Theory of generalized measures and its applications in decision theory and image processing.

Andrey G. Bronevich

JSC "Research, Development and Planning Institute for Railway Information Technology, Automation and Telecommunication"

Nizhegorodskaya 27, building 1, 109029, Moscow, Russia

brone@mail.ru

european social fund in the czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION, YOUTH AND SPORTS

OP Education for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

Monotone (generalized) measures

Let X be a finite set. A set function $\mu : 2^X \to [0, 1]$ is called a monotone measure if

1. $\mu(\emptyset) = 0$, $\mu(X) = 1$ (norming);
2. $A \subseteq B$ implies $\mu(A) \leqslant \mu(B)$ (monotonicity).

Notation:

- $M_{mon}(X)$ is the set of all monotone measures on 2^X;
- $\mu_1 \leqslant \mu_2$ for $\mu_1, \mu_2 \in M_{mon}(X)$ if $\mu_1(A) \leqslant \mu_2(A)$ for all $A \in 2^X$.

Examples of monotone measures

Example 1.

Let us assume that we have public opinion poll before elections and some persons do not determined about their voting.

Any person can choose a non-empty set of candidates for which she (he) can vote.

Let we have three candidates 1,2,3 and the information about the future voting is represented in Table 1.

1	2	3	m	Bel	Pl	P
0	0	0	0	0	0	0
1	0	0	0.1	0.1	0.7	1/3
0	1	0	0.05	0.05	0.75	1/3
1	1	0	0.1	0.25	0.95	2/3
0	0	1	0.05	0.05	0.75	1/3
1	0	1	0.1	0.25	0.95	2/3
0	1	1	0.2	0.3	0.9	2/3
1	1	1	0.4	1	1	1

Table 1: Results of public opinion poll.

For estimating these results, we can introduce the following functions:

$Bel(A) = \sum_{B|B \subseteq A} m(A)$ is the pessimistic estimate of percentage of voting for candidates from the set A.

$Pl(A) = \sum_{B|B \cap A \neq \emptyset} m(A)$ is the optimistic estimate of percentage of voting for candidates from the set A.

P is the probability measure, where

$$P(\{i\}) = \sum_{A|i \in A} m(A)/|A|$$

is the average estimate of percentage of voting for candidate i.

Let M_{pr} be the set of possible probability measures on the powerset of $\{1, 2, 3\}$. Then the set of probability measures

$$\mathbf{P} = \{P \in M_{pr} | Bel(A) \leqslant P(A) \leqslant Pl(A) \text{ for all } A\}$$

describes all predictable results of voting (see Fig. 1).

Figure 1: Set of probability measures

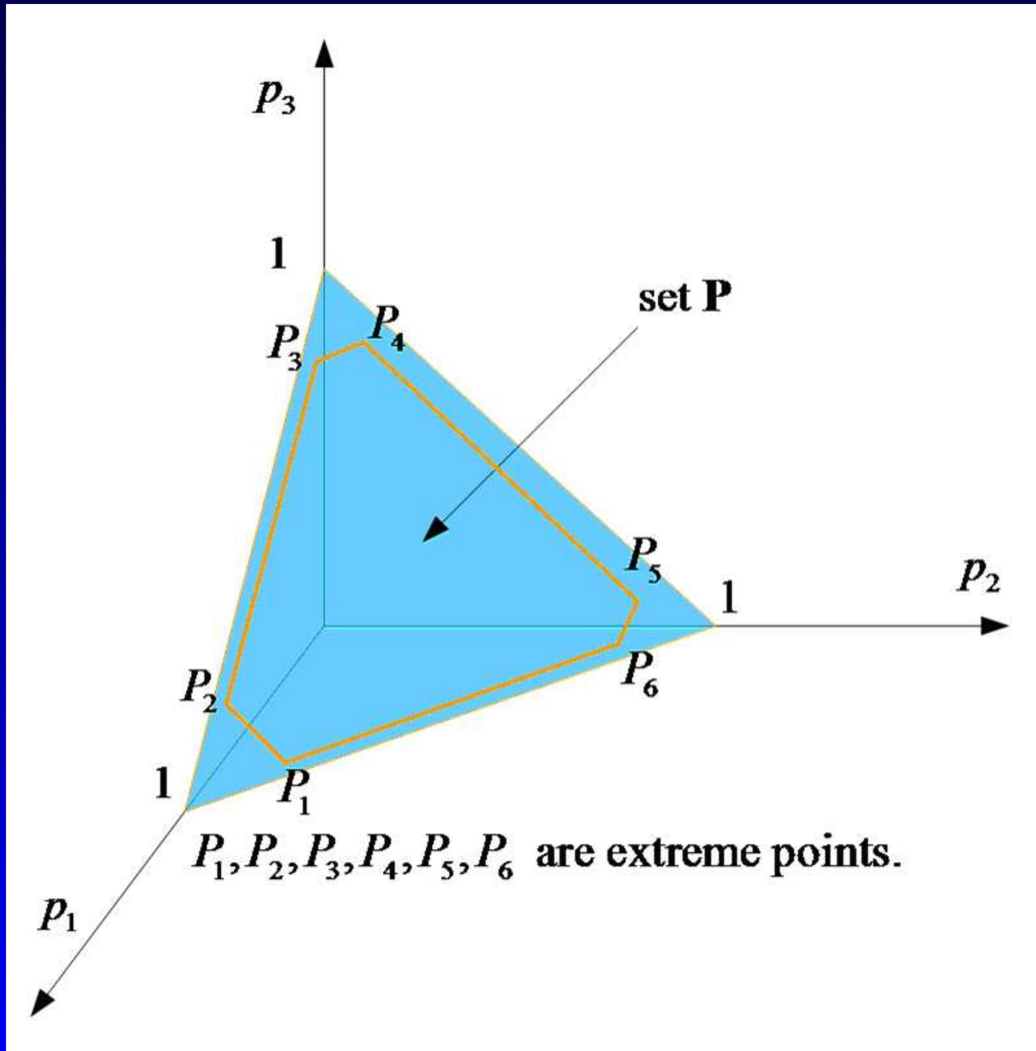
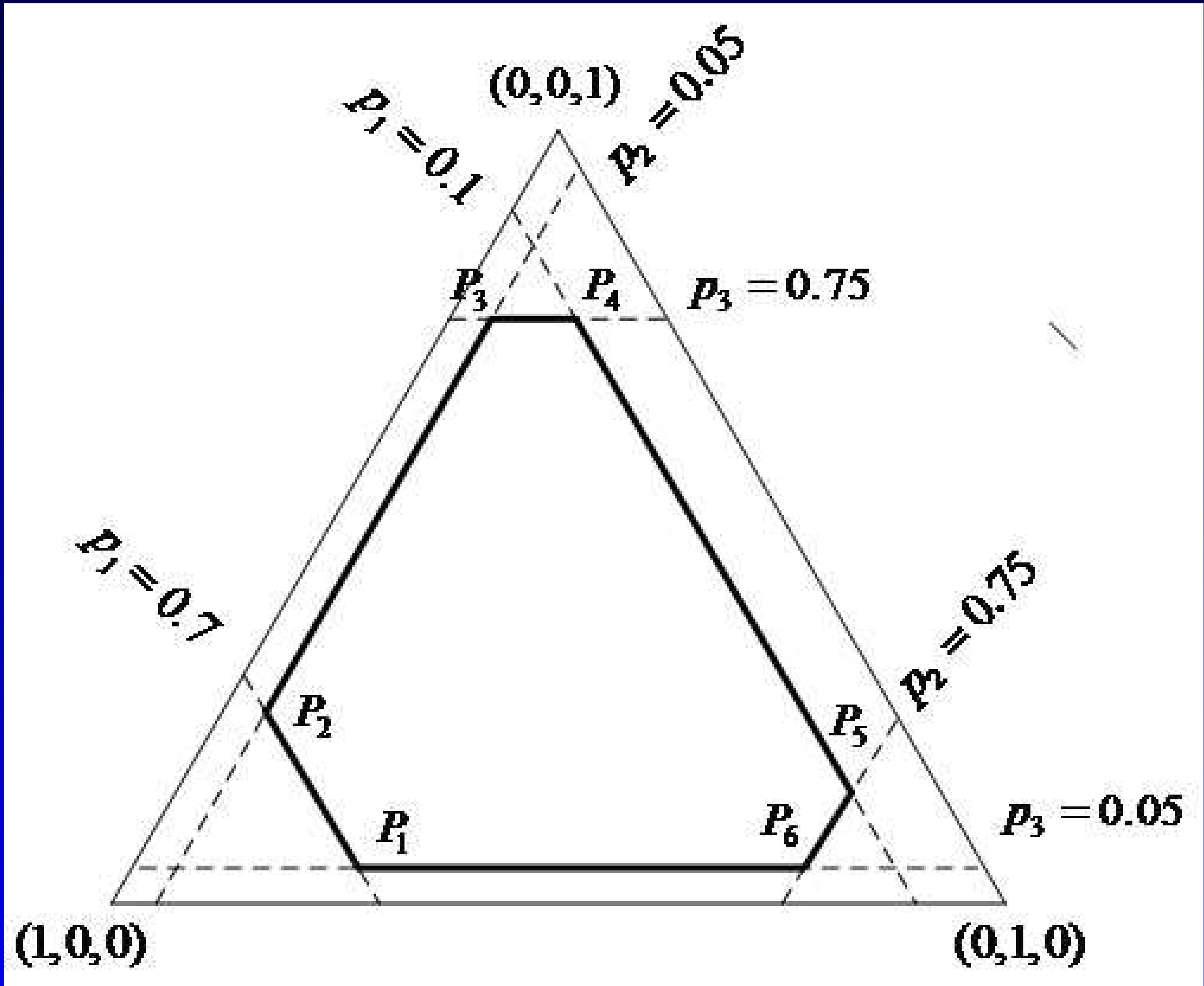


