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

Th. Augustin
Ludwig-Maximilians University, Munich
thomas@stat.uni-muenchen.de

Abstract

The paper studies the Generalized Neyman-Pearson problem where both hypotheses are described by interval probability. First the Huber-Strassen theorem and the literature based on it is reviewed. Then some recent 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 ([29]).

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

1 Introduction

Interval probability is a substantial extension of the usual concept of probability. By assigning intervals $[L(A), U(A)]$ instead of single real numbers $p(A)$ an appropriate modeling of more general aspects of uncertainty is provided. Though the idea to use interval-valued probabilities has a long history,[1] the main steps towards a comprehensive theory have been achieved only recently.

In the last years the interest has mainly concentrated on generalized Bayesian inference. Nevertheless, interval probability also proves to be quite important for the non-Bayesian ('objectivistic/ frequentist') point of view. The present paper reviews and extends results on the generalization of the Neyman-Pearson alternative testing problem to the situation where both hypotheses are described by interval probability. Section 2 states some basic definitions from Weichselberger's interpretation-independent concept of general interval probability (F-probability), which is underlying this work. Also the special case of C-probability is considered, which contains two monotone capacities and pseudo-capacities and therefore the neighborhood models commonly used in robust statistics. It turns out that nevertheless the condition of being C-probability would be 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 it, is reviewed in section 4. — The next two sections show that it is not necessary to restrict the consideration on 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 nevertheless globally least favorable pairs exist. Furthermore a 'decomposition-theorem' is developed allowing to consider more complex neighborhood models. Section 6 briefly sketches the concept of locally least favorable pairs. Some directions of further research are indicated in the concluding remarks in section 7.

[1] For reviews and references on the development of interval probability see especially [24, Chapter 1 and 5], [29, Chapter 1] and [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 ([29]).[2] His concept is founded on the following generalization of Kolmogorov's axioms ([26],[29, Chapter 2]).

Definition 1 (The Axioms of (totally-determinated) interval probability) Let $(\Omega, \mathcal{A})$ be a measurable space.

[2] For selected aspects cf. also [30], [26], [27], [28].

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

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

  2. The set
  $$\mathcal{M} := \{\, p(\cdot) \in \mathcal{K}(\Omega, \mathcal{A}) \mid L(A) \le p(A) \le U(A),\ \forall A \in \mathcal{A}\} \qquad (1)$$
  is not empty.

- R-probability with structure $\mathcal{M}$ is called *F-probability*, if
$$\left.\begin{array}{l} \inf\limits_{p(\cdot)\in\mathcal{M}} p(A) = L(A) \\ \sup\limits_{p(\cdot)\in\mathcal{M}} p(A) = U(A) \end{array}\right\} \quad \forall A \in \mathcal{A}\,. \qquad (2)$$

For every F-probability $L(\cdot)$ and $U(\cdot)$ are *conjugated*, i.e. $L(A) = 1 - U(A^C)$, $\forall A \in \mathcal{A}$. Therefore every F-probability is uniquely determined either by $L(\cdot)$ or either by $U(\cdot)$ alone. Here $L(\cdot)$ is used, and $\mathcal{F} = (\Omega, \mathcal{A}, L(\cdot))$ is called an *F-probability field.* Weichselberger's theory relies on countable additive classical probability. So F-probability is equivalent to the concept of lower/upper probability in the sense of [13], but the latter term is avoided here because it is also used in the literature in several other meanings. F-probability is also strongly related to coherence in the setting of Walley ([24]) and to the concept of envelopes in the frequentist theory developed by Fine and students (e.g. [25], [17]).

