# Data-Based Decisions under Imprecise Probability and Least Favorable Models

**Robert Hable**
Department of Statistics, LMU Munich
Robert.Hable@stat.uni-muenchen.de

## Abstract

Data-based decision theory under imprecise probability has to deal with optimisation 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. In 1984, A. Buja 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 L. 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 assumed to be imprecise.

**Keywords.** Decision theory, robust statistics, imprecise probability, coherent upper previsions, Le Cam, equivalence of models, least favorable models.

## 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, for example. However, a serious problem in practical applications of decision theory is that the uncertainty often is too complex to be adequately described by a classical, i.e. 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, generalisations of the concept of probability have been developed, among others, by [24] (imprecise probability) and [25] (interval probability). 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. [3], [21] and [22].

Generalisations of probabilities as in [24] and [25] have a strong relationship with some concepts of robust statistics (cf. e.g. [20, §3.1.7]) - a fact which is frequently disregarded. Actually, [6] develops a concept of robust statistics (named "upper expectations") which lies between the concepts of [24] and [25]. [6] considers decision making which is explicitly data-based. This can be understood as a matter of its own as has been pointed out by [3]. In the spirit of the celebrated article [14], [6] characterises the existence of precise models which are simultaneously least favorable for a class of loss functions (or for a class of prior distributions):

[14] 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 $\mathcal{M}_1$. [14] shows that there is a pair $(p_0, p_1) \in \mathcal{M}_0 \times \mathcal{M}_1$ which is least favorable: Testing between $p_0$ and $p_1$ is as hard as testing between $\mathcal{M}_0$ and $\mathcal{M}_1$ and, as a consequence, there is an optimal test between $p_0$ and $p_1$ which is also an optimal test between $\mathcal{M}_0$ and $\mathcal{M}_1$. That way, testing between $\mathcal{M}_0$ and $\mathcal{M}_1$ 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 favorablility has attracted enormous attention after the publication of [14]. For a review of [14] and the work following [14], confer [2]. 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 [14] is: Does there exist a model $(q_1, q_2, \ldots, q_n) \in \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_n$ which is simultaneously least favorable for a class of loss functions? This is not always the case but [6] proves a

necessary and sufficient condition for the existence of such simultaneously least favorable models.

Unfortunately, [6] contains an error which reduces its applicability significantly. The validity of the conclusions in [6] can only be guaranteed by adding a restrictive assumption on the involved upper previsions (cf. [10]).

The present article follows the lines of [6] - but within the concept of [24] which dispenses with $\sigma$-additivity. It is shown that the same result as in [6] is possible without any additional assumption on the involved (coherent) upper previsions. This demonstrates that, in [6], insistence on $\sigma$-additivity of probabilities happens to be an unnecessary burden (cf. also Remark 2.2).

By ignoring $\sigma$-additivity, we are in line with Le Cam's decision theoretic framework (cf. [15] and [16]), which provides us with some effective methods. Within this framework some terms (e.g. randomisations) are slightly generalised (cf. [16, §1] and [9, §4]).

Sections 2 and 3 develop the decision theoretic framework. Section 4 contains a generalisation of the LeCam-Blackwell-Sherman-Stein-Theorem which plays an important role in Section 5. In Section 5, the analogue to [6, Theorem 8.2] is proven which characterises 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 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 $\Theta$ 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 $\theta$ 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 $\pi$ for the states of nature $\theta \in \Theta$. Then, we can choose the decision that minimises the expected loss

$$\int_\Theta W_\theta(t)\,\pi(dt)$$

Quite often, we can choose our decision on the base of an observation $y \in \mathcal{Y}$. For example, the observation $y$ may be the outcome of an experiment. The distribution of the observation $y$ might be a precise expectation $q_\theta$ which depends on the state of nature $\theta$. 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 formalised by choosing a decision function $\delta : \mathcal{Y} \to \mathbb{D}, \quad x \mapsto \delta(y)$ which minimises

$$\int_\Theta \int_\mathcal{Y} W_\theta(\delta(y))\, q_\theta(dy)\, \pi(dt)$$

Decision theory commonly also deals with randomised decisions. Randomised decision procedures (randomisations) are defined in Subsection 2.1. Confer [4] 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 $\mathcal{P}$ of possible precise expectations $\pi$. Also the knowledge about the distribution of the observation may only be imprecise so that there are sets $\mathcal{M}_\theta$ of possible precise expectations $q_\theta$. While minimising the expected loss in case of precise expectations is widely accepted, there are several reasonable optimality criteria in case of imprecise expectations, confer [21] for a discussion of the most important ones. In the present article the so-called $\Gamma$-minimax criterion is used which represents a worst case consideration.[1] That is we choose a decision function $\delta$ (or rather a randomisation later on) which minimises the twofold upper expectation

$$\sup_{\pi\in\mathcal{P}} \int_\Theta \sup_{q_\theta\in\mathcal{M}_\theta} \int_\mathcal{Y} W_\theta(\delta(y))\, q_\theta(dy)\, \pi(dt)$$

