Augustin:

Neyman-Pearson Testing under Interval Probability by Globally Least Favorable Pairs: A Survey of Huber-Strassen Theory and Some Results on its Extension to General Interval Probability

Sonderforschungsbereich 386, Paper 172 (1999)

Online unter: http://epub.ub.uni-muenchen.de/

Neyman-Pearson Testing under Interval Probability by Globally Least Favorable Pairs

A Survey of Huber-Strassen Theory and Some Results on its Extension to General Interval Probability

Th. Augustin*

Abstract

The paper studies the extension of one of the basic issues of classical statistics to interval probability. It is concerned with the Generalized Neyman-Pearson problem, i.e. an alternative testing problem where both hypotheses are described by interval probability. First the Huber-Strassen theorem and the literature based on it is reviewed. Then some results are presented indicating that the restrictive assumption of C-probability (two-monotonicity) underlying all that work can be overcome in favor of considering general interval probability in the sense of Weichselberger (1999A). So the full expressive power, which is provided by interval probability, can also be utilized in testing hypotheses.

AMS classification: primary 62A20; secondary 60A05; 62F35

Keywords: Interval probability, F-probability, capacities, Neyman-Pearson testing, Huber-Strassen theorem, generalized neighborhood models, least favorable pseudo-capacities

*Thomas Augustin, Department of Statistics, Ludwig-Maximilians University, Munich, D-80799 Munich. thomas@stat.uni-muenchen.de

1 Introduction

Objections have been raised to the paradigm, all situations under uncertainty could adequately be described by a single *classical probability*, i.e. by a ($\sigma-$)additive and normalized set-function. In theoretical, as well as in applied work, a frequent complaint is that the concept of classical probability requires a much higher degree of precision and internal consistency of the information than that which is available in many situations. Modeling such situations nevertheless with classical probability may lead to deceptive conclusions. Therefore, the need for a generalized concept of probability has become more and more obvious.

Such a substantial extension of classical probability is provided by the theory of interval probability. In this concept intervals $[L(A), U(A)]$ are used instead of single real numbers $p(A)$ to describe the probability of an event A. This allows for an appropriate modeling of more general aspects of uncertainty. Though the idea to use interval-valued probabilities has a long history and can be traced back at least to the middle of the nineteenth century, the main steps towards a comprehensive theory have been achieved only recently. Based on a certain generalized betting interpretation, Walley (1991) developed a Neobayesian theory extending de Finetti's concept of probability to imprecise previsions. An interpretation-independent theory of interval probability generalizing Kolmogorov's axioms has been developed by Weichselberger (1999A). — For reviews and references on the emergence of interval probability see especially Walley, 1991, chapters 1 and 5, Weichselberger, 1999A, chapter 1, and de Cooman and Walley, 1999.

In the last years the interest in statistical application of interval probability and related concepts has mainly concentrated on generalized Bayesian inference, cf. for instance Walley, 1996, Bernard, 1999, and the survey by Wassermann, 1997. Nevertheless, interval probability also proves to be quite important for the non-Bayesian ('objective/ frequentist') point of view. Moreover, this topic has a – nowadays mostly forgotten – prominent tradition. Some twenty-five years ago, Huber and Strassen (1973) studied the generalization of the Neyman-Pearson alternative testing problem to the situation where both hypotheses are described by interval probability. Their results as well as all the work following their seminal paper, however, kept restricted to quite a special type of interval probability. Moreover, a so-called 'necessity-theorem' was – erroneously (see below) – understood to make it impossible to allow for more expressive classes of interval probabilities.

This paper surveys the 'Huber-Strassen theory' and then extends some of the main results to general interval probability. Section 2 states some basic definitions from Weichselberger's concept of interval probability (F-probability), which underlies this work. Also the special case of C-probability is considered, which contains two-monotone capacities and pseudo-capacities and therefore also the neighborhood models commonly used in robust statistics. It turns out that nevertheless the condition of being C-probability is too restrictive to serve as basis for a general theory. Section 3 looks at the generalized testing problem and the concept of (globally) least favorable least pairs to construct (level-α-)maximin-tests. The Huber-Strassen theorem, which ensures the existence of globally least favorable pairs for typical C-probabilities, as well as the literature following that result, is reviewed in section 4. — The next two sections show that it is not necessary to restrict the consideration to the narrow class of C-probabilities as has been done up to now. In section 5 the so-called 'necessity theorem for C-probability' is toned down by characterizing some situations of F-probability not being C-probability, where globally least favorable pairs exist. Furthermore a 'decomposition-theorem' is developed allowing for handling more complex neighborhood models by studying 'least favorable neighborhoods'. Section 6 briefly sketches the concept of locally least favorable pairs. Some directions for further research are indicated in the concluding remarks in section 7.

2 Some basic aspects of interval probability

2.1 F-probability, structure

This paper is based on the interpretation-independent theory of interval probability developed by Weichselberger (1999A), for selected aspects cf. also Weichselberger and Pöhlmann, 1990, Weichselberger, 1995A, 1995B, 1996, 1999B. His concept is founded on the following generalization of Kolmogorov's axioms:

Definition 2.1 (The axioms of interval probability) *Let* $(\Omega, \mathcal{A})$ *be a measurable space.*

- *A function* $p(\cdot)$ *on* $\mathcal{A}$ *fulfilling the axioms of Kolmogorov is called* K-probability *or* classical probability. *The set of all classical probabilities on* $(\Omega, \mathcal{A})$ *will be denoted by* $\mathcal{K}(\Omega, \mathcal{A})$.

- *A function* $P(\cdot)$ *on* $\mathcal{A}$ *is called* R-probability *with* structure $\mathcal{M}$*, if*

  1. $P(\cdot)$ *is of the form*

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

  2. *The set*

  $$\mathcal{M} \stackrel{def}{=} \{\,p(\cdot) \in \mathcal{K}(\Omega, \mathcal{A}) \mid L(A) \leq p(A) \leq U(A),\ \forall A \in \mathcal{A}\} \qquad (1)$$

  is not empty.

- *R-probability with structure* $\mathcal{M}$ *is called* F-probability*, if*

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

The following convention is made for the sake of clarity: Throughout the paper capital-letter P is used for interval-valued assignments, while small letters $(p, q, \ldots)$ stand for classical probability.

For every F-probability $L(\cdot)$ and $U(\cdot)$ are *conjugate*, i.e. $L(A) = 1 - U(A^C)$, $\forall A \in \mathcal{A}$. The other way round, presupposing conjugacy, every F-probability is uniquely determined either by $L(\cdot)$ or either by $U(\cdot)$ alone. Therefore one obtains for its structure $\mathcal{M}$:

$$\begin{array}{rcl} \mathcal{M} & = & \{p(\cdot) \in \mathcal{K}(\Omega, \mathcal{A}) \mid L(A) \leq p(A)\,,\ \forall A \in \mathcal{A}\} = \\ & = & \{p(\cdot) \in \mathcal{K}(\Omega, \mathcal{A}) \mid p(A) \leq U(A)\,,\ \forall A \in \mathcal{A}\}\,. \end{array} \qquad (3)$$

Here $L(\cdot)$ is used throughout, and $\mathcal{F} = (\Omega, \mathcal{A}, L(\cdot))$ is called an *F-probability field.* Specifying an F-probability field $(\Omega, \mathcal{A}, L(\cdot))$, it is implicitly assumed that the conjugate set-function $U(\cdot) = 1 - L(\cdot^C)$ describes the upper interval-limit.

Weichselberger's theory relies on countable additive classical probability. So the interval-limits of an F-probability are lower and upper probabilities in the sense of Huber and Strassen (1973), but these terms are avoided here because they are also used in the literature in several other meanings. Furthermore, F-probability is strongly related to coherence in the setting of Walley (1991) and to the concept of envelopes in the frequentist theory developed by Fine and students (e.g. Walley and Fine, 1982, or Papamarcou and Fine, 1991).

The relation between interval probabilities and non-empty sets of classical probabilities expressed by the concept of the structure (see (1)) proves to be quite important for the whole theory. It indicates how to extend concepts of classical probability to interval probability. For instance, generalizing expectation is straightforward:

Definition 2.2 (Expectation with respect to an F-probability field) *For every F-probability field $\mathcal{F} = (\Omega, \mathcal{A}, L(\cdot))$ with structure $\mathcal{M}$ a random variable X on $(\Omega, \mathcal{A})$ is called* $\mathcal{M}$-integrable, *if X is p-integrable for each element $p(\cdot)$ of $\mathcal{M}$. Then*

$$\begin{aligned} \mathbb{E}_{\mathcal{M}}X &\stackrel{def}{=} [L\mathbb{E}_{\mathcal{M}}X, U\mathbb{E}_{\mathcal{M}}X] \\ &\stackrel{def}{=} \big[\inf_{p(\cdot)\in\mathcal{M}} \mathbb{E}_pX\,,\ \sup_{p(\cdot)\in\mathcal{M}} \mathbb{E}_pX\big] \\ &\subseteq [-\infty, \infty] \end{aligned} \tag{4}$$

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

2.2 Prestructures

Not every set of classical probabilities is a structure of an F-probability, but every non-empty set of classical probabilities can be used to construct an unique, narrowest F-probability field corresponding to it.

