaneously least favorable model of $(\mathscr{M}_\theta)_{\theta\in\Theta}$ for $\mathscr{P}$, there is a decision procedure $\tilde{\sigma} \in \mathscr{T}(\mathscr{Y}, \mathbb{D})$ which minimizes*

$$R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) \qquad \text{and also} \qquad R_{\overline{\Pi}}((\tilde{q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$$

over $\mathscr{T}(\mathscr{Y}, \mathbb{D})$.

Proof. For every $\sigma \in \mathscr{T}(\mathscr{Y}, \mathbb{D})$ and $\pi \in \mathscr{P}$, put

$$\Gamma_1(\sigma, \pi) = R_\pi((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) \qquad \text{and} \qquad \Gamma_2(\sigma, \pi) = R_\pi((\tilde{q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$$

It is easy to see that $\sigma \mapsto \Gamma_j(\sigma, \pi)$ is convex and lower semicontinuous for every $\pi \in \mathscr{P}$ and $j \in \{1, 2\}$. Then, [20, Theorem 2] and simultaneous least favorability implies

$$\begin{aligned}\inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) &= \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} \sup_{\pi \in \mathscr{P}} \Gamma_1(\sigma, \pi) = \sup_{\pi \in \mathscr{P}} \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} \Gamma_1(\sigma, \pi) = \sup_{\pi \in \mathscr{P}} \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} \Gamma_2(\sigma, \pi) \\ &= \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} \sup_{\pi \in \mathscr{P}} \Gamma_2(\sigma, \pi) = \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} R_{\overline{\Pi}}((\tilde{q}_\theta)_\theta, \sigma, (W_\theta)_\theta)\end{aligned} \tag{9}$$

Lower semicontinuity of

$$\sigma \mapsto R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$$

and compactness of $\mathscr{T}(\mathscr{Y}, \mathbb{D})$ ensure existence of some $\tilde{\sigma}$ which minimizes $R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$ (cf. [21, Theorem 3.7]). Additionally,

$$R_{\overline{\Pi}}((\tilde{q}_\theta)_\theta, \tilde{\sigma}, (W_\theta)_\theta) \leqslant R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \tilde{\sigma}, (W_\theta)_\theta) = \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta) \overset{(9)}{=} \inf_{\sigma \in \mathscr{T}(\mathscr{Y},\mathbb{D})} R_{\overline{\Pi}}((\tilde{q}_\theta)_\theta, \sigma, (W_\theta)_\theta) \quad \square$$

Remark 6.2. It can easily be read off from the above proof that a decision procedure $\tilde{\sigma}$ which minimizes $R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$ minimizes $R_{\overline{\Pi}}((\tilde{q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$, too. However, the reverse statement will not always be true.[3] So, it does not suffice to find a decision procedure $\hat{\sigma}$ which minimizes $R_{\overline{\Pi}}((\tilde{q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$. It still has to be checked that $\hat{\sigma}$ really minimizes $R_{\overline{\Pi}}((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta)$. Theorem 6.1 only states that there is a decision procedure which solves both minimization problems.

7. Concluding remarks

In decision theory, straightforward updating may lead to decisions with a too high risk if the data is distributed according to imprecise probabilities (cf. [4]). Therefore, data-based decision theory can be seen as a matter of its own. One of the major problems in data-based decision theory is that direct solutions of the involved optimization problems are quite often computationally intractable. Theorem 6.1 offers an opportunity to reduce the computational effort significantly if the imprecise model admits a least favorable (precise) model. Therefore, it is important to know for a given decision problem if such a least favorable model exists or not.

This question has been addressed by Buja [7]. The concept of imprecise probability developed in [7] is very close to that one developed in [1]. From a mathematical point of view, the only difference is that [7] assumes that precise probabilities (i.e. linear previsions) have to be σ-additive. Surprisingly, this turns out to be a burden which significantly reduces the applicability of [7]; cf. Remark 2.2 and Section 8.1. The present article shows that the same result as in [7] is possible without any assumption on the involved (coherent) upper previsions if we dispense with σ-additivity.