The interrelation 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 (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
$$\begin{array}{rcl} \mathbb{E}_{\mathcal{M}}X & := & [L\mathbb{E}_{\mathcal{M}}X, U\mathbb{E}_{\mathcal{M}}X] \\ & := & \left[\inf\limits_{p(\cdot)\in\mathcal{M}} \mathbb{E}_p X\,,\ \sup\limits_{p(\cdot)\in\mathcal{M}} \mathbb{E}_p X\right] \\ & \subseteq & [-\infty, \infty] \end{array} \qquad (3)$$
is called *(interval-valued) expectation of X (with respect to $\mathcal{F}$).*

If one gives up the requirement underlying Weichselberger's 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 especially the theory of Walley ([24])[4]). Proceeding along these lines sets $\check{\mathcal{M}}$ arising from an analogy to (1) may be of a more general form than structures can be (see [24, section 2.7.3, p. 82ff.] for an example), leading by some scholars to the conclusion that F-probability might not always be general enough. — But every such set $\check{\mathcal{M}}$ of classical probabilities can be incorporated into the framework considered here by using $\check{\mathcal{M}}$ as a prestructure (see remark 1 below) and transferring it to the structure of the corresponding F-probability. This, however, may lead to some enlargement of $\check{\mathcal{M}}$. If one is worried by this, one will try to use such sets gained from starting with interval-valued expectation directly, i.e. without enlarging them to structures. Some aspects concerning the question how to do so in the generalized Neyman-Pearson problem considered here, are briefly sketched in Section 7.

2.2 Prestructures

As just exemplified, by far not every set of classical probabilities is a structure of an F-probability underlying, but every non-empty set of classical probabilities can be used to construct an unique, narrowest F-probability field corresponding to it.

Remark 1 (Prestructure) *([29, Chapters 2.1 and 2.6])* Let $\mathcal{V}$ be a non-empty set of classical probabilities on a measurable space $(\Omega, \mathcal{A})$. Then $P_{\mathcal{V}}(\cdot) := [L_{\mathcal{V}}(\cdot), U_{\mathcal{V}}(\cdot)]$ with
$$L_{\mathcal{V}}(A) := \inf_{p(\cdot)\in\mathcal{V}} p(A) \quad \wedge \quad U_{\mathcal{V}}(A) := \sup_{p(\cdot)\in\mathcal{V}} p(A) \quad (4)$$
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 $\mathcal{F} =$

[3] Like in this definition, throughout the paper capital-letter P is used for interval-valued assignments, while small letters $(p, q, \ldots)$ stand for classical probability.

[4] The difference between countable and finite additivity should be neglected for a moment.

$(\Omega, \mathcal{A}, L(\cdot))$ with structure $\mathcal{M}$ also fulfilling $\mathcal{M} \supseteq \mathcal{V}$ is *weaker* than $\mathcal{F}_{\mathcal{V}} = (\Omega, \mathcal{A}, L_{\mathcal{V}}(\cdot))$, i.e.: $L(A) \leq L_{\mathcal{V}}(A)$, for all $A \in \mathcal{A}$.

Mainly two applications of this concept will be used in what follows:[5]

Definition 3 (Independent product of F-probability fields)[6] Let a finite number of F-probability fields $\mathcal{F}_l = (\Omega_l, \mathcal{A}_l, L_l(\cdot))$ with structures $\mathcal{M}_l$, $l = 1, \ldots, n$, be given. Then the F-probability field $\otimes_{l=1}^n \mathcal{F}_l := (\times_{l=1}^n \Omega_l, \otimes_{l=1}^n \mathcal{A}_l, L(\cdot))$, which has $\times_{l=1}^n \mathcal{M}_l$ as its prestructure, is called the *independent product of the F-probability fields* $\mathcal{F}_l, l = 1, \ldots n$.

Definition 4 (Parametricly constructed F-probability fields) Consider a (strictly) parametric set $\mathcal{Q} = \{p_\theta(\cdot) \mid \theta \in \Theta\}$, $\Theta \subseteq I\!R^m$, of K-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

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

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

2.3 C-probability

In this subsection a special case of F-probability is considered, called C-probability with [29, 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. [32]).

