The Description of Least Favorable Pairs in Huber-Strassen Theory, Finite Case

Andrey G. Bronevich
JSC "Research, Development and Planning Institute for Railway Information Technology, Automation and Telecommunication"
brone@mail.ru

Abstract

In this paper we provide the algebraic description of the minmax problem solutions, which are considered in Huber-Strassen theory providing effective algorithms of searching least favorable pairs. This investigation gives also new insights to understanding well-known algorithms for maximizing Shannon entropy and other functionals.

Keywords. 2-monotone capacities, least favorable pairs, Huber-Strassen theory, Kullback–Leibler distance.

1 Introduction

In 1973 Huber and Strassen [13] have published their prodigious paper showing that the optimal test between composite hypotheses described by 2-alternative capacities can be reduced to testing two simple hypotheses described by usual probability measures called a least favorable pair. This result was derived for Polish spaces and supplied with other remarkable results. In particular, the case of 2-alternative capacities cannot be extended for a wider class of coherent upper probabilities, the likelihood ratio does not depend on the chosen least favorable pair, i.e. this likelihood ratio is unique; any pair of probability measures minimizing a functional of a certain type has to be least favorable. In addition, they have shown the way of constructing the optimal test for independent experiments with observations described by 2-alternative capacities. After this famous work, there were some works generalizing Huber-Strassen results by using other topological assumptions [7,17], for special neighborhood models [12,16], or even for more general theories of imprecise probabilities [2,3,11] (see the overview of the results in [3]). However, most of the results are not constructive: they are based on the theorems establishing existence of least favorable pairs without showing how to obtain them. However, for some special neighborhood models there are explicit solutions for finding least favorable pairs (see results obtained by Österreicher [15], Rieder [16], and Bednarski [4]). Augustin [2] proposed a method for finding least favorable pairs for models on finite spaces, based on linear programming.

On the other hand, we can observe that recently developed algorithms [1,14] for computing the maximum entropy functional for 2-monotone capacities are evidently based on recovering the likelihood ratio between 2-monotone capacity and equiprobable probability distribution. But this fact has not been recognized yet.

In the paper we try to get more explicit expressions of Huber-Strassen results for a finite case. This allows us to get the description of all possible least favorable pairs and to construct the algorithm for searching them. This algorithm generalizes the procedure for computing the maximal entropy functional considered in [1].

2 Technical preliminaries

Let X be a finite universal set and $\mathfrak{A} = 2^X$ is the algebra consisting of all subsets of X. A set function $\mu : \mathfrak{A} \to [0,1]$ is called a *monotone measure* [9] or *capacity* [8] if 1) $\mu(\varnothing) = 0$, $\mu(X) = 1$; and 2) $A, B \in \mathfrak{A}$, $A \subseteq B$ implies $\mu(A) \le \mu(B)$. We write $\mu_1 \le \mu_2$ for monotone measures μ_1, μ_2 on $\mathfrak{A}$ if $\mu_1(A) \le \mu_2(A)$ for all $A \in \mathfrak{A}$. In this paper we consider the following families of monotone measures:

1) M_{mon} is the set of all monotone measures on $\mathfrak{A}$;

2) M_{pr} is the set of all probability measures on $\mathfrak{A}$, i.e. $M_{pr} \subseteq M_{mon}$ and additionally $\mu(A \cup B) = \mu(A) + \mu(B)$ for disjoint sets $A, B \in \mathfrak{A}$;

3) M_{low} is the set of all *lower probabilities* [18] on $\mathfrak{A}$, i.e. $M_{low} \subseteq M_{mon}$ and for any $\mu \in M_{low}$ there exists $P \in M_{pr}$ such that $\mu \le P$;

4) M_{coh} is the set of all *coherent lower probabilities* [18] on $\mathfrak{A}$, i.e. for any $\mu \in M_{coh}$ and $B \in \mathfrak{A}$ there exists $P \in M_{pr}$ such that $\mu \le P$ and $\mu(B) = P(B)$;

5) M_{2-mon} is the set of all *2-monotone measures* [8] on $\mathfrak{A}$, i.e. $M_{2-mon} \subseteq M_{mon}$ and $\mu(A)+\mu(B) \le \mu(A \cup B)+\mu(A \cap B)$ for any $A, B \in \mathfrak{A}$.

For any $\mu \in M_{mon}$ we define the set $core(\mu) = \{P \in M_{pr} \mid P \ge \mu\}$, Clearly $core(\mu) \ne \varnothing$ if $\mu \in M_{low}$. We also remind that $M_{mon} \supset M_{low} \supset M_{coh} \supset M_{2-mon} \supset M_{pr}$. In the sequel we use also the upper probability measures that can be got from lower probability measures using dual relation. A dual monotone measure μ^d of μ is computed by $\mu^d = 1-\mu(A^c)$, where $A \in \mathfrak{A}$ and $A^c = X \setminus A$ is the complement of A. Let us remind also that measures, which are dual to coherent lower probabilities, are called coherent upper probabilities [18], and also if $\mu \in M_{2-mon}$ then μ is 2-alternative monotone measure [8], i.e. it characterizes by the following inequality: $\mu^d(A)+\mu^d(B) \ge \mu^d(A \cup B)+\mu^d(A \cap B)$.

3 Huber-Strassen theory, finite case

In this section we consider the Huber-Strassen theory for the finite case: we establish connections between Huber-Strassen theory and canonical sequences of monotone measures [5] and provide an effective algorithm for finding least favorable pairs.

Let us remind that the Huber-Strassen theory solves the problem of the Neymann-Pearson testing between two hypotheses H_0 and H_1 described by 2-monotone measures μ_0 and μ_1 on an algebra $\mathfrak{A}$. We assume here that $\mathfrak{A}$ is the powerset of some nonempty set X. According to this theory the testing problem can be reduced to the classical case, i.e. there exist probability measures (a *least favorable pair*) $P_0 \in core(\mu_0)$ and $P_1 \in core(\mu_1)$ such that the optimal test for any level of significance can be obtained by using P_0 and P_1. The searching of P_0 and P_1 is closely connected to the following optimization problem:

$$q_{\mu_0^d,\mu_1^d}(t) = \min_{A \in 2^X} \left\{(1-t)\mu_0^d(A) + t\mu_1^d(A^c)\right\},$$

where $t \in [0,1]$. Obviously, the value $q_{\mu_0^d,\mu_1^d}(t)$ gives us the exact upper probability of error if we use the Bayesian classifier and the prior probability of H_0 is $(1-t)$ and the prior probability of H_1 is t. Hence, the expression for $q_{\mu_0^d,\mu_1^d}$ can be rewritten as

$$q_{\mu_0^d,\mu_1^d}(t) = \min_{A \in 2^X} \max_{\substack{P_0 \in core(\mu_0),\\ P_1 \in core(\mu_1)}} (1-t)P_0(A) + tP_1(A^c).$$

This optimization problem can be considered also for coherent lower probabilities, but for this case it is impossible to choose $P_0 \in core(\mu_0), P_1 \in core(\mu_1)$ for any $\mu_0, \mu_1 \in M_{coh}$, such that

$$q_{\mu_0^d,\mu_1^d}(t) = \min_{A \in 2^X} \left\{(1-t)\mu_0^d(A) + t\mu_1^d(A^c)\right\} = \min_{A \in 2^X} \left\{(1-t)P_0(A) + tP_1(A^c)\right\},$$

in general. Let us denote

$$\mathcal{L}_{\mu_0,\mu_1}(t) = \left\{A \in 2^X \,\middle|\, (1-t)\mu_0^d(A) + t\mu_1^d(A^c) = q(t)\right\}$$

for $t \in [0,1]$ and $\mathcal{L}_{\mu_0,\mu_1} = \bigcup_{t \in (0,1)} \mathcal{L}_{\mu_0,\mu_1}(t)$.

We analyze first the properties of $\mathcal{L}_{\mu_0,\mu_1}$. We will show that $\mathcal{L}_{\mu_0,\mu_1}$ is a lattice and measures μ_0^d and μ_1 are additive on it.