This offers a general tool which makes it possible to reduce the computational effort in data-based decision theory under imprecision. However, further research has to be done for using it in concrete problems: As in [8], Theorem 5.4 is only concerned with the *existence* of a least favorable model but an algorithm for calculating least favorable models has not yet been developed.

After [8], a lot of work was done to construct least favorable pairs in hypothesis testing for special cases (e.g. [23–25,22]). In the much more general case of the present article, this is a matter of further research.

The present article might not only be interesting because of its results but also because of the applied tools: Getting around σ-additivity in the proofs of the present paper was possible by the use of notions and methods of [11]. This article is probably the first one which explicitly uses concepts of [11] in the theory of imprecise probability. Since these concepts were especially developed for large models, it is most likely that they can profitably be used in the theory of imprecise probability further on. Additionally, a theory of "sufficiency" is used which is not formulated in terms of conditional probabilities. In this way, a sufficiency theory for imprecise probabilities may be possible which is not affected by the problems which arise for conditional imprecise probabilities.

8. Appendix

8.1. About an incorrect statement in [7]

This subsection deals with classical probability theory. Here, $\mathscr{Y}$ is a Polish space, $\mathscr{B}$ is the Borel-σ-algebra of $\mathscr{Y}$ and $\mathrm{ca}_1^+(\mathscr{Y}, \mathscr{B})$ denotes the set of all probability measures on $(\mathscr{Y}, \mathscr{B})$. Hence, the elements of $\mathrm{ca}_1^+(\mathscr{Y}, \mathscr{B})$ are σ-additive. For classical probability theory, confer e.g. [26].

Let $\mathscr{C}^{\mathrm{b}}(\mathscr{Y})$ be the set of all continuous, bounded functions $g : Y \to \mathbb{R}$. As usual in this context, $\mathrm{ca}_1^+(\mathscr{Y}, \mathscr{B})$ is endowed with the weak topology of probability measures, i.e. the topology of pointwise convergence on $\mathscr{C}^{\mathrm{b}}(\mathscr{Y})$.

Proposition 2.1 in [7] contains the following statement:

Let $\mathscr{Q}$ be a *tight* subset of $\mathrm{ca}_1^+(\mathscr{Y}, \mathscr{B})$ and define

[3] In case of hypothesis testing, for example, this follows from [22, p. 162ff].

$$\overline{\mathrm{Q}}[g] = \sup\{q[g] | q \in \mathcal{Q}\} \qquad \forall g \in \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B})$$

Then, $(g_n)_{n\in\mathbb{N}} \subset \mathscr{C}^{\mathrm{b}}(\mathcal{Y}),\ g \in \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B}),\ g_n \searrow g(\text{pointwise})$ implies

$$\overline{\mathrm{Q}}[g_n] \searrow \overline{\mathrm{Q}}[g] \tag{10}$$

In general, this statement is not correct. As a counterexample, take $\mathcal{Y} = \mathbb{R},\ \mathcal{Q} = \{q \in \mathrm{ca}_1^+(\mathcal{Y}, \mathcal{B}) | q[I_{[0,1)}] = 1\}$ and $g_n : x \mapsto x^n I_{[0,1]}(x) + I_{(1,\infty)}(x)$.

In general, (10) follows if $g \in \mathscr{C}^{\mathrm{b}}(\mathcal{Y})$; cf. [27, Theorem II.25].