Definition 5 (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*[7], i.e., if

$$L(A \cup B) + L(A \cap B) \geq L(A) + L(B), \quad \forall A, B \in \mathcal{A}. \quad (5)$$

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

Two related classes of C-probabilities have been extensively studied in the literature.

Remark 2 (Typical examples of C-probability fields) Assume $(\Omega, \mathcal{A})$ to be a Polish measurable space.[8]

- *Pseudo-capacities:*[9] Let $f(\cdot) : [0,1] \to [0,1]$ be a convex function with $f(0) = 0$ and let $p(\cdot)$ be a K-probability on $(\Omega, \mathcal{A})$ (called *central distribution* in this context). Then

  $$P(\cdot) := \left[(f \otimes p)(\cdot), 1 - (f \otimes p)(\cdot^C)\right]$$

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

  $$(f \otimes p)(A) := f\left(\bar{p}(A)\right)$$

  , for all $A \in \mathcal{A} \setminus \{\Omega\}$, is C-probability. Its structure will be denoted by $\mathcal{M}(f \otimes p)$.

- *Two-monotone (Choquet)-capacities:* Every set-function $L(\cdot)$ on $\mathcal{A}$ with $L(\Omega) = 1$ and (5) additionally obeying the condition

  $$(A_n)_{n \in I\!N} \uparrow A, \ A_n \text{ open}, \ n \in I\!N \Longrightarrow \lim_{n \to \infty} L(A_n) = L(A)\,, \quad (6)$$

  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 conjugated upper limit $U(\cdot) \geq L(\cdot)$, to F-probability and therefore to C-probability. ([13, lemma 2.5,p. 254)])

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 $f(y) := \max\left((1 - \epsilon) \cdot y - \delta\,, 0\right)$.

[5] 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 all the mixtures of the distributions corresponding to a parameter inside $\boldsymbol{\theta}$ belong to the structure as well.

[6] Compare [25, p. 745], [29, Chapter 7] and the 'sensitivity analysis definition' in [24, Chapter 9.1]

[7] A lot of different names are common for this property, especially it is also called '(strong) superadditivity', 'supermodularity' or 'convexity'.

[8] Ω has to be a complete, separable and metrizable space and $\mathcal{A}$ the corresponding Borel-σ-algebra (e.g. $\Omega = I\!R^n, \mathcal{A} = \mathcal{B}$).

[9] The term 'pseudo-capacity' is due to [6]. Also the name 'special capacity' used by [18] and [3] is common.

C-probability is a distinguished special case of F-probability possessing some mathematical peculiarity and elegance, but it is not comprehensive enough to provide an exclusive, neat basis for a theory of interval-valued probability. In the meanwhile Walley's conclusion that there isn't "[...] any 'rationality' argument for 2-monotonicity, beyond its computational convenience" ([23, p. 51]) has experienced a lot of additional support.[10] 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.

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

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 two is the true one. But now the hypotheses may consist of F-probabilities instead of classical probabilities: 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[11] $(\Omega, \mathcal{A})$ with disjoint[12] structures $\mathcal{M}_0$ and $\mathcal{M}_1$. After observing a singleton $\{\omega\} =: E$,[13] 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 $\varphi^*(\cdot) \in \Phi$ has to be made between the two hypotheses H_i : "The 'true' probability field is $\mathcal{F}_i$", $i \in \{0, 1\}$.

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 – consequent to allow only for precise (i.e. not interval-valued) probabilities for rejecting H_0. So the set Φ of all tests still has to consist of all $\mathcal{A} - \mathcal{B}$-measurable functions $\varphi(\cdot) : \Omega \to [0, 1]$.

