Frameworks of Imprecision and Uncertainty

Problems with Probability Theory

Representation of Ignorance (dt. Unwissen)

- We are given a die with faces $1, \ldots, 6$
  What is the certainty of showing up face i ?
  - Conduct a statistical survey (roll the die 10000 times) and estimate the relative frequency: $P(\{i\}) = \frac{1}{6}$
  - Use subjective probabilities (which is often the normal case): We do not know anything (especially and explicitly we do not have any reason to assign unequal probabilities), so the most plausible distribution is a uniform one.

  ⇒ Problem: Uniform distribution because of ignorance or extensive statistical tests

- Experts analyze aircraft shapes: 3 aircraft types A, B, C
  "It is type A or B with 90% certainty. About C, I don't have any clue and I do not want to commit myself. No preferences for A or B."

⇒ Problem: Propositions hard to handle with Bayesian theory

Modeling Imprecise Data

"$A \subseteq X$ being an imprecise date" means: the true value x_0 lies in A but there are no preferences on A.

Ω	set of possible elementary events
$\Theta = \{\xi\}$	set of observers
$\lambda(\xi)$	importance of observer ξ

Some elementary event from Ω occurs and every observer $\xi \in O$ shall announce which elementary events she personally considers possible. This set is denoted by $\Gamma(\xi) \subseteq \Omega$. $\Gamma(\xi)$ is then an imprecise date.

$\lambda : 2^\Theta \to [0, 1]$	probability measure (interpreted as importance measure)
$(\Theta, 2^\Theta, \lambda)$	probability space
$\Gamma : \Theta \to 2^\Omega$	set-valued mapping

Imprecise Data (2)

Let $A \subseteq \Omega$:

a) $\Gamma^*(A) \stackrel{\text{Def}}{=} \{\xi \in \Theta \mid \Gamma(\xi) \cap A \neq \emptyset\}$

b) $\Gamma_*(A) \stackrel{\text{Def}}{=} \{\xi \in \Theta \mid \Gamma(\xi) \neq \emptyset \text{ and } \Gamma(\xi) \subseteq A\}$

Remarks:

a) If $\xi \in \Gamma^*(A)$, then it is *plausible* for ξ that the occurred elementary event lies in A.

b) If $\xi \in \Gamma_*(A)$, then it is *certain* for ξ that the event lies in A.

c) $\{\xi \mid \Gamma(\xi) \neq \emptyset\} \;=\; \Gamma^*(\Omega) \;=\; \Gamma_*(\Omega)$

Let $\lambda(\Gamma^*(\Omega)) > 0$. Then we call

$$P^*(A) = \frac{\lambda(\Gamma^*(A))}{\lambda(\Gamma^*(\Omega))} \quad \text{the upper, and} \quad P_*(A) = \frac{\lambda(\Gamma_*(A))}{\lambda(\Gamma_*(\Omega))} \quad \text{the lower}$$

probability w. r. t. λ and Γ.

Example

$$\Theta = \{a, b, c, d\} \qquad \lambda: \ a \mapsto {}^1\!/_6 \qquad \Gamma: \ a \mapsto \{1\}$$
$$\Omega = \{1, 2, 3\} \qquad b \mapsto {}^1\!/_6 \qquad b \mapsto \{2\}$$
$$\Gamma^*(\Omega) = \{a, b, d\} \qquad c \mapsto {}^2\!/_6 \qquad c \mapsto \emptyset$$
$$\lambda(\Gamma^*(\Omega)) = {}^4\!/_6 \qquad d \mapsto {}^2\!/_6 \qquad d \mapsto \{2, 3\}$$

A	$\Gamma^*(A)$	$\Gamma_*(A)$	$P^*(A)$	$P_*(A)$
$\emptyset$	$\emptyset$	$\emptyset$	0	0
$\{1\}$	$\{a\}$	$\{a\}$	$\frac{1}{4}$	$\frac{1}{4}$
$\{2\}$	$\{b, d\}$	$\{b\}$	$\frac{3}{4}$	$\frac{1}{4}$
$\{3\}$	$\{d\}$	$\emptyset$	$\frac{1}{2}$	0
$\{1, 2\}$	$\{a, b, d\}$	$\{a, b\}$	1	$\frac{1}{2}$
$\{1, 3\}$	$\{a, d\}$	$\{a\}$	$\frac{3}{4}$	$\frac{1}{4}$
$\{2, 3\}$	$\{b, d\}$	$\{b, d\}$	$\frac{3}{4}$	$\frac{3}{4}$
$\{1, 2, 3\}$	$\{a, b, d\}$	$\{a, b, d\}$	1	1

One can consider $P^*(A)$ and $P_*(A)$ as upper and lower probability bounds.

Imprecise Data (3)

Some properties of probability bounds:

a) $P^*\colon\ 2^\Omega \to [0,1]$

b) $0 \leq P_* \leq P^* \leq 1, \qquad P_*(\emptyset) = P^*(\emptyset) = 0, \qquad P_*(\Omega) = P^*(\Omega) = 1$

c) $A \subseteq B \quad \Rightarrow \quad P^*(A) \leq P^*(B) \quad \text{and} \quad P_*(A) \leq P_*(B)$

d) $A \cap B = \emptyset \quad \not\Rightarrow \quad P^*(A) + P^*(B) = P^*(A \cup B)$

e) $P_*(A \cup B) \geq P_*(A) + P_*(B) - P_*(A \cap B)$

f) $P^*(A \cup B) \leq P^*(A) + P^*(B) - P^*(A \cap B)$

g) $P_*(A) = 1 - P^*(\Omega \backslash A)$

Imprecise Data (4)

One can prove the following generalized equation:

$$P_*(\bigcup_{i=1}^{n} A_i) \quad \geq \sum_{\emptyset \neq I : I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \cdot P_*(\bigcap_{i \in I} A_i)$$

These set functions also play an important role in theoretical physics (capacities, Choquet, 1955). Shafer did generalize these thoughts and developed a theory of belief functions.