Remark 2.3 (Prestructure)*(Weichselberger, 1999A) Let $\mathcal{V} \neq \emptyset$ be a set of classical probabilities on a measurable space $(\Omega, \mathcal{A})$. Then $P_{\mathcal{V}}(\cdot) \stackrel{def}{=} [L_{\mathcal{V}}(\cdot), U_{\mathcal{V}}(\cdot)]$ with*

$$L_{\mathcal{V}}(A) \stackrel{def}{=} \inf_{p(\cdot)\in\mathcal{V}} p(A) \quad \wedge \quad U_{\mathcal{V}}(A) \stackrel{def}{=} \sup_{p(\cdot)\in\mathcal{V}} p(A) \tag{5}$$

is F-probability, and $\mathcal{V}$ is called prestructure *of $\mathcal{F}_{\mathcal{V}} = (\Omega, \mathcal{A}, L_{\mathcal{V}}(\cdot))$.*

For the structure $\mathcal{M}_{\mathcal{V}}$ of $\mathcal{F}_{\mathcal{V}}$ the relation $\mathcal{M}_{\mathcal{V}} \supseteq \mathcal{V}$ holds. Furthermore, every F-probability field $\mathcal{F} = (\Omega, \mathcal{A}, L(\cdot))$ with structure $\mathcal{M}$ also fulfilling $\mathcal{M} \supseteq \mathcal{V}$ is weaker *than $\mathcal{F}_{\mathcal{V}} = (\Omega, \mathcal{A}, L_{\mathcal{V}}(\cdot))$ in the sense that $L(A) \leq L_{\mathcal{V}}(A)$ for all $A \in \mathcal{A}$.*

Mainly two applications of this concept will be used in what follows:

Definition 2.4 (Independent product of F-probability fields, *compare Walley and Fine, 1982, p. 745, Weichselberger, 1999A, chapter 7, and the 'sensitivity analysis definition' in Walley, 1991, chapter 9.1.***)** *Let a finite number of F-probability fields* $\mathcal{F}_l = (\Omega_l, \mathcal{A}_l, L_l(\cdot))$ *with structures* $\mathcal{M}_l$, $l = 1, \dots, n$, *be given. Then the F-probability field*

$$\bigotimes_{l=1}^{n} \mathcal{F}_l \stackrel{def}{=} \left(\mathop{\times}_{l=1}^{n} \Omega_l, \bigotimes_{l=1}^{n} \mathcal{A}_l, L(\cdot) \right),$$

which has

$$\mathop{\times}_{l=1}^{n} \mathcal{M}_l$$

as its prestructure, is called the independent product of the F-probability fields $\mathcal{F}_l, l = 1, \dots, n$.

Definition 2.5 (Parametricly constructed F-probability fields) *Consider a (strictly) parametric set* $\mathcal{Q} = \{p_\theta(\cdot) \mid \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^m$, *of classical probabilities on a measurable space* $(\Omega, \mathcal{A})$. *An F-probability field* $\mathcal{F}(\boldsymbol{\theta}) = (\Omega, \mathcal{A}, L(\cdot))$ *with structure* $\mathcal{M}$ *is called* parametricly constructed with respect to $\mathcal{Q}$, *if there exists a*

$$\boldsymbol{\theta} = [\theta_1^L, \theta_1^U] \times [\theta_2^L, \theta_2^U] \times \ldots \times [\theta_m^L, \theta_m^U] \subseteq \Theta$$

in such a way, that

$$\mathcal{Q}(\boldsymbol{\theta}) \stackrel{def}{=} \{p_\theta(\cdot) \mid \theta \in \boldsymbol{\theta}\}$$

is a prestructure of $\mathcal{M}$. *Then* $\boldsymbol{\theta}$ *is called* parameter of $\mathcal{F}(\boldsymbol{\theta})$ (with respect to $\mathcal{Q}$).

One interpretation of such concepts defined via prestructures is to take them as a robustification of the classical concepts (e.g. of independence or parametric distributions). In general, the structures of the resulting F-probability fields are richer than the sets used for construction. The structure of the independent product also contains 'slightly dependent' classical probabilities 'lying between' the independent ones. In the second case in particular all the mixtures of the distributions corresponding to a parameter value inside $\boldsymbol{\theta}$ belong to the structure as well.

Prestructures are also often helpful in easily verifying certain properties. In section 5 it will be used that the existence of a dominated prestructure is sufficient for the structure to be dominated. — To exclude misunderstandings it may be noted at this point that the expression 'dominated' is used here in its measure-theoretic meaning and not in the sense of 'dominated lower and upper probabilities' as for instance in Papamarcou and Fine, 1986.

Lemma 2.6 (Dominated (pre)structures) *The structure $\mathcal{M}$ of an F-probability field is dominated by a $\sigma-$finite measure $\lambda(\cdot)$, iff there exists any prestructure $\mathcal{V}$, which is dominated by $\lambda(\cdot)$.*

Proof: Since according to remark 2.3 $\mathcal{M} \supseteq \mathcal{V}$ one direction is trivially fulfilled. Now let $\lambda(\cdot)$ be a $\sigma-$finite measure dominating $\mathcal{V}$. For every $A \in \mathcal{A}$ with $\lambda(A) = 0$ one has $p(A) = 0$ for all $p(\cdot) \in \mathcal{V}$ implying $sup_{p(\cdot)\in\mathcal{V}}\, p(A) = 0$. The claim now immediately follows from the fact that for every prestructure $sup_{p(\cdot)\in\mathcal{V}}\, p(A) = sup_{p(\cdot)\in\mathcal{M}}\, p(A)$. ☐

2.3 C-probability

In this subsection a special case of F-probability is considered, called C-probability with Weichselberger (1999A, chapter 5). It provides a superstructure upon neighborhood models commonly used in robust statistics (see below) and additionally contains the so-called Dempster-Shafer belief-functions (e.g. Yager, Fedrizzi and Kacprzyk, 1994.).

Definition 2.7 (C-probability) *Let $(\Omega, \mathcal{A})$ be a measurable space. F-probability $P(\cdot) = [L(\cdot), U(\cdot)]$ is called* C-probability, *if $L(\cdot)$ is* two-monotone, *i.e., if*

$$L(A \cup B) + L(A \cap B) \geq L(A) + L(B), \quad \forall A, B \in \mathcal{A}\,. \tag{6}$$

Then the F-probability field $\mathcal{C} = (\Omega, \mathcal{A}, L(\cdot))$ is called a C-probability field.

For the property of two-monotonicity many different names are common, too. In particular it is also called '(strong) superadditivity', 'supermodularity' or 'convexity'.

Two related types of set-functions leading to C-probability have been extensively studied in the literature. The first one, the class of two-monotone capacities, was introduced by Choquet (1954) in the context of potential theory. The members of the second branch share most of the properties and are therefore called 'pseudo-capacities' (Buja, 1986) or 'special capacities' (e.g. Bednarski, 1981).

Remark 2.8 (Typical examples of C-probability fields) *Assume $(\Omega, \mathcal{A})$ to be a Polish measurable space, i.e. Ω is a complete, separable and metrizable space, and $\mathcal{A}$ the corresponding Borel σ-field (e.g. $\Omega = I\!R^n, \mathcal{A} = \mathcal{B}$).*

- Two-monotone (Choquet-)capacities: *Every set-function $L(\cdot)$ on $\mathcal{A}$ with $L(\Omega) = 1$ and (6) additionally obeying the condition*

$$(A_n)_{n\in I\!N} \uparrow A,\ A_n \text{ open},\ n \in I\!N \Longrightarrow \lim_{n\to\infty} L(A_n) = L(A)\,, \qquad (7)$$

  and the condition

$$(A_n)_{n\in I\!N} \downarrow A,\ A_n \in \mathcal{A},\ n \in I\!N \Longrightarrow \lim_{n\to\infty} L(A_n) = L(A)$$

  leads, together with the corresponding conjugate upper limit $U(\cdot) \geq L(\cdot)$, to F-probability and therefore to C-probability (Huber and Strassen, 1973, lemma 2.5, p. 254).

- Pseudo-capacities: *Let $f(\cdot) : [0,1] \to [0,1]$ be a convex function with $f(0) = 0$ and let $p(\cdot)$ be a classical probability on $(\Omega, \mathcal{A})$ (called* central distribution *in this context). Then*

$$P(\cdot) \stackrel{def}{=} \left[(f \otimes p)(\cdot), 1 - (f \otimes p)(\cdot^{C})\right]$$

  with $(f \otimes p)(\Omega) = 1$ and

$$(f \otimes p)(A) \stackrel{def}{=} f\left(p(A)\right)\,, \quad \forall A \in \mathcal{A} \setminus \{\Omega\}\,,$$

  is C-probability. The corresponding structure will be denoted by $\mathcal{M}(f \otimes p)$.