Huber and Strassen and also this paper exclusively consider 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. $\mathbb{E}_{\mathcal{M}_1}(1-\varphi)$) while controlling for the error of the first kind (i.e. $\mathbb{E}_{\mathcal{M}_0}\varphi$)' leads to a complex, non-parametric (not 'easily parametrizable') (level-α-)maximin-problem between the structures:

Definition 6 (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. $U\mathbb{E}\varphi^* \leq \alpha$, and $\varphi^*(\cdot)$ has maximal power among all tests under consideration, i.e.

$$\forall \psi \in \Phi \ \left[U\mathbb{E}_{\mathcal{M}_0}\psi \leq \alpha \Rightarrow L\mathbb{E}_{\mathcal{M}_1}\psi \leq L\mathbb{E}_{\mathcal{M}_1}\varphi^* \right]. \quad (7)$$

3.2 (Globally) least favorable pairs

Based on the idea 'If one succeeds in convincing the hardliner of two parties one has convinced all their members' 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. This "being least favorable" can be formalized as follows.

Definition 7 (Globally least favorable pairs) A pair $(q_0(\cdot), q_1(\cdot))$ of K-probabilities is called *a globally least favorable pair*[14] (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 is a version $\pi(\cdot)$ of the likelihood ratio of $q_0(\cdot)$ and $q_1(\cdot)$ with

$$\begin{aligned} &\forall t \geq 0, \forall p_0(\cdot) \in \mathcal{M}_0 : \\ &\quad p_0(\{\omega \,|\pi(\omega) > t\}) \leq q_0(\{\omega \,|\pi(\omega) > t\}) \end{aligned} \quad (8)$$

and

$$\begin{aligned} &\forall t \geq 0, \forall p_1(\cdot) \in \mathcal{M}_1 : \\ &\quad p_1(\{\omega \,|\pi(\omega) > t\}) \geq q_1(\{\omega \,|\pi(\omega) > t\}) . \end{aligned} \quad (9)$$

Proposition 1 (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 hypotheses $\overline{H_0} : \{q_0(\cdot)\}$ versus $\overline{H_1} : \{q_1(\cdot)\}$, which is a level-$\alpha$-maximin-test for $\mathcal{F}_0$ versus $\mathcal{F}_1$, too.

The following lemma shows that it is sufficient to check the conditions for globally least favorable pairs

[10] For a detailed argumentation see:[1, Chapter 1.2.3]). To mention just one argument, on which it will be recurred later: The generalization of the usual parametric families to interval probability (like the F-normal distribution) along the lines of definition 4 leads to F- but not C-probability.

[11] Throughout the paper it is assumed that the set $\{\omega\}$ is measurable for every $\omega \in \Omega$. (This condition is very mild. It is, in particular, fulfilled by all Polish spaces.)

[12] To have an alternative-problem in the narrow sense it is implicitly assumed that $\mathcal{M}_0$ and $\mathcal{M}_1$ have a positive distance with respect to an appropriate metric.

[13] For technical reasons the formulation uses sample-size 1. Situations with sample-size n are included by considering the independent products (see definition 3).

[14] In Huber-Strassen theory it is usual to call $(q_0(\cdot), q_1(\cdot))$ a "least favorable pair". The term "globally" is added here to make a distinction to "locally least favorable pairs", which will be introduced later.

on any arbitrary prestructure. This helps to systematize some well known results and plays an important role in the proofs of the extensions discussed in section 5.

Lemma 1 (Globally least favorable pairs and prestructures) A pair $(q_0(\cdot), q_1(\cdot))$ of K-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 (8) and (9) hold for $\mathcal{V}_i$ instead of $\mathcal{M}_i$, $i \in \{0,1\}$.

It should be noted that the property of being globally least favorable does not depend on the level of significance making that property also independent from the sample size:[15]

Proposition 2 (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 $\mathcal{F}_1^{(n)}$.