Proposition 1. *Let* $\mu_0, \mu_1 \in M_{2-mon}$ *and assume that* $A \in \mathcal{L}_{\mu_0,\mu_1}(t)$, $B \in \mathcal{L}_{\mu_0,\mu_1}(s)$, *where* $t \le s$. *Then* $A \cap B \in \mathcal{L}_{\mu_0,\mu_1}(t)$ *and* $A \cup B \in \mathcal{L}_{\mu_0,\mu_1}(s)$. *In addition,* $\mu_0^d(A) = \mu_0^d(A \cap B)$ *and* $\mu_1(B) = \mu_1(A \cup B)$ *if* $t < s$.

Proof. Let $A \in \mathcal{L}_{\mu_0,\mu_1}(t)$, $B \in \mathcal{L}_{\mu_0,\mu_1}(s)$ and $t \le s$. Then

$$\begin{gathered} q_{\mu_0^d,\mu_1^d}(t) + q_{\mu_0^d,\mu_1^d}(s) = \\ (1-t)\mu_0^d(A) + t\mu_1^d(A^c) + (1-s)\mu_0^d(B) + s\mu_1^d(B^c) = \\ (1-s)\left(\mu_0^d(A) + \mu_0^d(B)\right) + (s-t)\mu_0^d(A) + \\ t\left(\mu_1^d(A^c) + \mu_1^d(B^c)\right) + (s-t)\mu_1^d(B^c). \end{gathered}$$

Because μ_0^d, μ_1^d are 2-alternative, we get the following inequality:

$$\begin{gathered} q_{\mu_0^d,\mu_1^d}(t) + q_{\mu_0^d,\mu_1^d}(s) \ge \\ (1-s)\left(\mu_0^d(A \cap B) + \mu_0^d(A \cup B)\right) + (s-t)\mu_0^d(A) + \\ t\left(\mu_1^d\left((A \cap B)^c\right) + \mu_1^d\left((A \cup B)^c\right)\right) + (s-t)\mu_1^d(B^c) = \\ \left[(1-s)\mu_0^d(A \cup B) + s\mu_1^d\left((A \cup B)^c\right)\right] + \\ \left[(1-t)\mu_0^d(A \cap B) + t\mu_1^d\left((A \cap B)^c\right)\right] + \\ (s-t)\left[\mu_1^d(B^c) - \mu_1^d\left((A \cup B)^c\right)\right] + \\ (s-t)\left[\mu_0^d(A) - \mu_0^d(A \cap B)\right]. \end{gathered}$$

Since $\mu_1^d(B^c) - \mu_1^d\left((A \cup B)^c\right) \ge 0$, $\mu_0^d(A) - \mu_0^d(A \cap B) \ge 0$, and, by our assumption,

$$(1-s)\mu_0^d(A\cup B)+ s\mu_1^d\left((A\cup B)^c\right)\geq q_{\mu_0^d,\mu_1^d}(s),$$

$$(1-t)\mu_0^d(A\cap B)+ t\mu_1^d\left((A\cap B)^c\right)\geq q_{\mu_0^d,\mu_1^d}(t),$$

we get that there is the only possibility that

$$(1-s)\mu_0^d(A\cup B)+s\mu_1^d\left((A\cup B)^c\right)= q_{\mu_0^d,\mu_1^d}(s),$$

$$(1-t)\mu_0^d(A\cap B)+t\mu_1^d\left((A\cap B)^c\right)= q_{\mu_0^d,\mu_1^d}(t),$$

and if $s>t$, then

$$\mu_1^d(B^c)-\mu_1^d\left((A\cup B)^c\right)=0,$$

$$\mu_0^d(A)-\mu_0^d(A\cap B)=0,$$

i.e. $A\cap B\in\mathcal{L}_{\mu_0,\mu_1}(t)$ and $A\cup B\in\mathcal{L}_{\mu_0,\mu_1}(s)$.

Corollary 1. *$\mathcal{L}_{\mu_0,\mu_1}$ is a lattice, and monotone measures μ_0^d and μ_1 are additive on $\mathcal{L}_{\mu_0,\mu_1}$.*

Proof. It is clear that $\mathcal{L}_{\mu_0,\mu_1}$ is a lattice. It follows from Proposition 1. Let $A\in\mathcal{L}_{\mu_0,\mu_1}(t)$, $B\in\mathcal{L}_{\mu_0,\mu_1}(s)$ and $t<s$. Then, by Proposition 1, $\mu_0^d(A)=\mu_0^d(A\cap B)$ and $\mu_1(B)=\mu_1(A\cup B)$. Since μ_0^d is 2-alternative,

$$\mu_0^d(A)+\mu_0^d(B)\geq\mu_0^d(A\cap B)+\mu_0^d(A\cup B),$$

i.e. $\mu_0^d(B)\geq\mu_0^d(A\cup B)$, and this is possible if $\mu_0^d(B)=\mu_0^d(A\cup B)$, i.e.

$$\mu_0^d(A)+\mu_0^d(B)=\mu_0^d(A\cap B)+\mu_0^d(A\cup B).$$

Analogously, since μ_1 is 2-monotone,

$$\mu_1(A)+\mu_1(B)\leq\mu_1(A\cap B)+\mu_1(A\cup B),$$

i.e. $\mu_1(A)\leq\mu_1(A\cap B)$, and this is possible if $\mu_1(A)=\mu_1(A\cap B)$, i.e.

$$\mu_1(A)+\mu_1(B)=\mu_1(A\cap B)+\mu_1(A\cup B).$$

Consider the case, when $s=t\in(0,1)$. Then

$$(1-t)\left(\mu_0^d(A)+\mu_0^d(B)\right)+t\left(\mu_1^d(A^c)+\mu_1^d(B^c)\right)=$$
$$(1-t)\left(\mu_0^d(A\cap B)+\mu_0^d(A\cup B)\right)+$$
$$t\left(\mu_1^d\left((A\cap B)^c\right)+\mu_1^d\left((A\cup B)^c\right)\right).$$

Because $\mu_0^d(A)+\mu_0^d(B)\geq\mu_0^d(A\cap B)+\mu_0^d(A\cup B)$ and $\mu_1^d(A^c)+\mu_1^d(B^c)\geq\mu_1^d((A\cap B)^c)+\mu_1^d((A\cup B)^c)$, this is possible if $\mu_0^d(A)+\mu_0^d(B)=\mu_0^d(A\cap B)+\mu_0^d(A\cup B)$ and $\mu_1^d(A^c)+\mu_1^d(B^c)=\mu_1^d((A\cap B)^c)+\mu_1^d((A\cup B)^c)$, i.e. monotone measures μ_0^d and μ_1 are additive on $\mathcal{L}_{\mu_0,\mu_1}$.

Further we will consider sublattices $\mathcal{L}$ of $\mathcal{L}_{\mu_0,\mu_1}$ such that $\mathcal{L}\cap\mathcal{L}_{\mu_0,\mu_1}(t)\neq\varnothing$ for any $t\in(0,1)$. The notable examples of these lattices are $\{\underline{A}_t\}_{t\in(0,1)}$ and $\{\overline{A}_t\}_{t\in(0,1)}$, where $\underline{A}_t=\bigcap_{A\in\mathcal{L}_{\mu_0,\mu_1}(t)}A$, $\overline{A}_t=\bigcup_{A\in\mathcal{L}_{\mu_0,\mu_1}(t)}A$. Notice that $\underline{A}_t,\overline{A}_t\in\mathcal{L}_{\mu_0,\mu_1}(t)$ by Proposition 1.

Consider now the case, when $\mu_0,\mu_1\in M_{pr}$.

Proposition 2. *Let $P_0,P_1\in M_{pr}$. Then for any $t\in[0,1]$*

$$\mathcal{L}_{P_0,P_1}(t)=\left\{A\in 2^X \mid \underline{A}_t\subseteq A\subseteq\overline{A}_t\right\},$$

where $\underline{A}_t=\left\{x\in X \mid (1-t)P_0(\{x\})<tP_1(\{x\})\right\}$ *and* $\overline{A}_t=\left\{x\in X \mid (1-t)P_0(\{x\})\leq tP_1(\{x\})\right\}$.

