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", Moscow, RUSSIA, brone@mail.ru

Objectives of the paper:

a) Algebraic description of least favorable pairs in Huber-Strassen theory;

b) To construct a simple algorithm for searching least favorable pairs;

c) To find connections between least favorable pairs and functionals in probability theory (Shannon entropy and Kullback–Leibler distance).

Notation and Definitions

X is a finite set and $\mathfrak{A} = 2^X$ is an algebra of its subsets.

D1. $\mu : \mathfrak{A} \to [0,1]$ is a *monotone measure* if

1) $\mu(\varnothing) = 0,\ \mu(X) = 1$;

2) $\mu(A) \le \mu(B)$ if $A \subseteq B$ for $A, B \in \mathfrak{A}$.

Notation.

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

$\mu_1 \le \mu_2$ for $\mu_1, \mu_2 \in M_{mon}$ if $\mu_1(A) \le \mu_2(A)$ for all $A \in \mathfrak{A}$.

ν is dual to μ if $\nu(A) = 1 - \mu(A^c),\ A \in \mathfrak{A}$ $(\nu = \mu^d)$;

M_{pr} is the set of all probability measures on $\mathfrak{A}$.

$M_{low} = \left\{ \mu \in M_{mon} \mid \exists P \in M_{pr} : \mu \le P \right\}$ is the set of all *lower probabilities* on $\mathfrak{A}$.

$M_{coh} = \left\{ \mu \in M_{mon} \mid \forall B \in \mathfrak{A}, \exists P \in M_{pr} : \mu \le P, \mu(B) = P(B) \right\}$ is the set of all *coherent lower probabilities* on $\mathfrak{A}$.

$\mu \in M_{mon}$ is 2-*monotone* if $\mu(A) + \mu(B) \le \mu(A \cup B) + \mu(A \cap B)$ for all $A, B \in \mathfrak{A}$; M_{2-mon} is the set of all 2-monotone measures on $\mathfrak{A}$.

Neymann-Pearson testing for 2-monotone measures

Let two hypotheses H_0 and H_1 be described by 2-monotone measures μ_0 and μ_1. Then any optimal test between them can be found by solving the following optimization problem:

$$q_{\mu_0^d,\mu_1^d}(t)=\min_{A\in 2^X}\left\{(1-t)\mu_0^d(A)+t\mu_1^d(A^c)\right\}, \text{.where } t\in[0,1].$$

$q_{\mu_0^d,\mu_1^d}(t)$ is 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.

$q_{\mu_0^d,\mu_1^d}(t)$ can be rewritten as

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

By Huber-Strassen theory, there is a pair $(P_0,P_1)\in core(\mu_0)\times core(\mu_1)$ (least favorable) such that

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

Algebraic description of the optimization problem

Notation.

$$\mathcal{L}_{\mu_0,\mu_1}(t)=\left\{A\in 2^X \,\middle|\, (1-t)\mu_0^d(A)+t\mu_1^d(A^c)=q_{\mu_0^d,\mu_1^d}(t)\right\},\ t\in[0,1];$$

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

Proposition 1. *Let* $\mu_0,\mu_1\in M_{2-mon}$, $A\in\mathcal{L}_{\mu_0,\mu_1}(t)$, $B\in\mathcal{L}_{\mu_0,\mu_1}(s)$, *and* $t\le s$. *Then* $A\cap B\in\mathcal{L}_{\mu_0,\mu_1}(t)$ *and* $A\cup B\in\mathcal{L}_{\mu_0,\mu_1}(s)$.

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

Notation. $\underline{A}_t=\bigcap_{A\in\mathcal{L}_{\mu_0,\mu_1}(t)}A$, $\overline{A}_t=\bigcup_{A\in\mathcal{L}_{\mu_0,\mu_1}(t)}A$.

P1. $\underline{A}_t,\overline{A}_t\in\mathcal{L}_{\mu_0,\mu_1}(t)$ by Proposition 1.