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

4.1 The main theorems

The fame of the work of Huber and Strassen ([13]) is mainly based on the fact that they succeeded in showing that a globally least favorable pair always exists for two-monotone capacities:

Proposition 3 (Huber-Strassen theorem) *([13, p. 257, theorem 4.1])* Let $\mathcal{F}_0$ and $\mathcal{F}_1$ be C-probability fields on a Polish space $(\Omega, \mathcal{A})$ fulfilling (6). 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, because the following result was understood to show the impossibility of a generalization:

Proposition 4 ("Necessity theorem")[16] 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 K-probability $p_1(\cdot)$ with $p_1(\cdot) \notin \mathcal{M}_0$ a K-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 $\mathcal{F}_1 := (\Omega, \mathcal{P}(\Omega), p_1(\cdot))$, 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 entitled his article ([15]), where he introduced his generalization of proposition 4, "The necessity of [...C-probability] for Neyman-Pearson minimax tests". 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 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 3 for totally monotone C-probability on finite sample-spaces ([21, p. 282, Satz 2.1)]), and recognized one year later([22, p. 431)]) that indeed two-monotonicity is sufficient. On the other side, Huber ([12]) managed to "guess" a globally least favorable pair for contamination neighborhood models.

The synthesis leading 1973 to proposition 3 induced a lot of work, which is mainly concentrated on two aspects. Since on non-compact Ω, e.g. $\Omega = I\!R$, the usual neighborhood models do not fulfill (6) (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 [18] and [3], while [6] 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 (6) for the usual topology must nevertheless obey this condition for some non-standard topology. Conditions, which are sufficient to extent proposition 3 to non-Polish spaces, are given in [14, see especially p. 30, Satz 6.4].

The other main topic is initiated by the fact that the Huber & Strassen theorem is a general existence result without providing a method for constructing least favorable pairs. Rieder ([18]) presented a solution for the (ϵ, δ)-contamination model. Bednarski ([3, p. 402f.]) derived sufficient conditions for pseudo-capacities under which the likelihood-ratio of the glob-

[15] In principle, this property was already mentioned – quite informally – by Huber & Strassen ([13, Corollary 4.2, p. 257]) without having a clear independence concept for interval probability. A proof of proposition 2 is given in [1, p. 223ff.], which mainly is based on results from [31, p. 237f., Satz 2.57] and lemma 1.

[16] After [13, p. 262, theorem 7.1] (for finite spaces). [15, p. 123, theorem 2.3] extended this proposition to arbitrary Polish spaces.

ally 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 [16] the leitmotif of their work is the idea to use model-specific characteristics of the generalized risk-function for constructing the likelihood-ratio of the globally least favorable pair.[17] For several neighborhood models they managed to find that transformation of the central distribution, which leads to the risk-function of the globally least favorable pair (e.g. for the Prohorov model see: [8]). 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 ([9],[11]). — All these methods elegantly use particular properties of the special models considered. Additionally, as a side-product of [1, Chapter 5], for models on finite sample 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 underlying (see [1, p. 196f]).

There are a lot of other problems, where the solutions essentially are based on the Huber-Strassen theorem. Two examples for this are the extension to dependent random variables ([5]) and the development of asymptotic (level-α-)maximin-tests under sequences of shrinking neighborhood models[18].

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 conditional 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 might be toned down! Its premise is rather artificial. If the existence of a globally least favorable pair for *all* possible alternative hypotheses should 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 parametric families with monotone likelihood-ratio ([1, proposition 6.4, p. 227f.])

Proposition 5 (Existence in the case of monotone likelihood-ratio) Consider a (strict) parametric family $\mathcal{Q} = \{p_\theta(\cdot) \mid \theta \in \Theta\}$, $\Theta \subseteq I\!R$, of (mutually absolutely continuous) K-probabilities on $(I\!R, \mathcal{B})$ 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*. Pseudo-capacities are based on the 'convex distortion' of a single classical probability used as central distribution (see remark 2). This can substantially be extended by considering interval-valued central distributions. In particular this leads to neighborhood models of F-probabilities and it also prepares the study of testing problems with composite interval-valued hypotheses.[19]

Proposition 6 (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*

$$(f \otimes P)(\cdot) := [(f \otimes L)(\cdot), (1 - f \otimes L)(\cdot^C)]$$

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

$$(f \otimes L)(A) := f\left[L(A)\right], A \in \mathcal{A} \setminus \{\Omega\},$$

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, $\cup_{p(\cdot)\in\mathcal{M}} \mathcal{M}(f \otimes p)$ is a prestructure of the corresponding F-probability field $(f \otimes \mathcal{F}) := (\Omega, \mathcal{A}, (f \otimes L)(\cdot))$.