Figure 2: Set of probability measures on plane



The set P is described by the following system of inequalities:

$$\begin{cases} p_1 + p_2 + p_3 = 1, \\ 0.1 \leqslant p_1 \leqslant 0.7, \\ 0.05 \leqslant p_2 \leqslant 0.75, \\ 0.05 \leqslant p_3 \leqslant 0.75, \end{cases}$$

and it can be also represented as a convex closure of probability measures $P_1 = (0.7, 0.25, 0.05)$, $P_2 = (0.7, 0.05, 0.25)$, $P_3 = (0.25, 0.05, 0.75)$, $P_4 = (0.1, 0.15, 0.75)$, $P_5 = (0.1, 0.75, 0.15)$, $P_6 = (0.2, 0.75, 0.05)$.

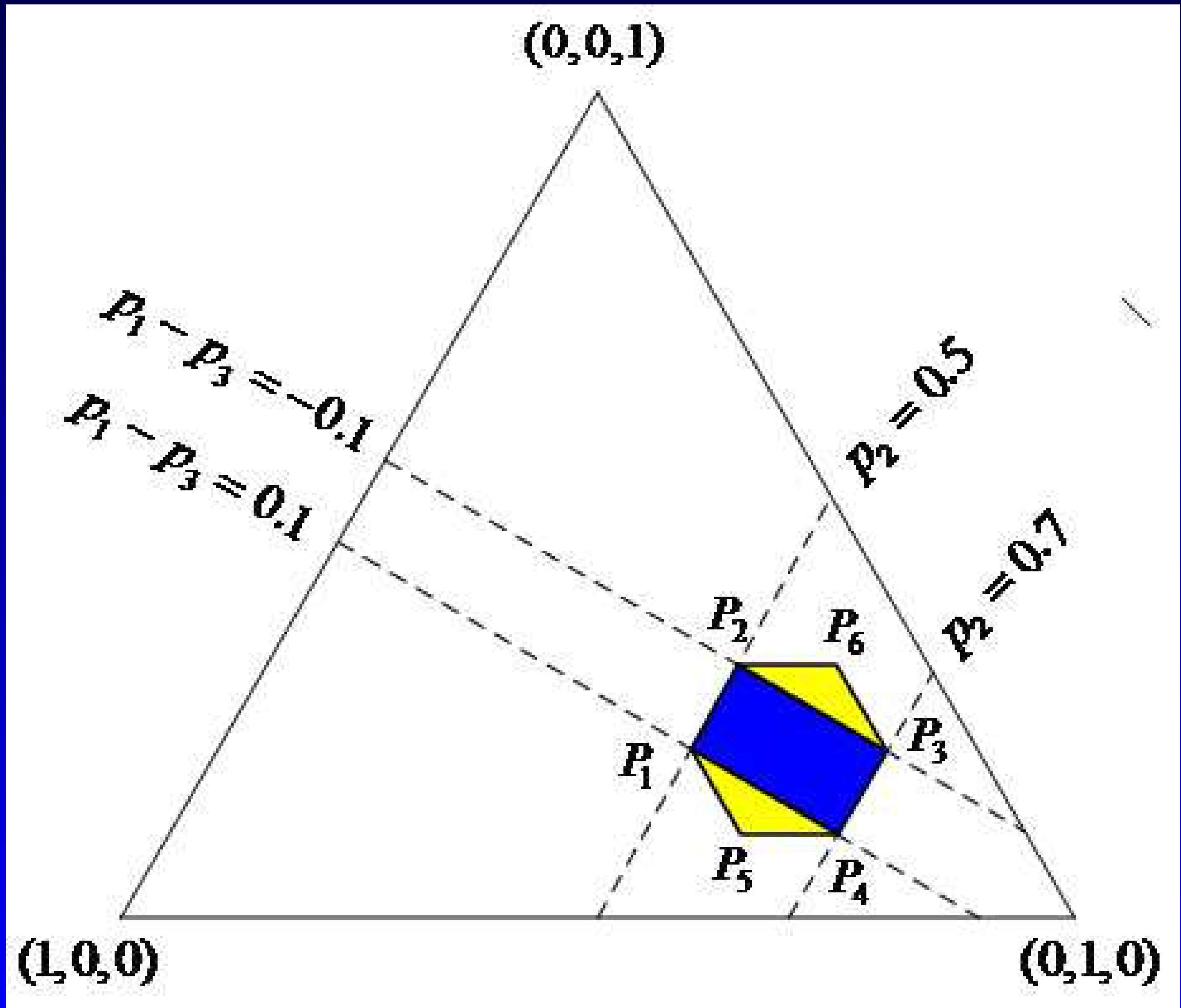
Example 2. It is known that the random value ξ takes its values in the set $\{-1, 0, 1\}$, and $-0.2 \leqslant E[\xi] \leqslant 0.2$, $0.3 \leqslant E[\xi^2] \leqslant 0.5$. What are the set $\mathbf{P}$ of permissible probability measures describing ξ?

Let $\Pr(\xi = -1) = p_1$, $\Pr(\xi = 0) = p_2$, $\Pr(\xi = 1) = p_3$.

Solution. $E[\xi] = -p_1 + p_3$, $E[\xi^2] = p_1 + p_3$. Therefore, the set $\mathbf{P}$ is described by the following system of linear inequalities:

$$\begin{cases} p_1 + p_2 + p_3 = 1, \\ -0.1 \leqslant p_1 - p_3 \leqslant 0.1, \\ 0.3 \leqslant p_1 + p_3 \leqslant 0.5. \end{cases}$$

Figure 3: Set of probability measures on plane



The set $\mathbf{P}$ is depicted on Fig. 3. It easy to see that $\mathbf{P}$ is a blue rectangle on Fig. 3. Its vertices are probability measures $P_1 = (0.3, 0.5, 0.2)$, $P_2 = (0.2, 0.5, 0.3)$, $P_3 = (0.1, 0.7, 0.2)$, $P_4 = (0.2, 0.7, 0.1)$. We can also represent the information about random variable ξ using monotone measures. In this case, measures giving us exact lower and boundaries of probabilities are defined as

$$\underline{\mu}(A) = \min_{i=1,2,3,4} P_i(A),$$

$$\bar{\mu}(A) = \max_{i=1,2,3,4} P_i(A).$$

1	2	3	$\underline{\mu}$	$\bar{\mu}$
0	0	0	0	0
1	0	0	0.1	0.3
0	1	0	0.5	0.7
1	1	0	0.7	0.9
0	0	1	0.1	0.3
1	0	1	0.3	0.5
0	1	1	0.7	0.9
1	1	1	1	1

Table 2: Values of measures $\underline{\mu}$ and $\bar{\mu}$.