As a consequence, tightness of $\mathcal{Q}$ does not generally imply compactness (in the weak topology of probability measures) of

$$\mathcal{M} = \{q \in \mathrm{ca}_1^+(\mathcal{Y}, \mathcal{B}) | q[g] \leqslant \overline{\mathrm{Q}}[g]\ \forall\ g \in \mathcal{L}_\infty(\mathcal{Y}, \mathcal{B})\}$$

So, compactness of $\mathcal{M}$ is an additional assumption in [7] which is restrictive because it implies that

$$\mathcal{M} = \{q \in \mathrm{ca}_1^+(\mathcal{Y}, \mathcal{B}) |\ q[g] \leqslant \overline{\mathrm{Q}}[g]\ \forall\ g \in \mathscr{C}^{\mathrm{b}}(\mathcal{Y})\}$$

(The last assertion is a consequence of the Separation Theorem [14, Corollary V.2.13].)

8.2. Some lemmas

Lemma 8.1. *For $i \in \{1;2\}$, let $(\mathcal{X}_i, \mathcal{A}_i)$ be a measurable space, $\mathcal{E}_i = (p_{i,\theta})_{\theta\in\Theta}$ a precise model on $(\mathcal{X}_i, \mathcal{A}_i)$ and $L(\mathcal{E}_i)$ the smallest band in the L-space $\mathrm{ba}(\mathcal{X}_i, \mathcal{A}_i)$. Analogously to $\mathcal{T}(\mathcal{X}_1, \mathcal{X}_2)$, let $\mathcal{T}(\mathcal{E}_1, \mathcal{E}_2)$ be the set of all linear, positive, normalized maps $\widehat{T} : L(\mathcal{E}_1) \to L(\mathcal{E}_2)$. Then: There is some $T \in \mathcal{T}(\mathcal{X}_1, \mathcal{X}_2)$ such that $T(p_{1,\theta}) = p_{2,\theta}\ \forall\ \theta \in \Theta$ if and only if*

$$\inf_{\widehat{T}\in\mathcal{T}(\mathcal{E}_1,\mathcal{E}_2)} \sup_{\theta\in\Theta} \|\widehat{T}(p_{1,\theta}) - p_{2,\theta}\| = 0 \tag{11}$$

Proof. According to [11, Theorem 2.3.2], the existence of some $\widehat{T} \in \mathcal{T}(\mathcal{E}_1, \mathcal{E}_2)$ so that $\widehat{T}(p_{1,\theta}) = p_{2,\theta}\ \forall \theta \in \Theta$ is equivalent to (11).

Let $\widehat{T}$ be such an element of $\mathcal{T}(\mathcal{E}_1, \mathcal{E}_2)$. For any fixed $x_2 \in \mathcal{X}_2$, $S(\mu)[g] := g(x_2)\mu[I_{\mathcal{X}_1}]$ defines an element of $\mathcal{T}(\mathcal{X}_1, \mathcal{X}_2)$. Let $\Pi_{L(\mathcal{E}_1)}$ be the projection of $\mathrm{ba}(\mathcal{X}_1, \mathcal{A}_1)$ on the band $L(\mathcal{E}_1)$. Some simple calculations show that

$$T(\mu) = (\widehat{T} \circ \Pi_{L(\mathcal{E})})(\mu) + (S - S \circ \Pi_{L(\mathcal{E})})(\mu)$$

defines an extension of $\widehat{T}$ to an element of $\mathcal{T}(\mathcal{X}_1, \mathcal{X}_2)$.

The converse statement follows from [10, Proposition 7]. □

Lemma 8.2. *Assume that s is a linear prevision on $\mathcal{L}_\infty(\mathcal{U}, \mathcal{C})$ so that $s[\imath_\theta] = \frac{1}{n}\ \forall\ \theta \in \Theta$. Then, $s_\theta :\ h \mapsto s[n\imath_\theta h]$ defines a precise model $(s_\theta)_{\theta\in\Theta}$ on $(\mathcal{U}, \mathcal{C})$ and*

$$\inf_{\rho\in\mathcal{T}(\mathcal{U},\mathbb{D})} R((s_\theta)_\theta, \rho, (W_\theta)_\theta) = s[K((W_\theta)_\theta)] \tag{12}$$

for every decision space $(\mathbb{D}, \mathcal{D})$ and every $(W_\theta)_\theta \subset \mathcal{L}_\infty(\mathbb{D}, \mathcal{D})$. $K((W_\theta)_\theta)$ is defined as in (2).

