Approximation of coherent lower probabilities by 2-monotone measures

Andrey G. Bronevich
Technological Institute of Southern Federal University,
Taganrog, RUSSIA, brone@mail.ru

Thomas Augustin
Department of Statistics, Ludwig-Maximilians
University (LMU), Munich, GERMANY
thomas@stat.uni-muenchen.de

1. Personal Background: Andrey Bronevich

Associate Professor at the
Technological Institute,
Taganrog, Russia

- imprecise probabilities
- non-additive measure theory
- possibility theory
- multi-criteria decision-making under risk and uncertainty
- pattern recognition and image processing.

Personal Background: Thomas Augustin

- Department of Statistics, University of Munich, Germany

- Kurt Weichselberger
- Marco Cattaneo, Gero Walter; Andrea Wiencierz, Carolin Strobl; Robert Hable, Martin Gümbel

- IP in statistics and decision making; foundations of statistics (poster on handling unobserved data heterogeneity by credal maximum likelihood)

- Statistics in the social sciences; handling of deficient data (measurement error, misclassification) --?--> partial identification (poster)

- Nonparametric predictive inference
- Classification trees

2. Background of the paper

Why can approximation of coherent lower probabilities by 2-monotone measures be reasonable?

I do not „…know any ‘rationality’ argument

for two-monotonicity, beyond its computational convenience.”

Walley (1981, p. 51)

On the computational convenience of two-monotonicity:

“Lower and upper distribution functions fit”:

For any underlying order of the elements of the sample space corresponding lower and upper distribution function are attained *simultaneously* by a certain element of the structure (core, creedal set)

- Closed form expressions for natural extension/expectation: Choquet integral

- Closed from expressions for conditional probabilities (c.p.)

- Closeness of GBR (intuitive concept of c.p.) and Dempster’s rule of conditioning (maximum likelihood updating)

- Statistical hypotheses testing (Huber-Strassen theory)

Statistical hypotheses testing (Huber-Strassen theory)

- Which distribution governs the data?

- (Level-alpha-)Maximin testing to decide between two hypotheses

- Two-monotonicity of the underlying hypotheses is sufficient to guaranty the existence of a globally least favorable pair

  à big sample sizes are no computational problem

  à just use the product measure of the globally least favorable pair of sample size 1

  à k-dimensional problem, instead of a k^n-dimensional problem

3. **Find optimal outer approximation by two-monotone measure!**

Not unique

uniformly		criterion-based
		linear imprecision index or metric
set of Pareto optimal solutions	improve given approximation	optimal solution via linear programming
• characterization • approximate calculation via linear programming *		*

---

* methods also applicable for approximation by completely-monotone measures

Notation and Definitions

X is a measurable space with a σ-algebra $\mathcal{A}$.

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

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

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

Notation.

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

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

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

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

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

$\mu\in M_{mon}$ is 2-*monotone* if $\mu(A)+\mu(B)\le\mu(A\cup B)+\mu(A\cap B)$ for all $A,B\in\mathcal{A}$.

M_{2-mon} is the set of all 2-monotone measures on $\mathcal{A}$.

Description of Pareto optimal 2-monotone measures (finite case)

X is a finite set; $\mathsf{A} = 2^X$.

D2. $\nu \in M_{mon}$ is a *Pareto optimal approximation* of $\mu \in M_{low}$ if

a) $\nu \le \mu$; b) $\nu' \in M_{mon}$, $\nu \le \nu' \le \mu \Rightarrow \nu' = \nu$.

$\nu \le \nu' \le \mu$ $\nu' \in M_{mon}$ implies that $\nu' = \nu$. For any $\mu \in M_{low}$, we denote $M_{2-mon\le\mu} = \{\nu \in M_{2-mon} | \nu \le \mu\}$.

Notation. $M_{2-mon\le\mu} = \{\nu \in M_{2-mon} | \nu \le \mu\}$ for $\mu \in M_{mon}$.

Lemma 1. *Any Pareto optimal 2-monotone measure for a* $\mu \in M_{coh}$ *can be represented as a convex linear combination of Pareto optimal extreme points of* $M_{2-mon\le\mu}$.

D3. $\Lambda \subseteq \mathsf{A}$ is a *lattice* if $A, B \in \Lambda \Rightarrow A \cap B, A \cup B \in \Lambda$.
D4. $\mu \in M_{2-mon}$ is *additive* on L if
$\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$ for any $A, B \in \Lambda$.