Convex sets

Any convex set in our context can be represented as a convex set in $\mathbb{R}^n$.

By definition a set $M \subseteq \mathbb{R}^n$ is *convex* if $\mathbf{x} \in M$ and $\mathbf{y} \in M$ implies $a\mathbf{x} + (1-a)\mathbf{y} \in M$ for any $a \in [0,1]$.

The point $a\mathbf{x} + (1-a)\mathbf{y}$ is called a *convex sum* (or a *convex linear combination*) of $\mathbf{x}$ and $\mathbf{y}$.

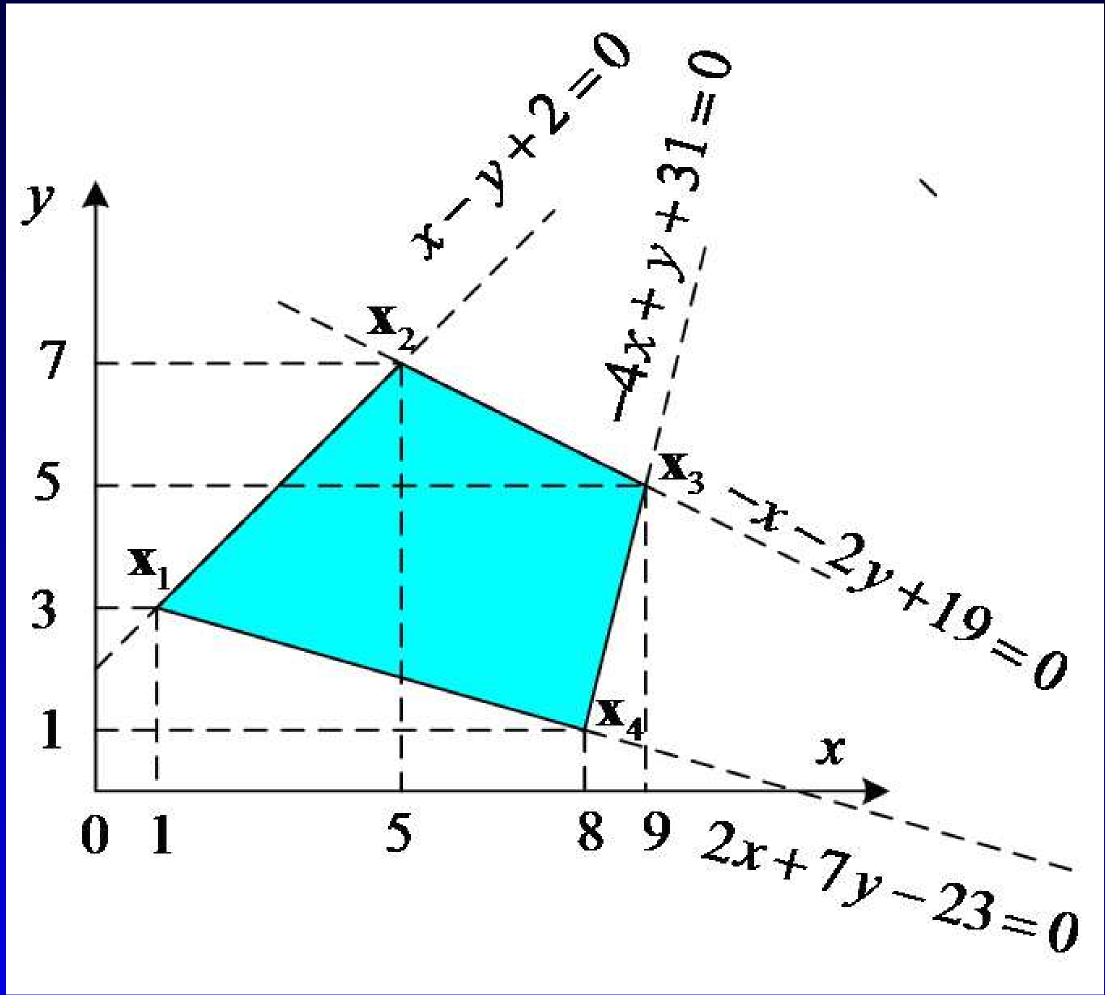
A point $\mathbf{z} \in M$ is called *extreme* for a convex set M if $\mathbf{z}$ cannot be represented as a convex sum

$$\mathbf{z} = a\mathbf{x} + (1-a)\mathbf{y}$$

of any points $\mathbf{x}, \mathbf{y} \in M$ with $\mathbf{x} \neq \mathbf{z}$ and $\mathbf{y} \neq \mathbf{z}$.

On Fig. 1 you can see a convex set in $\mathbb{R}^2$.

Figure 3: An example of a convex set in $\mathbb{R}^n$



In this case $\mathbf{x}_1 = (1, 3)$, $\mathbf{x}_2 = (5, 7)$, $\mathbf{x}_3 = (9, 5)$, $\mathbf{x}_4 = (8, 1)$ are extreme points, and the convex set can be described by a linear system of inequalities:

$$\begin{cases} x - y + 2 \geqslant 0, \\ -x - 2y + 19 \geqslant 0, \\ -4x + y + 31 \geqslant 0, \\ 2x + 7y - 23 \geqslant 0. \end{cases}$$

If the convex set $M \subseteq \mathbb{R}^n$ is closed and bounded and contains a finite number of extreme points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$, then any $\mathbf{x} \in M$ can be represented as a convex sum of $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$, i.e.

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_m\mathbf{x}_m,$$

where $a_i \geqslant 0$, $i = 1, ..., m$ and $\sum_{i=1}^{m} a_i = 1$.

Let us consider our example. In this case, any point $\mathbf{x} \in M$ can be represented as a convex linear combination of extreme points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$. Assume that $\mathbf{x} = (5, 5)$, then $\mathbf{x}$ can be represented as

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + a_4\mathbf{x}_4,$$

where

$$\begin{cases} a_1 = (1/3) - (3/4)a_4, \\ a_2 = (1/3) + (5/4)a_4, \\ a_3 = (1/3) - (3/2)a_4, \\ \quad 0 \leqslant a_4 \leqslant (2/9). \end{cases}$$

It is easy to see that the representation $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_m\mathbf{x}_m$ is not unique in general. For our case, we have different representations changing a_4 in $[0, (2/9)]$.

The structure of monotone measures

Any monotone measure $\mu \in 2^X$ can be considered as a point in $\mathbb{R}^{2^n}$, because values of μ can be represented as a vector

$$(\mu(\emptyset), \mu(\{x_1\}), ..., \mu(\{x_n\}), \mu(\{x_1, x_2\}), ..., \mu(X)) .$$

Therefore, we can define a convex sum of monotone measures μ_1 and μ_2 as

$$\mu = a\mu_1 + (1-a)\mu_2 \text{ for } a \in [0,1] \text{ if}$$

$$\mu(A) = a\mu_1(A) + (1-a)\mu_2(A) \text{ for all } A \in 2^X.$$

It is easy to see that M_{mon} is a convex set.

Question: What are the extreme points of M_{mon}?

Extreme points of M_{mon}

A $\mu \in M_{mon}$ is called $\{0,1\}$-*valued* iff $\mu(A) \in \{0,1\}$ for all $A \in 2^X$.