Proof. According to the definition $A\in\mathcal{L}_{P_0,P_1}(t)$ iff it minimizes the value

$$(1-t)P_0(A)+t\left(1-P_1(A)\right)=t+\sum_{x\in A}\left[(1-t)P_0(\{x\})-tP_1(\{x\})\right],$$

and we get the minimum if $(1-t)P_0(\{x\})-tP_1(\{x\})\leq 0$ for all $x\in A$ and $(1-t)P_0(\{x\})-tP_1(\{x\})\geq 0$ for all $x\notin A$, and the above condition implies the statement of the proposition.

Remark 1. We can express the statement of Proposition 2 using the likelihood ratio of probability measures P_0 and P_1. For this reason, let us introduce two functions $\underline{\pi}:X\to[0,+\infty]$ and $\overline{\pi}:X\to[0,+\infty]$ by

1) $\underline{\pi}(x)=\overline{\pi}(x)=P_0(\{x\})/P_1(\{x\})$ if at least one of the values $P_0(\{x\})$ and $P_1(\{x\})$ is greater than zero (we define $\underline{\pi}(x)=\overline{\pi}(x)=+\infty$ if $P_0(\{x\})>0$ and $P_1(\{x\})=0$);

2) $\underline{\pi}(x)=0$ and $\overline{\pi}(x)=+\infty$ if $P_0(\{x\})=0$ and $P_1(\{x\})=0$.

Then

$$\underline{A}_t=\left\{x\in X \mid \overline{\pi}(x)<t/(1-t)\right\}$$

and

$$\overline{A}_t=\left\{x\in X \mid \underline{\pi}(x)\leq t/(1-t)\right\}.$$

Our next problem is to find sufficient and necessary conditions, under which $q_{\mu_0^d,\mu_1^d}(t)=q_{P_0,P_1}(t)$, $t\in[0,1]$, for probability measures $P_0\in core(\mu_0)$ and $P_1\in core(\mu_1)$. The solution of this problem is presented below.

Lemma 1. *Let $\mu_0,\mu_1\in M_{2-mon}$. Assume also that $P_0\in core(\mu_0)$ and $P_1\in core(\mu_1)$ such that*

$q_{\mu_0^d,\mu_1^d}(t)=q_{P_0,P_1}(t)$ *for all* $t\in[0,1]$. *Then* $\mathcal{L}_{\mu_0,\mu_1}(t)\subseteq\mathcal{L}_{P_0,P_1}(t)$ *for all* $t\in[0,1]$.

Proof. Assume that the conditions of the lemma are fulfilled, and $B\in\mathcal{L}_{\mu_0,\mu_1}(t)$. Then

$$q_{\mu_0^d,\mu_1^d}(t)=(1-t)\mu_0^d(B)+t\mu_1^d(B^c)\geq(1-t)P_0(B)+tP_1(B^c).$$

Because $q_{\mu_0^d,\mu_1^d}(t)=q_{P_0,P_1}(t)$ by our assumption, we have

$$q_{P_0,P_1}(t)=(1-t)P_0(B)+tP_1(B^c),$$

i.e. $B\in\mathcal{L}_{P_0,P_1}(t)$. The lemma is proved.

Obviously, there are some cases, when $\mathcal{L}_{\mu_0,\mu_1}(t)=\mathcal{L}_{P_0,P_1}(t)$ for probability measures from Lemma 1. In these cases it is possible to recover the likelihood ratio of (P_0,P_1) as shown in the following lemma.

Lemma 2. *Let the conditions of Lemma 1 be fulfilled and* $\mathcal{L}_{P_0,P_1}(t)=\mathcal{L}_{\mu_0,\mu_1}(t)$ *for all* $t\in[0,1]$. *Then the likelihood ratio of* (P_0,P_1) *is uniquely defined on* X *by*

1) $\underline{\pi}(x)=0$ *and* $\overline{\pi}(x)=+\infty$ *if* $x\in\overline{A}_0\setminus\underline{A}_1$;

2) $\underline{\pi}(x)=\overline{\pi}(x)=\sup\{t/(1-t)\mid x\in\overline{A}_t,t\in[0,1]\}$ *if* $x\in\underline{A}_1$;

3) $\underline{\pi}(x)=\overline{\pi}(x)=+\infty$ *if* $x\in X\setminus(\overline{A}_0\cup\underline{A}_1)$

Proof. Let us show that the formulas for $\underline{\pi}$ is valid. We see that $\overline{A}_0=\{x\in X\mid P_0(\{x\})=0\}$ and $\underline{A}_1=\{x\in X\mid P_1(\{x\})>0\}$. Therefore, formulas 1) and 3) are valid. Let $t\in(0,1)$, then

$$\{x\in X\mid\underline{\pi}(x)=t/(1-t)\}=\bigcap_{t>s}(\overline{A}_t\setminus\overline{A}_s),$$

i.e. the formula 2) is also valid. The lemma is proved.

Remark 2. Let us notice that functions $\underline{\pi}(x)$ and $\overline{\pi}(x)$, considered in Lemma 2 can be computed for every pair of 2-monotone measures $\mu_0,\mu_1\in M_{2-mon}$. We call these functions as in Huber-Strassen theory, a likelihood ratio of 2-monotone measures μ_0,μ_1.

Lemma 3. *Let* $\mu_0,\mu_1\in M_{2-mon}$ *and let* $P_0\in core(\mu_0)$ *and* $P_1\in core(\mu_1)$ *be such that* $q_{\mu_0^d,\mu_1^d}(t)=q_{P_0,P_1}(t)$ *for all* $t\in[0,1]$. *Then the likelihood ratio of* (P_0,P_1) *is equal to the likelihood ratio of* (μ_0,μ_1) *in all points, where at least* $P_0(\{x\})>0$ *or* $P_1(\{x\})>0$.

Proof. For these points the likelihood ratio of (P_0,P_1) can be computed by

$$\underline{\pi}(x)=\overline{\pi}(x)=\sup\{t/(1-t)\mid x\in\overline{B}_t,t\in[0,1]\},$$

where $\overline{B}_t$ is the maximal element of $\mathcal{L}_{P_0,P_1}(t)$. Consider also maximal elements $\overline{A}_t$ of lattices $\mathcal{L}_{\mu_0,\mu_1}(t)$, $t\in[0,1]$. Because $P_0(\overline{B}_t)=P_0(\overline{A}_t)$, $P_0(\overline{B}_t)=P_0(\overline{A}_t)$, and $\overline{A}_t\subseteq\overline{B}_t$ by Lemma 1 and Proposition 2, we find that $P_0(\{x\})=P_1(\{x\})=0$ for all $x\in\overline{B}_t\setminus\overline{A}_t$. Therefore,

$$\underline{\pi}(x)=\overline{\pi}(x)=\sup\{t/(1-t)\mid x\in\overline{A}_t,t\in[0,1]\}.$$

The proposition is proved.

Proposition 3. *Let* $\mu_0,\mu_1\in M_{2-mon}$, *and let* $\underline{A}_t$ *be the minimal elements of* $\mathcal{L}_{\mu_0,\mu_1}(t)$, $t\in[0,1]$. *Assume also that* $P_0\in core(\mu_0)$ *and* $P_1\in core(\mu_1)$. *Then* $q_{\mu_0^d,\mu_1^d}(t)=q_{P_0,P_1}(t)$ *for all* $t\in[0,1]$ *iff*

1) $P_0(\underline{A}_t)=\mu_0^d(\underline{A}_t)$, $P_1(\underline{A}_t)=\mu_1(\underline{A}_t)$ *for all* $t\in[0,1]$;

2) *the likelihood ratio of* (P_0,P_1) *is equal to the likelihood ratio of* (μ_0,μ_1) *in all points, where at least* $P_0(\{x\})>0$ *or* $P_1(\{x\})>0$.