Belief Revision

How is new knowledge incoporated?

Every observer announces the location of the ship in form of a subset of all possible ship locations. Given these set-valued mappings, we can derive upper and lower probabilities with the help of the observer importance measure. Let us assume the ship is certainly at sea.

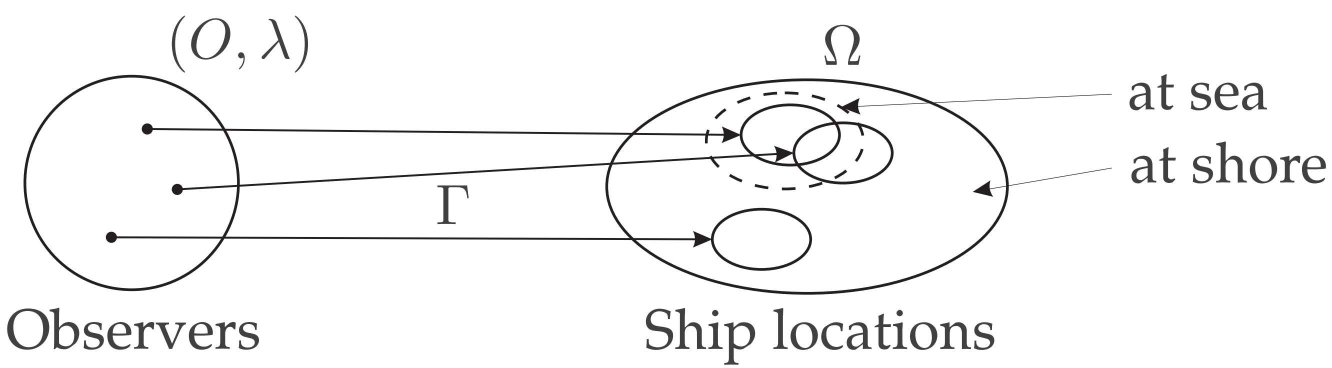
How do the upper/lower probabilities change?

Example

a) *Geometric Conditioning*
(observers that give partial or full wrong information are discarded)

$$P_*(A \mid B) = \frac{\lambda(\{\xi \in \Theta \mid \Gamma(\xi) \subseteq A \text{ and } \Gamma(\xi) \subseteq B\})}{\lambda(\{\xi \in \Theta \mid \Gamma(\xi) \subseteq B\})} = \frac{P_*(A \cap B)}{P_*(B)}$$

$$P^*(A \mid B) = \frac{\lambda(\{\xi \in \Theta \mid \Gamma(\xi) \subseteq B \text{ and } \Gamma(\xi) \cap A \neq \emptyset\})}{\lambda(\{\xi \in \Theta \mid \Gamma(\xi) \subseteq B\})} = \frac{P^*(A \cup \overline{B}) - P^*(\overline{B})}{1 - P^*(\overline{B})}$$

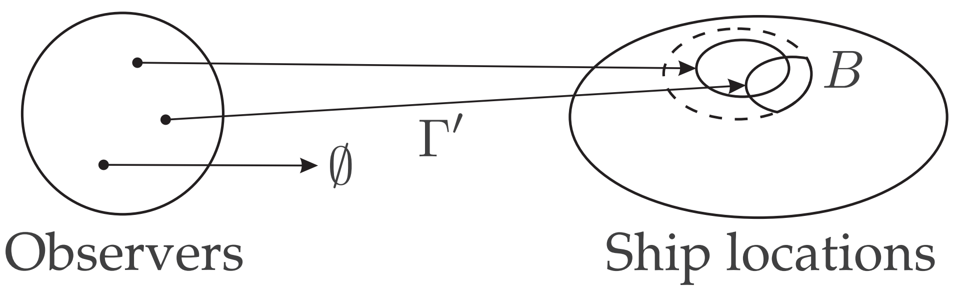


Belief Revision (2)

b) *Data Revision*
(the observed data is modified such that they fit the certain information)

$$(P_*)_B(A) = \frac{P_*(A \cup \overline{B}) - P_*(\overline{B})}{1 - P_*(B)}$$

$$(P^*)_B(A) = \frac{P^*(A \cap B)}{P^*(B)}$$



These two concepts have different semantics. There are several more belief revision concepts.

Imprecise Probabilities

Let x_0 be the true value but assume there is no information about $P(A)$ to decide whether $x_0 \in A$. There are only probability boundaries.

Let $\mathcal{L}$ be a set of probability measures. Then we call

$$(P_{\mathcal{L}})_* : 2^{\Omega} \to [0,1]\,, A \mapsto \inf\{P(A) \mid P \in \mathcal{L}\} \qquad \text{the lower and}$$

$$(P_{\mathcal{L}})^* : 2^{\Omega} \to [0,1]\,, A \mapsto \sup\{P(A) \mid P \in \mathcal{L}\} \qquad \text{the upper}$$

probability of A w. r. t. $\mathcal{L}$.

a) $(P_{\mathcal{L}})_*(\emptyset) = (P_{\mathcal{L}})^*(\emptyset) = 0; \quad (P_{\mathcal{L}})_*(\Omega) = (P_{\mathcal{L}})^*(\Omega) = 1$

b) $0 \leq (P_{\mathcal{L}})_*(A) \leq (P_{\mathcal{L}})^*(A) \leq 1$

c) $(P_{\mathcal{L}})^*(A) = 1 - (P_{\mathcal{L}})_*(\overline{A})$

d) $(P_{\mathcal{L}})_*(A) + (P_{\mathcal{L}})_*(B) \leq (P_{\mathcal{L}})_*(A \cup B)$

e) $(P_{\mathcal{L}})_*(A \cap B) + (P_{\mathcal{L}})_*(A \cup B) \not\geq (P_{\mathcal{L}})_*(A) + (P_{\mathcal{L}})_*(B)$