Some models often used in robust statistics naturally fit into this framework. Perhaps the most prominent pseudo-capacity – which is also a two-monotone capacity, if Ω is compact – is the (ϵ, δ)-*contamination-model* $(0 < \epsilon, \delta, \epsilon + \delta < 1)$ containing the *contamination model* in the narrow sense $(\delta = \epsilon)$ and the *total-variation model* $(\epsilon = 0)$. There $f(\cdot)$ has the form $f(y) = \max\left((1-\epsilon)\cdot y - \delta\,, 0\right)$.

C-probability is a distinguished special case of F-probability possessing some mathematical elegance, but it is not comprehensive enough to provide an exclusive, neat basis for a theory of interval-valued probability.

In the meantime Walley's conclusion that there is not "[...] any 'rationality' argument for 2-monotonicity, beyond its computational convenience" (Walley, 1981, p. 51) has experienced a good deal of additional support. It turned out that the expressive power of the concept of interval probability is mainly due to the extension of the calculus to arbitrary F-probability fields. To mention just one argument, on which it will be recurred later (for a detailed argumentation on this topic see: Augustin, 1998, chapter 1.2.3.): The generalization of the usual parametric families to interval probability (like the F-normal distribution) along the lines of definition 2.5 leads to F- but not to C-probability.

3 (Level-α-)Maximin tests and (globally) least favorable pairs

In this section the testing problem studied will be precisely stated and a method for obtaining optimal tests will be described.

3.1 Neyman-Pearson testing between interval probabilities

Formulating the *Generalized Neyman-Pearson Problem* is straightforward. Just as in classical Neyman-Pearson theory, one probability is tested versus another one, without any (non-vacuous) prior knowledge which one of the hypotheses is the true one. But now the hypotheses may consist of F-probabilities instead of classical probabilities:

Problem 3.1 (Generalized Neyman-Pearson Problem) *Consider two F-probability fields $\mathcal{F}_0 = (\Omega, \mathcal{A}, L_0(\cdot))$ and $\mathcal{F}_1 = (\Omega, \mathcal{A}, L_1(\cdot))$ on a measurable space $(\Omega, \mathcal{A})$ with disjoint structures $\mathcal{M}_0$ and $\mathcal{M}_1$. Based on the observation of a certain singleton $\{\omega\} =: E$, which has the probability $P_0(E) = [L_0(E), U_0(E)]$ or $P_1(E) = [L_1(E), U_1(E)]$ to occur, an optimal decision via a test has to be made between the two hypotheses H_i : "The 'true' probability field is $\mathcal{F}_i$", $i \in \{0, 1\}$.*

For technical reasons the formulation of problem 3.1 uses sample size 1. Of course, situations with sample size n are included by considering the independent products (see definition 2.4) of the F-probability fields describing the hypotheses. By stating problem 3.1 also two regularity conditions are implied, which are assumed to hold throughout the paper: To have an alternative-problem in the narrow sense, it is supposed that $\mathcal{M}_0$ and $\mathcal{M}_1$ have a positive distance with respect to an appropriate metric. Furthermore, to have the problem well defined, the set $\{\omega\}$ is taken to be measurable for every $\omega \in \Omega$. This condition is very mild. It is, in particular, fulfilled by all Polish spaces.

Since the concept of randomization is based on an idealized random-experiment without any non-probabilistic uncertainty, it should be described by classical probability. Therefore, it is – even in the area of interval probability – consistent to allow only for precise (i.e. not interval-valued) probabilities for rejecting H_0. So the concept of a test remains the same as in classical statistics, the set Φ of all tests still is the set of all measurable functions $\varphi(\cdot) : \Omega \to [0,1]$.

As Huber and Strassen (1973) and the work following them did, also the present paper exclusively considers the case where only the upper limits of the error probability are taken into account. Then the Neyman-Pearson principle 'Minimize the probability of the error of the second kind (i.e. $I\!E_{\mathcal{M}_1}(1-\varphi)$) while controlling for the error of the first kind (i.e. $I\!E_{\mathcal{M}_0}\varphi$)' leads to a complex, nonparametric (not 'easily parametrizable') (level-α-)maximin-problem between the structures.

Definition 3.2 (Level-α-maximin-criterion under F-probability) *Let a level of significance $\alpha \in (0,1)$ be given. A test $\varphi^*(\cdot) \in \Phi$ is called a* level-α-maximin-test *(for $\mathcal{F}_0$ versus $\mathcal{F}_1$), if $\varphi^*(\cdot)$ respects the level of significance, i.e.*

$$UI\!E\varphi^* \leq \alpha\,, \tag{8}$$

and $\varphi^*(\cdot)$ has maximal power among all tests under consideration, i.e.

$$\forall \psi \in \Phi \ \left[UI\!E_{\mathcal{M}_0}\psi \leq \alpha \Rightarrow LI\!E_{\mathcal{M}_1}\psi \leq LI\!E_{\mathcal{M}_1}\varphi^* \right]. \tag{9}$$

3.2 (Globally) least favorable pairs

To get a proper principle for constructing optimal tests, the following idea is helpful: "If one succeeds in convincing the hardliners of two parties, one has convinced all their members". Therefore, one may try to construct level-α-maximin-tests by searching for two elements $q_0(\cdot)$ and $q_1(\cdot)$ of the structures, where the testing is most difficult. In the spirit of Huber and Strassen (1973) and the work related to it this 'being least favorable' can be formalized as follows.

Definition 3.3 (Globally least favorable pairs) *Consider problem 3.1. A pair $(q_0(\cdot), q_1(\cdot))$ of classical probabilities is called* a globally least favorable pair *(for $\mathcal{F}_0$ against $\mathcal{F}_1$), if*

1. $$(q_0(\cdot), q_1(\cdot)) \in \mathcal{M}_0 \times \mathcal{M}_1 \, , \tag{10}$$

2. *there is a version $\pi(\cdot)$ of the likelihood ratio of $q_0(\cdot)$ and $q_1(\cdot)$ with*

$$\forall t \geq 0, \forall p_0(\cdot) \in \mathcal{M}_0 : \; p_0(\{\omega \,|\pi(\omega) > t\}) \leq q_0(\{\omega \,|\pi(\omega) > t\}) \tag{11}$$

and

$$\forall t \geq 0, \forall p_1(\cdot) \in \mathcal{M}_1 : \; p_1(\{\omega \,|\pi(\omega) > t\}) \geq q_1(\{\omega \,|\pi(\omega) > t\}) \, . \tag{12}$$

The heuristics given above can be formally supported. It is not hard to prove that globally least favorable pairs indeed lead to level-α-maximin-tests:

Proposition 3.4 (Globally least favorable pairs and level-α-maximin-tests) *If $(q_0(\cdot), q_1(\cdot))$ is a globally least favorable pair for $\mathcal{F}_0$ versus $\mathcal{F}_1$, then there exists a best level-α-test for testing the hypothesis $\overline{H_0}$: $\{q_0(\cdot)\}$ versus the hypothesis $\overline{H_1}$: $\{q_1(\cdot)\}$, which is a level-α-maximin-test for $\mathcal{F}_0$ versus $\mathcal{F}_1$, too.*

It should be remarked that in the literature it is usage to call $(q_0(\cdot), q_1(\cdot))$ just a "least favorable pair". The term 'globally' is added here to make a distinction to 'locally least favorable pairs', which will be introduced later.

Definition 3.3 immediately implies that for every globally least favorable pair $(q_0(\cdot), q_1(\cdot))$ for every nonnegative t the following holds:

$$\begin{aligned} q_0(\{\omega \,|\pi(\omega) > t\}) &= U_0(\{\omega \,|\pi(\omega) > t\}) \\ q_1(\{\omega \,|\pi(\omega) > t\}) &= L_1(\{\omega \,|\pi(\omega) > t\}) \, . \end{aligned} \tag{13}$$

From this it is directly deduced that it is suffices to check the conditions for globally least favorable pairs on any arbitrary prestructure. Since this conclusion helps to systematize some well-known results and plays an important role in the proofs of the extensions discussed in section 5 it is for ease of reference formulated as a separate lemma.

Lemma 3.5 (Globally least favorable pairs and prestructures) *A pair* $(q_0(\cdot), q_1(\cdot))$ *of classical probabilities with* $(q_0(\cdot), q_1(\cdot)) \in \mathcal{M}_0 \times \mathcal{M}_1$ *is a globally least favorable pair for* $\mathcal{F}_0$ *versus* $\mathcal{F}_1$*, if there exist prestructures* $\mathcal{V}_0$ *and* $\mathcal{V}_1$ *of* $\mathcal{F}_0$ *and* $\mathcal{F}_1$ *in such a way that for a suitable version* $\pi(\cdot)$ *of the likelihood-ratio relations (11) and (12) hold for* $\mathcal{V}_i$ *instead of* $\mathcal{M}_i$*,* $i \in \{0,1\}$.