$core(\mu) = \{P \in M_{pr} \mid P \geq \mu\}$, where $\mu \in M_{low}$.
S_{μ} is the covering of A: $\Lambda \in \mathsf{S}_{\mu}$ if Λ is a maximal lattice, on which $\mu \in M_{2-mon}$ is additive.

Proposition 1. *There is the one-to-one correspondence between maximal lattices in* S_{μ} *and extreme points of* $core(\mu)$ *for every extreme point* P *defined by* $\Lambda = \{A \in \mathsf{A} \mid P(A) = \mu(A)\}$.

Proposition 1 is the generalization of the following result:

Let $X=\{x_1,x_2,...,x_n\}$ *and* $\mu\in M_{2-mon}$, *then every extreme point* P_γ *of* *core*(μ) *correspond to a maximal chain* $\gamma=\{B_0,B_1,...,B_n\}$ *of* $\mathsf{A}=2^X$, *where* $\gamma=\{B_0,B_1,...,B_n\}$, $\varnothing=B_0\subset B_1\subset...\subset B_n=X$, *and* $|B_k\setminus B_{k-1}|=1$, $k=1,...,n$, *by the rule* $P_\gamma(B_k)=\mu(B_k)$, $k=1,...,n$.

Proposition 2. *Let* $\mu\in M_{coh}, \nu\in M_{2-mon\le\mu}, S_{\nu=0}=\{A\in\mathsf{A}\mid\nu(A)=0\}$, $S_{\nu=\mu}=\{A\in\mathsf{A}\mid\nu(A)=\mu(A)\}$. *Then* ν *is an extreme point of* $M_{2-mon\le\mu}$ *iff its values are defined by the sets* $S_{\nu=\mu}$, $S_{\nu=0}$, S_ν *uniquely.*

Necessary and sufficient condition of 2-monotonicity

$\mu : 2^X \to [0,1]$ is in M_{2-mon} iff

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

2) $\mu(A) \le \mu\left(A \cup \{x_i\}\right)$ for all possible $A \in 2^X$ and $x_i \notin A$;

3) $\mu\left(A \cup \{x_i\}\right) + \mu\left(A \cup \{x_j\}\right) \le \mu\left(A\right) + \mu\left(A \cup \{x_i\} \cup \{x_j\}\right)$ for all possible $A \in 2^X$ and $x_i, x_j \notin A$.

This result can be reformulated trough elementary lattices:

Type 1 elementary lattices:

$\left\{A, A \cup \{x_i\}\right\}$, where $A \in 2^X$ and $x_i \notin A$,

Type 2 elementary lattices:

$\left\{A, A \cup \{x_i\}, A \cup \{x_j\}, A \cup \{x_i\} \cup \{x_j\}\right\}$, where $A \in 2^X,\ x_i, x_j \notin A$.

Proposition 3. $\mu : 2^X \to [0,1]$ *is in* M_{2-mon} *iff*

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

2) μ *is monotone on all lattices in* 2^X *of the first type;*

3) μ *is 2-monotone on all lattices in* 2^X *of the second type.*

Proposition 4. *Let* $\nu \in M_{2-mon\le\mu}$, L_1 *be the set of all elementary lattices of the first type on which* ν *is constant, and* L_2 *be the set of all elementary lattices of the second type, on which* ν *is additive. Then* ν *is not Pareto optimal iff there is a non-identical zero, non-negative set function* $\Delta\nu : 2^X \to \mathbb{R}_+$ *such that*

1) $\Delta\nu(A) = 0$ *if* $A \in S_{\nu=\mu}$*;*

2) $\Delta\nu$ *is monotone on all lattices in* L_1*;*

3) $\Delta\nu$ *is 2-monotone on all lattices in* L_2*.*

Algorithms for finding Pareto optimal 2-monotone measures

Algorithm I.

Input data: coherent lower probability μ on 2^X.

First step. Searching a 2-monotone measure ν_0 with $\nu_0 \le \mu$.

1) Compute 2-monotone set function g on 2^X:

a) $g(A) = \mu(A)$ for all $A \in 2^X$ with $|A| \le 1$;

b) for sets A with cardinality $|A| = k$, $k = 1, 2, \ldots$:

$$g(A) = \max\left\{\mu(A), \max_{x_i, x_j \in A} g\left(A \setminus \{x_i\}\right) + g\left(A \setminus \{x_j\}\right) - g\left(A \setminus \{x_i, x_j\}\right)\right\}.$$

2) $\nu_0 = \varphi \circ g$, where $\varphi : [0, g(X)] \to [0,1]$ is a convex distortion function such that:

(i) $\varphi(0) = 0$, $\varphi(g(X)) = 1$;

(ii) $\varphi(g(A)) \le \mu(A)$ for all $A \in 2^X$.