Proof. The necessary statement of the proposition follows from Lemma 1 and Lemma 3. Let us show sufficiency. Let $\underline{\pi}(x)$ and $\overline{\pi}(x)$ define the likelihood ratio of (P_0,P_1), then the maximal and minimal elements of $\mathcal{L}_{P_0,P_1}(t)$ are defined by

$$\underline{B}_t=\{x\in X\mid\overline{\pi}(x)<t/(1-t)\},$$

$$\overline{B}_t=\{x\in X\mid\underline{\pi}(x)\leq t/(1-t)\}.$$

Obviously, 2) implies that $\underline{B}_t\subseteq\underline{A}_t\subseteq\overline{B}_t$, i.e. $\underline{A}_t\in\mathcal{L}_{P_0,P_1}(t)$ by Proposition 2, and

$$q_{P_0,P_1}(t)=(1-t)P_0(\underline{A}_t)+t(1-P_1(\underline{A}_t)=$$

$$(1-t)\mu_0^d(\underline{A}_t)+t(1-\mu_1(\underline{A}_t))=q_{\mu_0^d,\mu_1^d}(t).$$

The proposition is proved.

The existence of probability measures, considered in Proposition 3, is shown in the next proposition.

Proposition 4. *Let* $\mu_0,\mu_1\in M_{2-mon}$, $\overline{A}_t$ *and* $\underline{A}_t$ *be maximal and minimal elements of* $\mathcal{L}_{\mu_0,\mu_1}(t)$, *respectively. Then there are* $P_0\in core(\mu_0)$ *and* $P_1\in core(\mu_1)$ *such that*

1) $P_0(\underline{A}_t)=\mu_0^d(\underline{A}_t)$, $P_1(\underline{A}_t)=\mu_1(\underline{A}_t)$ *for all* $t\in[0,1]$;

2) $q_{\mu_0^d,\mu_1^d}(t)=q_{P_0,P_1}(t)$ *for all* $t\in[0,1]$.

Proof. Let functions $\underline{\pi}(x)$ and $\overline{\pi}(x)$ define the likelihood ratio of (μ_0,μ_1). Because we consider the finite case, there is an increasing sequence $\{t_1,t_2,...,t_{m-1}\}$, such that $0<t_1<t_2<...<t_{m-1}=1$ and $\underline{\pi}(x)=t_k/(1-t_k)$ if $x\in \underline{A}_{t_{k+1}}\setminus \underline{A}_{t_k}$, $k=1,...,m-2$. Let us denote $B_k=\underline{A}_{t_k}$, $k=1,...,m-1$, $B_m=\left(X\setminus \overline{A}_0\right)\cup \underline{A}_1$. Observe that $\underline{\pi}(x)=\overline{\pi}(x)=t_k/(1-t_k)$ if $x\in B_{k+1}\setminus B_k$, $k=1,...,m-1$, and also $\underline{\pi}(x)=\overline{\pi}(x)=0$ if $x\in B_1$, and $\underline{\pi}(x)=0$ and $\overline{\pi}(x)=+\infty$ if $x\in X\setminus B_m$. According to the condition 1), we should find a pair of probability measures (P_0,P_1) satisfying the conditions:

$$P_0\in\left\{P\in M_{pr}\mid P\le\mu_0^d, P(B_k)=\mu_0^d(B_k), k=1,...,m\right\}, \quad (1)$$

$$P_1\in\left\{P\in M_{pr}\mid P\ge\mu_1, P(B_k)=\mu_1(B_k), k=1,...,m\right\}. \quad (2)$$

These families of probability measures can be obviously described by canonical sequences of monotone measures[1]. For this purpose, let us consider a sequence of sets $\Gamma=\{B_k\}_{k=1}^m$ and limit measures $\left(\mu_0^d\right)_\Gamma$ and $(\mu_1)_\Gamma$ generated by canonical sequences of monotone measures by the sequence Γ. Then obviously, the conditions (1) and (2) are equivalent to $P_0\in\left\{P\in M_{pr}\mid P\le\left(\mu_0^d\right)_\Gamma\right\}$ and $P_1\in\left\{P\in M_{pr}\mid P\ge(\mu_1)_\Gamma\right\}$.

Because $\Gamma=\{B_k\}_{k=1}^m$ is an increasing sequence, we can use the explicit expressions for limit measures $\left(\mu_0^d\right)_\Gamma$ and $(\mu_1)_\Gamma$ as follows:

$$\left(\mu_0^d\right)_\Gamma(A)=\sum_{k=1}^{m+1}\left[\mu_0^d((A\cap B_k)\cup B_{k-1})-\mu_0^d(B_{k-1})\right],$$

$$(\mu_1)_\Gamma(A)=\sum_{k=1}^{m+1}\left[\mu_1((A\cap B_k)\cup B_{k-1})-\mu_1(B_{k-1})\right],$$

where $B_0=\varnothing$ and $B_{m+1}=X$. Let us analyze what kind of additional conditions for a pair of probability measures P_0 and P_1 should be fulfilled. According to condition 2) of Proposition 3 a pair (P_0,P_1) should have the same likelihood ratio as (μ_0,μ_1) in all points, where at least $P_0(\{x\})>0$ or $P_1(\{x\})>0$. In other words, P_0 and P_1 have the constant positive likelihood ratio on sets $B_k\setminus B_{k-1}$, $k=2,...,m-1$, excluding points, where $P_0(\{x\})=0$ or $P_1(\{x\})=0$. This means that conditional probability measures $(P_0)_{B_k\setminus B_{k-1}}$ and $(P_1)_{B_k\setminus B_{k-1}}$ defined by

$$(P_0)_{B_k\setminus B_{k-1}}(A)=\frac{P_0((A\cap B_k)\cup B_{k-1})-P_0(B_{k-1})}{P_0(B_k)-P_0(B_{k-1})},$$

$$(P_1)_{B_k\setminus B_{k-1}}(A)=\frac{P_1((A\cap B_k)\cup B_{k-1})-P_1(B_{k-1})}{P_1(B_k)-P_1(B_{k-1})}$$

are the same. Introduce also into consideration monotone measures

$$\left(\mu_0^d\right)_{B_k\setminus B_{k-1}}(A)=\frac{\mu_0^d((A\cap B_k)\cup B_{k-1})-\mu_0^d(B_{k-1})}{\mu_0^d(B_k)-\mu_0^d(B_{k-1})},$$

$$(\mu_1)_{B_k\setminus B_{k-1}}(A)=\frac{\mu_1((A\cap B_k)\cup B_{k-1})-\mu_1(B_{k-1})}{\mu_1(B_k)-\mu_1(B_{k-1})}.$$

Assume that measures $\left(\mu_0^d\right)_{B_k\setminus B_{k-1}}$, $(\mu_1)_{B_k\setminus B_{k-1}}$ are defined for $k=1,...,m$ if the corresponding divisor $\mu_0^d(B_k)-\mu_0^d(B_{k-1})$ or $\mu_1(B_k)-\mu_1(B_{k-1})$ is not equal to zero. Then using expressions for $\left(\mu_0^d\right)_\Gamma$ and $(\mu_1)_\Gamma$, we have $(P_0)_{B_k\setminus B_{k-1}}\le\left(\mu_0^d\right)_{B_k\setminus B_{k-1}}$ and $(P_1)_{B_k\setminus B_{k-1}}\ge(\mu_1)_{B_k\setminus B_{k-1}}$, or combining with $(P_0)_{B_k\setminus B_{k-1}}$ and $(P_1)_{B_k\setminus B_{k-1}}$, we get

$$(\mu_1)_{B_k\setminus B_{k-1}}\le(P_1)_{B_k\setminus B_{k-1}}=(P_0)_{B_k\setminus B_{k-1}}\le\left(\mu_0^d\right)_{B_k\setminus B_{k-1}},$$

where $k=2,...,m-1$. To prove that a probability measure $P\in M_{pr}$ with $(\mu_1)_{B_k\setminus B_{k-1}}\le P\le\left(\mu_0^d\right)_{B_k\setminus B_{k-1}}$ exists, let us show first that the inequality $(\mu_1)_{B_k\setminus B_{k-1}}\le\left(\mu_0^d\right)_{B_k\setminus B_{k-1}}$ holds, i.e.