Theorem. *The set of all extreme points for a convex set M_{mon} consists of $\{0,1\}$-valued measures.*

The description of $\{0,1\}$-valued measures

Algebra 2^X can be considered as a partially ordered set w.r.t. the set-theoretical inclusion.

A subset f of 2^X is *filter* if $A \in$ f and $A \subseteq B$ implies $B \in$ f. By definition, no filter contains $\emptyset$.

Proposition. *Any $\{0,1\}$-valued measure is a charactectic function of some filter.*

A set A is *minimal* in $\mathbf{f}$ if $\{B \in \mathbf{f} | B \subseteq A\} = \{A\}$. Each filter $\mathbf{f}$ is generated by a set of its minimal elements $\{A_1, ..., A_k\}$ in a way

$$\mathbf{f} = \left\{ A \in 2^X \,\middle|\, \exists A_i \subseteq A \right\}.$$

This fact is denoted by $\mathbf{f} = \langle A_1, ..., A_k \rangle$.

The algorithm of finding representation $\mu = \sum_i a_i \eta_i$ of $\mu \in M_{mon}$ through $\{0, 1\}$-valued measures η_i.

0. $\nu := \mu$, $i = 1$.

1. $\eta_i(A) = \begin{cases} 1, & \mu(A) > 0, \\ 0, & \mu(A) = 0, \end{cases}$ $a_i = \min\limits_{A|\mu(A)>0} \mu(A)$.

2. $\mu := \mu - a_i \eta_i$, $i := i + 1$.

3. if $\mu \equiv 0$, then $\mu = \sum_i a_i \eta_i$; else go to 1.

Example. Let P be a probability measure on 2^X, where $X = \{x_1, x_2, x_3\}$, such that $P(\{x_i\}) = 1/3$, $i = 1, 2, 3$. Then

$$P = \frac{1}{3}\eta_{\langle\{x_1\}\rangle} + \frac{1}{3}\eta_{\langle\{x_2\}\rangle} + \frac{1}{3}\eta_{\langle\{x_3\}\rangle}.$$

If we use the algorithm, then

$$P = \frac{1}{3}\eta_{\langle\{x_1\},\{x_2\},\{x_3\}\rangle} + \frac{1}{3}\eta_{\langle\{x_1,x_2\},\{x_1,x_3\},\{x_2,x_3\}\rangle} + \frac{1}{3}\eta_{\langle\{x_2,x_3,x_3\}\rangle}$$

Basic concepts of imprecise probabilities

- Classical probability theory works with single probability measures.
- The theory of imprecise probabilities works with sets of probability measures.

In this lecture we consider probability measures defined on the powerset 2^X of a finite set $X = \{x_1, ..., x_n\}$.

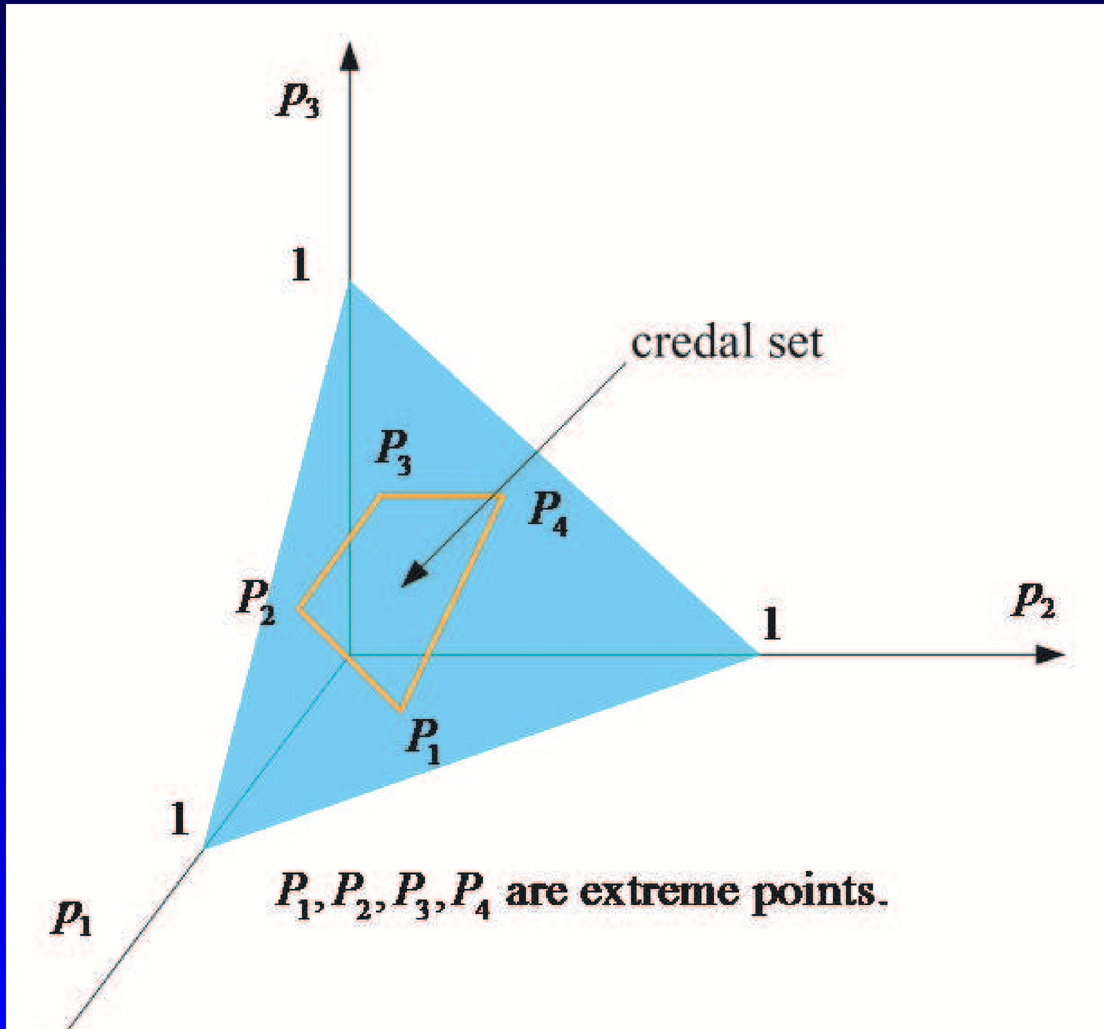
$M_{pr}(X)$ is the set of all probability measures on 2^X.

Credal sets

In this lecture a credal set is understood as a closed convex set of probability measures with a finite number of extreme points. If $\mathbf{P}$ is a credal set and $P_k \in M_{pr}(X)$, $k = 1, ..., m$, are its extreme points then

$$\mathbf{P} = \left\{ \sum_{k=1}^{m} a_i P_i | a_i \geqslant 0, \sum_{k=1}^{m} a_i = 1 \right\}.$$

Let $X = \{x_1, x_2, x_3\}$, then any credal set is convex subset of triangle consisting of points (p_1, p_2, p_3): $p_i \geqslant 0$, $p_1 + p_2 + p_3 = 1$.



Lower probabilities

A monotone measure μ is called a *lower probability* if there is a $P \in M_{pr}$ such that $\mu \leqslant P$.

Any lower probability μ defines a credal set

$$\mathbf{P}(\mu) = \{P \in M_{pr}(X) | P \geqslant \mu\}.$$

Let μ be a lower probability on 2^X, where $X = \{x_1, x_2, x_3\}$, then extreme points of $\mathbf{P}(\mu)$ can be found by solving the following inequalities:

$$\begin{cases} p_1 \geqslant \mu\left(\{x_1\}\right), \\ p_2 \geqslant \mu\left(\{x_2\}\right), \\ p_3 \geqslant \mu\left(\{x_3\}\right), \\ p_1 + p_2 \geqslant \mu\left(\{x_1, x_2\}\right), \\ p_1 + p_3 \geqslant \mu\left(\{x_1, x_3\}\right), \\ p_2 + p_3 \geqslant \mu\left(\{x_2, x_3\}\right), \\ p_1 + p_2 + p_3 = 1. \end{cases}$$

Clearly lower probabilities are less general than credal sets.