Belief Revision

Let $B \subseteq \Omega$ and $\mathcal{L}$ a class of probabilities. The we call

$$A \subseteq \Omega : (P_{\mathcal{L}})_*(A \mid B) = \inf\{P(A \mid B) \mid P \in \mathcal{L} \wedge P(B) > 0\} \qquad \text{the lower and}$$
$$A \subseteq \Omega : (P_{\mathcal{L}})^*(A \mid B) = \sup\{P(A \mid B) \mid P \in \mathcal{L} \wedge P(B) > 0\} \qquad \text{the upper}$$

conditional probability of A given B.

A class $\mathcal{L}$ of probability measures on $\Omega = \{\omega_1, \ldots, \omega_n\}$ is of type 1, iff there exist functions R_1 and R_2 from 2^Ω into $[0,1]$ with:

$$\mathcal{L} = \{P \mid \forall A \subseteq \Omega : R_1(A) \leq P(A) \leq R_2(A)\}$$

Belief Revision (2)

Intuition: P is determined by $P(\{\omega_i\})$, $i = 1, \ldots, n$ which corresponds to a point in $\mathbb{R}^n$ with coordinates $\big(P(\{\omega_1\}), \ldots, P(\{\omega_n\})\big)$.

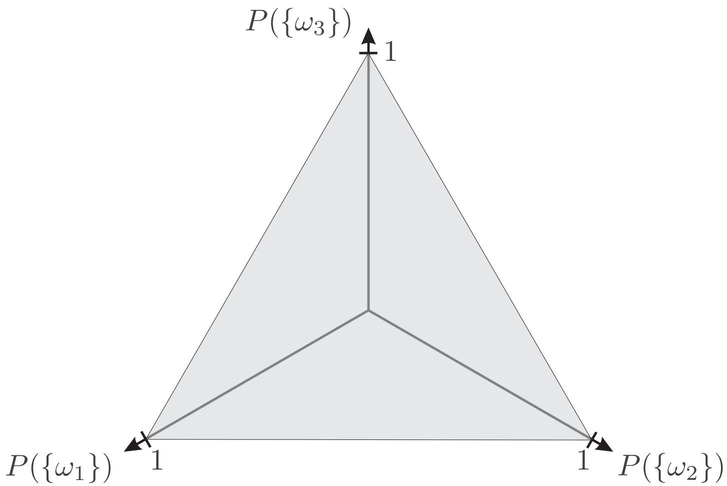
If $\mathcal{L}$ is type 1, it holds true that:

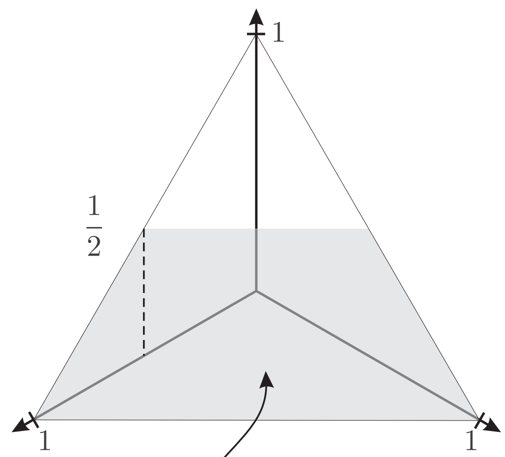
$$\mathcal{L} \Leftrightarrow \Big\{(r_1, \ldots, r_n) \in \mathbb{R}^n \mid \exists P \colon \forall A \subseteq \Omega \colon \\ (P_{\mathcal{L}})_*(A) \leq P(A) \leq (P_{\mathcal{L}})^*(A) \\ \text{and } \; r_i = P(\{\omega_i\}),\; i = 1, \ldots, n\Big\}$$

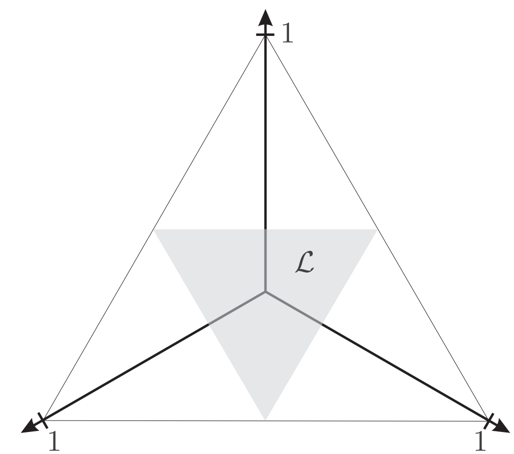
Example

$\Omega = \{\omega_1, \omega_2, \omega_3\}$

$\mathcal{L} = \{P \mid \frac{1}{2} \leq P(\{\omega_1, \omega_2\}) \leq 1, \quad \frac{1}{2} \leq P(\{\omega_2, \omega_3\}) \leq 1, \quad \frac{1}{2} \leq P(\{\omega_1, \omega_3\}) \leq 1\}$







general restriction:
$0 \leq P(\{\omega_i\}) \leq 1$
$P(\{\omega_1\}) + P(\{\omega_2\}) + P(\{\omega_3\}) = 1$