The property of being globally least favorable does not depend on the level of significance. This makes globally least favorable pairs 'independent of the sample size':

Proposition 3.6 (Product-theorem for globally least favorable pairs) *Let* $q_i^{(n)}(\cdot)$ *and* $\mathcal{F}_i^{(n)}$*,* $i \in \{0,1\}$*, denote the* n*-dimensional independent products of* $q_i(\cdot)$ *and* $\mathcal{F}_i$*, respectively. If* $(q_0(\cdot), q_1(\cdot))$ *is globally least favorable for* $\mathcal{F}_0$ *versus* $\mathcal{F}_1$*, then* $(q_0^{(n)}(\cdot), q_1^{(n)}(\cdot))$ *is globally least favorable for* $\mathcal{F}_0^{(n)}$ *versus* $\mathcal{F}_1^{(n)}$*.*

Sketch of the proof: In principle, the core of this proposition was already mentioned informally by Huber and Strassen (1973, corollary 4.2, p. 257) without having a clear independence concept for interval probability. A proof of proposition 3.6 is given in Augustin (1998, p. 223ff). It is mainly based on results taken from Witting (1985, Satz 2.57, p. 237f) and on lemma 3.5. □

4 Huber-Strassen theorem and the 'necessity' of C-probability

4.1 The main theorems

The fame of the work of Huber and Strassen (1973) is first and foremost due to the fact that they succeeded in showing that a globally least favorable pair always exists for two-monotone capacities:

Proposition 4.1 (Huber-Strassen theorem) *(Huber and Strassen, 1973, theorem 4.1, p. 257) Consider problem 3.1 on a Polish space* $(\Omega, \mathcal{A})$*, and let* $\mathcal{F}_0$ *and* $\mathcal{F}_1$ *be C-probability fields fulfilling (7). Then there exists a globally least favorable pair for* $\mathcal{F}_0$ *versus* $\mathcal{F}_1$*.*

An extension to F-probability had not been considered so far, apparently because the following result was misunderstood to show the impossibility of a generalization:

Proposition 4.2 ('Necessity theorem') *Consider a finite space* Ω*, the corresponding power set* $\mathcal{P}(\Omega)$*, and an F-probability field* $\mathcal{F}_0 = (\Omega, \mathcal{P}(\Omega), L_0(\cdot))$ *with structure* $\mathcal{M}_0$*. If there exists for any classical probability* $p_1(\cdot)$ *with* $p_1(\cdot) \notin \mathcal{M}_0$ *a classical probability* $p_0(\cdot) \in \mathcal{M}_0$ *in such a way that* $(p_0(\cdot), p_1(\cdot))$ *is a globally least favorable pair for* $\mathcal{F}_0$ *versus* $\mathcal{F}_1 \stackrel{def}{=} (\Omega, \mathcal{P}(\Omega), p_1(\cdot))$*, then* $\mathcal{F}_0$ *must be a C-probability field.*

The consequence has been an exclusive concentration on models producing C-probability. For instance, Lembcke (1988) entitles his article, where he introduces his generalization of proposition 4.2, "The necessity of [...C-probability] for Neyman-Pearson minimax tests", and Huber himself calls relation (6) "the crucial [property...] to obtain a neat theory" (Huber, 1973, p. 182). Though – as mentioned in section 2.3 – this is rather unsatisfactory with regard to the expressive power of modeling, the restriction on C-probability has seemed to be the inevitable price one has to pay for Neyman-Pearson testing under interval probability.

4.2 A short survey of the work following the Huber-Strassen theorem

The Huber-Strassen theorem has two different roots, each connected with one of the authors. Already in 1964 Strassen formulated proposition 4.1 for totally monotone C-probability on finite spaces (Strassen, 1964, Satz 2.1, p. 282). One year later he recognized (Strassen, 1965, p. 431) that indeed two-monotonicity is sufficient. On the other side, based on heuristic arguments Huber (1965) managed to derive a globally least favorable pair for contamination neighborhood models.

The synthesis leading 1973 to proposition 4.1 induced a great deal of work, which is mainly concentrated on two aspects. Since on non-compact Ω, e.g. $\Omega = \mathbb{R}$, the usual neighborhood models do not fulfill (7) (with respect to the standard topology), the first branch was concerned with the existence of globally least favorable pairs in such situations. Important steps towards a solution were obtained among others by Rieder (1977) and by Bednarski (1981), while Buja (1986) succeeded in giving a general and comprehensive answer. Using a general result from topological measure theory (Kuratowski isomorphism theorem) he showed that on Polish spaces pseudo-capacities not fulfilling (7) for the usual topology must nevertheless obey this condition for some non-standard topology. Conditions, which are sufficient to extent proposition 4.1 to non-Polish spaces, are given in Hummitzsch (1978, see especially Satz 6.4, p. 30).

The other main topic is initiated by the fact that the Huber-Strassen theorem is a general existence result without providing a method for the construction of least favorable pairs. Rieder (1977) presented a solution for most of the (ϵ, δ)-contamination models. Bednarski (1981, p. 402f) derived sufficient conditions for pseudo-capacities under which the likelihood-ratio of a globally least favorable pair is a monotone function of the likelihood-ratio of the central distributions and described special cases, where the construction can be done by differentiating. Of particular interest in this context are the contributions of Österreicher and Hafner. Starting with Österreicher (1978) the leitmotif of their work is to use model-specific characteristics of the generalized risk-function for constructing the likelihood-ratio of a globally least favorable pair. (Cf. Hafner, 1992, for summarizing some aspects.) For several neighborhood models they managed to find that transformation, which leads to the risk-function of a globally least favorable pair (e.g. for the Prohorov model see: Hafner, 1982). Furthermore, Hafner was able to show that similar methods can also be used for models defined via lower and upper density functions or via lower and upper distribution functions (Hafner, 1987, 1993). — All these methods elegantly use particular properties of the special models considered. Additionally, as a byproduct of the considerations in Augustin (1998, chapter 5), for models on finite spaces a general algorithm for calculating globally least favorable pairs via linear programming can be developed, which does not assume a certain type of models (see Augustin, 1998, Korollar 5.12, p. 196f).

There are many other problems, where the solutions essentially are based on the Huber-Strassen theorem. Two examples for this are the extension to dependent random variables (Bednarski, 1986) and the development of asymptotic (level-α-)maximin-tests under sequences of shrinking neighborhood models (cf. for instance: Rieder, 1978, Bednarski, 1985, and Rieder, 1994, chapter 5.4 and the references therein).

5 Globally least favorable pairs under general interval probability

In section 2.3 it was discussed that C-probability is too restrictive to serve as an exclusive and indispensable minimal condition for a powerful theory of interval probability. For an extension of Neyman-Pearson theory to F-probability allowing for much more flexible and comprehensive models firstly note that the 'necessity' stated in proposition 4.2 might be toned down! Its premise is rather artificial. If the existence of a globally least favorable pair for *all* possible alternative hypotheses is to be guaranteed, then C-probability is indeed necessary. Surprisingly it has often been overlooked that this does not exclude the existence of a globally least favorable pair in *one concrete* testing problem, where neither $\mathcal{F}_0$ nor $\mathcal{F}_1$ are C-probability fields. This is of particular interest, because it will turn out that there is a plenty of relevant models, where both hypotheses are not described by C-probability fields, but nevertheless globally least favorable pairs exist. One example is provided by F-probability fields, which are parametricly constructed (cf. definition 2.5) with respect to a family with monotone likelihood-ratio. Using lemma 3.5 and the relation between monotone likelihood-ratio and stochastically ordered classes of distributions one obtains

Proposition 5.1 (Existence in the case of monotone likelihood-ratio) *Consider a (strict) parametric family* $\mathcal{Q} = \{p_\theta(\cdot) \mid \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}$, *of (mutually absolutely continuous) classical probabilities on* $(\mathbb{R}, \mathcal{B})$ *with monotone likelihood-ratio in* θ. *If the F-probability fields* $\mathcal{F}_0(\boldsymbol{\theta}_0)$ *and* $\mathcal{F}_1(\boldsymbol{\theta}_1)$ *are parametricly constructed with respect to* $\mathcal{Q}$ *with parameters* $\boldsymbol{\theta}_0 = [\theta_0^L, \theta_0^U] \subset \Theta$ *and* $\boldsymbol{\theta}_1 = [\theta_1^L, \theta_1^U] \subset \Theta$, $\theta_0^U < \theta_1^L$, *then*

$$(p_{\theta_0^U}(\cdot), p_{\theta_1^L}(\cdot))$$

is a globally least favorable pair for testing $\mathcal{F}_0(\boldsymbol{\theta}_0)$ *versus* $\mathcal{F}_1(\boldsymbol{\theta}_1)$.