Upper probabilities

A monotone measure μ is called an *upper probability* if there is a $P \in M_{pr}$ such that $\mu \geqslant P$.

Any upper probability generate a credal set $\{P \in M_{pr}(X) | P \leqslant \mu\}$.

It is possible to consider only lower probabilities. Let μ be an upper probability. Introduce into consideration a measure $\mu^d(A) = 1 - \mu(A^c)$. The measure μ^d is called dual of μ. Clearly μ^d and μ generate the same credal set

Coherent lower probabilities

A lower probability μ is called a *coherent lower probability* if for any $A \in 2^X$ there is a $P \in M_{pr}$ such that $\mu \leqslant P$ and $\mu(A) = P(A)$.

Any coherent lower probability can be generated as follows: if $\mathbf{P}$ is a credal set then

$$\mu(A) = \min_{P \in \mathbf{P}} P(A),\ A \in 2^X,$$

is a coherent lower probability.

Coherent upper probabilities

An upper probability μ is called a *coherent upper probability* if for any $A \in 2^X$ there is a $P \in M_{pr}$ such that $\mu \geqslant P$ and $\mu(A) = P(A)$.

Any coherent upper probability can be generated as follows: if $\mathbf{P}$ is a credal set then

$$\mu(A) = \max_{P \in \mathbf{P}} P(A),\ A \in 2^X,$$

is a coherent upper probability.

2-monotone measures

A monotone measure is called 2-*monotone* if the following inequality holds:

$$\mu(A) + \mu(B) \leqslant \mu(A \cap B) + \mu(A \cup B).$$

for the dual measure the following inequality holds:

$$\mu^d(A) + \mu^d(B) \geqslant \mu^d(A \cap B) + \mu^d(A \cup B).$$

This measure is called 2-*alternative*. It is known that any 2-monotone measure is a coherent lower probability, and any 2-alternative measure is a coherent upper probability.

Example. Let μ is a lower envelope of probability measures P_1 and P_2 with values

$$P_1(\{x_1\}) = 1/4,\ P_1(\{x_2\}) = 0,\ P_1(\{x_3\}) = 3/4,$$
$$P_1(\{x_4\}) = 0,$$
$$P_2(\{x_1\}) = 0,\ P_2(\{x_2\}) = 1/2,\ P_2(\{x_3\}) = 0,$$
$$P_2(\{x_4\}) = 1/2,$$

i.e. $\mu(A) = \min\limits_{i=1,2} P_i(A)$. Then

$$\underbrace{\mu(\{x_1, x_4\})}_{1/4} + \underbrace{\mu(\{x_3, x_4\})}_{1/2} > \underbrace{\mu(\{x_4\})}_{0} + \underbrace{\mu(\{x_1, x_3, x_4\})}_{1/2}.$$

Therefore, μ is a coherent lower probability, but it is not 2-monotone.

Properties of 2-monotone measures

Theorem 1. A $\mu \in M_{mon}$ on 2^X ($|X| = n$) is 2-monotone if for any $B \in 2^X$ ($|B| \leqslant n-2$) and any $x, y \in X \backslash B$ ($x \neq y$) the following inequality is fulfilled:
$\mu(B \cup \{x\}) + \mu(B \cup \{y\}) \leqslant \mu(B \cup \{x, y\}) + \mu(B)$.

Theorem 2. Let $\mu \in M_{mon}$, $|X| = n$, and let $\{B_0, B_1, ..., B_n\}$ be a complete chain in 2^X, i.e. $\emptyset = B_0 \subset B_1 \subset ... \subset B_n = X$ and $|B_{i+1} \backslash B_i| = 1$, $i = 0, ..., n-1$.
Then a $P \in M_{pr}$ with $P(B_i) = \mu(B_i)$, $i = 0, ..., n$, is an extreme point of $\mathbf{P}(\mu)$ and any extreme point of $\mathbf{P}(\mu)$ is described as above.

Example. Consider a 2-monotone measure Bel, which we got for describing a public opinion poll. Let us find the extreme points of Bel, using Theorem 2:

if $B_1 = \{1\}$, $B_2 = \{2, 3\}$, $B_3 = \{1, 2, 3\}$, then

$$P_1(\{1\}) = \mu(B_1) = 0.1,$$
$$P_1(\{2\}) = \mu(B_2) - \mu(B_1) = 0.25 - 0.1 = 0.15,$$
$$P_1(\{3\}) = \mu(B_3) - \mu(B_2) = 1 - 0.25 = 0.75.$$

if $B_1 = \{2\}$, $B_2 = \{2, 3\}$, $B_3 = \{1, 2, 3\}$, then

$$P_2(\{2\}) = \mu(B_2) = 0.05,$$
$$P_2(\{3\}) = \mu(B_2) - \mu(B_1) = 0.2 - 0.05 = 0.15,$$
$$P_2(\{1\}) = \mu(B_3) - \mu(B_2) = 1 - 0.2 = 0.8.$$

$P_1 = (0.1, 0.15, 0.75)$, $P_2 = (0.8, 0.05, 0.15)$.

Conditioning of coherent lower probabilities

Let $P \in M_{pr}$ and an event B occurred. Then the probability of A given B is calculated as

$$P(A|B) = P(A \cap B)/P(B).$$

Let us denote the conditional probability measure P_B.

In case of imprecise probabilities we have a credal set $\mathbf{P}$ and the same conditioning leads to a credal set

$$\mathbf{P}_B = \{P_B | P \in \mathbf{P}\}.$$

Let μ be a coherent lower probability. Then the conditional measure μ_B can be calculated as

$$\mu_B(A) = \inf \{P_B(A) | P \in \mathbf{P}(\mu)\}.$$

Conditioning of 2-monotone measures

Theorem. Let μ be 2-monotone. Then

$$\mu_B(A) = \frac{\mu(A \cap B)}{\mu(A \cap B) + \mu^d(B \backslash A)}.$$

Proof. By definition

$$\mu_B(A) = \inf_{P \geqslant \mu} \frac{P(A \cap B)}{P(A \cap B) + P(B \cap \bar{A})}. \qquad (1)$$

Notice that the function $f(\alpha, \beta) = \frac{\alpha}{\alpha+\beta}$ is increasing w.r.t. α and decreasing w.r.t. β. Therefore, in the right part of (1) we must choose the smallest α and the biggest β.

It is so, if $P(A \cap B) = \mu(A \cap B)$ and $P(B \cap \bar{A}) = \mu^d(B \cap \bar{A})$. Therefore,

$$\mu_B(A) \geqslant \frac{\mu(A \cap B)}{\mu(A \cap B) + \mu^d(B \backslash A)}.$$

Let us show that $P \in \mathbf{P}(\mu)$ with $P(A \cap B) = \mu(A \cap B)$ and $P(B \cap \bar{A}) = \mu^d(B \cap \bar{A})$ exists.

Notice that $\mu^d(B \cap \bar{A}) = 1 - \mu(A \cup \bar{B})$ and $A \cap B \subseteq A \cup \bar{B}$. Thus, 2-monotonicity of μ implies that $\exists P \in \mathbf{P}(\mu)$ with $P(A \cap B) = \mu(A \cap B)$ and $P(A \cup \bar{B}) = \mu(A \cup \bar{B})$. $\blacksquare$

Upper and lower expectations (previsions)

Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set and P is a probability measure on 2^X. In classical probability theory a random value ξ is a mapping $\xi : X \to \mathbb{R}$ and its expectation is defined by

$$E_P[\xi] = \sum_{i=1}^{n} \xi(x_i) P(\{x_i\}).$$

If probabilities are defined imprecise, then instead of a probability measure we have a credal set $\mathbf{P}$ and we can define lower and upper expectations using formulas:

$$\underline{E}_{\mathbf{P}}[\xi] = \inf_{P \in \mathbf{P}} E_P[\xi], \; \bar{E}_{\mathbf{P}}[\xi] = \sup_{P \in \mathbf{P}} E_P[\xi].$$

Clearly, the functionals $\underline{E}_{\mathbf{P}}$ and $\bar{E}_{\mathbf{P}}$ can be defined on the set $\mathcal{F}$ of all bounded functions on X.