$\{P \mid \frac{1}{2} \leq P(\{\omega_1, \omega_2\}) \leq 1\}$

Let $A_1 = \{\omega_1, \omega_2\}$, $A_2 = \{\omega_2, \omega_3\}$, $A_3 = \{\omega_1, \omega_3\}$

$$P_*(A_1) + P_*(A_2) + P_*(A_3) - P_*(A_1 \cap A_2) - P_*(A_2 \cap A_3) - P_*(A_1 \cap A_3) + P_*(A_1 \cap A_2 \cap A_3)$$
$$= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} - 0 - 0 - 0 + 0 = \frac{3}{2} > 1 = P(A_1 \cup A_2 \cup A_3)$$

Belief Revision (3)

If $\mathcal{L}$ is type 1 and $(P_{\mathcal{L}})^*(A \cup B) \geq (P_{\mathcal{L}})^*(A) + (P_{\mathcal{L}})^*(B) - (P_{\mathcal{L}})^*(A \cap B)$, then

$$(P_{\mathcal{L}})^*(A \mid B) = \frac{(P_{\mathcal{L}})^*(A \cap B)}{(P_{\mathcal{L}})^*(A \cap B) + (P_{\mathcal{L}})_*(B \cap \overline{A})}$$

and

$$(P_{\mathcal{L}})_*(A \mid B) = \frac{(P_{\mathcal{L}})_*(A \cap B)}{(P_{\mathcal{L}})_*(A \cap B) + (P_{\mathcal{L}})^*(B \cap \overline{A})}$$

Let $\mathcal{L}$ be a class of type 1. $\mathcal{L}$ is of type 2, iff

$$(P_{\mathcal{L}})_*(A_1 \cup \cdots \cup A_n) \geq \sum_{I:\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \cdot (P_{\mathcal{L}})_*(\bigcap_{i \in I} A_i)$$

Belief Functions

Motivation

(Θ, Q)	Sensors	
Ω	possible results, $\Gamma : \Theta \to 2^{\Omega}$	
Γ, Q	induce a probability m on 2^{Ω}	
$m :$	$A \mapsto Q(\{\theta \in \Theta \mid \Gamma(\theta) = A\})$	mass distribution
$\text{Bel} :$	$A \mapsto \sum_{B:B\subseteq A} m(B)$	Belief (lower probability)
$\text{Pl} :$	$A \mapsto \sum_{B:B\cap A\neq\emptyset} m(B)$	Plausibility (upper probability)

- Random sets: Dempster (1968)
- Belief functions: Shafer (1974)
  Development of a completely new uncertainty calculus

Belief Functions (2)

The function $\text{Bel} : 2^\Omega \to [0,1]$ is called *belief function*, if it possesses the following properties:

- $\text{Bel}(\emptyset) = 0$
- $\text{Bel}(\Omega) = 1$
- $\forall n \in \mathbb{N}\colon\ \forall A_1, \ldots, A_n \in 2^\Omega :$
  $\text{Bel}(A_1 \cup \cdots \cup A_n) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \cdot \text{Bel}(\bigcap_{i \in I} A_i)$

If Bel is a belief function then for $m : 2^\Omega \to \mathbb{R}$ with $m(A) = \sum_{B : B \subseteq A} (-1)^{|A \setminus B|} \cdot \text{Bel}(B)$ the following properties hold:

- $0 \leq m(A) \leq 1$
- $m(\emptyset) = 0$
- $\sum_{A \subseteq \Omega} m(A) = 1$

Belief Functions (3)

Let $|\Omega| < \infty$ and $f, g : 2^\Omega \to [0, 1]$.

$$\forall A \subseteq \Omega\colon \left(f(A) = \sum_{B:B\subseteq A} g(B)\right)$$

$$\Leftrightarrow$$

$$\forall A \subseteq \Omega\colon \left(g(A) = \sum_{B:B\subseteq A} (-1)^{|A\setminus B|} \cdot f(B)\right)$$

(g is called the *Möbius transformed* of f)

The mapping $m : 2^\Omega \to [0, 1]$ is called a *mass distribution*, if the following properties hold:

- $m(\emptyset) = 0$
- $\sum_{A\subseteq\Omega} m(A) = 1$

Example

A	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{2,3\}$	$\{1,3\}$	$\{1,2,3\}$
$m(A)$	0	$1/4$	$1/4$	0	0	0	$2/4$	0
$\mathrm{Bel}(A)$	0	$1/4$	$1/4$	0	$2/4$	$1/4$	$3/4$	1

Belief $\widehat{=}$ lower probability with modified semantic

$$\mathrm{Bel}(\{1,3\}) = m(\emptyset) + m(\{1\}) + m(\{3\}) + m(\{1,3\})$$
$$m(\{1,3\}) = \mathrm{Bel}(\{1,3\}) - \mathrm{Bel}(\{1\}) - \mathrm{Bel}(\{3\})$$

$m(A)$ — measure of the trust/belief that exactly A occurs

$\mathrm{Bel}_m(A)$ — measure of total belief that A occurs

$\mathrm{Pl}_m(A)$ — measure of not being able to disprove A (plausibility)

$$\mathrm{Pl}_m(A) = \sum_{B: A\cap B\neq\emptyset} m(B) = 1 - \mathrm{Bel}(\overline{A})$$

Given one of m, Bel or Pl, the other two can be efficiently computed.

Knowledge Representation

$m(\Omega) = 1$, $m(A) = 0$ else	total ignorance
$m(\{\omega_0\}) = 1$, $m(A) = 0$ else	value (ω_0) known
$m(\{\omega_i\}) = p_i, \sum_{i=1}^{n} p_i = 1$	Bayesian analysis

Further intermediate steps can be modeled.