An enrichment gained by allowing for F-probability is the study of *generalized pseudo-capacities*. Usual, (not generalized) pseudo-capacities are based on the 'convex distortion' of a single classical probability used as central distribution (see remark 2.8). This can substantially be extended by considering interval-valued central distributions. In particular this leads to neighborhood models of F-probabilities.

Proposition 5.2 (Generalized pseudo-capacities) *Let* $\mathcal{F} = (\Omega, \mathcal{A}, L(\cdot))$ *be an F-probability field with structure* $\mathcal{M}$ *and* $f(\cdot) : [0,1] \to [0,1]$ *a convex function with* $f(0) = 0$*. Then the* generalized pseudo-capacity

$$(f \otimes P)(\cdot) \stackrel{def}{=} [(f \otimes L)(\cdot), (1 - f \otimes L)(\cdot^C)]$$

with $(f \otimes L)(\Omega) = 1$ *and*

$$(f \otimes L)(A) \stackrel{def}{=} f\,(L(A))\,, \quad A \in \mathcal{A} \setminus \{\Omega\}\,,$$

is F-probability (with its structure denoted by $\mathcal{M}(f \otimes L)$*).*

Furthermore, with $\mathcal{M}(f \otimes p)$ *as defined in remark 2.8,*

$$\tilde{\mathcal{M}} \stackrel{def}{=} \bigcup_{p(\cdot) \in \mathcal{M}} \mathcal{M}(f \otimes p) \tag{14}$$

is a prestructure of the F-probability field $(f \otimes \mathcal{F}) \stackrel{def}{=} (\Omega, \mathcal{A}, (f \otimes L)(\cdot))$.

Proof: First of all it should be noted that the conditions imposed on $f(\cdot)$ guarantee that $f(\cdot)$ is not decreasing. This implies

$$\forall p(\cdot) \in \mathcal{K}(\Omega, \mathcal{A})\,, \forall A \in \mathcal{A}\; [\,p(A) \geq L(A) \Rightarrow (f \otimes p)(A) \geq (f \otimes L)(A)\,]\,. \tag{15}$$

The proof now consists of three steps:

i) Firstly it is shown that $\tilde{\mathcal{M}}$ *is not empty*: For this one uses that the structure $\mathcal{M}$ of the central distribution is not empty and takes any $p(\cdot) \in \mathcal{M}$. The corresponding set-function $(f \otimes p)(\cdot)$ is the lower interval limit of a pseudo-capacity with a non-empty structure $\mathcal{M}(f \otimes p)(\cdot)$.

ii) The second step proves the relation $\tilde{\mathcal{M}} \subseteq \mathcal{M}(f \otimes L)$ where

$$\begin{aligned} \mathcal{M}(f \otimes L) &= \{p(\cdot) \in \mathcal{K}(\Omega, \mathcal{A}) \mid \\ &\qquad (f \otimes L)(A) \leq p(A) \leq (1 - f \otimes L)(A^C)\,, \forall A \in \mathcal{A}\} \\ &= \{p(\cdot) \in \mathcal{K}(\Omega, \mathcal{A}) \mid (f \otimes L)(A) \leq p(A)\,, \forall A \in \mathcal{A}\}\,. \end{aligned}$$

By the definition of $\tilde{\mathcal{M}}$ according to (14) there is for every $\tilde{p}(\cdot) \in \tilde{\mathcal{M}}$ a $p(\cdot)$ in $\mathcal{M}$ so that $\tilde{p}(\cdot)$ is an element of the structure $\mathcal{M}(f \otimes p)$ of the (not generalized) pseudo-capacity $[(f \otimes p)(\cdot)\,, (1 - f \otimes p)(\cdot^C)]$. Since this is F-probability one has $\tilde{p}(A) \geq (f \otimes p)(A)$ for all $A \in \mathcal{A}$. Using the monotonicity formulated in (15) and the relation $p(A) \geq L(A)$, $A \in \mathcal{A}$ gained from the fact that $\mathcal{F}$ is an F-probability field, it can be concluded that $\tilde{p}(A) \geq (f \otimes L)(A)$ for every $A \in \mathcal{A}$, i.e. $\tilde{p}(\cdot) \in \mathcal{M}(f \otimes L)$. In particular $\mathcal{M}(f \otimes L) \neq \emptyset$ and $(f \otimes P)(\cdot)$ is R-probability with structure $\mathcal{M}(f \otimes L)$.

iii) To conclude the proof it is shown that for every $A \in \mathcal{A}$ and every $\epsilon > 0$ there is a classical probability $\tilde{p}(\cdot) \in \tilde{\mathcal{M}}$ with $\tilde{p}(A) < (f \otimes L)(A) + \epsilon$. From this it is deduced together with the results from ii) and the conjugacy in the definition of $(f \otimes P(\cdot))$ that $(f \otimes P)(\cdot)$ *is F-probability* and $\tilde{\mathcal{M}}$ *is a prestructure* of the corresponding F-probability field.

First consider an event $A \in \mathcal{A}$ with $L(A) = 1$. $\mathcal{F}$ is an F-probability field. Its structure $\mathcal{M}$ is not empty, and for every $p(\cdot) \in \mathcal{M}$ one obtains $p(A) = 1 = L(A)$. Now fix one such $p(\cdot)$. The pseudo-capacity corresponding to $(f \otimes p)(\cdot)$ is F-probability. Therefore there is an $\tilde{p}(\cdot) \in \mathcal{M}(f \otimes p)$ with $\tilde{p}(A) < (f \otimes p)(A) + \epsilon = (f \otimes L)(A) + \epsilon$.

In the case of an event $A \in \mathcal{A}$ with $L(A) < 1$ one uses the continuity of $f(\cdot)$ on $[0, 1)$, which can be deduced from the upper semicontinuity of $f(\cdot)$ on $[0, 1]$ (cf. e.g. Rockafellar (1972, theorem 10.2, p. 84)) and from $f(0) = 0$ and $f(\cdot) \geq 0$. Therefore, for every $\epsilon > 0$ there exists a $\delta > 0$ so that $|f(y) - f(L(A))| < \epsilon/2$ for all $y \in [0, 1)$ with $|y - L(A)| < \delta$. Since $\mathcal{F}$ is an F-probability field there is a $p(\cdot) \in \mathcal{M}$ with $p(A) < L(A) + \delta$, which implies $(f \otimes p)(A) < (f \otimes L)(A) + \epsilon/2$. Consider the pseudo-capacity corresponding to $(f \otimes p)(\cdot)$. It is F-probability. So there is a $\tilde{p}(\cdot) \in \mathcal{M}(f \otimes p)$ with $\tilde{p}(A) < (f \otimes p)(A) + \epsilon/2 < (f \otimes L)(A) + \epsilon$. □

To find least favorable pairs for testing two generalized pseudo-capacities $(f_0 \otimes \mathcal{F}_0)$ and $(f_1 \otimes \mathcal{F}_1)$ it may be promising to proceed in two steps. One may try to reduce the testing problem firstly to a testing problem between a so-to-say *'least favorable pair of F-probabilities'*. These should be more easy to handle but nevertheless represent the whole testing problem in the sense that globally least favorable pairs for testing between them are also globally least favorable pairs for the complex problem.

It would be most functional, if one could calculate the least favorable F-probabilities from the testing problem between the central-distributions. In the rest of this section situations are described, where this convenient way of splitting up the complex problem into simpler ones proves to be successful. Theorem 5.3 and theorem 5.5 will state sufficient conditions under which the pair of 'least favorable F-probabilites' consists just of the (usual, not generalized) pseudo-capacities around the elements of the globally least favorable pair for testing between the central distributions with the same 'distortion functions' $f_0(\cdot)$ and $f_1(\cdot)$. Then the following procedure can be recommended to obtain globally least favorable pairs for testing between generalized pseudo-capacities:

- Firstly search for a globally least favorable pair $(q_0(\cdot), q_1(\cdot))$ for testing between the F-probability fields forming the central distributions.
- Secondly, if the first step proved successful, determine a globally least favorable pair for testing between the (not generalized) pseudo-capacities around $q_0(\cdot)$ and $q_1(\cdot)$. Then it is a globally least favorable pair for the problem $(f_0 \otimes \mathcal{F}_0)$ versus $(f_1 \otimes \mathcal{F}_1)$, too.

In these situations the efficient construction methods for (usual, not generalized) pseudo-capacities cited in section 4.2 can also be used without any additional complication to obtain least favorable pairs in the complex situations.