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 $(\tilde{q}_\theta)_{\theta\in\Theta} \in (\mathcal{M}_\theta)_{\theta\in\Theta}$ which is simultaneously least favorable for $\mathcal{P}$ (or for a corresponding set of loss functions) the above minimisation problem may be solved by minimising

$$\sup_{\pi\in\mathcal{P}} \int_\Theta \int_\mathcal{Y} W_\theta(\delta(y))\, \tilde{q}_\theta(dy)\, \pi(dt)$$

However, such a least favorable model $(\tilde{q}_\theta)_{\theta\in\Theta}$ need not exist. In Section 5, a necessary and sufficient con-

[1] For the use of the $\Gamma$-minimax criterion in Bayesian analysis, cf. [23] and the literature cited therein.

dition for existence is proven (Theorem 5.4). This condition is formulated in terms of standard models.

Standard models are our main tool. They are introduced in Subsection 2.3. An important fact is that every model (consisting of precise expectations) is equivalent to a standard model. In Subsection 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 randomisation) from observations of model $(q_\theta)_{\theta\in\Theta}$ and vice versa. Here and also as decision procedures, randomisations become important. By topological reasons, the term "randomisation" has to be slightly generalised in the present article (cf. Subsection 2.1). All these tools from decision theory (namely randomisations, 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, which in fact is the reason why we use standard models.

Section 4 contains a generalisation of the LeCam-Blackwell-Sherman-Stein-Theorem, which is important in the proof of the main theorem, Theorem 5.4. Theorem 5.4 characterises the existence of simultaneously least favorable models.

### 1.3 Some Notation

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

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

The set of all finitely additive signed measures $\mathrm{ba}(\mathcal{Y},\mathcal{B})$ can be identified with the dual space of $\mathcal{L}_\infty(\mathcal{Y},\mathcal{B})$, i.e. the Banach space of all linear continuous real functionals on $\mathcal{L}_\infty(\mathcal{Y},\mathcal{B})$ where $\|\mu\| = \sup\big\{|\mu[g]| \,\big|\, g \in \mathcal{L}_\infty(\mathcal{Y},\mathcal{B}),\ \|g\| \le 1\big\}$ for all $\mu \in \mathrm{ba}(\mathcal{Y},\mathcal{B})$ (cf. [7, Theorem IV.5.1]). $\mu \in \mathrm{ba}(\mathcal{Y},\mathcal{B})$ is called *positive* if $\mu[g] \ge 0$ for every $g \ge 0$. This is denoted by $\mu \ge 0$.

Let $\Theta$ be an index set. Throughout the article, $(\overline{Q}_\theta)_{\theta\in\Theta}$ is a family of coherent upper previsions $\overline{Q}_\theta : \mathcal{L}_\infty(\mathcal{Y},\mathcal{B}) \to \mathbb{R}$ (cf. [24]). The corresponding sets of majorised linear previsions are denoted by $\mathcal{M}_\theta := \big\{q_\theta \in \mathrm{ba}(\mathcal{Y},\mathcal{B}) \,\big|\, q_\theta[g] \le \overline{Q}_\theta[g]\ \forall\, g \in \mathcal{L}_\infty(\mathcal{Y},\mathcal{B})\big\}$. Analogously to [25], $\mathcal{M}_\theta$ is called *structure*. $(\overline{Q}_\theta)_{\theta\in\Theta}$ is called *imprecise model* on $(\mathcal{Y},\mathcal{B})$. A family $(q_\theta)_{\theta\in\Theta}$ of linear previsions $q_\theta : \mathcal{L}_\infty(\mathcal{Y},\mathcal{B}) \to \mathbb{R}$ is called *precise model* on $(\mathcal{Y},\mathcal{B})$. These terms are adapted from the notion "statistical model". [6] and [15] use the term "experiment" instead of "model".