$$\frac{\mu_1(A\cup B_{k-1})-\mu_1(B_{k-1})}{\mu_1(B_k)-\mu_1(B_{k-1})}\le\frac{\mu_0^d(A\cup B_{k-1})-\mu_0^d(B_{k-1})}{\mu_0^d(B_k)-\mu_0^d(B_{k-1})} \quad (3)$$

for any $A\subseteq B_k\setminus B_{k-1}$. Let us notice that the choice of sets B_k implies that

$$(1-t_{k-1})\mu_0^d(B_{k-1})+t_{k-1}(1-\mu_1(B_{k-1}))=$$
$$(1-t_{k-1})\mu_0^d(B_k)+t_{k-1}(1-\mu_1(B_k)).$$

Or equivalently,

$$\frac{\mu_0^d(B_k)-\mu_0^d(B_{k-1})}{\mu_1(B_k)-\mu_1(B_{k-1})}=\frac{t_{k-1}}{1-t_{k-1}}.$$

Therefore, if the inequality (3) is not fulfilled for some $A\subseteq B_k\setminus B_{k-1}$, then

$$\frac{\mu_0^d(A\cup B_{k-1})-\mu_0^d(B_{k-1})}{\mu_1(A\cup B_{k-1})-\mu_1(B_{k-1})}<\frac{t_{k-1}}{1-t_{k-1}},$$

but this contradicts to the choice of B_{k-1}, because in this case

$$(1-t_{k-1})\mu_0^d(B_{k-1})-\mu_1(B_{k-1})>$$
$$(1-t_{k-1})\mu_0^d(A\cup B_{k-1})-\mu_1(A\cup B_{k-1}).$$

[1] A reader can find the main results on canonical sequences of monotone measures in [5] and a brief description of them in [6].

Because $(\mu_1)_{B_k \setminus B_{k-1}}$ is 2-monotone and $(\mu_0^d)_{B_k \setminus B_{k-1}}$ is 2-alternative, the existence of P follows from [10, Lemma 4.3].

It remains to show how to define values of probability measures $P_0(\{x\})$ and $P_1(\{x\})$ if $x \in B_1$, $x \in B_m \setminus B_{m-1}$, or $x \in X \setminus B_m$. (It is worth to keep in mind that it is often $B_1 = \varnothing$ or $B_m \setminus B_{m-1} = \varnothing$.)

If $x \in B_1$, then $P_0(\{x\}) = 0$ and values $P_1(\{x\})$ should be chosen such that $(\mu_1)_{B_1} \le (P_1)_{B_1}$.

If $x \in B_m \setminus B_{m-1}$, then $P_1(\{x\}) = 0$ and values $P_0(\{x\})$ should be chosen such that $(P_0)_{B_m \setminus B_{m-1}} \le (\mu_0^d)_{B_m \setminus B_{m-1}}$.

If $x \in X \setminus B_m$, then $P_0(\{x\}) = 0$ and $P_1(\{x\}) = 0$.

Let us notice that the constructed pair of probability measures (P_0, P_1) has the same likelihood ratio as (μ_0, μ_1) in all points, where at least $P_0(\{x\}) > 0$ or $P_1(\{x\}) > 0$. It means that $q_{\mu_0^d, \mu_1^d}(t) = q_{P_0, P_1}(t)$ for all $t \in [0,1]$ by Proposition 3. The proposition is proved.

Remark 3. Let us notice that Proposition 4 establishes the existence of a least favorable pair, because the optimal test for any level of significance can be obtained by probability measures P_0 and P_1, considered in Proposition 4. In some sense, a least favorable pair (P_0, P_1) gives an approximation of (μ_0, μ_1) and its exactness depends on the chosen (P_0, P_1). The exact approximation should give the best approximation of sets $\mathcal{L}_{\mu_0,\mu_1}(t)$, $t \in [0,1]$ in a sense that the cardinality of $\mathcal{L}_{P_0,P_1}(t) \setminus \mathcal{L}_{\mu_0,\mu_1}(t)$ should be minimal. It happens if $\mathcal{L}_{P_0,P_1}(t) = \{A \mid \underline{A}_t \subseteq A \subseteq \overline{A}_t\}$ for all $t \in [0,1]$, where $\overline{A}_t$ and $\underline{A}_t$ are maximal and minimal elements of $\mathcal{L}_{\mu_0,\mu_1}(t)$, respectively. In this case the likelihood ratios of (P_0, P_1) and (μ_0, μ_1) coincide.

Corollary 2. *Let $\mu_0, \mu_1 \in M_{2-mon}$ and we use the notation from Proposition 4 and its proof. Then every least favorable pair of probability measures P_0 and P_1 can be represented as*

$$P_0 = \sum_{k=2}^{m+1} \left(\mu_0^d(B_k) - \mu_0^d(B_{k-1})\right)(P_0)_{B_k \setminus B_{k-1}},$$

$$P_1 = \sum_{k=1}^{m} \left(\mu_1(B_k) - \mu_1(B_{k-1})\right)(P_1)_{B_k \setminus B_{k-1}},$$

where conditional probability measures satisfy the following inequalities:

$$(\mu_1)_{B_1} \le (P_1)_{B_1};$$

$$(\mu_1)_{B_k \setminus B_{k-1}} \le (P_1)_{B_k \setminus B_{k-1}} = (P_0)_{B_k \setminus B_{k-1}} \le (\mu_0^d)_{B_k \setminus B_{k-1}},$$
$$k = 2, \ldots, m-1;$$

$$(P_0)_{B_m \setminus B_{m-1}} \le (\mu_0^d)_{B_m \setminus B_{m-1}}.$$

The algorithm for searching sets B_k, $k = 2, \ldots, m-1$, is based on the following lemma.

Lemma 4. *The construction of sets B_k, $k = 1, 2, \ldots, m$, can be based on the following:*

a) B_1 is the set with the smallest cardinality such that

$$\mu_1(B_1) = \max\{\mu_1(B) \mid \mu_0^d(B) = 0\};$$

b) Let us assume that sets $B_0 = \varnothing$, B_1, ..., B_{k-1}, $k \ge 2$, have been constructed. Then if $\mu_1(B_{k-1}) < 1$ the next B_k should be chosen from the set Ω of possible solutions of the following optimization problem

$$\min_{B \left| \substack{B_{k-1} \subset B \\ \mu_1(B) > \mu_1(B_{k-1})} \right.} \frac{\mu_0^d(B) - \mu_0^d(B_{k-1})}{\mu_1(B) - \mu_1(B_{k-1})}.$$

If the set Ω is not singleton, then we should choose set B_k with the smallest cardinality such that

$$\mu_1(B_k) = \max_{B \in \Omega} \mu_1(B).$$

c) the set B_m ($\mu_1(B_{m-1}) = 1$) is the set with the smallest cardinality from the family

$$\{B \in \mathfrak{A} \mid B \supseteq B_{m-1}, \mu_0{}^d(B) = 1\}.$$

The above conditions define sets B_k, $k = 1, 2, \ldots, m$, uniquely.

Proof. The conditions a) and c) follow easily from the definition of sets B_1 and B_m. Let us show that b) is true. We prove first that the set B_k should be chosen from Ω. Assume to the contrary $B_k \notin \Omega$. Then there is a $B \in 2^X$ such that

$$\frac{\mu_0^d(B_k) - \mu_0^d(B_{k-1})}{\mu_1(B_k) - \mu_1(B_{k-1})} > \frac{\mu_0^d(B) - \mu_0^d(B_{k-1})}{\mu_1(B) - \mu_1(B_{k-1})}.$$

Let us notice that $\frac{\mu_0^d(B_k) - \mu_0^d(B_{k-1})}{\mu_1(B_k) - \mu_1(B_{k-1})} = \frac{t_{k-1}}{1 - t_{k-1}}$, i.e. we can rewrite the last inequality as

$$\frac{\mu_0^d(B) - \mu_0^d(B_{k-1})}{\mu_1(B) - \mu_1(B_{k-1})} < \frac{t_{k-1}}{1 - t_{k-1}},$$

or

$$(1 - t_{k-1})\mu_0^d(B) - t_{k-1}\mu_1(B) <$$
$$(1 - t_{k-1})\mu_0^d(B_{k-1}) - t_{k-1}\mu_1(B_{k-1}).$$

However, the last inequality contradicts to the choice of $B_{k-1} = \underline{A}_{t_{k-1}}$.