Theorem 5.3 (Least favorable pseudo-capacities I) *Let $(f_0 \otimes \mathcal{F}_0)$ and $(f_1 \otimes \mathcal{F}_1)$ be two generalized pseudo-capacities on a Polish space with $f_i(x_0) = 0$ for an $x_0 \in (0, 1)$, $i \in \{0, 1\}$. If there exists a globally least favorable pair $(q_0(\cdot), q_1(\cdot))$ for $\mathcal{F}_0$ versus $\mathcal{F}_1$ fulfilling the three conditions*

a) $q_0(\cdot)$ and $q_1(\cdot)$ are mutually absolute continuous,

b) for every $\alpha \in (0, 1)$ there is a best level-α-test for $\{q_0(\cdot)\}$ versus $\{q_1(\cdot)\}$, which is non-randomized,

c) there are prestructures $\mathcal{V}_0$ and $\mathcal{V}_1$ of $\mathcal{F}_0$ and $\mathcal{F}_1$, which are dominated by $q_0(\cdot)$,

then the following holds:

1) There exists a globally least favorable pair for $(f_0 \otimes \mathcal{F}_0)$ versus $(f_1 \otimes \mathcal{F}_1)$.

2) $((f_0 \otimes q_0)(\cdot), (f_1 \otimes q_1)(\cdot))$ is a pair of least favorable pseudo-capacities *in the following sense: If $(\bar{q}_0(\cdot), \bar{q}_1(\cdot))$ is a globally least favorable* *pair for* $(\Omega, \mathcal{A}, (f_0 \otimes q_0)(\cdot))$ *versus* $(\Omega, \mathcal{A}, (f_1 \otimes q_1)(\cdot))$*, then it is a globally least favorable pair for* $(f_0 \otimes \mathcal{F}_0)$ *versus* $(f_1 \otimes \mathcal{F}_1)$*, too.*

Proof: According to Buja (1986) (cf. section 4.2) there always exists a globally least favorable pair for testing the two (not generalized) pseudo-capacities $(\Omega, \mathcal{A}, (f_0 \otimes q_0)(\cdot))$ and $(\Omega, \mathcal{A}, (f_1 \otimes q_1)(\cdot))$. So the first statement follows from the second one.

To verify that every globally least favorable pair $(\bar{q}_0, \bar{q}_1)$ for testing $(\Omega, \mathcal{A}, (f_0 \otimes q_0)(\cdot))$ and $(\Omega, \mathcal{A}, (f_1 \otimes q_1)(\cdot))$ is also a globally least favorable pair for the complex problem $(f_0 \otimes \mathcal{F}_0)$ versus $(f_1 \otimes \mathcal{F}_1)$, only condition (12) will be shown more detailed. The proof of (11) is analogous, and (10) is clear.

Firstly one follows the lines of the proof of theorem 5.1 in Bednarski (1981, p. 402) to confirm that – under the condition a) and b) – there is for every $t \in I\!R_0^+$ an $w(t)$ with

$$\{\omega \mid \bar{\pi}(\omega) > t\} = \{\omega \mid \pi(\omega) > w(t)\}, \quad q_0(\cdot) + q_1(\cdot) \ a.s.$$

Here $\bar{\pi}(\cdot)$ and $\pi(\cdot)$ denote suitable versions of the likelihood-ratios of $\bar{q}_0(\cdot)$ and $\bar{q}_1(\cdot)$ and of $q_0(\cdot)$ and $q_1(\cdot)$, respectively.

Then one uses that $q_0(\cdot)$ and – because of a) also – $q_1(\cdot)$ dominate the prestructures $\mathcal{V}_0$ and $\mathcal{V}_1$ and therefore according to lemma 2.6 also the structures $\mathcal{M}_0$ and $\mathcal{M}_1$. So the relation $\{\omega \mid \bar{\pi}(\omega) > t\} = \{\omega \mid \pi(\omega) > w(t)\}$ also holds $p_1(\cdot)$ almost surely for any $p_1(\cdot)$ in $\mathcal{M}_1$.

Together with the fact that $(q_0(\cdot), q_1(\cdot))$ is a globally least favorable pair for $\mathcal{F}_0$ versus $\mathcal{F}_1$ this leads to

$$\begin{aligned} q_1(\{\omega \mid \bar{\pi}(\omega) > t\}) &= q_1(\{\omega \mid \pi(\omega) > w(t)\}) \leq \\ &\leq p_1(\{\omega \mid \pi(\omega) > w(t)\}) = p_1(\{\omega \mid \bar{\pi}(\omega) > t\}) \end{aligned} \tag{16}$$

for every element $p_1(\cdot)$ of the structure $\mathcal{M}_1$ of $\mathcal{F}_1$ and every $t \in I\!R_0^+$.

The function $f_1(\cdot)$ is not decreasing (cf. the proof of proposition 5.2). From this one concludes

$$(f_1 \otimes q_1)(\{\omega \mid \bar{\pi}(\omega) > t\}) \leq (f_1 \otimes p_1)(\{\omega \mid \bar{\pi}(\omega) > t\}), \qquad \forall p_1(\cdot) \in \mathcal{M}_1 .$$

The expression $(f_1 \otimes q_1)(\{\omega \mid \bar{\pi}(\omega) > t\})$ on the left hand side of this inequality is – according to (13) – equal to $\bar{q}_1(\{\omega \mid \bar{\pi}(\omega) > t\})$, because $(\bar{q}_0(\cdot), \bar{q}_1(\cdot))$ is a globally least favorable pair for $(\Omega, \mathcal{A}, (f_0 \otimes q_0)(\cdot))$ versus $(\Omega, \mathcal{A}, (f_1 \otimes q_1)(\cdot))$. Therefore

$$\bar{q}_1(\{\omega \mid \bar{\pi}(\omega) > t\}) \leq (f_1 \otimes p_1)(\{\omega \mid \bar{\pi}(\omega) > t\}), \qquad \forall p_1(\cdot) \in \mathcal{M}_1 .$$

Looking on the right hand side one uses that $(f_1 \otimes p_1)(\cdot)$ is a pseudo-capacity for every $p_1(\cdot) \in \mathcal{M}_1$. So for every element $s_1(\cdot)$ in the corresponding structure $\mathcal{M}(f_1 \otimes p_1)$ one obtains

$$(f_1 \otimes p_1)(\{\omega \mid \bar{\pi}(\omega) > t\}) \le s_1(\{\omega \mid \bar{\pi}(\omega) > t\}) \ .$$

Finally, the last two relations are brought together and the union of all $p_1(\cdot) \in \mathcal{M}_1$ is taken. This leads to

$$\bar{q}_1(\{\omega \mid \bar{\pi}(\omega) > t\}) \le s_1(\{\omega \mid \bar{\pi}(\omega) > t\}) \ , \ \forall s_1(\cdot) \in \tilde{\mathcal{M}}_1 = \bigcup_{p_1 \in \mathcal{M}_1} \mathcal{M}(f_1 \otimes p_1) \, .$$

Since, according to proposition 5.2, the set $\tilde{\mathcal{M}}_1$ is a prestructure of $(f_1 \otimes \mathcal{F}_1)$, the claim is gained via lemma 3.5. $\square$

Example 5.4 (Generalized pseudo-capacities around F-normal distributions) *An important class of interval probabilities, where the requirements of theorem 5.3 are usually fulfilled, consists of parametricly constructed interval probabilities with respect to a family, which is absolutely continuous with respect to the Lebesgue-measure and possesses monotone likelihood ratio. (These conditions, however, are not always sufficient. A well-known counterexample is provided by the family of continuous uniform distributions.) In particular theorem 5.3 applies to generalized pseudo-capacities around F-normal distributions: According to proposition 5.1 the classical probabilities $q_0(\cdot)$ and $q_1(\cdot)$ forming the globally least favorable pair are the normal-distributions corresponding to the right and to the left border of the parameter intervals, respectively. They are mutually absolutely continuous, for every level of significance the optimal test between them can be chosen to be non-randomized, and they dominate the prestructures consisting of all normal distributions in the parameter interval and therefore (cf. lemma 2.6) both structures.*

Discussing the area of application of theorem 5.3 it should be stressed that this theorem only is valid when a globally least favorable pair exists for testing between the central distributions – indeed, this condition seems to be indispensable for any related form of splitting up the testing problem. Given a globally least favorable pair, condition a) of theorem 5.3 is fulfilled in most situations. From a practical point of view also the requirement b) is by far not so restrictive as it might look at a first glance. The nonexistence of non-randomized optimal tests usually occurs on discrete spaces. But this is only a problem if the discrete space is really of infinite cardinality. When the space is finite the reduction to least favorable pairs of F-probabilities aimed at in theorem 5.3 is not necessary. In this situation globally least favorable pairs can be obtained from parametric linear optimization (Augustin, 1998, Proposition 5.11, p. 192f).

However, for not parametricly constructed interval probability the third condition, which in particular forces the structures of the central distributions to be dominated, is to some degree demanding. In several situations modeling with interval probability naturally leads to undominated structures. (For example most neighborhood-models can produce undominated structures.)

Especially for such cases a variate of theorem 5.3 is of interest, which implicitly slightly restricts the form of the 'distortion functions' $f_0(\cdot)$ and $f_1(\cdot)$ in favor of the possibility to handle undominated structures and to do also without the other conditions just discussed.