Notation:

$\mathbf{0}$ is a function on X that is equal to zero;
$\mathbf{1}$ is a function on X that is equal to one.

Theorem. The functionals $\underline{E}_{\mathbf{P}}$ and $\bar{E}_{\mathbf{P}}$ are lower and upper expectations of some non-empty credal set $\mathbf{P}$ iff

1. $\underline{E}_{\mathbf{P}}[\mathbf{0}] = 0$, $\underline{E}_{\mathbf{P}}[\mathbf{1}] = 1$;
2. $\underline{E}_{\mathbf{P}}[af + c] = a\underline{E}_{\mathbf{P}}[f] + c$, $a, c \in \mathbb{R}$, $f \in \mathcal{F}$;
3. $\underline{E}_{\mathbf{P}}[f_1] + \underline{E}_{\mathbf{P}}[f_2] \leqslant \underline{E}_{\mathbf{P}}[f_1 + f_2]$ for any $f_1, f_2 \in \mathcal{F}$;
4. $\bar{E}_{\mathbf{P}} = -\underline{E}_{\mathbf{P}}(-f)$ for any $f \in \mathcal{F}$.

Upper and lower expectations for 2-monotone measures

Lemma. Let $P \in M_{pr}$ and let f a be non-negative function on X. Then

$$E_P[f] = \int\limits_0^{\infty} P(\{x \in X | f(x) > t\})\,dt \quad (1)$$

Proof. Let $\chi(y) = \begin{cases} 1, & y > 0, \\ 0, & y \leqslant 0. \end{cases}$ Then the right part of (1) can be transformed to

$$\int\limits_0^{\infty} \left(\int\limits_X \chi(f(x) - t) dP \right) dt.$$

Changing the order of integration we get

$$\int_X \left(\int_0^\infty \chi(f(x) - t)dt \right) dP = \int_X f(x)dP = E_P[f].$$

Theorem 1. Let $\mu \in M_{2-mon}$ and let f be a non-negative function on X. Then

$$\underline{E}_{\mathbf{P}(\mu)}[f] = \int_0^\infty \mu(\{x \in X | f(x) > t\})\, dt.$$

Proof. Clearly,

$$\underline{E}_{\mathbf{P}(\mu)}[f] = \inf_{P \in \mathbf{P}(\mu)} \int_0^\infty P(\{x \in X | f(x) > t\})\, dt \geqslant$$

$$\int_0^\infty \mu(\{x \in X | f(x) > t\})\, dt.$$

Since sets $\{x \in X | f(x) > t\}$ form a chain and $\mu \in M_{2-mon}$, $\exists P \in \mathbf{P}(\mu)$ with

$$\mu\left(\{x \in X | f(x) > t\}\right) = P\left(\{x \in X | f(x) > t\}\right).$$

for all $t \in [0, +\infty)$. This implies the result.

Theorem 2. Let $\mu \in M_{2-mon}$ and $f \in \mathcal{F}$. Let $f = f_+ - f_-$, where

$$f_+(x) = \begin{cases} f(x), & x > 0, \\ 0, & x \leqslant 0, \end{cases}$$

Then

$$\underline{E}_{\mathbf{P}(\mu)}[f] = \int\limits_0^\infty \mu\left(\{x \in X | f_+(x) > t\}\right) dt - \int\limits_0^\infty \mu^d\left(\{x \in X | f_-(x) > t\}\right) dt.$$

Choquet Integral

Let $\mu \in M_{mon}$, then the Choquet integral is a functional defined by

$$(Choquet)\int_X f(x)d\mu =$$

$$\int_0^\infty \mu\left(\{x \in X | f_+(x) > t\}\right) dt -$$

$$\int_0^\infty \mu^d\left(\{x \in X | f_-(x) > t\}\right) dt.$$

Definition. Functions $f_1, f_2 \in \mathcal{F}$ are called *comonotonic* if $f_1(x) \leqslant f_1(y)$ implies $f_2(x) \leqslant f_2(y)$ for any $x, y \in X$.

Properties of Choquet Integral

1. $\int\limits_X 1_A(x)d\mu = \mu(A)$.
2. $\int\limits_X cf(x)d\mu = c\int\limits_X f(x)d\mu$, $c \in \mathbb{R}$.
3. $\int\limits_X -f(x)d\mu = -\int\limits_X f(x)d\mu^d$.
4. $\int\limits_X (f(x)+c)d\mu = \int\limits_X f(x)d\mu + c$, $c \in \mathbb{R}$.
5. $f(x) \leqslant g(x)$ for all $x \in X$ implies $\int\limits_X f(x)d\mu \leqslant \int\limits_X g(x)d\mu$.
6. if f and g are comonotonic, then $\int\limits_X (f(x)+g(x)d\mu = \int\limits_X f(x)d\mu + \int\limits_X g(x)d\mu$.

7. if $\mu \in M_{2-mon}$, then $\underline{E}_{\mathbf{P}(\mu)}[f] = \int\limits_X f(x)d\mu$ and $\bar{E}_{\mathbf{P}(\mu)}[f] = \int\limits_X f(x)d\mu$.

8. $\int\limits_X f(x)d(a\mu_1 + b\mu_2) = a\int\limits_X f(x)d\mu_1 + b\int\limits_X f(x)d\mu_2$, $a, b \geq 0$, $a + b = 1$.

Theorem. A functional F on $\mathcal{F}$ is a Choquet integral for some $\mu \in M_{mon}$ iff

1) $F(1_X) = 1$;
2) $F(c1_A) = cF(1_A)$ for any $c \in \mathbb{R}$;
3) $f \leqslant g$ implies $F(f) \leqslant F(g)$;
4) if f and g are comonotonic, then $F(f+g) = F(f) + F(g)$.

Computing of Choquet Integral

Let $f \in \mathcal{F}$ and we order the elements of X such that

$$f\left(x_{i_1}\right) \geqslant f\left(x_{i_2}\right) \geqslant ... \geqslant f\left(x_{i_n}\right).$$

Let us consider the sequence of sets

$$B_0 = \emptyset,\, B_1 = \left\{x_{i_1}\right\}, B_1 = \left\{x_{i_1}, x_{i_2}\right\}, ...,$$
$$B_n = \left\{x_{i_1}, ..., x_{i_n}\right\} = X.$$

Then

$$\int\limits_X f(x) d\mu = \sum_{k=1}^{n} f\left(x_{i_k}\right)\left(\mu(B_k) - \mu(B_{k-1})\right).$$

Example. $X = \{x_1, x_2, x_3\}$, $f(x_1) = 0.25$, $f(x_2) = 1$, $f(x_3) = 0.5$.

$$f(x_2) > f(x_3) > f(x_1).$$

$$B_1 = \{x_2\},\ B_2 = \{x_2, x_3\},\ B_3 = \{x_1, x_2, x_3\}.$$

$$Bel(B_1) = 0.05,\ Bel(B_2) = 0.3,\ Bel(B_3) = 1.$$

$$Pl(B_1) = 0.75,\ Pl(B_2) = 0.9,\ Pl(B_3) = 1.$$

$$\int_X f(x)dBel = 1 \cdot 0.1 + 0.5 \cdot 0.25 + 0.25 \cdot 0.7 = 0.4.$$

$$\int_X f(x)dPl = 1 \cdot 0.75 + 0.5 \cdot 0.25 + 0.25 \cdot 0.05 = 0.8875.$$

Sugeno Integral

Let $\mu \in M_{mon}$, then the Sugeno integral is a functional on the set of all functions $f : X \to [0,1]$ defined by

$$(Sug)\int_X f(x)d\mu = \bigvee_{t\in[0,1]} (t \wedge \mu(\{x \in X | f(x) > t\}))\,.$$

Theorem. A functional F on the set of all functions $f : X \to [0,1]$ is a Sugeno integral for some $\mu \in M_{mon}$ iff it satisfies the following properties:

1. $F(1_X) = 1$.
2. $F(c \wedge 1_A) = c \wedge F(1_A)$.
3. $f \leqslant g$ implies $F(f) \leqslant F(g)$.
4. if f and g are comonotonic, then $F(f \vee g) = F(f) \vee F(g)$.