Belief Revision

- Data Revision:
  - Mass of A flows onto $A \cap B$.
  - Masses are normalized to 1 ($\emptyset$-mass is destroyed)
- Geometric Conditioning:
  - Masses that do not lie completely inside B, flow off
  - Normalize

There is a mass flow from t to s (written: $s \sqsubseteq t$) iff for every $A \subseteq \Omega$ there exist functions $h_A : 2^\Omega \to [0,1]$ such that the following properties hold:

- $\sum_{B:B\subseteq\Omega} h_A(B) = t(A)$ for all A
- $h(A(B) \neq 0 \;\Rightarrow\; B \subseteq A$ for all A, B
- $s(B) = \frac{\sum_{A:A\subseteq\Omega} h_A(B)}{1-\sum_{A:A\subseteq\Omega} h_A(\emptyset)}$

Example

A	$s(A)$	$t(A)$	$u(A)$
$\emptyset$	0	0	0
$\{1\}$	0	0	0.1
$\{2\}$	0.4	0.4	0
$\{3\}$	0.1	0	0
$\{1,2\}$	0.2	0.5	0.1
$\{1,3\}$	0	0	0.4
$\{2,3\}$	0.3	0.1	0.4
Ω	0	0	0

The following relations hold:
$s \sqsubseteq t, t \sqsubseteq s, s \sqsubseteq u, t \sqsubseteq u, t \sqsubseteq t, u \not\sqsubseteq s$

Combination of Random Sets

Let $(\Omega, 2^\Omega)$ be a space of events. Further be $(O_1, 2^{O_1}, \lambda_1)$ and $(O_2, 2^{O_2}, \lambda_2)$ spaces of independent observers.

We call $(O_1 \times O_2, \lambda_1 \cdot \lambda_2)$ the product space of observers and

$$\Gamma : O_1 \times O_2 \to 2^\Omega, \Gamma(x_1, x_2) = \Gamma_1(x_1) \cap \Gamma_2(x_2)$$

the combined observer function.

We obtain with

$$(P_L)_*(A) = \frac{(\lambda_1 \cdot \lambda_2)(\{(x_1, x_2) \mid \Gamma(x_1, x_2) \neq \emptyset \wedge \Gamma(x_1, x_2) \sqsubseteq A\})}{(\lambda_1 \cdot \lambda_2)(\{(x_1, x_2 \mid \Gamma(x_1, x_2) \neq \emptyset)\})}$$

the lower probability of A that respects both observations.

Example

$$\Omega = \{1,2,3\} \qquad \lambda_1\colon\ \{a\} \mapsto {}^1\!/_3 \qquad \lambda_2\colon\ \{c\} \mapsto {}^1\!/_2$$

$$\{b\} \mapsto {}^2\!/_3 \qquad \lambda_2\colon\ \{d\} \mapsto {}^1\!/_2$$

$$O_1 = \{a,b\} \qquad \Gamma_1\colon\ a \mapsto \{1,2\} \qquad \Gamma_2\colon\ c \mapsto \{1\}$$

$$O_2 = \{c,d\} \qquad b \mapsto \{2,3\} \qquad d \mapsto \{2,3\}$$

Combination:

$$O_1 \times O_2 \ = \ \{\overline{ac}, \overline{bc}, \overline{ad}, \overline{bd}\}$$

$$\begin{aligned}
\lambda\colon\ \{\overline{ac}\} &\mapsto {}^1\!/_6 & \Gamma\colon\ \overline{ac} &\mapsto \{1\} \\
\{\overline{ad}\} &\mapsto {}^1\!/_6 & \overline{ad} &\mapsto \{2\} \\
\{\overline{bc}\} &\mapsto {}^2\!/_6 & \overline{bc} &\mapsto \emptyset \\
\{\overline{bd}\} &\mapsto {}^2\!/_6 & \overline{bd} &\mapsto \{2,3\}
\end{aligned}$$

$$\begin{aligned}
\Gamma_*(\Omega) &= \{(x_1,x_2) \mid \Gamma(x_1,x_2) \neq \emptyset\} \\
&= \{\overline{ac}, \overline{ad}, \overline{bd}\}
\end{aligned}$$

$$\lambda(\Gamma_*(\Omega)) = {}^4\!/_6$$

Example (2)

A	$m_1(A)$	$(P_*)_{\Gamma_1}(A)$	$m_2(A)$	$(P_*)_{\Gamma_2}(A)$	$m(A)$	$(P_*)_{\Gamma}(A)$
$\emptyset$	0	0	0	0	0	0
$\{1\}$	0	0	$1/2$	$1/2$	$1/4 = \frac{1/6}{4/6}$	$1/4$
$\{2\}$	0	0	0	0	$1/4$	$1/4$
$\{3\}$	0	0	0	0	0	0
$\{1,2\}$	$1/3$	$1/3$	0	$1/2$	0	$1/2$
$\{1,3\}$	0	0	0	$1/2$	0	$1/4$
$\{2,3\}$	$2/3$	$2/3$	$1/2$	$1/2$	$1/2$	$3/4$
$\{1,2,3\}$	0	1	0	1	0	1

Combinations of Mass Distributions

Motivation: Combination of m_1 and m_2

$m_1(A_i) \cdot m_2(B_j)$:	Mass attached to $A_i \cap B_j$, if only A_i or B_j are concerned
$\sum_{i,j:A_i \cap B_j = A} m_1(A_i) \cdot m_2(B_j)$:	Mass attached to A (after combination)

This consideration only leads to a mass distribution,
if $\sum_{i,j:A_i \cap B_j = \emptyset} m_1(A_i) \cdot m_2(B_j) = 0$.

If this sum is > 0 normalization takes place.

Combination Rule

If m_1 and m_2 are mass distributions over Ω with belief functions Bel_1 and Bel_2 and does further hold $\sum_{i,j:A_i \cap B_j = \emptyset} m_1(A_i) \cdot m_2(B_j) < 1$, then the function $m : 2^\Omega \to [0,1]\,, m(\emptyset) = 0$

$$m(A) = \frac{\sum_{B,C:B \cap C = A} m_1(B) \cdot m_2(C)}{1 - \sum_{B,C:B \cap C = \emptyset} m_1(B) \cdot m_2(C)}$$

is a mass distribution. The belief function of m is denoted as $\mathrm{comb}(\mathrm{Bel}_1, \mathrm{Bel}_2)$ or $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$. The above formula is called the combination rule.