Studying the results of the construction techniques reviewed in section 4.2 one recognizes that – in the notation needed below – nearly for all common pseudo-capacities around two classical probabilities $q_0(\cdot)$ and $q_1(\cdot)$ with likelihood ratio $\pi(\cdot)$ the likelihood ratio $\bar{\pi}(\cdot)$ of every globally least favorable pair is of the form

$$\bar{\pi}(\omega) = \begin{cases} \bar{t} & \pi(\omega) > t_{\max} \\ h(\pi(\omega)) \quad \text{if} & \pi(\omega) \in [t_{\min}, t_{\max}] \\ \underline{t} & \pi(\omega) < t_{\min} \end{cases} \tag{17}$$

with a non-decreasing function $h(\cdot)$ and appropriate quantities $\bar{t}, t_{max}, \underline{t}$, t_{min}. Additional support for this form can be gained from Bednarski (1981, theorem 5.1, p. 402) and also by following the original heuristic (cf. e.g. Huber, 1965) behind robust testing of the two classical probabilities $q_0(\cdot)$ and $q_1(\cdot)$: Their likelihood ratio $\pi(\cdot)$ is very sensitive to observations with very small likelihood so that one outlier can nearly determine the whole product $\Pi_{i=1}^{n}\pi(x_i)$ of a sample $(x_1, \ldots, x_n)$. To alleviate this unwanted effect one transforms and truncates the likelihood ratio to restrict the influence of extreme observations.

If $h(\cdot)$ is not only non-decreasing but strictly increasing, the two (not generalized) pseudo-capacities around the components of the least favorable pair for testing the central distributions can be again shown to form a pair of least favorable F-probabilities.

Theorem 5.5 (Least favorable pseudo-capacities II) *Let* $(f_0 \otimes \mathcal{F}_0)$ *and* $(f_1 \otimes \mathcal{F}_1)$ *be two generalized pseudo-capacities on a Polish space and assume that there exists a globally least favorable pair* $(q_0(\cdot), q_1(\cdot))$ *for* $\mathcal{F}_0$ *versus* $\mathcal{F}_1$*. If* $(\bar{q}_0(\cdot), \bar{q}_1(\cdot))$ *is a globally least favorable pair for* $(\Omega, \mathcal{A}, (f_0 \otimes q_0)(\cdot))$ *versus* $(\Omega, \mathcal{A}, (f_1 \otimes q_1)(\cdot))$*, whose likelihood ratio* $\pi(\cdot)$ *is of the form (17) with a strictly increasing function* $h(\cdot)$*, then the following holds:*

1) *There exists a globally least favorable pair for* $(f_0 \otimes \mathcal{F}_0)$ *versus* $(f_1 \otimes \mathcal{F}_1)$*.*

2) $(\bar{q}_0(\cdot), \bar{q}_1(\cdot))$ *is a globally least favorable pair for* $(f_0 \otimes \mathcal{F}_0)$ *versus* $(f_1 \otimes \mathcal{F}_1)$*, too.*

Proof: Just as in the proof of theorem 5.3 according to Buja's (1986) general existence result the first statement follows from the second one.

For the second part one observes that the additional assumption $f_i(x_0) = 0$ for a number $x_0 \in (0,1)$, $i \in \{0,1\}$ and the conditions a), b), and c) formulated in theorem 5.3 were only needed to establish relation (16). The rest of the proof of theorem 5.3 does not recur on them. So that proof literally carries over to any situation, where one succeeds in deducing (16) in a different way.

To verify (16) in the case considered here one uses the special form of $\bar{\pi}(\cdot)$ and the fact that, because $h(\cdot)$ is strictly monotone, its inverse $h^{-1}(\cdot)$ exists. Therefore for every $t \in \mathbb{R}_0^+$

$$\{\omega \mid \bar{\pi}(\omega) > t\} = \begin{cases} \emptyset & t \geq \bar{t} \\ \Omega \quad \text{if} & t < \underline{t} \\ \left\{\omega \mid \pi(\omega) > h^{-1}(t)\right\} & t \in [\underline{t}\,,\,\bar{t}) \end{cases} .$$

For every $t \in [\underline{t}\,,\,\bar{t})$ and for every element $p_1(\cdot)$ of the structure $\mathcal{M}_1$ of $\mathcal{F}_1$ one therefore obtains, using the fact that $(q_0(\cdot), q_1(\cdot))$ is a globally least favorable pair for $\mathcal{F}_0$ versus $\mathcal{F}_1$,

$$\begin{aligned} q_1\left(\{\omega \mid \bar{\pi}(\omega) > t\}\right) &= q_1\left(\left\{\omega \mid \pi(\omega) > h^{-1}(t)\right\}\right) \leq \\ &\leq p_1\left(\left\{\omega \mid \pi(\omega) > h^{-1}(t)\right\}\right) = p_1\left(\{\omega \mid \bar{\pi}(\omega) > t\}\right). \end{aligned}$$

Since $q_1(\{\omega \mid \bar{\pi}(\omega) > t\})$ trivially equals $p_1(\{\omega \mid \bar{\pi}(\omega) > t\})$ for t outside the interval $[\underline{t}\,,\,\bar{t})$, relation (16) is confirmed and the proof is complete. □

It remains to emphasize that requiring $h(\cdot)$ to be strictly monotone is not very restrictive. For instance for the practically important class of proper (ϵ, δ)−contamination models it is well-known that this is indeed always guaranteed (see e.g. Rieder, 1977, p. 914f., or Hafner 1992, p. 153, who also give formulas for calculating the quantities $\overline{t}, t_{max}, \underline{t}, t_{min}$). Therefore in this case, as long as there exists a globally least favorable pair for testing the central distributions, there always exists a pair of least favorable F-probabilities for testing between the generalized pseudo-capacities – not depending on whether the structures of the central distributions are dominated or not.

6 Locally least favorable pairs

It turned out that for many situations of practical interest with F- but not C-probability underlying globally least favorable pairs exist. Even in situations where no globally least favorable pair exists, one can often profit from the vivid possibility of a reduction to least favorable elements of the structure. The concept of globally least favorable pairs can be modified in a way that the main argument of the proof of proposition 3.4 remains valid. If the level of significance α is given and fixed (as usual in Neyman-Pearson-theory), it is only of importance to find classical probabilities, which are least favorable for that concrete level of significance *('locally')*. This is a much weaker condition, but it will nevertheless prove to be sufficient for formulating equivalents to the propositions 3.4 and 4.1.

Definition 6.1 (Locally least favorable pairs) *Consider problem 3.1, and let a level of significance* $\alpha \in (0,1)$ *be given. A pair* $(q_0(\cdot), q_1(\cdot))$ *of classical probabilities is called a* (level-α-)locally least favorable pair *(for* $\mathcal{F}_0$ *versus* $\mathcal{F}_1$*), if* $(q_0(\cdot), q_1(\cdot)) \in \mathcal{M}_0 \times \mathcal{M}_1$*, and there exists a best test* $\varphi^*(\cdot)$ *for* $\{q_0(\cdot)\}$ *versus* $\{q_1(\cdot)\}$ *with* $U\mathbb{E}_{\mathcal{M}_0}\varphi^* \leq \alpha$ *and* $\mathbb{E}_{q_1}\varphi^* = L\mathbb{E}_{\mathcal{M}_1}\varphi^*$.

It is straightforward to prove that locally least favorable pairs still lead to level-α-maximin-tests.

Proposition 6.2 (Locally least favorable pairs and level-α-maximin-tests) *If* $(q_0(\cdot), q_1(\cdot))$ *is a level-α-locally least favorable pair for* $\mathcal{F}_0$ *versus* $\mathcal{F}_1$*, then there exists a best level-α test for testing the hypothesis* $\overline{H_0} : \{q_0(\cdot)\}$ *versus the hypothesis* $\overline{H_1} : \{q_1(\cdot)\}$*, which is a level-$\alpha$-maximin-test for* $\mathcal{F}_0$ *versus* $\mathcal{F}_1$*, too.*

Also the existence of locally least favorable pairs can be guaranteed under quite general conditions. (Note, however, that condition (18) is a bit stronger than its pendant (7), because it refers to arbitrary measurable sets, and not only to the open ones. The topological properties, which are induced by these conditions on the structures, are compared in Augustin, 1998, chapter 3.3.4.).

Theorem 6.3 (Existence of locally least favorable pairs) *If $\mathcal{F}_0$ and $\mathcal{F}_1$ are fulfilling the condition*

$$(A_n)_{n\in I\!N} \uparrow A,\, n \in I\!N \Longrightarrow \lim_{n\to\infty} L_i(A_n) = L_i(A)\,, \quad i \in \{0,1\}\,, \qquad (18)$$

then there exists for every given $\alpha \in (0,1)$ a level-α-locally least favorable pair for $\mathcal{F}_0$ versus $\mathcal{F}_1$.