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

For some fixed $n \in \mathbb{N}$, put $\mathcal{U} := \big\{u \in \mathbb{R}^n \,\big|\, u = (u_{\theta_1},\ldots,u_{\theta_n})',\ u_\theta \ge 0\ \forall\,\theta\in\Theta,\ u_{\theta_1}+\cdots+u_{\theta_n} = 1\big\}$ and $\mathcal{C} := \mathbb{B}^{\otimes n} \cap \mathcal{U}$ where $\mathbb{B}^{\otimes n}$ is the Borel-$\sigma$-algebra of $\mathbb{R}^n$. For $\theta \in \Theta$, put $\iota_\theta : \mathcal{U} \to [0,1],\ u \mapsto u_\theta$ where $u_\theta$ is the $\theta$-component of $u$.

## 2 Some Tools from Decision Theory

### 2.1 Randomisations

#### 2.1.1 Introduction

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

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

$$\sigma : \mathrm{ba}(\mathcal{X},\mathcal{A}) \to \mathrm{ba}(\mathbb{D},\mathcal{D}),\ \ \mu \mapsto \sigma(\mu)$$

via

$$\sigma(\mu) : \ \ h \ \ \mapsto \ \ \sigma(\mu)[h] \ = \ \mu\big[\tau_\cdot[h]\big] \qquad (1)$$

It is easy to see that $\sigma$ is

- linear
- positive: $\sigma(\mu) \ge 0$ for every $\mu \ge 0$
- normalised: $\|\sigma(\mu)\| = \|\mu\|$ for every $\mu \ge 0$

#### 2.1.2 Definition

Let $(\mathcal{X},\mathcal{A})$ and $(\mathcal{Y},\mathcal{B})$ be measurable spaces. According to [15], a *randomisation* from $\mathcal{X}$ to $\mathcal{Y}$ is a linear, positive and normalised map

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

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

We also mark a class of randomisations of a very simple form: To this end, let $\kappa$ 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 $\alpha_y \in \mathcal{L}_\infty(\mathcal{X},\mathcal{A})\ \forall\, y \in S$, $\alpha_y \geq 0\ \forall\, y \in S$ 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 randomisation*. It is easy to see that every restricted randomisation is generated by a (very simple) randomised decision function via (1). Every restricted randomisation is in fact a randomisation , 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 randomisations.

#### 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.

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

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

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

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

**Theorem 2.1** *$\mathcal{M}$ is $\sigma(ba,\mathcal{L}_\infty)$-compact.*
(Cf. [24, §3.6.1].)

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

$\mathcal{T}(\mathcal{X},\mathcal{Y})$ can be provided with the topology of pointwise convergence on $\mathrm{ba}(\mathcal{X},\mathcal{A})\times\mathcal{L}_\infty(\mathcal{Y},\mathcal{B})$. 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 generalisation of randomised procedures:

**Theorem 2.3** *$\mathcal{T}(\mathcal{X},\mathcal{Y})$ is a compact Hausdorff space.* (Cf. [16, Theorem 1.4.2].)

The following theorem indicates that the term "randomisation" has only been slightly generalised:

**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 [15, Theorem 1]. □

Especially, Theorem 2.4 implies that the randomised procedures defined via (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 [6], $(p_\theta)_{\theta\in\Theta}$ is called *sufficient* for $(q_\theta)_{\theta\in\Theta}$ if there is a randomisation $T \in \mathcal{T}(\mathcal{X},\mathcal{Y})$ so that $T(p_\theta) = q_\theta\ \ \forall\, \theta \in \Theta$.

This definition of "sufficiency" essentially goes back to [5]. 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. [13]). At least, if the randomisation $T$ is generated by a randomised function $x \mapsto \tau_x$ via (1), the above definition has a very descriptive interpretation:

Let $x$ be an observation distributed according to $p_\theta$. After observing $x$, start an auxiliary random experiment according to $\tau_x$. Then, the outcome $y$ of the auxiliary random experiment is distributed according to $q_\theta$. 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 $T_2(q_\theta) = p_\theta\ \forall\, \theta \in \Theta$.

The descriptive interpretation of sufficiency already indicates that equivalent models essentially coincide from a decision theoretic point of view. Our definition of equivalence is in accordance with Le Cam's definition (cf. [9, §5.2]).

Let $(\overline{Q}_\theta)_{\theta\in\Theta}$ be an imprecise model with corresponding structures $\mathcal{M}_\theta$, $\theta \in \Theta$.

Analogously to [6], $(p_\theta)_{\theta\in\Theta}$ 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 $\forall\,\theta\in\Theta$

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

### 2.3 Standard Models

Let the index set $\Theta$ be finite with cardinality $n$.