Example

$$m_1(\{1,2\}) = {}^1\!/_3 \qquad m_2(\{1\}) = {}^1\!/_2$$

$$m_1(\{2,3\}) = {}^2\!/_3 \qquad m_2(\{2,3\}) = {}^1\!/_2$$

$$m = m_1 \oplus m_2 :$$

$$\begin{aligned}
\{1\} &\mapsto \frac{{}^1\!/_6}{{}^4\!/_6} = {}^1\!/_4 \\
\{2\} &\mapsto \frac{{}^1\!/_6}{{}^4\!/_6} = {}^1\!/_4 \\
\emptyset &\mapsto 0 \\
\{2,3\} &\mapsto \frac{{}^2\!/_6}{{}^4\!/_6} = {}^1\!/_2
\end{aligned}$$

Combination Rule (2)

Remarks:

a) The result from the combination rule and the analysis of random sets is identical

b) There are more efficient ways of combination

c) $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2 = \mathrm{Bel}_2 \oplus \mathrm{Bel}_1$

d) $\oplus$ is associative

e) $\mathrm{Bel}_1 \oplus \mathrm{Bel}_1 \neq \mathrm{Bel}_1$ (in general)

f) $\mathrm{Bel}_2 : 2^\Omega \to [0,1]\,, m_2(B) = 1$

$$\mathrm{Bel}_2(A) = \begin{cases} 1 & \text{if} B \subseteq A \\ 0 & \text{otherwise} \end{cases}$$

The combination of Bel_1 and Bel_2 yields the data revision of m_1 with B.

Possibility Theory

- The best-known calculus for handling uncertainty is, of course, **probability theory**. [Laplace 1812]
- An less well-known, but noteworthy alternative is **possibility theory**. [Dubois and Prade 1988]
- In the interpretation we consider here, possibility theory can handle **uncertain and imprecise information**, while probability theory, at least in its basic form, was only designed to handle *uncertain information.*
- Types of **imperfect information**:
  - **Imprecision:** disjunctive or set-valued information about the obtaining state, which is certain: the true state is contained in the disjunction or set.
  - **Uncertainty:** precise information about the obtaining state (single case), which is not certain: the true state may differ from the stated one.
  - **Vagueness:** meaning of the information is in doubt: the interpretation of the given statements about the obtaining state may depend on the user.

Possibility Theory: Axiomatic Approach

Definition: Let Ω be a (finite) sample space.
A **possibility measure** Π on Ω is a function $\Pi : 2^{\Omega} \to [0, 1]$ satisfying

1. $\Pi(\emptyset) = 0$ and
2. $\forall E_1, E_2 \subseteq \Omega : \Pi(E_1 \cup E_2) = \max\{\Pi(E_1), \Pi(E_2)\}$.

- Similar to Kolmogorov's axioms of probability theory.
- From the axioms follows $\Pi(E_1 \cap E_2) \leq \min\{\Pi(E_1), \Pi(E_2)\}$.
- Attributes are introduced as random variables (as in probability theory).
- $\Pi(A = a)$ is an abbreviation of $\Pi(\{\omega \in \Omega \mid A(\omega) = a\})$
- If an event E is possible without restriction, then $\Pi(E) = 1$.
  If an event E is impossible, then $\Pi(E) = 0$.

Interpretation of Degrees of Possibility [Gebhardt and Kruse 1993]

- Let Ω be the (nonempty) set of all possible states of the world, ω_0 the actual (but unknown) state.

- Let $C = \{c_1, \ldots, c_n\}$ be a set of contexts (observers, frame conditions etc.) and $(C, 2^C, P)$ a finite probability space (context weights).

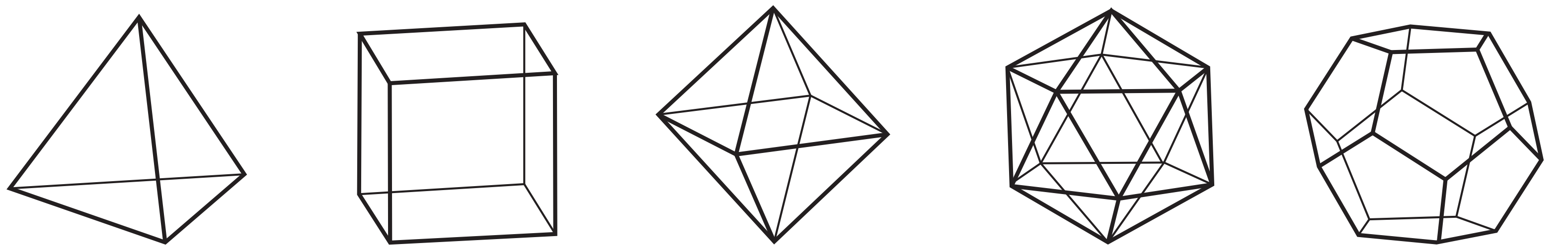
- Let $\Gamma : C \to 2^\Omega$ be a set-valued mapping, which assigns to each context the **most specific correct set-valued specification of** ω_0. The sets $\Gamma(c)$ are called the **focal sets** of Γ.