Proof. Obviously, $(s_\theta)_{\theta\in\Theta}$ is a precise model on $(\mathcal{U}, \mathcal{C})$. Statement (12) is proven by two steps:

1. Let $(\widehat{W}_\theta)_{\theta\in\Theta} \subset \mathcal{L}_\infty(\mathbb{D}, \mathcal{D})$ be a family of simple functions. Since Θ is finite, there is a finite subset $\widehat{S} := \{t_1, \ldots, t_m\} \subset \mathbb{D}$ so that

$$\{(\widehat{W}_\theta(t))_{\theta\in\Theta} | t \in \widehat{S}\} = \{(\widehat{W}_\theta(t))_{\theta\in\Theta} | t \in \mathbb{D}\}$$

Let the elements of the set A be the families $(\alpha_t)_{t\in S} \subset \mathcal{L}_\infty(\mathcal{U}, \mathcal{C})$ where S is a finite subset of $\mathbb{D}$, $\alpha_t \geqslant 0\ \forall\ t \in S$ and $\sum_{t\in S}\alpha_t \equiv 1$.

Put $\Gamma_t(u) = \sum_{\theta\in\Theta} n\pi_\theta \widehat{W}_\theta(t)\imath_\theta(u)$, thus $\inf_{\tau\in\mathbb{D}}\Gamma_\tau = K((\widehat{W}_\theta)_\theta)$.

For $j \in \{1, \ldots, m\}$, let V_j be the set of elements $u \in \mathcal{U}$ so that $\Gamma_{t_j}(u) = \inf_{\tau\in\mathbb{D}}\Gamma_\tau(u)$,

$$U_j := V_j \setminus (\cup_{l=1}^{j-1} V_l) \qquad \text{and} \qquad \hat{\alpha}_{t_j} = I_{U_j}, \qquad j = 1, \ldots, m$$

Note that $U_j \in \mathcal{C}$. The definition of $\{t_1, \ldots, t_m\}$ ensures that $(U_j)_{j=1,\ldots,m}$ is a partition of $\mathcal{U}$. Hence, $\sum_{t\in\widehat{S}}\hat{\alpha}_t \equiv 1$ and $(\hat{\alpha}_t)_{t\in\widehat{S}} \in A$. Furthermore,

$$\sum_{t\in\widehat{S}} \hat{\alpha}_t(u)\Gamma_t(u) = \inf_{\tau\in\mathbb{D}} \Gamma_\tau(u) \tag{13}$$

Let $\hat{\rho}$ be the restricted randomization which corresponds to $(\hat{\alpha}_t)_{t\in\widehat{S}} \in A$. Then,

$$\sum_{\theta\in\Theta} \pi_\theta \hat{\rho}(s_\theta)[\widehat{W}_\theta] \overset{(13)}{=} \int \inf_{\tau\in\mathbb{D}} \Gamma_\tau(u) s(\mathrm{d}u) = s[K((W_\theta)_\theta)] \tag{14}$$

So, (12) follows from (14) and

$$\inf_{\rho\in\mathscr{T}(\mathscr{U},\mathbb{D})}\sum_{\theta\in\Theta}\pi_\theta\rho(s_\theta)[\widehat{W}_\theta] \overset{\text{Proposition 3.1}}{=} \inf_{(\alpha_t)_{t\in S}\in A}\sum_{\theta\in\Theta}\pi_\theta s_\theta\left[\sum_{t\in S}\widehat{W}_\theta(t)\alpha_t\right] = \inf_{(\alpha_t)_{t\in S}\in A}\int\sum_{t\in S}\alpha_t(u)\Gamma_t(u)\ s(\mathrm{d}u)$$
$$\geqslant \inf_{(\alpha_t)_{t\in S}\in A}\int\inf_{\tau\in\mathbb{D}}\Gamma_\tau(u)\underbrace{\textstyle\sum_{t\in S}\alpha_t(u)}_{=1}s(\mathrm{d}u) = \int\inf_{\tau\in\mathbb{D}}\Gamma_\tau(u)\ s(\mathrm{d}u)$$