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 so-to-say a pair of 'least favorable F-probabilites', which are more easy to handle but nevertheless represent the whole testing problem in

[17] Cf. [10] for summarizing some aspects.

[18] For instance: [19], [4], [20, Chapter 5.4 and the references therein]

[19] See: [2], where also proofs of proposition 6 and of theorem 1 are given.

the sense that a level-α-maximin-test between them is also a level-α-maximin-test for the complex problem.

In [1, proposition 6.6, p. 231f.] the case of *generalized* (ϵ, δ)*-contaminations* was considered. Under some mild conditions it was shown that if there exists a globally least favorable pair $(q_0(\cdot), q_1(\cdot))$ for testing between the F-probability fields forming the central distributions, then there exists a least favorable pair of F-probabilites, namely just the pseudo-capacities around $q_0(\cdot)$ and $q_1(\cdot)$ with the same 'distortion functions' $f_0(\cdot)$ and $f_1(\cdot)$.

To extend this result to other generalized pseudo-capacities a regularity condition already needed in [3, theorem 5.1, p. 402] has to be added. For this call a pair $(p_0(\cdot), p_1(\cdot))$ of classical probabilities *randomization circumventing*, if $p_0(\cdot)$ and $p_1(\cdot)$ are mutually absolute continuous and for every $\alpha \in (0, 1)$ there is a best level-α-test for $\{p_0(\cdot)\}$ versus $\{p_1(\cdot)\}$ which is non-randomized.

Theorem 1 (Least favorable pseudo-capacities) 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 $\mathcal{F}_1$, which is randomization circumventing, 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 $(\Omega, \mathcal{A}, (f \otimes q_0)(\cdot))$ versus $(\Omega, \mathcal{A}, (f \otimes q_1)(\cdot))$, then it is a globally least favorable pair for $(f_0 \otimes \mathcal{F}_0)$ versus $(f_1 \otimes \mathcal{F}_1)$, too.

Therefore, often the following procedure to construct globally least favorable pairs for testing between generalized pseudo-capacities can be applied:

- 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 the globally least favorable pair for testing between the (non-generalized) pseudo-capacities around $q_0(\cdot)$ and $q_1(\cdot)$. According to theorem 1 it is the globally least favorable pair for the complex problem.

Then the efficient construction methods for (non-generalized) pseudo-capacities cited in section 4.2 can also be used to construct the least favorable pairs in these complexer situations. An immediate application of theorem 1 is the study of neighborhood models of parametricly constructed F-probability fields with monotone likelihood ratio.

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 1 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 K-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 1 and 3 ([1, Chapter 3.3)]).

Definition 8 (Locally least favorable pairs) Let a level of significance $\alpha \in (0, 1)$ be given. A pair $(q_0(\cdot), q_1(\cdot))$ of K-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 $U\!E_{\mathcal{M}_0}\varphi^* \leq \alpha$ and $E_{q_1}\varphi^* = L\!E_{\mathcal{M}_1}\varphi^*$.

Proposition 7 (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 hypotheses $\overline{H_0} : \{q_0(\cdot)\}$ versus $\overline{H_1} : \{q_1(\cdot)\}$, which is a level-$\alpha$-maximin-test for $\mathcal{F}_0$ versus $\mathcal{F}_1$, too.