- Γ is a **random set** (i.e., a set-valued random variable) [Nguyen 1978]. The **basic possibility assignment** induced by Γ is the mapping

$$\begin{aligned} \pi : \Omega &\to [0, 1] \\ \pi(\omega) &\mapsto P(\{c \in C \mid \omega \in \Gamma(c)\}). \end{aligned}$$

Example: Dice and Shakers

shaker 1	shaker 2	shaker 3	shaker 4	shaker 5



tetrahedron	hexahedron	octahedron	icosahedron	dodecahedron
1 – 4	1 – 6	1 – 8	1 – 10	1 – 12

numbers	degree of possibility
1 – 4	$\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{1}{5} = 1$
5 – 6	$\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{1}{5} = \frac{4}{5}$
7 – 8	$\frac{1}{5}+\frac{1}{5}+\frac{1}{5} = \frac{3}{5}$
9 – 10	$\frac{1}{5}+\frac{1}{5} = \frac{2}{5}$
11 – 12	$\frac{1}{5} = \frac{1}{5}$

Definition: Let $\Gamma : C \to 2^\Omega$ be a random set.
The **possibility measure** induced by Γ is the mapping

$$\begin{aligned} \Pi : 2^\Omega &\to [0,1], \\ E &\mapsto P(\{c \in C \mid E \cap \Gamma(c) \neq \emptyset\}). \end{aligned}$$

Problem: From the given interpretation it follows only:

$$\forall E \subseteq \Omega : \quad \max_{\omega \in E} \pi(\omega) \;\leq\; \Pi(E) \;\leq\; \min\left\{1, \sum_{\omega \in E} \pi(\omega)\right\}.$$

	1	2	3	4	5
$c_1 : \frac{1}{2}$			●		
$c_2 : \frac{1}{4}$		●	●	●	
$c_3 : \frac{1}{4}$	●	●	●	●	●
π	0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{4}$

	1	2	3	4	5
$c_1 : \frac{1}{2}$			●		
$c_2 : \frac{1}{4}$	●	●			
$c_3 : \frac{1}{4}$				●	●
π	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

From the Context Model to Possibility Measures (cont.)

Attempts to solve the indicated problem:

- Require the focal sets to be **consonant**:
  Definition: Let $\Gamma : C \to 2^\Omega$ be a random set with $C = \{c_1, \ldots, c_n\}$. The focal sets $\Gamma(c_i)$, $1 \leq i \leq n$, are called **consonant**, iff there exists a sequence $c_{i_1}, c_{i_2}, \ldots, c_{i_n}$, $1 \leq i_1, \ldots, i_n \leq n$, $\forall 1 \leq j < k \leq n : i_j \neq i_k$, so that

  $$\Gamma(c_{i_1}) \subseteq \Gamma(c_{i_2}) \subseteq \ldots \subseteq \Gamma(c_{i_n}).$$

  $\to$ mass assignment theory [Baldwin *et al.* 1995]

  Problem: The "voting model" is not sufficient to justify consonance.

- Use the lower bound as the "most pessimistic" choice. [Gebhardt 1997]

  Problem: Basic possibility assignments represent negative information, the lower bound is actually the *most optimistic* choice.

- Justify the lower bound from decision making purposes.

- Assume that in the end we have to decide on a single event.
- Each event is described by the values of a set of attributes.
- Then it can be useful to assign to a set of events the degree of possibility of the "most possible" event in the set.

Example:

$\sum$	36	18	18	28	
28	0	0	0	28	28
18	18	0	0	0	18
18	18	0	0	0	18
36	0	18	18	0	18
	18	18	18	28	max