Second step. Finding a Pareto optimal 2-monotone measure ν with $\nu_0 \le \nu \le \mu$.

Let $\nu_k \in M_{2-mon\le\mu}$. Let exist $A \in 2^X$ such that

$$\Delta_1 = \mu(A) - \nu_k(A) > 0,$$

$$\Delta_2 = \min_{x_i \in X \setminus A} \left(\nu_k \left(A \cup \{x_i\} \right) - \nu_k(A) \right) > 0,$$

$$\Delta_3 = \min_{x_i \in X \setminus A, x_j \in A} \left(\nu_k \left(A \cup \{x_i\} \right) - \nu_k(A) - \right.$$

$$\left. \nu_k \left(\left(A \setminus \{x_j\} \right) \cup \{x_i\} \right) + \nu_k \left(A \setminus \{x_j\} \right) \right) > 0.$$

Then

$$\nu_{k+1}(B) = \begin{cases} \nu_k(B) + d, & B = A, \\ \nu_k(B), & otherwise. \end{cases}$$

where $d = \min\{\Delta_1, \Delta_2, \Delta_3\}$.

Algorithm II.
Based on a linear imprecision index.
D1. $f : M_{low} \to [0,1]$ is a linear imprecision index if
1) $f(P) = 0$ for any $P \in M_{pr}$;
2) $f\left(\eta_{\langle X \rangle}\right) = 1$, where $\eta_{\langle X \rangle}(A) = 1$ if $A = X$, $\eta_{\langle X \rangle}(A) = 0$ otherwise;
3) $f\left(\nu_1\right) \le f\left(\nu_2\right)$ for any $\nu_1, \nu_2 \in M_{low}$ such that $\nu_1 \ge \nu_2$;
4) $f\left(a\nu_1 + (1-a)\nu_2\right) = af\left(\nu_1\right) + (1-a) f\left(\nu_2\right)$ for arbitrary $a \in [0,1]$ and $\nu_1, \nu_2 \in M_{low}$.

A Pareto optimal $\nu \in M_{2-mon \le \mu}$, $\mu \in M_{coh}$, is the solution of the linear programming problem:

find $\nu \in M_{2-mon \le \mu}$ such that $f(\nu) \to \min$.

Examples of imprecision indices

a) *the generalized Hartley measure*: $GH(\nu) = \frac{1}{\ln|X|} \sum_{A \in 2^X} m(A) \ln|A|$,

where m is the Möbius transform of $\nu \in M_{low}$;

b) the imprecision index based on L_1 distance:

$$f_{L_1}(\nu) = \frac{1}{2^{|X|} - 2} \sum_{A \in 2^X} \left| \overline{\nu}(A) - \nu(A) \right|$$

Approximate the set of Pareto-optimal approximations by linear programming

A two-monotone set-function $\nu(\cdot)$ is not a Pareto-optimal approximation of $\mu(.)$ iff $\exists \nu'(\cdot)$ two-monotone, $\varepsilon > 0$:

$$L(\cdot) \leq L'(\cdot) \leq \mu(.) \qquad \text{and} \qquad \sum_{A \subseteq \omega} L'(A) - L(A) \geq \varepsilon$$

Therefore, for arbitrary small $\varepsilon > 0$, consider the following optimization problem:

$$\tilde{\varepsilon} \longrightarrow \min$$

subject to the constraints

$$\begin{aligned}
\nu'(A) - \nu(A) &\geq 0 \quad \forall A \\
\mu(A) - \nu'(A) &\geq 0 \quad \forall A \\
\sum_A \nu'(A) - \nu(A) &\geq \tilde{\varepsilon} \geq \varepsilon
\end{aligned}$$

$\nu(.)$ two-monotone $\qquad$ $\nu'(.)$ two-monotone

Then the set of optimal solutions has the form $\begin{pmatrix} L^*(.) \\ L'^*(.) \end{pmatrix}$, projection on the first part approximates the set of not-Pareto optimal approximations, subtracting it from the set of all 2-monotone measures dominated by $\mu(\cdot)$ gives an approximation of the set of all Pareto-optimal approximations.