In Subsection 2.2, we have defined an equivalence relation on the precise models with a fixed index set $\Theta$. 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 ($\sigma$-additive) probability measure $s^{\mathcal{F}}$ on $(\mathcal{U},\mathcal{C})$ so that $ds_\theta^{\mathcal{F}} = n\iota_\theta\, ds^{\mathcal{F}}$ defines a precise model $(s_\theta^{\mathcal{F}})_{\theta\in\Theta}$ on $(\mathcal{U},\mathcal{C})$ which is equivalent to $\mathcal{F}$.* (Cf. [9, Theorem 6.5].)

Analogously to [6], $s^{\mathcal{F}}$ 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:

- They are defined on the very nice measurable space $(\mathcal{U},\mathcal{C})$ (cf. Subsection 1.3).
- They consist of linear previsions $s_\theta$ which are $\sigma$-additive probability measures.

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

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

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

$\overline{S}$ 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 $\pi$ be a prior distribution on $(\Theta, 2^\Theta)$, i.e. $\pi$ is a linear prevision on $\mathcal{L}_\infty(\Theta,2^\Theta)$. Put $\pi_\theta := \pi[I_{\{\theta\}}]$.

[2] As stated in Subsection 2.2, equivalent models essentially coincide from a decision theoretic point of view. Therefore, every decision problem coincides with a "standard decision problem" where a standard model is involved. We will deduce properties of the original decision problem from the corresponding "standard decision problem" later on.

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 [15], a *decision procedure* is a randomisation

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

i.e. $\sigma\in\mathcal{T}(\mathcal{Y},\mathcal{D})$.

Now, Bayes risks can be defined for precise models (Subsection 3.1) and for imprecise models (Subsection 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

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

The *Bayes risk* is

$$\begin{aligned} R\big((q_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta}\big) &= \pi\big[\sigma(q)[W]\big] = \\ &= \sum_{\theta\in\Theta}\pi_\theta\sigma(q_\theta)[W_\theta] \end{aligned}$$

Note that this definition coincides with the usual one if $\sigma$ is defined by a randomised decision function via (1).

The minimal Bayes risk is the same if we let $\sigma$ vary among the randomisations or the restricted randomisations:

**Proposition 3.1**

$$\begin{aligned} &\inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R\big((q_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta}\big) = \\ &\qquad = \inf_{\sigma\in\mathcal{T}_r(\mathcal{Y},\mathbb{D})} R\big((q_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta}\big) \end{aligned}$$

*Proof:* The definition of the topology of pointwise convergence implies continuity of the map

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

and, therefore, continuity of

$$\sigma \;\mapsto\; R\big((q_\theta)_{\theta\in\Theta},\sigma,(W_\theta)_{\theta\in\Theta}\big)$$

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

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

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

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

**Theorem 3.2**

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

*Proof:* According to Theorem 2.5, the standard model $\big(s_\theta^{\mathcal{F}}\big)_{\theta\in\Theta}$ is equivalent to $\mathcal{F} := (q_\theta)_{\theta\in\Theta}$. That is $\big(s_\theta^{\mathcal{F}}\big)_{\theta\in\Theta}$ and $\mathcal{F}$ are mutual sufficient. So, a twofold application of Lemma 8.2 yields

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

and an application of Lemma 8.1 closes the proof. □

### 3.2 Imprecise Models

Let $(\overline{Q}_\theta)_{\theta\in\Theta}$ be an imprecise model on $(\mathcal{Y},\mathcal{B})$ with corresponding structures $\mathcal{M}_\theta$, $\theta\in\Theta$, and standard upper prevision $\overline{S}$. 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 $(\overline{Q}_\theta)_{\theta\in\Theta}$ is

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

and the Bayes risk is

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

Hence,

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

These definitions includes that we have chosen the Γ-minimax optimality criterion which represents a worst case consideration (cf. Subsection 1.2) - as done in [14] and [6].

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

**Proposition 3.3**

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

*Proof:* This is a direct consequence of Lemma 8.3 (a), Proposition 3.1 and Lemma 8.3 (b). □

**Theorem 3.4**

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

*Proof:* This is a direct consequence of Lemma 8.3, Theorem 3.2 and the definition of the standard upper prevision. □

## 4 The General LeCam-Blackwell-Sherman-Stein-Theorem

This section contains a generalisation of the LeCam-Blackwell-Sherman-Stein-Theorem. We need this theorem in the proof of our main theorem, Theorem 5.4.

Let $\Theta$ be a finite index set. Let $\pi$ 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 $(\mathcal{X},\mathcal{A})$ and $(\overline{Q}_\theta)_{\theta\in\Theta}$ an imprecise model on $(\mathcal{Y},\mathcal{B})$ where $(\mathcal{M}_\theta)_{\theta\in\Theta}$ 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 $(\mathcal{U},\mathcal{C})$.