2. Fix any $\varepsilon > 0$. Then, for every $\theta\in\Theta$, there is a simple function $\widehat{W}_\theta\in\mathscr{L}_\infty(\mathbb{D},\mathscr{D})$ so that $\widehat{W}_\theta-\varepsilon\leqslant W_\theta\leqslant\widehat{W}_\theta+\varepsilon\quad\forall\theta\in\Theta$ (cf. [26, p. 86]). Hence,

$$\inf_{\rho\in\mathscr{T}(\mathscr{U},\mathbb{D})}\sum_{\theta\in\Theta}\pi_\theta\rho(s_\theta)[W_\theta]\leqslant\left(\inf_{\rho\in\mathscr{T}(\mathscr{U},\mathbb{D})}\sum_{\theta\in\Theta}\pi_\theta\rho(s_\theta)\left[\widehat{W}_\theta\right]\right)+\varepsilon\overset{1.}{=}s\left[\widehat{W}_\theta)_\theta)\right]+\varepsilon = s\left[\inf_{\tau\in\mathbb{D}}\sum_{\theta\in\Theta}n\pi_\theta\widehat{W}_\theta(\tau)\imath_\theta\right]$$
$$+\varepsilon\leqslant s\left[\inf_{\tau\in\mathbb{D}}\sum_{\theta\in\Theta}n\pi_\theta W_\theta(\tau)\imath_\theta\right]+2\varepsilon = s[K((W_\theta)_\theta)]+2\varepsilon$$

and, analogously, $\inf_{\rho\in\mathscr{T}(\mathscr{U},\mathbb{D})}\sum_{\theta\in\Theta}\pi_\theta\rho(s_\theta)[W_\theta]\geqslant s[K((W_\theta)_\theta)]-2\varepsilon$.

Since $\varepsilon>0$ was arbitrarily chosen, (12) follows. □

Lemma 8.3. *If a precise model $(p_\theta)_{\theta\in\Theta}$ on $(\mathscr{X},\mathscr{A})$ is sufficient for the precise model $(q_\theta)_{\theta\in\Theta}$ on $(\mathscr{Y},\mathscr{B})$, then*

$$\inf_{\rho\in\mathscr{T}(\mathscr{X},\mathbb{D})}R((p_\theta)_\theta,\rho,(W_\theta)_\theta)\leqslant\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}R((q_\theta)_\theta,\sigma,(W_\theta)_\theta)$$

for every decision space $(\mathbb{D},\mathscr{D})$ and every $(W_\theta)_\theta\subset\mathscr{L}_\infty(\mathbb{D},\mathscr{D})$.

Proof. There is some $T\in\mathscr{T}(\mathscr{X},\mathscr{Y})$ so that $T(p_\theta)=q_\theta\ \forall\ \theta\in\Theta$. Therefore,

$$\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}\sum_{\theta\in\Theta}\pi_\theta\sigma(q_\theta)[W_\theta]=\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}\sum_{\theta\in\Theta}\pi_\theta\sigma(T(p_\theta))[W_\theta]=\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}\sum_{\theta\in\Theta}\pi_\theta(\sigma\circ T)(p_\theta)[W_\theta]\geqslant\inf_{\rho\in\mathscr{T}(\mathscr{X},\mathbb{D})}\sum_{\theta\in\Theta}\pi_\theta\rho(p_\theta)[W_\theta]$$

because $\sigma\circ T\in\mathscr{T}(\mathscr{X},\mathbb{D})\forall\sigma\in\mathscr{T}(\mathscr{X},\mathbb{D})$. □

Lemma 8.4.