Also the existence can be guaranteed under general regularity conditions:

Theorem 2 (Existence of locally least favorable pairs) If $\mathcal{F}_0$ and $\mathcal{F}_1$ are fulfilling the condition[20]

$$(A_n)_{n \in I\!N} \uparrow A,\, n \in I\!N \Longrightarrow \lim_{n \to \infty} L_i(A_n) = L_i(A)\,, \quad i \in \{0, 1\}\,, \tag{10}$$

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

[20] Note that condition (10) is a bit stronger than its pendant (6), because it refers to arbitrary measurable sets, and not only to the open ones. — The topological properties of the structures, which are caused by these conditions, are compared in [1, Chapter 3.3.4]

This "local perspective", which concentrates on a fixed level of significance, allows to develop universally applicable algorithms for calculating level-α-maximin-tests and locally least favorable pairs on finite sample-spaces ([1, Chapter 4 and 5]).

7 Concluding remarks

In general, 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 2 has been proved. For its proof it is essential that the relations (8) and (9) hold for *every* t, i.e. 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.

The level-α-maximin-criterion considers exclusively the maximal error. Research on equivalents to least favorable pairs in situations, where other interval orderings are used to judge the probability of error is still in its infanty.

Another topic deserving detailed investigation should only be sketched informally here: As mentioned at the end of section 2.2., by considering interval-valued expectation as the basic concept one arrives at sets $\check{\mathcal{M}}_0$ and $\check{\mathcal{M}}_1$ of classical probabilities, which may be more general than structures. The definitions given in section 3 to generalize the Neyman-Pearson problem can be straightforwardly extended to the problem of testing $\check{\mathcal{M}}_0$ versus $\check{\mathcal{M}}_1$, by replacing $\mathcal{M}_i$ (and $\mathcal{F}_i$) with $\check{\mathcal{M}}_i$, $i \in \{0,1\}$.

In the case of a finite sample space results on testing $\check{\mathcal{M}}_0$ versus $\check{\mathcal{M}}_1$ can be obtained by noticing that the calculation of level-α-maximin-tests and least favorable pairs for testing F-probability fields on finite sample spaces according to [1, Chapter 4 and 5] is only based on the fact that then structures are convex polyhedrons in the space of all classical probabilities. Therefore, all the results gained there can be immediately generalized to all those sets $\check{\mathcal{M}}_0$ and $\check{\mathcal{M}}_1$, which arise from a *finite* number of restrictions on the expectations. The same is true for the extension of the Generalized Neyman-Pearson Lemma given there ([1, Satz 5.1, p. 169ff.]).

For infinite spaces, however, general conditions for the existence of (globally or locally) least favorable pairs for testing $\check{\mathcal{M}}_0$ versus $\check{\mathcal{M}}_1$ might be difficult to obtain. It may be more promising to use the considerations of section 3 to 6 to derive sufficient conditions. This means that one takes the indirect way via the structures $\mathcal{M}_i$, $i \in \{0,1\}$, of the F-probability fields $\mathcal{F}_i$ gained from using $\check{\mathcal{M}}_i$ as prestructures. If $(q_0(\cdot), q_1(\cdot))$ is (globally or locally) least favorable for $\hat{\mathcal{F}}_0$ versus $\hat{\mathcal{F}}_1$ and if additionally $(q_0(\cdot), q_1(\cdot))$ is in $\check{\mathcal{M}}_0 \times \check{\mathcal{M}}_1$, then $(q_0(\cdot), q_1(\cdot))$ is least favorable for $\check{\mathcal{M}}_0$ versus $\check{\mathcal{M}}_1$, too.

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 possibly 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 6 and theorem 1?

Acknowledgements

I am grateful to Kurt Weichselberger, Toni Wallner and Peter Walley for their valuable comments.

References

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

[2] T. Augustin. Generalized Neighborhood Models and Least Favorable Pseudo-capacities. In preparation, 1999.

[3] T. Bednarski. On the Solution of Minimax Problems for Special Capacities. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete,* 58:397-405, 1981.