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

**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)*** $s^{(p_\theta)_\theta}[k] \ \le \ \overline{S}[k] \qquad \forall\, k\in\Psi$

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

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

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

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

The proof of Theorem 4.1 is located in [11].

## 5 Least Favorable Models

Let the index set $\Theta$ be finite with cardinality $n$. Let $\pi$ 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 $(\mathcal{Y}, \mathcal{B})$ where $(\mathcal{M}_\theta)_{\theta\in\Theta}$ is the corresponding family of structures. Let $(\mathbb{D}, \mathcal{D})$ be a fixed decision space and let $\mathcal{W}$ be a set of loss functions $(W_\theta)_{\theta\in\Theta} \subset \mathcal{L}_\infty(\mathbb{D}, \mathcal{D})$.

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

$$\inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R\big((q_\theta)_\theta, \sigma, (W_\theta)_\theta\big) = \; = \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R\big((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta\big)$$

*for every* $(W_\theta)_{\theta\in\Theta} \in \mathcal{W}$. [3]

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

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

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

The following lemma is an easy consequence of the definitions. A written proof may be found in [11].

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

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

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

$\tilde{\Psi}_\mathcal{W}$ denotes the smallest norm-closed convex cone in $\mathcal{L}_\infty(\mathcal{U}, \mathcal{C})$ which contains $\Psi_\mathcal{W}$. The following lemma is a direct consequence of Theorem 3.2 and Theorem 3.4:

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

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

Theorem 5.4 is the analogue to [6, Theorem 8.2]. It characterises the existence of least favorable models in full generality.

[3] 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.

**Theorem 5.4** *The following statements are equivalent:*

***(a)*** *There is some $\mathcal{F} := (q_\theta)_{\theta\in\Theta} \in (\mathcal{M}_\theta)_{\theta\in\Theta}$ which is least favorable for $\mathcal{W}$.*

***(b)*** $\overline{S}[k_1 + k_2] = \overline{S}[k_1] + \overline{S}[k_2] \qquad \forall\, k_1, k_2 \in \tilde{\Psi}_\mathcal{W}$

*Proof:*
*(a)⇒(b):* Statement (a) and Lemma 5.3 imply $\Psi_\mathcal{W} \subset \Phi_\mathcal{F}$. According to Lemma 5.2, $\tilde{\Psi}_\mathcal{W} \subset \Phi_\mathcal{F}$ and $k_1 + k_2 \in \Phi_\mathcal{F} \;\; \forall\, k_1, k_2 \in \tilde{\Psi}_\mathcal{W}$. Hence, for every $k_1, k_2 \in \tilde{\Psi}_\mathcal{W}$

$$\begin{aligned}\overline{S}[k_1 + k_2] &= s^\mathcal{F}[k_1 + k_2] = s^\mathcal{F}[k_1] + s^\mathcal{F}[k_2] = \\ &= \overline{S}[k_1] + \overline{S}[k_2]\end{aligned}$$

*(a)⇐(b):* Put $s[k] := \overline{S}[k] \;\; \forall\, k \in \tilde{\Psi}_\mathcal{W}$ and

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

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

$$\begin{aligned}s[k] &= \overline{S}[k_2 + k_1 - k_2] - \overline{S}[k_2] \le \\ &\le \overline{S}[k_2] + \overline{S}[k_1 - k_2] - \overline{S}[k_2] = \overline{S}[k]\end{aligned}$$

According to the Hahn-Banach-Theorem ([7, Theorem II.3.10]), $s$ can be extended to a linear functional on $\mathcal{L}_\infty(\mathcal{U}, \mathcal{C})$ (again denoted by $s$) so that

$$s[h] \le \overline{S}[h] \qquad \forall\, h \in \mathcal{L}_\infty(\mathcal{U}, \mathcal{C}) \tag{3}$$