Let us show that we should choose set $B \in \Omega$ with the maximal value $\mu_1(B) - \mu_1(B_{k-1})$ and with the smallest cardinality. Assume that $B'_k, B''_k \in \Omega$. Then

$$\frac{\mu_0^d(B'_k) - \mu_0^d(B_{k-1})}{\mu_1(B'_k) - \mu_1(B_{k-1})} = \frac{\mu_0^d(B''_k) - \mu_0^d(B_{k-1})}{\mu_1(B''_k) - \mu_1(B_{k-1})} = \frac{t_{k-1}}{1-t_{k-1}},$$

and, obviously, $B'_k, B''_k \in \mathcal{L}_{\mu_0,\mu_1}(t_{k-1})$.

Let us check the sign of the following difference:

$$\Delta = (1-t_k)\mu_0^d(B'_k) + t_k(1-\mu_1(B'_k)) -$$
$$\left[(1-t_k)\mu_0^d(B''_k) - t_k(1-\mu_1(B''_k))\right] =$$
$$(1-t_k)\left(\mu_0^d(B'_k) - \mu_0^d(B_k)\right) - t_k\left(\mu_1(B'_k) - \mu_1(B_k)\right) -$$
$$\left[(1-t_k)\left(\mu_0^d(B''_k) - \mu_0^d(B_k)\right) - t_k\left(\mu_1(B''_k) - \mu_1(B_k)\right)\right].$$

Assuming that $\mu_1(B'_k) - \mu_1(B_k) > 0$, we get

$$\Delta = \left[\frac{(1-t_k)t_{k-1}}{1-t_{k-1}} - t_k\right]\left(\mu_1(B'_k) - \mu_1(B''_k)\right),$$

i.e. $\Delta < 0$, and we should choose the set B in Ω with the largest value $\mu_1(B)$ and, of course, with the smallest cardinality, because this follows from the definition of the set $\underline{A}_{t_k}$. The set B_k is defined uniquely by above conditions, because Ω coincides with the lattice $\mathcal{L}_{\mu_0,\mu_1}(t_{k-1})$. The lemma is proved.

Example 1. Consider 2-monotone measures μ_0 and μ_1 defined on the algebra 2^X, where $X = \{x_1, x_2, x_3, x_4\}$, with values given in Table 1. Let us describe the algorithm proposed for searching favorable pairs of probability measures if the first hypothesis is described by μ_0, and the second hypothesis is described by μ_1. Applying the algorithm we get $B_1 = \varnothing$; the set B_1 should be chosen to minimize the value $\dfrac{\mu_0^d(B)}{\mu_1(B)}$. Clearly, $B_2 = \{x_1\}$ and $\dfrac{\mu_0^d(B_2)}{\mu_1(B_2)} = \dfrac{0.02}{0.3} = \dfrac{1}{15}$. Then minimizing the value $\dfrac{\mu_0^d(B) - \mu_0^d(B_2)}{\mu_1(B) - \mu_1(B_2)}$ for $B \supset B_2$, we get that $B_3 = \{x_1, x_3\}$ and $\dfrac{\mu_0^d(B_3) - \mu_0^d(B_2)}{\mu_1(B_3) - \mu_1(B_2)} = \dfrac{0.3-0.02}{0.7-0.3} = \dfrac{7}{10}$. By analogy, $B_4 = X$ and $\dfrac{\mu_0^d(B_4) - \mu_0^d(B_3)}{\mu_1(B_4) - \mu_1(B_3)} = \dfrac{1-0.3}{1-0.7} = \dfrac{7}{3}$. Now it is easy to calculate the likelihood ratio. In our case, $\underline{\pi}(x) = \overline{\pi}(x) = \pi(x)$, and $\pi(x_1) = 1/15$; $\pi(x_3) = 7/10$, and $\pi(x_2) = \pi(x_4) = 7/3$. We see that in our case $(P_0)_{B_k \setminus B_{k-1}}$, $(P_1)_{B_k \setminus B_{k-1}}$, $k = 2,3$, are Dirac measures. For searching $(P_0)_{B_4 \setminus B_3}$, $(P_1)_{B_4 \setminus B_3}$, we need to solve the following inequalities:

x_1	x_2	x_3	x_4	μ_0	μ_0^d	μ_1
0	0	0	0	0	0	0
1	0	0	0	0	0.02	0.3
0	1	0	0	0	0.2	0
1	1	0	0	0	0.2	0.3
0	0	1	0	0	0.3	0
1	0	1	0	0	0.3	0.7
0	1	1	0	0	0.45	0
1	1	1	0	0	0.45	0.7
0	0	0	1	0.55	1	0
1	0	0	1	0.55	1	0.3
0	1	0	1	0.7	1	0
1	1	0	1	0.7	1	0.6
0	0	1	1	0.8	1	0
1	0	1	1	0.8	1	0.7
0	1	1	1	0.98	1	0
1	1	1	1	1	1	1

Table 1: Values of monotone measures.

$$(\mu_1)_{B_4 \setminus B_3} \le (P_1)_{B_4 \setminus B_3} = (P_0)_{B_4 \setminus B_3} \le (\mu_0^d)_{B_4 \setminus B_3}.$$

This system of inequalities can be rewritten as

$$\begin{cases} 0 \le (P_0)_{B_4 \setminus B_3}(\{x_2\}) \le \dfrac{15}{70}, \\ 0 \le (P_0)_{B_4 \setminus B_3}(\{x_4\}) \le 1, \\ (P_0)_{B_4 \setminus B_3}(\{x_2\}) + (P_0)_{B_4 \setminus B_3}(\{x_4\}) = 1. \end{cases}$$

It is easy to find any solution of this system of inequalities can be represented as

$$\begin{pmatrix} (P_0)_{B_4 \setminus B_3}(\{x_2\}) \\ (P_0)_{B_4 \setminus B_3}(\{x_4\}) \end{pmatrix} = a \begin{pmatrix} 0 \\ 1 \end{pmatrix} + (1-a) \begin{pmatrix} 15/70 \\ 55/70 \end{pmatrix},$$

where $a \in [0,1]$. Thus, we have the following solution for favorable pairs:

$$\begin{pmatrix} P_0(\{x_1\}) \\ P_0(\{x_2\}) \\ P_0(\{x_3\}) \\ P_0(\{x_4\}) \end{pmatrix} = a \begin{pmatrix} 0.02 \\ 0 \\ 0.28 \\ 0.7 \end{pmatrix} + (1-a) \begin{pmatrix} 0.02 \\ 0.15 \\ 0.28 \\ 0.55 \end{pmatrix},$$

$$\begin{pmatrix} P_1(\{x_1\}) \\ P_1(\{x_2\}) \\ P_1(\{x_3\}) \\ P_1(\{x_4\}) \end{pmatrix} = a \begin{pmatrix} 0.3 \\ 0 \\ 0.4 \\ 0.3 \end{pmatrix} + (1-a) \begin{pmatrix} 0.3 \\ 45/700 \\ 0.4 \\ 165/700 \end{pmatrix},$$

where $a \in [0,1]$. Let us notice that if $a \in [0,1)$ the least favorable pair (P_0, P_1) has the same likelihood ratio as (μ_0, μ_1), but $P_0(\{x_2\}) = P_1(\{x_2\}) = 0$, when $a = 1$, this pair does not give us sufficient information how to construct the optimal test, when we observe the outcome x_2.