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 $\underline{A}_t = \left\{ x \mid (1-t)P_0\left(\{x\}\right) < tP_1\left(\{x\}\right) \right\}$, $\overline{A}_t = \left\{ x \mid (1-t)P_0\left(\{x\}\right) \le tP_1\left(\{x\}\right) \right\}$.

The *likelihood ratio* of probability measures P_0 and P_1.can be defined by

$\underline{\pi} : X \to [0,+\infty]$, $\overline{\pi} : X \to [0,+\infty]$ and

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

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

P2. $\underline{A}_t = \left\{ x \in X \mid \overline{\pi}(x) < t/(1-t) \right\}$,

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

Necessary and sufficient conditions for least favorable pairs

Lemma 1. *Let* $\mu_0,\mu_1 \in M_{2-mon}$, $P_0 \in core(\mu_0)$ *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* $t \in [0,1]$.

Lemma 2. *Let the conditions of Lemma 1 hold 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* $x \in X \setminus (\overline{A}_0 \cup \underline{A}_1)$.

D2. Functions $\underline{\pi}(x)$ and $\overline{\pi}(x)$ from Lemma 2 are called 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* $P_1(\{x\}) > 0$.

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* $P_1(\{x\}) > 0$.

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* $t \in [0,1]$.

The algorithm for searching least favorable pairs

Description of sets $\{\underline{A}_t\}$ (finite case)

$\exists\{t_1, t_2, ..., t_{m-1}\}$ such that $0 < t_1 < t_2 < ... < t_{m-1} = 1$, $\underline{\pi}(x) = t_k/(1-t_k)$ if $x \in \underline{A}_{t_{k+1}} \setminus \underline{A}_{t_k}$, $k = 1,...,m-2$.

Notation:

$B_k = \underline{A}_{t_k}$, $k = 1,...,m-1$, $B_m = \left(X \setminus \overline{A}_0\right) \cup \underline{A}_1$.

P3. 1) $\underline{\pi}(x) = \overline{\pi}(x) = t_k/(1-t_k)$ if $x \in B_{k+1} \setminus B_k$, $k = 1,...,m-1$,

2) $\underline{\pi}(x) = \overline{\pi}(x) = 0$ if $x \in B_1$,

3) $\underline{\pi}(x) = 0$ and $\overline{\pi}(x) = +\infty$ if $x \in X \setminus B_m$.

Corollary 2. *Let $\mu_0, \mu_1 \in M_{2-mon}$. Then every least favorable pair (P_0, 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:

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

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

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

Lemma 4. *The choice of sets* B_k, $k = 1,\dots,m$, *is produced as follows:*

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

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

b) If sets $B_0 = \varnothing, B_1, \dots, B_{k-1}$, $k \ge 2$, *are known and* $\mu_1(B_{k-1}) < 1$. *Then* B_k *should be chosen from the set* Ω *of possible solutions of the following optimization problem*

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

If $|\Omega| \neq 1$, *then the set* B_k *should be 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*

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

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

Characterization of least favorable pairs by functionals

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* $t \in [0,1]$.

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

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

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

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* $\Phi''(y) \ge 0$ *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 $(P_0, P_1) \in core(\mu_0) \times core(\mu_1)$.

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

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

Example 2. The Kullback–Leibler distance is $D_{KL}(P_1, P_0) = \int_X \ln\left(\frac{dP_1}{dP_0}\right) dP_1$.

It can be rewritten as $D_{KL}(P_1, P_0) = \int_X \Phi(y) d\left(P_1 + P_0\right)$, where

$\Phi(y) = (1-y)\ln\left(\frac{1-y}{y}\right)$. Observe that $\varphi(y) = \Phi''(y) = \frac{1}{y^2(1-y)} \geq 0$ 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 $\left(P_0, P_1\right) \in core\left(\mu_0\right) \times core\left(\mu_1\right)$.

Conclusion

1. The algebraic description of least favorable pairs is given.
2. An effective algorithm for searching least favorable pairs is constructed.
3. It is established the connection between computing functionals (Shannon entropy and Kullback–Leibler distance) and searching least favorable pairs.