(3) implies, that $s\big[I_\mathcal{U}\big] = 1$ 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 $(\mathcal{U}, \mathcal{C})$. For every decision space $(\hat{\mathbb{D}}, \hat{\mathcal{D}})$ and every $(\hat{W}_\theta)_\theta \subset \mathcal{L}_\infty(\hat{\mathbb{D}}, \hat{\mathcal{D}})$,

$$\inf_{\rho\in\mathcal{T}(\mathcal{U},\hat{\mathbb{D}})} R\big((s_\theta)_\theta, \rho, (\hat{W}_\theta)_\theta\big) = s\big[K\big((\hat{W}_\theta)_\theta\big)\big] \tag{4}$$

according to Lemma 8.1 and

$$\begin{aligned}\inf_{\rho\in\mathcal{T}(\mathcal{U},\hat{\mathbb{D}})} R\big((s_\theta)_\theta, \rho, (\hat{W}_\theta)_\theta\big) &\overset{(4)}{=} s\big[K\big((\hat{W}_\theta)_\theta\big)\big] \le \\ &\overset{(3)}{\le} \overline{S}\big[K\big((\hat{W}_\theta)_\theta\big)\big] = \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\hat{\mathbb{D}})} R\big((\overline{Q}_\theta)_\theta, \sigma, (\hat{W}_\theta)_\theta\big)\end{aligned}$$

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 $T \in \mathcal{T}(\mathcal{U}, \mathcal{Y})$ so that $q_\theta := T(s_\theta) \in \mathcal{M}_\theta \;\; \forall \theta \in \Theta$. Finally for all $(W_\theta)_{\theta\in\Theta} \in \mathcal{W}$,

$$\begin{aligned}\inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R\big((\overline{Q}_\theta)_\theta, \sigma, (W_\theta)_\theta\big) &= \\ &= \overline{S}\Big[K\big((W_\theta)_\theta\big)\Big] = s\Big[K\big((W_\theta)_\theta\big)\Big] = \\ &\overset{(4)}{=} \inf_{\rho\in\mathcal{T}(\mathcal{U},\mathbb{D})} R\big((s_\theta)_\theta, \rho, (W_\theta)_\theta\big) \le \\ &\le \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R\big((q_\theta)_\theta, \sigma, (W_\theta)_\theta\big)\end{aligned}$$

where the last inequality follows from Lemma 8.2. □

## 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 $\Theta$ be a finite index set with cardinality $n$ and $(W_\theta)_{\theta\in\Theta} \subset \mathcal{L}_\infty(\mathbb{D},\mathcal{D})$ be a loss function. Let $(\overline{Q}_\theta)_{\theta\in\Theta}$ be an imprecise model on $(\mathcal{Y},\mathcal{B})$ where $(\mathcal{M}_\theta)_{\theta\in\Theta}$ is the corresponding family of structures. Let $\overline{\Pi}$ be a coherent upper prevision on $\mathcal{L}_\infty(\Theta,2^\Theta)$ i.e. $\overline{\Pi}$ corresponds to a set of prior distributions $\mathcal{P} := \big\{\pi \in \mathrm{ba}(\Theta,2^\Theta) \,\big|\, \pi[a] \le \overline{\Pi}[a]\ \forall\, a \in \mathcal{L}_\infty(\Theta,2^\Theta)\big\}$.
For some $\pi \in \mathcal{P}$, put $\pi_\theta := \pi[I_{\{\theta\}}]\ \ \forall\,\theta\in\Theta$. Let $\sigma$ be a randomisation. For the prior $\pi$, the Bayes risk is

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

where $R_0\big((\overline{Q}_\theta)_\theta,\sigma,(n\pi_\theta W_\theta)_\theta\big)$ denotes the Bayes risk for the uniform prior $\pi_0$ 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 $\mathcal{P}$ of priors $\pi$ into a set $\mathcal{W}$ 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 $\mathcal{W}$. We may also say that such a precise model is *simultaneously least favorable for the set of priors* $\mathcal{P}$.

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 minimises the upper Bayes risk

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

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

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

*and also*

$$R_{\overline{\Pi}}\big((\tilde{q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$$

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

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

$$\Gamma_1(\sigma,\pi) = R_\pi\big((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$$

and

$$\Gamma_2(\sigma,\pi) = R_\pi\big((\tilde{q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$$

It is easy to see that $\sigma \mapsto \Gamma_j(\sigma,\pi)$ is convex and lower semicontinuous for every $\pi\in\mathcal{P}$ and $j\in\{1,2\}$. Then, [8, Theorem 2] and simultaneous least favorability implies

$$\begin{aligned}
&\inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R_{\overline{\Pi}}\big((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big) = \\
&= \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})}\sup_{\pi\in\mathcal{P}}\Gamma_1(\sigma,\pi) = \sup_{\pi\in\mathcal{P}}\inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})}\Gamma_1(\sigma,\pi)\\
&= \sup_{\pi\in\mathcal{P}}\inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})}\Gamma_2(\sigma,\pi) = \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})}\sup_{\pi\in\mathcal{P}}\Gamma_2(\sigma,\pi)\\
&= \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R_{\overline{\Pi}}\big((\tilde{q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big) \qquad (5)
\end{aligned}$$