Computing of Sugeno Integral

Let $f \in \mathcal{F}$ and we order the elements of X such that

$$f\left(x_{i_1}\right) \geqslant f\left(x_{i_2}\right) \geqslant \ldots \geqslant f\left(x_{i_n}\right).$$

Let us consider the sequence of sets

$$B_1 = \left\{x_{i_1}\right\}, B_1 = \left\{x_{i_1}, x_{i_2}\right\}, \ldots.,$$
$$B_n = \left\{x_{i_1}, \ldots, x_{i_n}\right\} = X.$$

Then $f = \bigvee\limits_{k=1}^{n} \left(f(x_{i_k}) \wedge 1_{B_k}\right)$ and

$$\int\limits_X f(x) d\mu = \bigvee\limits_{k=1}^{n} \left(f(x_{i_k}) \wedge \mu(B_k)\right).$$

Example. $X = \{x_1, x_2, x_3\}$, $f(x_1) = 0.25$, $f(x_2) = 1$, $f(x_3) = 0.5$.

$$f(x_2) > f(x_3) > f(x_1).$$

$$B_1 = \{x_2\},\ B_2 = \{x_2, x_3\},\ B_3 = \{x_1, x_2, x_3\}.$$

$$Bel(B_1) = 0.05,\ Bel(B_2) = 0.3,\ Bel(B_3) = 1.$$

$$Pl(B_1) = 0.75,\ Pl(B_2) = 0.9,\ Pl(B_3) = 1.$$

$$\int_X f(x)dBel = (1 \wedge 0.1) \vee (0.5 \wedge 0.3) \vee (0.25 \cdot 1) = 0.3.$$

$$\int_X f(x)dPl = (1 \wedge 0.75) \vee (0.5 \wedge 0.9) \vee (0.25 \wedge 1) = 0.75.$$

Belief and plausibility measures

Belief and plausibility measures are defined by means of a basic probability assignment. A basic probability assignment m is a non-negative set function on 2^X such that

1. $m(\emptyset) = 0$;
2. $\sum_{A \in 2^X} m(A) = 1$ (norming).

Then

$$Bel(A) = \sum_{B \subseteq A} m(B) \text{ and } Pl(B) = \sum_{B \cap A \neq \emptyset} m(A).$$

The set A is called focal if $m(A) > 0$.

Some times, it is useful to represent belief functions using $\{0,1\}$-valued measures:

$$\eta_{\langle B\rangle}(A)=\begin{cases}1, & B\subseteq A,\\ 0, & otherwise.\end{cases}$$

Then

$$Bel(A)=\sum_{B\in 2^X} m(B)\eta_{\langle B\rangle}(A).$$

The sense of $\eta_{\langle B\rangle}$ is the following. It describes the situation when we know that the random variable definitely takes values from the set B, but we don't know any additional information.

Clearly, $Pl=Bel^d$.

Dempster-Shafer theory

Let we have an experiment in which we cannot fix elementary events and the only information is that some event occurred (say $B \in 2^X$). In this case a monotone measure

$$\eta_{\langle B \rangle}(A) = \begin{cases} 1, & B \subseteq A, \\ 0, & otherwise. \end{cases}$$

describes events that occured necessarily, a measure

$$\eta^d_{\langle B \rangle}(A) = \begin{cases} 1, & A \cap B \neq \emptyset, \\ 0, & A \cap B = \emptyset, \end{cases}$$

describes events that occurred possibly.

Let $m(B)$ be a frequency of fixing $B \in 2^X$. Then a measure

$$Bel(B) = \sum_{B \in 2^X} m(B)\eta(B)$$

gives an exact lower bound of the probability of event B, and a measure

$$Pl(B) = \sum_{B \in 2^X} m(B)\eta^d(B)$$

gives an exact upper bound of the probability of event B.

Possibility and necessity measures

A possibility measure Pos is such that $Pos \in M_{mon}$,

$$Pos(A \cup B) = \max\{Pos(A), Pos(B)\} \; A, B \in 2^X.$$

A necessity measure Nec is such that $Nec \in M_{mon}$,

$$Nec(A \cap B) = \min\{Nec(A), Nec(B)\} \; A, B \in 2^X.$$

The dual of a necessity measure is a possibility measure. Any necessity measure is a belief measure. A belief measure is a necessity measure if focal elements form a chain.

Möbius transform

The set of all set functions on 2^X is a linear space and the system of set functions $\left\{\eta_{\langle B\rangle}\right\}_{B\in 2^X}$ is the basis of it. We can find the representation

$$\mu = \sum_{B\in 2^X} m(B)\eta_{\langle B\rangle}$$

of any $\mu : 2^X \to \mathbb{R}$ using the Möbius transform:

$$m(B) = \sum_{A\subseteq B} (-1)^{|B\setminus A|}\mu(A).$$

The representation of the Choquet integral through the Möbius transform

Let m be the Möbius transform of $\mu \in M_{mon}$. Then

$$\mu = \sum_{B \in 2^X} m(B) \eta_{\langle B \rangle},$$

$$(Ch) \int_X f(x) d\mu = \sum_{B \in 2^X} m(B) \int_X f(x) \eta_{\langle B \rangle},$$

Since $\int_X f(x) d\eta_{\langle B \rangle} = \bigwedge_{x_i \in B} f(x_i)$,

$$(Ch) \int_X f(x) d\mu = \sum_{B \in 2^X} m(B) \left(\bigwedge_{x_i \in B} f(x_i) \right).$$

Decision theory

Assume that you decided to invest money for election campaign and your gain depends on the candidate $i = 1, 2, 3$, who will win. These gains are described in the following table.

	State 1	State 2	State 3
Act 1	$u(A_1\|S_1)$	$u(A_1\|S_2)$	$u(A_1\|S_3)$
Act 2	$u(A_2\|S_1)$	$u(A_2\|S_2)$	$u(A_2\|S_3)$
Act 3	$u(A_3\|S_1)$	$u(A_3\|S_2)$	$u(A_3\|S_3)$

Act i (A_i), $i = 1, 2, 3$ means that you choose to invest the campaign of candidate i, and State i (S_i) means that candidate i will win.

Let us assume that information about future elections is presented by probabilities $P(S_i)$. Then you choose the act A_i with the highest expected gain:

$$u(A_i) = \sum_{k=1}^{3} u(A_i|S_k)P(S_k).$$

Imprecise probability model

If your beliefs cannot be expressed by a probability measure, we can add uncertainty assuming that the future elections results are described by a credal set $\mathbf{P}$. Then we have lower and upper bounds of the expected gain:

$$\underline{u}(A_i) = \inf_{P \in \mathbf{P}} \sum_{k=1}^{3} u(A_i|S_k)P(S_k),$$
$$\bar{u}(A_i) = \sup_{P \in \mathbf{P}} \sum_{k=1}^{3} u(A_i|S_k)P(S_k).$$

In this case, if $\underline{u}(A_i) \geqslant \underline{u}(A_j)$ and $\bar{u}(A_i) \geqslant \bar{u}(A_j)$, $j = 1, 2, 3$, you should definitely choose the act A_i. In other situations it depends on behavior of a decision maker. For example, the cautious behavior means that he choose act A_i with $\underline{u}(A_i) \geqslant \underline{u}(A_j)$, $j = 1, 2, 3$.