$$(\mathbf{a})\ \inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})}R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta)=\sup_{(q_\theta)_\theta\in(\mathscr{M}_\theta)_\theta}\ \inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})}R((q_\theta)_\theta,\sigma,(W_\theta)_\theta)$$
$$(\mathbf{b})\ \inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta)=\sup_{(q_\theta)_\theta\in(\mathscr{M}_\theta)_\theta}\ \inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}R((q_\theta)_\theta,\sigma,(W_\theta)_\theta)$$

Proof. (*a*) Theorem 2.1 and [14, Lemma V.3.3, Lemma I.8.2 and Theorem I.8.5] imply that $\prod_{\theta\in\Theta}\mathscr{M}_\theta$ is a compact Hausdorff space. For every $\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})$ there is some $\kappa:\mathscr{L}_\infty(\mathscr{Y},\mathscr{B})\to\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ so that $\sigma(\mu)[g]=\mu[\kappa(g)]$ for every $g\in\mathscr{L}_\infty(\mathscr{Y},\mathscr{B}),\mu\in\mathrm{ba}(\mathscr{Y},\mathscr{B})$. Hence,

$$\mathscr{M}_\theta\to\mathbb{R},\qquad q_\theta\mapsto\sigma(q_\theta)[W_\theta]$$

is continuous for every $\theta\in\Theta$ and this implies continuity of the map

$$(q_\theta)_\theta\mapsto-\sum_{\theta\in\Theta}\pi_\theta\sigma(q_\theta)[W_\theta]=:\Gamma((q_\theta)_\theta,\sigma)$$

on $\prod_{\theta\in\Theta}\mathscr{M}_\theta$ for every $\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})$. $(q_\theta)_\theta\mapsto\Gamma((q_\theta)_\theta,\sigma)$ is convex on $\prod_{\theta\in\Theta}\mathscr{M}_\theta$ for every $\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})$ and $\sigma\mapsto\Gamma((q_\theta)_\theta,\sigma)$ is concave on $\mathscr{T}_r(\mathscr{Y},\mathbb{D})$ for every $(q_\theta)_\theta\in\prod_{\theta\in\Theta}\mathscr{M}_\theta$. Then, the minimax theorem [20, Theorem 2] yields

$$\inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})}R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta)=-\sup_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})}\ \inf_{(q_\theta)_\theta\in(\mathscr{M}_\theta)_\theta}\Gamma((q_\theta)_\theta,\sigma)=-\inf_{(q_\theta)_\theta\in(\mathscr{M}_\theta)_\theta}\ \sup_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})}\Gamma((q_\theta)_\theta,\sigma)$$
$$=\sup_{(q_\theta)_\theta\in(\mathscr{M}_\theta)_\theta}\ \inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})}R((q_\theta)_\theta,\sigma,(W_\theta)_\theta)$$

(*b*) Proposition 3.1 and part (a) of the present lemma yield

$$\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta)\geqslant\sup_{(q_\theta)_\theta\in(\mathscr{M}_\theta)_\theta}\ \inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}R((q_\theta)_\theta,\sigma,(W_\theta)_\theta)$$
$$=\sup_{(q_\theta)_\theta\in(\mathscr{M}_\theta)_\theta}\ \inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})}R((q_\theta)_\theta,\sigma,(W_\theta)_\theta)\overset{(a)}{=}\inf_{\sigma\in\mathscr{T}_r(\mathscr{Y},\mathbb{D})}R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta)$$
$$\geqslant\inf_{\sigma\in\mathscr{T}(\mathscr{Y},\mathbb{D})}R((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta)\quad\square$$

Acknowledgement

I thank Thomas Augustin for valuable discussions. His suggestions have also greatly improved the readability of the article. Furthermore, I thank Helmut Rieder and Peter Ruckdeschel, who have drawn my attention to the work of Buja and Le Cam. I am also much indebted to the Cusanuswerk (Foundation of the Roman Catholic Church) for a Ph.D. scholarship. This article is a revised version of an article presented at ISIPTA'07. I thank the reviewers of the ISIPTA'07 - article and of this revised version for their helpful comments.