Lower semicontinuity of

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

and compactness of $\mathcal{T}(\mathcal{Y},\mathbb{D})$ ensure existence of some $\tilde{\sigma}$ which minimises $R_{\overline{\Pi}}\big((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$ (cf. [17, Theorem 3.7]). Additionally,

$$\begin{aligned}
R_{\overline{\Pi}}\big((\tilde{q}_\theta)_\theta,\tilde{\sigma},(W_\theta)_\theta\big) \;&\le\; R_{\overline{\Pi}}\big((\overline{Q}_\theta)_\theta,\tilde{\sigma},(W_\theta)_\theta\big) = \\
&= \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R_{\overline{\Pi}}\big((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big) = \\
&\overset{(5)}{=} \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R_{\overline{\Pi}}\big((\tilde{q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)
\end{aligned}$$

□

**Remark 6.2** *It can easily be read off from the above proof that a decision procedure $\tilde{\sigma}$ which minimises $R_{\overline{\Pi}}\big((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$ also minimises $R_{\overline{\Pi}}\big((\tilde{q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$. However, the reverse statement will not always be true.*[4] *So, it does not suffice to find a decision procedure $\hat{\sigma}$ which minimises $R_{\overline{\Pi}}\big((\tilde{q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$. It still has to be checked that $\hat{\sigma}$ really minimises $R_{\overline{\Pi}}\big((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$. Theorem 6.1 only states that there is a decision procedure which solves both minimisation problems.*

## 7 Concluding Remarks

In decision theory, straightforward updating may lead to suboptimal decisions if the data is distributed according to imprecise probabilities (cf. [3]). 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

[4] In case of hypothesis testing, for example, this follows from [1, p. 162ff].

involved optimisation 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 [6]. The concept of imprecise probability developed in [6] is very close to that one developed in [24]. From a mathematical point of view, the only difference is that [6] assumes that precise probabilities (i.e. linear previsions) have to be $\sigma$-additive. Surprisingly, this appears to be a burden which significantly reduces the applicability of [6].[5] The present article shows that the same result as in [6] is possible without any assumption on the involved (coherent) upper previsions if we dispense with $\sigma$-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 [14], 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 [14], a lot of work was done to construct least favorable pairs in hypothesis testing for special cases (e.g. [19], [18], [12], [1]). 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 $\sigma$-additivity in the proofs of the present paper was possible by the use of notions and methods of [16]. This article is probably the first one which explicitly uses concepts of [16] 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

**Lemma 8.1** *Assume that $s$ is a linear prevision on $\mathcal{L}_\infty(\mathcal{U},\mathcal{C})$ so that $s[\iota_\theta] = \frac{1}{n}\ \forall\, \theta \in \Theta$. Then, $s_\theta : h \mapsto s[n\iota_\theta h]$ defines a precise model $(s_\theta)_{\theta\in\Theta}$ on $(\mathcal{U},\mathcal{C})$ and*

$$\inf_{\rho\in\mathcal{T}(\mathcal{U},\mathbb{D})} R\big((s_\theta)_\theta,\rho,(W_\theta)_\theta\big) = s\big[K\big((W_\theta)_\theta\big)\big] \qquad (6)$$

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

For a proof of Lemma 8.1, confer [9, §6.3].

**Lemma 8.2** *If a precise model $(p_\theta)_{\theta\in\Theta}$ on $(\mathcal{X},\mathcal{A})$ is sufficient for the precise model $(q_\theta)_{\theta\in\Theta}$ on $(\mathcal{Y},\mathcal{B})$, then*

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

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

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

$$\begin{aligned}\inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} \sum_{\theta\in\Theta} \pi_\theta\sigma(q_\theta)[W_\theta] &= \\ &= \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} \sum_{\theta\in\Theta} \pi_\theta\sigma\big(T(p_\theta)\big)[W_\theta] = \\ &= \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} \sum_{\theta\in\Theta} \pi_\theta\big(\sigma\circ T\big)(p_\theta)[W_\theta] \ge \\ &\ge \inf_{\rho\in\mathcal{T}(\mathcal{X},\mathbb{D})} \sum_{\theta\in\Theta} \pi_\theta\rho(p_\theta)[W_\theta]\end{aligned}$$

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

The following lemma is a consequence of the minimax theorem [8, Theorem 2]. Here, topological properties are crucial (Subsection 2.1.3). For a proof, confer [11].