For illustration, we can also calculate sets $\underline{A}_t$:

$$\underline{A}_t = B_1 = \varnothing \text{ if } \frac{t}{1-t} \le \frac{1}{15} \text{ (or } 0 \le t \le \frac{1}{16}\text{)};$$

$$\underline{A}_t = B_2 \text{ if } \frac{1}{16} < t \le \frac{7}{17};$$

$$\underline{A}_t = B_3 \text{ if } \frac{7}{17} < t \le \frac{7}{10};$$

$$\underline{A}_t = B_4 = X \text{ if } \frac{7}{10} < t \le 1.$$

4 Characterization of least favorable pairs by functionals

The next theorem is the analog of the result proved by Huber and Strassen [13]. Its proof gives us also some corollaries that are useful for computing functionals using least favorable pairs.

Theorem 1. *Let $\mu_0, \mu_1 \in M_{2-mon}$ and let Φ be any twice continuously differentiable function on $[0,1]$, such that $\Phi'' > 0$. Then the pair $(Q_0, Q_1) \in core(\mu_0) \times core(\mu_1)$ minimizes the functional*

$$H(P_0, P_1) = \int_X \Phi\left(\frac{dP_0}{dP_0 + dP_1}\right) d(P_0 + P_1)$$

among all $(P_0, P_1) \in core(\mu_0) \times core(\mu_1)$ iff $q_{P_0,P_1}(t) \le q_{Q_0,Q_1}(t)$ for all $t \in [0,1]$.

Proof. Let us denote by $y = \frac{\underline{\pi}(x)}{\underline{\pi}(x)+1} = \frac{dP_0}{dP_0 + dP_1}$, where $\underline{\pi}(x)$ is the likelihood ratio of probability measures P_0 and P_1. We derive first the explicit expression for $q_{P_0,P_1}(t)$ using the result formulated in Proposition 2:

$$q_{P_0,P_1}(t) = (1-t)P_0(\overline{A}_t) + t\left(1 - P_1(\overline{A}_t)\right) =$$

$$(1-t)\int_{y \le t} dP_0 + t\left(1 - \int_{y \le t} dP_1\right) =$$

$$(1-t)\int_{y \le t}\left(\frac{dP_0}{dP_0 + dP_1}\right) d(P_0 + P_1) -$$

$$t\int_{y \le t}\left(\frac{dP_1}{dP_0 + dP_1}\right) d(P_0 + P_1) + t =$$

$$(1-t)\int_{y \le t} y\, d(P_0 + P_1) - t\int_{y \le t}(1-y)\, d(P_0 + P_1) + t =$$

$$\int_{y \le t}(y - t)\, d(P_0 + P_1) + t.$$

Therefore, $q_{P_0,P_1}(t) \le q_{Q_0,Q_1}(t)$ for all $t \in [0,1]$ iff

$$\int_{\frac{dQ_0}{dQ_0 + dQ_1} \le t}\left[\frac{dQ_0}{dQ_0 + dQ_1} - t\right] d(Q_0 + Q_1) \ge$$

$$\int_{\frac{dP_0}{dP_0 + dP_1} \le t}\left[\frac{dP_0}{dP_0 + dP_1} - t\right] d(P_0 + P_1) \tag{4}$$

for all $t \in [0,1]$. Introduce into consideration an arbitrary positive integrable function φ on $[0,1]$. Then the condition (4) can be equivalently transformed to

$$\int_0^1 \varphi(t)\left(\int_{\frac{dQ_0}{dQ_0 + dQ_1} \le t}\left[t - \frac{dQ_0}{dQ_0 + dQ_1}\right] d(Q_0 + Q_1)\right) dt \le$$

$$\int_0^1 \varphi(t)\left(\int_{\frac{dP_0}{dP_0 + dP_1} \le t}\left[t - \frac{dP_0}{dP_0 + dP_1}\right] d(P_0 + P_1)\right) dt.$$

Let us denote

$$H(P_0, P_1) = \int_0^1 \varphi(t)\left(\int_{y \le t}(t - y)\, d(P_0 + P_1)\right) dt.$$

We transform next the functional H to the form, which is used in the theorem. After changing the order of integration, we get

$$H(P_0, P_1) = \int_X\left(\int_y^1 \varphi(t)(t - y)\, dt\right) d(P_0 + P_1).$$

Let $\Phi(y) = \int_y^1 \varphi(t)(y - t)\, dt$. Then $\Phi'(y) = -\int_y^1 \varphi(t)\, dt$ and $\Phi''(y) = \varphi(y) > 0$. Notice that for this function Φ: $\Phi(1) = 0$ and $\Phi'(1) = 0$. Let $\Phi_1(y) = \Phi(y) + by + a$, where $a, b \in \mathbb{R}$. Observe that this is the general possible choice of twice differentiable function with $\Phi'' = \varphi > 0$. Then

$$\int_X \Phi_1\left(\frac{dP_0}{dP_0 + dP_1}\right) d(P_0 + P_1) = H(P_0, P_1) +$$

$$b\int_X dP_0 + a\int_X d(P_0 + P_1) = H(P_0, P_1) + 2a + b.$$

The theorem is proved.

Corollary 3. *Let us use assumptions and notations from Theorem 1. Then*

$$H(P_0,P_1)=\int_X \Phi\left(\frac{dP_0}{dP_0+dP_1}\right)d(P_0+P_1)=$$

$$\int_0^1 \left(t-q_{P_0,P_1}(t)\right)\Phi''(t)dt-\Phi'(1)-\Phi(1).$$

Obviously we can use Theorem 1 for solving optimization problems using least favorable pairs. This result is formulated below.

Corollary 4. *Let* $\mu_0,\mu_1\in M_{2-mon}$ *and let* $\Phi:[0,1]\to(-\infty,+\infty]$ *be any twice continuously differentiable function on* $(0,1)$*, such that* $\Phi''(y)\geq 0$ *for all* $y\in(0,1)$*; in addition* $\Phi(0)=\lim_{y\to+0}\Phi(y)$ *and* $\Phi(1)=\lim_{y\to 1-0}\Phi(y)$*. Then any least favorable pair* $(Q_0,Q_1)\in core(\mu_0)\times core(\mu_1)$ *minimizes the functional*

$$H(P_0,P_1)=\int_X \Phi\left(\frac{dP_0}{dP_0+dP_1}\right)d(P_0+P_1)$$

among all $(P_0,P_1)\in core(\mu_0)\times core(\mu_1)$.

Proof. Let Φ be any twice continuously differentiable function on $[0,1]$ such that $\Phi''(y)\geq 0$ for all $y\in[0,1]$. Then this result obviously follows from Corollary 3, namely, from the formula:

$$H(P_0,P_1)=\int_0^1 \left(t-q_{P_0,P_1}(t)\right)\varphi(t)dt+C,$$

If $(Q_0,Q_1),(P_0,P_1)\in core(\mu_0)\times core(\mu_1)$ and (Q_0,Q_1) is a least favorable pair, then $q_{P_0,P_1}(t)\leq q_{Q_0,Q_1}(t)$ for all $t\in[0,1]$ and obviously $H(P_0,P_1)\geq H(Q_0,Q_1)$. If (P_0,P_1) is also a favorable pair, then $q_{P_0,P_1}(t)=q_{Q_0,Q_1}(t)$ for all $t\in[0,1]$ and $H(P_0,P_1)=H(Q_0,Q_1)$.

Consider now the general case, formulated in the corollary. For any $\varepsilon>0$ introduce the functional

$$H_\varepsilon(P_0,P_1)=\int_X \Phi_\varepsilon(y)d(P_0+P_1),$$

where

1) $\Phi_\varepsilon(y)=\Phi(\varepsilon)+\Phi'(\varepsilon)(y-\varepsilon)+0.5\Phi''(\varepsilon)(y-\varepsilon)^2$ if $y\in[0,\varepsilon]$;

2) $\Phi_\varepsilon(y)=\Phi(y)$ if $y\in(\varepsilon,1-\varepsilon)$;

3) $\Phi_\varepsilon(y)=\Phi(1-\varepsilon)+\Phi'(1-\varepsilon)(y+\varepsilon-1)+$ $0.5\Phi''(1-\varepsilon)(y+\varepsilon-1)^2$ if $y\in[1-\varepsilon,1]$.