References

[1] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman & Hall, London, 1991.
[2] K. Weichselberger, The theory of interval-probability as a unifying concept for uncertainty, International Journal of Approximate Reasoning 24 (2000) 149–170.
[3] K. Weichselberger, Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept, Physica, Heidelberg, 2001.
[4] 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. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), ISIPTA'03, Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications, Lugano, Carleton Scientific, Waterloo, 2003, pp. 31–45.
[5] M. Troffaes, Decision making under uncertainty using imprecise probabilities, International Journal of Approximate Reasoning 45 (2007) 17–29.
[6] L. Utkin, T. Augustin, Powerful algorithms for decision making under partial prior information and general ambiguity attitudes, in: F. Cozman, R. Nau, T. Seidenfeld (Eds.), ISIPTA'05, Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications, Pittsburg, SIPTA, Manno, 2005, pp. 349–358.
[7] A. Buja, Simultaneously least favorable experiments. I. Upper standard functionals and sufficiency, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 65 (1984) 367–384.
[8] P. Huber, V. Strassen, Minimax tests and the Neyman–Pearson lemma for capacities, The Annals of Statistics 1 (1973) 251–263.
[9] T. Augustin, Neyman–Pearson testing under interval probability by globally least favorable pairs reviewing Huber–Strassen theory and extending it to general interval probability, Journal of Statistical Planning and Inference 105 (2002) 149–173.
[10] L. Le Cam, Sufficiency and approximate sufficiency, Annals of Mathematical Statistics 35 (1964) 1419–1455.
[11] L. Le Cam, Asymptotic Methods in Statistical Decision Theory, Springer-Verlag, New York, 1986.
[12] J. Berger, Statistical Decision Theory and Bayesian Analysis, second ed., Springer-Verlag, New York, 1985.
[13] B. Vidakovic, Γ-Minimax: a paradigm for conservative robust Bayesians, in: Robust Bayesian Analysis, Springer-Verlag, New York, 2000, pp. 241–259.
[14] N. Dunford, J. Schwartz, Linear Operators. I. General Theory, Wiley Interscience Publishers, New York, 1958.
[15] D. Blackwell, Comparison of experiments, in: J. Neyman (Ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, University of California Press, Berkeley and Los Angeles, 1951, pp. 93–102.
[16] H. Heyer, Mathematische Theorie Statistischer Experimente, Springer-Verlag, Berlin, 1973.
[17] R. Hable, Decision theory: Some of Le Cam's concepts, E-print, 2007. <www.statistik.lmu.de/~hable/publications.html>.
[18] E. Siebert, Statistical experiments and their conical measures, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 46 (1979) 247–258.
[19] R. Hable, Data-based decisions under complex uncertainty, Ph.D. Thesis, Ludwig-Maximilians-Universitat (LMU) Munich, in preparation.
[20] K. Fan, Minimax theorems, Proceedings of the National Academy of Sciences of the United States of America 39 (1953) 42–47.
[21] E. McShane, T. Botts, Real Analysis, Van Nostrand, Princeton, 1959.
[22] T. Augustin, Optimale Tests bei Intervallwahrscheinlichkeit, Vandenhoeck & Ruprecht, Göttingen, 1998.
[23] H. Rieder, Least favorable pairs for special capacities, The Annals of Statistics 5 (1977) 909–921.
[24] F. Österreicher, On the construction of least favourable pairs of distributions, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 43 (1978) 49–55.
[25] R. Hafner, Konstruktion robuster Teststrategien, in: S. Schach, G. Trenkler (Eds.), Data Analysis and Statistical Inference, Eul, Bergisch Gladbach, 1992, pp. 145–160.
[26] P. Halmos, Measure Theory, second ed., Springer-Verlag, New York, 1974.
[27] V. Varadarajan, Measures on topological spaces, American Mathematical Society Translations 48 (1965) 161–228.