[4] T. Bednarski. A Robust Asymptotic Testing Model for Special Capacities. *Mathematische Operationsforschung und Statistik – Series Statistics.* 4:507-519, 1985.

[5] T. Bednarski. On the Extension of the Neyman-Pearson Lemma for Capacities to Dependent Random Variables. *Mathematische Operationsforschung und Statistik – Series Statistics.* 5:549-556, 1986.

[6] A. Buja. On the Huber-Strassen Theorem. *Probability Theory and Related Fields.* 73:149-152, 1986.

[7] G. de Cooman and P. Walley. *The Imprecise Probability Project.* `http://ensmain.rug.ac.be/~ipp`, 1999.

[8] R. Hafner. Simple Construction of Least Favorable Pairs of Distributions and of Robust Tests for Prohorov-neighborhoods. *Mathematische Operationsforschung und Statistik, Series Statistics.* 13:33-46, 1982.

[9] R. Hafner. Construction of Minimax-tests for Bounded Families of Distribution-Functions. In: W. Sendler (ed.). *Contributions to Stochastics.* Physika. Heidelberg, 1987.

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

[11] R. Hafner. Construction of Minimax-tests for Bounded Families of Probabilities-densities. *Metrika.* 40:1-23, 1993.

[12] P. J. Huber. A Robust Version of the Probability Ratio Test. *Annals of Mathematical Statistics.* 36:1753-1758, 1965.

[13] P. J. Huber, and V. Strassen. Minimax Tests and the Neyman-Pearson Lemma for Capacities. *Annals of Statistics.* 1:251-263, Correction: 2:223-224, 1973.

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

[15] J. Lembcke. The Necessity of Strongly Subadditive Capacities for Neyman-Pearson Minimax Tests. *Monatshefte für Mathematik.* 105:113-126, 1988.

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

[17] A. Papamarcou and T.L. Fine. Unstable Collectives and Envelopes of Probability Measures. *The Annals of Probability.* 19:893-906, 1991.

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

[19] H. Rieder. A Robust Asymptotic Testing Model. *The Annals of Statistics.* 6:1080-1094, 1978.

[20] H. Rieder. *Robust Asymptotic Statistics.* Springer. New York. 1994.

[21] V. Strassen. Meßfehler und Information. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete.* 2:273-305, 1964.

[22] V. Strassen. The Existence of Probability Measures with Given Marginals. *Annals of Mathematical Statistics.* 36:423-439, 1965.

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

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

[25] P. Walley and T.L. Fine. Towards a Frequentist Theory of Lower and Upper Probability. *The Annals of Statistics.* 10:741-761.

[26] K. Weichselberger. Axiomatic Foundations of the Theory of Interval-probability. In: V. Mammitzsch and H. Schneeweiß(ed.). *Proceedings of the Second Gauß Symposion, Section B.* de Gruyter. Berlin: 47-64, 1995a.

[27] K. Weichselberger. Stichproben und Intervallwahrscheinlichkeit. *ifo Studien* 41, 653-676, 1995b.

[28] K. Weichselberger. Interval-probability on Finite Sample-spaces. In: H. Rieder (ed.). *Robust Statistics, Data Analysis and Computer Intensive Methods.* Springer. New York: 391-409 (Lecture Notes in Statistics, 109), 1996.

[29] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung. Two volumes.* In preparation, 1999.

[30] K. Weichselberger and S. Pöhlmann. *A Methodology for Uncertainty in Knowledge-Based Systems.* Springer. Berlin. (Lecture Notes in Artificial Intelligence, 419), 1990.

[31] H. Witting. *Mathematische Statistik I.* Teubner. Stuttgart, 1985.

[32] R. R. Yager and M. Fedrizzi and J. Kacprzyk (ed.). *Advances in the Dempster-Shafer Theory of Evidence.* Wiley. New York, 1994.