Then Φ_ε is a twice continuously differentiable function on $[0,1]$ such that $\Phi''_\varepsilon(y)\geq 0$ for all $y\in[0,1]$. This implies that $H_\varepsilon(Q_0,Q_1)\leq H_\varepsilon(P_0,P_1)$ if $(Q_0,Q_1),(P_0,P_1)\in core(\mu_0)\times core(\mu_1)$ and (Q_0,Q_1) is a least favorable pair. Clearly that

$$H(Q_0,Q_1)=\lim_{\varepsilon\to 0}H_\varepsilon(Q_0,Q_1)\leq$$
$$\lim_{\varepsilon\to 0}H_\varepsilon(P_0,P_1)=H(P_0,P_1).$$

The corollary is proved.

Because the value $H(P_0,P_1)$ does not depend on a chosen favorable pair $(P_0,P_1)\in core(\mu_0)\times core(\mu_1)$, it can be expressed through the values of measures $\mu_0{}^d$ and μ_1 on the chain $\{B_0,B_1,...,B_m\}$. This result is given in the next corollary.

Corollary 5. *Assume that we use notations from Corollary 2 and* $(P_0,P_1)\in core(\mu_0)\times core(\mu_1)$ *be a least favorable pair. Let* $\nu=\mu_0{}^d+\mu_1$*. Then*

$$H(P_0,P_1)=\sum_{k=1}^m \Phi\left(\frac{\mu_0{}^d(B_k)-\mu_0{}^d(B_{k-1})}{\nu(B_k)-\nu(B_{k-1})}\right)\left(\nu(B_k)-\nu(B_{k-1})\right).$$

Proof. Notice that $P_0(\{x\})=P_1(\{x\})=0$ if $x\in X\setminus B_m$. Therefore,

$$H(P_0,P_1)=\sum_{x\in B_m}\Phi\left(\frac{P_0(\{x\})}{P_0(\{x\})+P_1(\{x\})}\right)\left(P_0(\{x\})+P_1(\{x\})\right),$$

and by Corollary 2 $\frac{P_0(\{x\})}{P_0(\{x\})+P_1(\{x\})}=$ $\frac{\mu_0{}^d(B_k)-\mu_0{}^d(B_{k-1})}{\mu_0{}^d(B_k)-\mu_0{}^d(B_{k-1})+\mu_1(B_k)-\mu_1(B_{k-1})}$ if $x\in B_k\setminus B_{k-1}$ and $P_0(\{x\})+P_1(\{x\})\neq 0$; $P_0(B_k\setminus B_{k-1})=\mu_0{}^d(B_k)-\mu_0{}^d(B_{k-1})$ and $P_1(B_k\setminus B_{k-1})=\mu_1(B_k)-\mu_1(B_{k-1})$. Hence, the formula in the corollary is true.

Example 2. The Kullback–Leibler divergence (distance) between probablity measures P_0 and P_1 is defined as

$$D_{KL}(P_1,P_0)=\int_X \ln\left(\frac{dP_1}{dP_0}\right)dP_1.$$

In applications, we need to minimize $D_{KL}(P_1,P_0)$ if $(P_0,P_1)\in core(\mu_0)\times core(\mu_1)$. In particular, if $X=\{x_1,...,x_n\}$ and $P_0(\{x_i\})=1/n$, then

$$D_{KL}(P_1,P_0)=\sum_{i=1}^n \ln\left(P_1\{x_i\}\right)P_1\{x_i\}+\ln(n)=$$
$$-S(P_1)+\ln(n),$$

where $S(P_1)=-\sum_{i=1}^n \ln\left(P_1\{x_i\}\right)P_1\{x_i\}$ is the Shannon entropy. Let us transform the functional D_{KL} to the form used in Theorem 1.

$$D_{KL}(P_1,P_0)=\int_X \ln\left(\frac{dP_1}{dP_0}\right)\frac{dP_1}{dP_1+dP_0}d\left(P_1+P_0\right).$$

Let $y=\frac{dP_0}{dP_0+dP_1}$. Then $D_{KL}(P_1,P_0)=\int_X \Phi(y)d\left(P_1+P_0\right)$, where $\Phi(y)=(1-y)\ln\left(\frac{1-y}{y}\right)$. Notice that in this case $\varphi(y)=\Phi''(y)=\frac{1}{y^2(1-y)}\geq 0$ for all $y\in(0,1)$, i.e. by Corollary 4 any least favorable pair $\left(Q_0,Q_1\right)\in core\left(\mu_0\right)\times core\left(\mu_1\right)$ minimizes the functional $D_{KL}(P_1,P_0)$ among all $\left(P_0,P_1\right)\in core\left(\mu_0\right)\times core\left(\mu_1\right)$. It is remarkable, that we can use the algorithm for finding least favorable pairs in the problem of maximizing the Shannon entropy functional. In this case, we get explicitly the same algorithm proposed firstly for belief measures [14] and then justified for 2-monotone measures [1].

5 Concluding remarks

This work gives a new look on Huber-Strassen results, presented here in the explicit form. Some of them are even strengthened (see Proposition 4 and Corollary 2) or clarified (see Theorem 1 and Corollaries 4-6). As a result we have an effective algorithm for searching least favorable pairs and also the way for minimizing functionals on 2-monotone measures described in Theorem 1 and its corollaries. As shown in [5], it is possible to generalize canonical sequences of 2-monotone measures generated by any chain of sets. This generalization can be useful for describing least favorable pairs for the general case of 2-monotone measures.

Acknowledgements

I am very grateful to the referees for many very helpful and detailed remarks. I express also my sincere thanks to Russian Foundation of Basic Research for support by grants 10-07-00135-a, 10-07-00478-a, 11-07-00591-a, and 11-07-08026-з.

References

[1] J. Abellán, S. Moral. An algorithm to compute the upper entropy for order-2 capacities. *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems* 14: 141 – 154, 2006.

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

[3] T. Augustin. Neyman-Pearson testing under interval probability by globally least favorable pairs. Reviewing Huber–Strassen theory and extending it to general interval probability. *Journal of Statistical Planning and Inference* 105: 149 – 173, 2002.

[4] T. Bednarski. On solutions of minimax test problems for special capacities. *Z. Wahrsch. Verw. Gebiete* 58: 397–405, 1981.

[5] A.G. Bronevich. Canonical sequences of fuzzy measures. In Proc. of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-2004), Perugia-Italy, 2004, 8 pp.

[6] A.G. Bronevich, T. Augustin. Approximation of coherent lower probabilities by 2-monotone measures. In Proc. of the 6th International Symposium on Imprecise Probability: Theories and Their Applications, Durham, United Kingdom, 2009, 9 pp.

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

[8] G. Choquet. Theory of capacities. *Ann. Inst. Fourier* 5: 131 – 295, 1954.

[9] D. Denneberg. Non-additive Measure and Integral. Dordrecht, Kluwer, 1997.

[10] D. Denneberg. Conditional expectation for monotone measures, the discrete case. *Journal of Mathematical Economics* 37: 105 – 121, 2002.

[11] R. Hable. Data-based decisions under imprecise probability and least favorable models. *International Journal of Approximate Reasoning* 50: 642 – 654, 2009.

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

[13] P. J. Huber, V. Strassen. Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.* 1: 251–263, 1973.

[14] G. J. Klir. *Uncertainty and Information: Foundations of Generalized Information Theory*. Hoboken, NJ: Wiley-Interscience, 2006.

[15] F. Österreicher. On the construction of least favorable pairs of distributions. *Z. Wahrsch. Verw. Gebiete* 43: 49–55, 1978.

[16] H. Rieder. Least favorable pairs for special capacities. *Ann. Statist.* 5: 909 – 921, 1977.

[17] H. Schwarte, J.S. Sadowsky. Revisiting the Huber-Strassen minimax theorem for capacities. In Proc. of IEEE International Symposium on Information Theory, Whistler, BC, Canada, 1995, 8 pp.

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