Multi-criteria decision making

In this case each act A_i is characterized by a vector $(u_1, ..., u_n)$ of criteria utilities. It is assumed that $u_i \in [0, 1]$, $i = 1, ..., n$. For decision making it is necessary to aggregate criteria in a one global criteria with the utility $u = \varphi(u_1, ..., u_n)$, where $\varphi : [0, 1]^n \to [0, 1]$ is an aggregation function with the following properties:

1) $\varphi\,(0, ..., 0) = 0$, $\varphi\,(1, ..., 1) = 1$;

2) $\mathbf{x} \leqslant \mathbf{y}$ implies $\varphi(\mathbf{x}) \leqslant \varphi(\mathbf{y})$.

It is well known that the aggregation function based on simple average

$$u = \sum_{i=1}^{n} a_i u_i, \text{ where } a_i \geqslant 0, \sum_{i=1}^{n} u_i = 1,$$

is not good if criteria interact to each other. Therefore, aggregation functions based on Choquet integral or Sugeno integral are used.

Choice of aggregation function

Given a learning sample $(\mathbf{x}_1, ..., \mathbf{x}_N)$, for which we know that for ideal aggregation function φ it should be fulfilled

$$\varphi(\mathbf{x}_i) < \varphi(\mathbf{x}_j) \text{ if } i < j.$$

Denote by $\varphi_\mu(f) = \int f d\mu$ the aggregation function based on Choquet integral or Sugeno integral. Then the optimal choice of aggregation function φ_μ is connected with seeking a monotone measure μ that properly classifies vectors in a sense that

$$\varphi_\mu(\mathbf{x}_i) < \varphi_\mu(\mathbf{x}_j) \text{ if } i < j.$$

It is easy to show that for the case of Choquet integral, this leads to solving the system of linear inequalities.

Image processing

Let we have a set of n classifiers that give us the answer that on the picture it is depicted a building or not. Assume that this answer they give by numbers

$$u_i,\ i = 1, ..., n, \text{ in } [0, 1]:$$

- if $u_i = 1$, then classifier i definitely says that there is a building on the picture;
- if $u_i = 0$ then there is no.

The problem is how to aggregate information from the set of all classifiers. This can be done by aggregation functions based on Choquet integral or Sugeno integral.

Solution

Given a learning sample $(\mathbf{x}_1, ..., \mathbf{x}_N)$, for which we know the right answer for any $\mathbf{x}_i$. Assume that

- $c(\mathbf{x}_i) = 1$ if the picture i contains a building;
- $c(\mathbf{x}_i) = -1$ otherwise.

Then the aggregation function φ_μ can be chosen such that it minimizes the number of false inequalities:

$$c(\mathbf{x}_i)(\varphi_\mu(\mathbf{x}_i) - 0.5) > 0, i = 1, ..., N.$$

Information measures in image processing

Let images be characterized by a set of features $X = \{x_1, x_2, ..., x_n\}$. Sometimes, these features are redundant, and it is necessary to choose a subset of X that characterize the image with the sufficient precision. For this purpose, the information measure $\mu : 2^X \to [0, +\infty)$ is introduced that reflects the amount of information of any subset $A \subseteq X$. Clearly, it should have the following properties:

1. $\mu(\emptyset) = 0$;
2. $A \subseteq B$ implies $\mu(A) \subseteq \mu(B)$.

An information measure based on the Shannon entropy

Assume that we can describe the occurrence of images and their features by a random vector $\xi = (\xi_1, ..., \xi_n)$. Denote a random vector consisting of random variables ξ_i, $i \in A$, by ξ_A. Then we introduce the information measure

$$\mu(A) = S\,(\xi_A),$$

where S is the Shannon entropy.

Proposition. The information measure based on the Shannon entropy is 2-monotone.

Information measures of polygonal representations

In image processing we need to analyze closed contours.

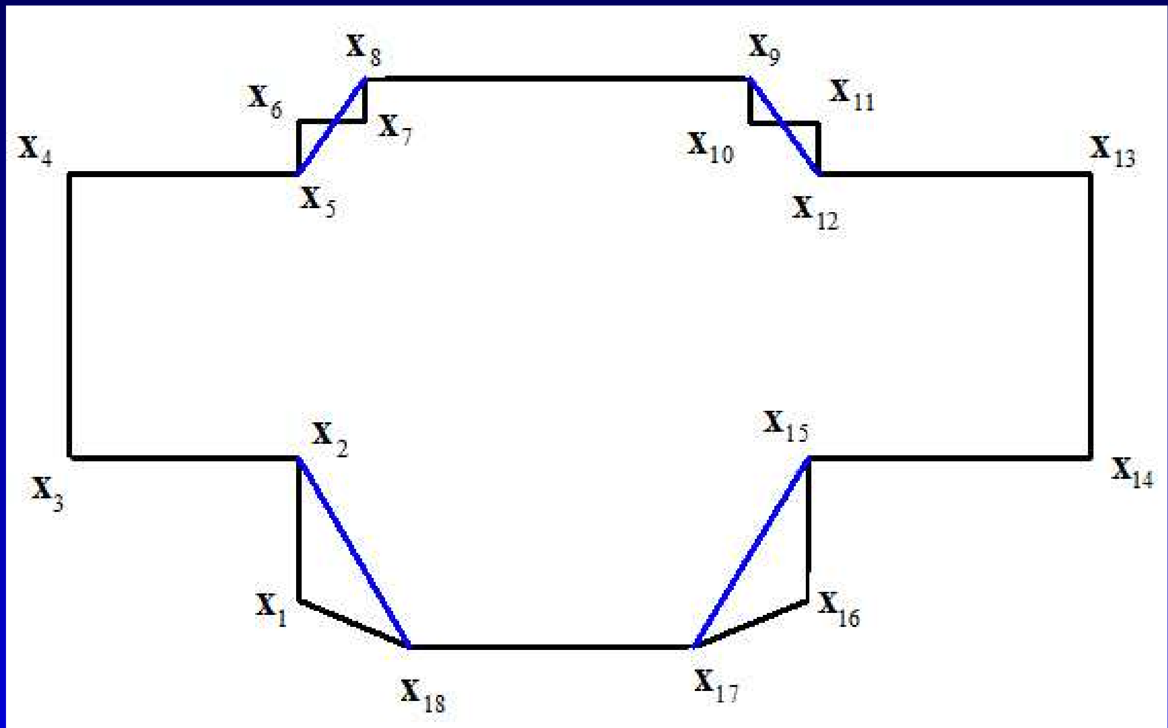
Their simplest representations are polygons.

Any polygon can be represented as a ordered set $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ of its vertices.

After contour extraction, each contour has a huge number of vertices.

Problem: how to reduce a number of vertices? $\Rightarrow$ to find subcontour B of X (see Fig. 1) that preserve information about the initial contour B.

Figure 1: Contour $X = \{\mathbf{x}_1, ..., \mathbf{x}_{18}\}$ and contour $B = \{\mathbf{x}_2, ..., \mathbf{x}_5, \mathbf{x}_8, \mathbf{x}_9, \mathbf{x}_{12}, ..., \mathbf{x}_{15}, \mathbf{x}_{17}, \mathbf{x}_{18}\}$



For choosing the optimal representation we can use information measures.

One way to use information measure based on contour length. It is defined as

$$\mu_L(B) = \sum_{i=0}^{m} |\mathbf{y}_i - \mathbf{y}_{i-1}|,$$

where $B = \{y_1, y_2, ..., y_m\}$ and $y_0 = y_m$.

The choice of optimal contour

Let $\mathcal{A}$ is the set of admissible polygonal representations.

For example, it can be a set of contours, in which a number of vertices is lower or equal than n (n is a parameter).

Then an $\mathcal{A}$-optimal contour B_{opt} is a solution of the following optimization problem:

$$\mu_L(B_{opt}) = \max_{A \in \mathcal{A}} \mu_L(A).$$