**Lemma 8.3**

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

**(b)**
$$\inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R\big((\overline{Q}_\theta)_\theta,\sigma,(W_\theta)_\theta\big) = = \sup_{(q_\theta)_\theta\in(\mathcal{M}_\theta)_\theta} \inf_{\sigma\in\mathcal{T}(\mathcal{Y},\mathbb{D})} R\big((q_\theta)_\theta,\sigma,(W_\theta)_\theta\big)$$


## Acknowledgements

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. Finally, I thank the reviewers for their helpful comments.


[5] By topological reasons, insistence on $\sigma$-additivity enforces an additional, restrictive assumption on the involved (coherent) upper previsions (cf. Remark 2.2 and [10]).

## References


[1] T. Augustin. *Optimale Tests bei Intervallwahrscheinlichkeit*. Vandenhoeck & Ruprecht, Göttingen, 1998.

[2] 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* 105:1–25, 2002.

[3] T. Augustin. On the Suboptimality of the Generalized Bayes Rule and Robust Bayesian Procedures from the Decision Theoretic Point of View: A Cautionary Note on Updating Imprecise Priors. In: J.M. Bernard, T. Seidenfeld, M. Zaffalon (eds.) *ISIPTA'03, Proceedings of the Third International Symposium on Imprecise Probabilities and their Applications, Lugano*. Carleton Scientific, Waterloo, 31–45, 2003.
www.carleton-scientific.com/isipta/2003-toc.html

[4] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis. 2nd edition*. Springer-Verlag, New York, 1985.

[5] D. Blackwell. Comparison of Experiments. In: J. Neyman (ed.) *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 93–102, 1951.

[6] A. Buja. Simultaneously Least Favorable Experiments, Part I: Upper Standard Functionals and Sufficiency. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 65:367–384, 1984.

[7] N. Dunford, J.T. Schwartz. *Linear Operators, Part I: General Theory*. Wiley-Interscience, New York, 1957.

[8] K. Fan. Minimax Theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39:42–47, 1953.

[9] R. Hable. Decision Theory: Some of Le Cam's Concepts. *E-print*.
www.statistik.lmu.de/~hable/publications.html

[10] R. Hable. A Note on an Error in Buja (1984). *E-print*.
www.statistik.lmu.de/~hable/publications.html

[11] R. Hable. Supplements to the Article "Data-Based Decisions under Imprecise Probability and Least Favorable Models". *E-print*.
www.statistik.lmu.de/~hable/publications.html

[12] R. Hafner. Konstruktion robuster Teststatistiken. In: S. Schach, G. Trenkler (eds.) *Data Analysis and Statistical Inference*, Eul, Bergisch Gladbach, 145–160, 1992.

[13] H. Heyer. *Mathematische Theorie statistischer Experimente*. Springer-Verlag, Berlin, 1973.

[14] P.J. Huber, V. Strassen. Minimax Tests and the Neyman-Pearson Lemma for Capacities. *The Annals of Statistics*, 1:251–263, 1973.

[15] L. Le Cam. Sufficiency and Approximate Sufficiency. *The Annals of Mathematical Statistics*, 35:1419–1455, 1964.

[16] L. Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York, 1986.

[17] E.J. McShane, T.A. Botts. *Reals Analysis*. Van Nostrand, Princeton, (1959).

[18] F. Österreicher. On the Construction of Least Favorable Pairs of Distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 43:49–55, 1978.

[19] H. Rieder. Least Favorable Pairs for Special Capacities. *The Annals of Statistics*, 5:909–921, 1977.

[20] B. Rüger. *Test- und Schätztheorie. Band II: Statistische Tests*. Oldenbourg-Verlag, München, 2002.

[21] M.C.M. Troffaes. Decision Making under Uncertainty Using Imprecise Probabilities *International Journal of Approximate Reasoning*, accepted, 2006.

[22] L.V. Utkin, T. Augustin. Powerful Algorithms for Decision Making under Partial Prior Information and General Ambiguity Attitudes. *4th International Symposium on Imprecise Probabilities and Their Applications*, Pittsburg, 2005.
www.sipta.org/isipta05/proceedings/046.html

[23] B. Vidakovic. Γ-minimax: a paradigm for conservative robust Bayesians. In: D.R. Insua, F. Ruggeri (eds.) *Robust Bayesian analysis*. Springer-Verlag, New York, 241–259, 2000.

[24] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London, 1991.

[25] K. Weichselberger. The Theory of Interval-Probability as a Unifying Concept for Uncertainty. *International Journal of Approximate Reasoning*, 24:149–170, 2000.