Sketch of the proof: (Cf. Augustin, 1998, Proposition 2.11, p. 76 and Satz 3.14, p. 113.) From Gänssler (1971) it can be derived that condition (18) is sufficient for the compactness of $\mathcal{M}_0$ and $\mathcal{M}_1$ (in an appropriate weak-star-topology). Since furthermore every structure is convex, theorem 6.3 can be shown by adopting results from Baumann (1968). □

7 Concluding remarks

The 'local perspective' of section 6, which concentrates on a fixed level of significance, enables one to develop universally applicable algorithms for calculating level-α-maximin-tests and locally least favorable pairs on finite sample spaces (Augustin, 1998, chapter 4 and 5). In general, by comparing globally and locally least favorable pairs one recognizes that both provide elegant means to reduce the testing problem to a much simpler one. The main disadvantage of locally least favorable pairs, however, is that no equivalent to proposition 3.6 has been proved. For the proof of proposition 3.6 it is essential that the relations (11) and (12) hold for *every* t meaning that the least favorable position is global. So it seems as if in the case of independent repetitions a reduction to the one-dimensional case were not possible. Efficient procedures for relating locally least favorable pairs for testing the n-dimensional product to simpler situations are open for further research.

In section 5 generalized pseudo-capacities were introduced as an extension of the usual neighborhood-models. The results on reducing the complex testing problem to the task of testing between least favorable pseudo-capacities around single classical probabilities allow for the study of neighborhood models of interval probabilities in an efficient way. Furthermore, generalized pseudo-capacities look quite promising to provide means for an extension of problem 3.1 to the Generalized Composite Neyman Pearson Problem, where H_0 and H_1 consist of several interval-valued hypotheses. — The level-α-maximin-criterion considers exclusively the maximal error. Research on optimal tests and equivalents to least favorable pairs in situations, where other interval orderings are used to judge the probability of error would be highly desirable, but is rather rare. For a symmetric Hurwitz-like criterion some substantial results have been gained by Jaffray and Said (1994).

Another topic deserving detailed investigation should only be briefly and informally mentioned here: If one gives up the requirement underlying the axioms that every (generalized) probability assignment should be completely described by the (interval-valued) probability of the events, one can go a step further and consider interval-valued *expectation* as the basic concept (see, under neglecting the difference between countable and finite additivity for a moment, especially the theory of Walley (1991). Proceeding along these lines sets of classical probabilities ('imprecise probabilities') arising from an equivalent to (1) may be of a more general form than structures can be (see Walley, 1991, section 2.7.3, p. 82ff, for an example). If one uses such sets as prestructures to obtain corresponding F-probability fields, one remains in the framework just described. However, if one is worried that (too) much information might be lost by proceeding from prestructures to structures, one can also try to study the Neyman Pearson problem between two imprecise probabilities themselves.

The present work was motivated by the insight that C-probability is too restrictive to serve as an exclusive basis for interval probability. Therefore further research should additionally provide some answer to the question, whether the results gained here are also of importance beyond the Neyman-Pearson approach, e.g. for robust Bayesian analysis. For instance, is it possible to extend the results gained there on pseudo-capacities as models for prior belief to the more flexible and expressive class of generalized pseudo-capacities along the lines of proposition 5.2 and theorem 5.3?

Acknowledgements

I am grateful to Kurt Weichselberger, Toni Wallner, Bernhard Rüger, Peter Walley, Astrid Heinicke and several discussants at ISIPTAA 99 for their valuable comments.

References

Augustin, T., 1998. Optimale Tests bei Intervallwahrscheinlichkeit. Vandenhoeck & Ruprecht, Göttingen.

Baumann, V., 1968. Eine parameterfreie Theorie der ungünstigsten Verteilungen für das Testen von Hypothesen. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 11, 41-60.

Bednarski, T., 1981. On the solution of minimax problems for special capacities. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 58, 397-405.

Bednarski, T., 1985. A robust asymptotic testing model for special capacities. Mathematische Operationsforschung und Statistik – Series Statistics. 4, 507-519.

Bednarski, T., 1986. On the extension of the Neyman-Pearson lemma for capacities to dependent random variables. Mathematische Operationsforschung und Statistik – Series Statistics. 5, 549-556.

Bernard, J.-M., 1999. Implicative analysis for multivariate binary data using an imprecise Dirichlet model. In: de Cooman, G., Cozman, F.G., Moral, S., Walley, P. (Eds.): ISIPTA'99: Proceedings of the First International Symposium on Imprecise Probabilities and their Applications, pp. 24-33.

Buja, A., 1986. On the Huber-Strassen theorem. Probability Theory and Related Fields. 73, 149-152.

Choquet, G., 1954. Theory of capacities. Annales de l'Institut Fourier. 5, 131-295.

de Cooman, G., Walley, P., 1999. The Imprecise Probability Project. http://ensmain.rug.ac.be/~ipp.

Gänssler, P., 1971. Compactness and sequential compactness in spaces of measures. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 17, 124-146.

Hafner, R., 1982. Simple construction of least favorable pairs of distributions and of robust tests for Prohorov-neighborhoods. Mathematische Operationsforschung und Statistik — Series Statistics. 13, 33-46.

Hafner, R., 1987. Construction of minimax-tests for bounded families of distribution-functions. In: Sendler, W. (Ed.), Contributions to Stochastics, Physika, Heidelberg.

Hafner, R., 1992. Konstruktion robuster Teststatistiken. In: Schach, S., Trenkler, G. (Eds.), Data Analysis and Statistical Inference, Eul, Bergisch Gladbach, pp. 145-160.

Hafner, R., 1993. Construction of minimax-tests for bounded families of probabilities-densities. Metrika. 40, 1-23.

Huber, P.J., 1965. A robust version of the probability ratio test. Annals of Mathematical Statistics. 36, 1753-1758.

Huber, P.J., 1973. The use of Choquet capacities in statistics. Bulletin of the International Statistical Institute. XLV, Book 4, 181-188.

Huber, P.J., Strassen, V., 1973. Minimax tests and the Neyman-Pearson lemma for capacities. Annals of Statistics. 1, 251-263, Correction, 2, 223-224.

Hummitzsch, W., 1978. Ungünstigste Verteilungen und Approximation von Minimax-Tests. Dissertation, Ruhr-University of Bochum. Bochum.

Jaffray, J.-Y., Said, T., 1994. Optimal hypothesis testing with a vague prior. In: Rios, S. (Ed.), Decision Theory and Decision Analysis: Trends and Challenges, Kluwer, Boston, pp. 207-222.

Lembcke, J., 1988. The necessity of strongly subadditive capacities for Neyman-Pearson minimax tests. Monatshefte für Mathematik. 105, 113-126.

Österreicher, F., 1978. On the construction of least favorable pairs of distributions. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 43, 49-55.

Papamarcou, A., Fine, T.L., 1986. A note on undominated lower probabilities. The Annals of Probability. 14, 710-723.

Papamarcou, A., Fine, T.L., 1991. Unstable collectives and envelopes of probability measures. The Annals of Probability. 19, 893-906.

Rieder, H., 1977. Least favorable pairs for special capacities. The Annals of Statistics. 5, 909-921.

Rieder, H., 1978. A robust asymptotic testing model. The Annals of Statistics. 6, 1080-1094.

Rieder, H., 1994. Robust Asymptotic Statistics. Springer, New York.

Rockafeller, R. T., 1972. Convex Analysis (2nd. edition). Princeton University Press, Princeton.

Strassen, V., 1964. Meßfehler und Information. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 2, 273-305.

Strassen, V., 1965. The existence of probability measures with given marginals. Annals of Mathematical Statistics. 36, 423-439.

Walley, P., 1981. Coherent Lower (and Upper) Probabilities. Research Report, Department of Statistics, University of Warwick. Warwick.

Walley, P., 1991. Statistical Reasoning with Imprecise Probabilities. Chapman & Hall, London.

Walley, P., 1996. Inferences from multinomial data: Learning from a bag of marbles (with discussion). Journal of the Royal Statistical Society, B, 58, 3-57.

Walley, P., Fine, T.L., 1982. Towards a frequentist theory of lower and upper probability. The Annals of Statistics. 10, 741-761.

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

Weichselberger, K., 1995A. Axiomatic foundations of the theory of interval-probability. In: Mammitzsch, V. and Schneeweiß, H. (Eds.). Proceedings of the Second Gauß Symposion, Section B, de Gruyter, Berlin, pp. 47-64.

Weichselberger, K., 1995B. Stichproben und Intervallwahrscheinlichkeit. ifo Studien 41, 653-676.

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

Weichselberger, K., 1999A. Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung. Two volumes. In preparation.

Weichselberger, K., 1999B. The theory of interval-probability as a unifying concept for uncertainty. International Journal of Approximate Reasoning, to appear.

Weichselberger, K., Pöhlmann, S., 1990. A Methodology for Uncertainty in Knowledge-Based Systems. Springer, Berlin.

Witting, H., 1985. Mathematische Statistik I. Teubner, Stuttgart.

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