0	40	0	40
40	0	0	40
0	0	20	20
40	40	20	max

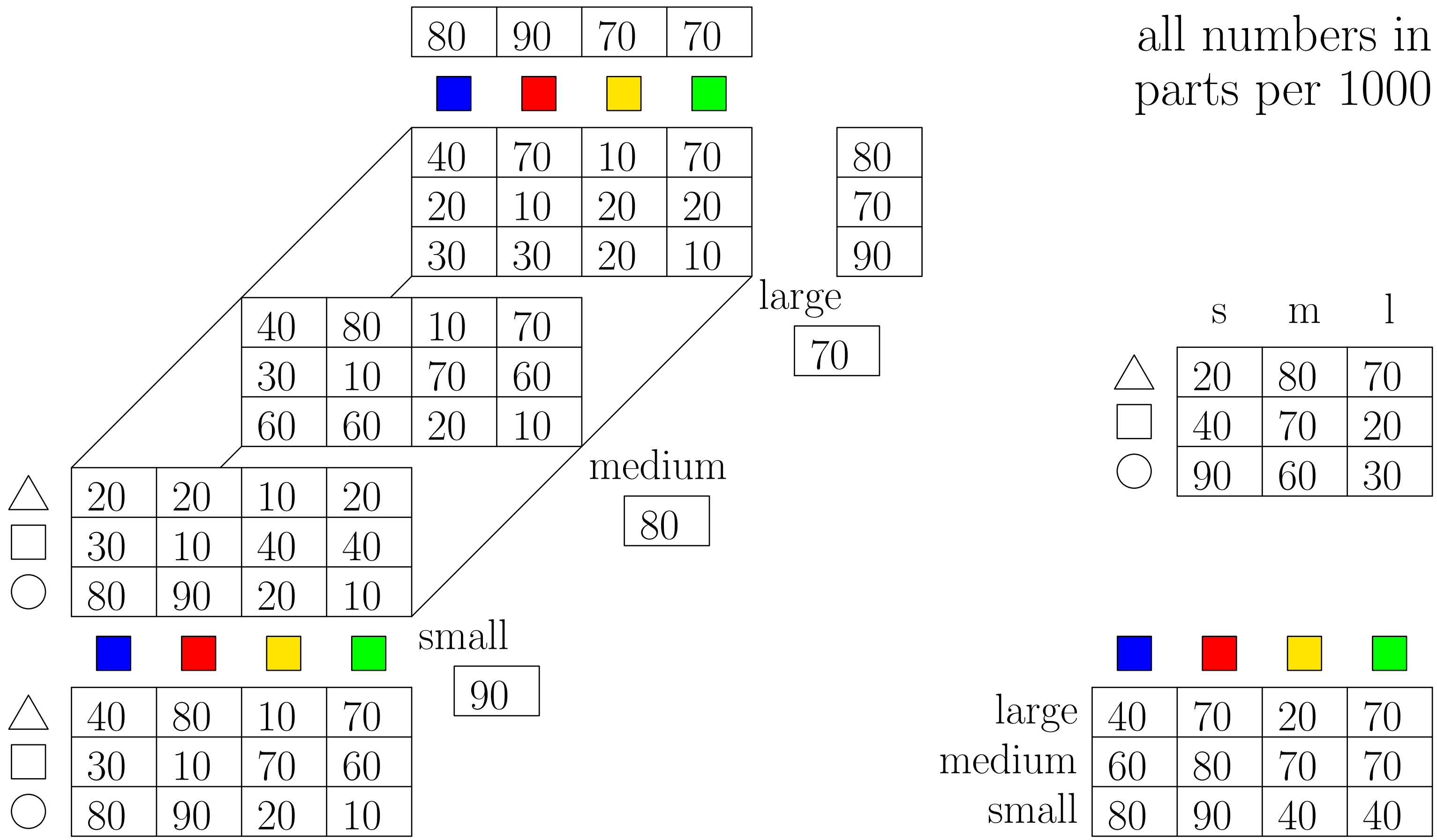
Possibility Distributions

Definition: Let $X = \{A_1, \ldots, A_n\}$ be a set of attributes defined on a (finite) sample space Ω with respective domains $\text{dom}(A_i)$, $i = 1, \ldots, n$. A **possibility distribution** π_X over X is the restriction of a possibility measure Π on Ω to the set of all events that can be defined by stating values for all attributes in X. That is, $\pi_X = \Pi|_{\mathcal{E}_X}$, where

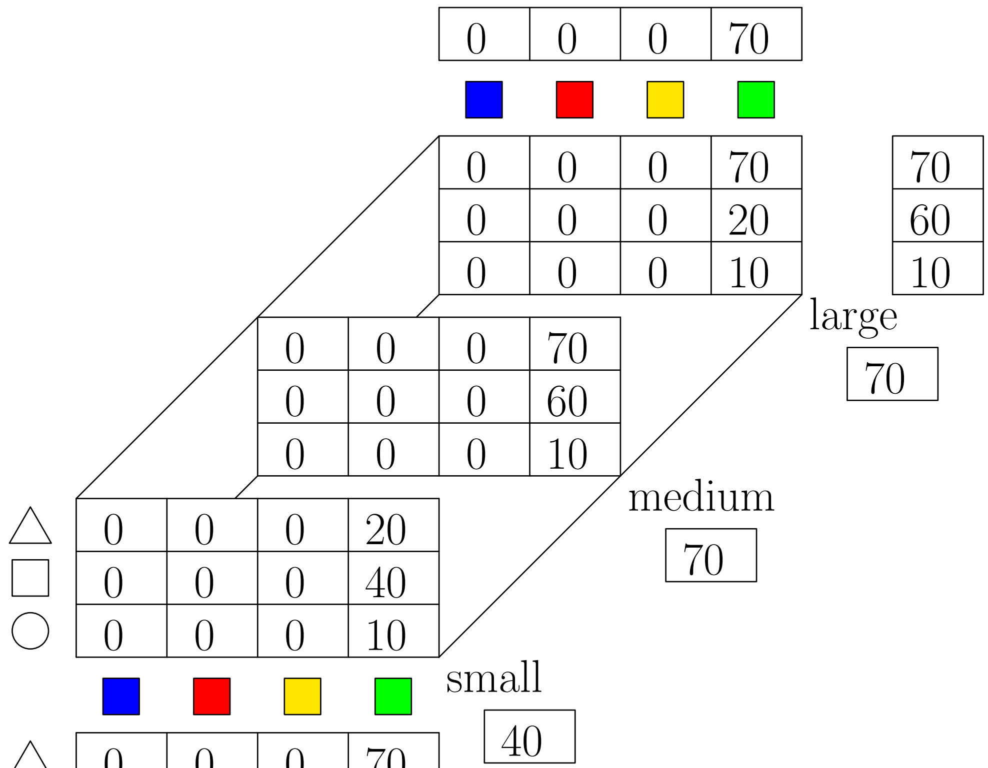
$$
\begin{aligned}
\mathcal{E}_X &= \left\{ E \in 2^\Omega \;\middle|\; \exists a_1 \in \text{dom}(A_1) : \ldots \exists a_n \in \text{dom}(A_n) : \right. \\
&\qquad \left. E \mathrel{\widehat{=}} \bigwedge_{A_j \in X} A_j = a_j \right\} \\
&= \left\{ E \in 2^\Omega \;\middle|\; \exists a_1 \in \text{dom}(A_1) : \ldots \exists a_n \in \text{dom}(A_n) : \right. \\
&\qquad \left. E = \left\{ \omega \in \Omega \;\middle|\; \bigwedge_{A_j \in X} A_j(\omega) = a_j \right\} \right\}.
\end{aligned}
$$

- Corresponds to the notion of a probability distribution.
- Advantage of this formalization: No index transformation functions are needed for projections, there are just fewer terms in the conjunctions.

A Possibility Distribution



- The numbers state the degrees of possibility of the corresp. value combination.



all numbers in parts per 1000

	blue	red	yellow	green
△	0	0	0	70
□	0	0	0	60
○	0	0	0	10

	s	m	l
△	20	70	70
□	40	60	20
○	10	10	10

	blue	red	yellow	green
large	0	0	0	70
medium	0	0	0	70
small	0	0	0	40

- Using the information that the given object is green.

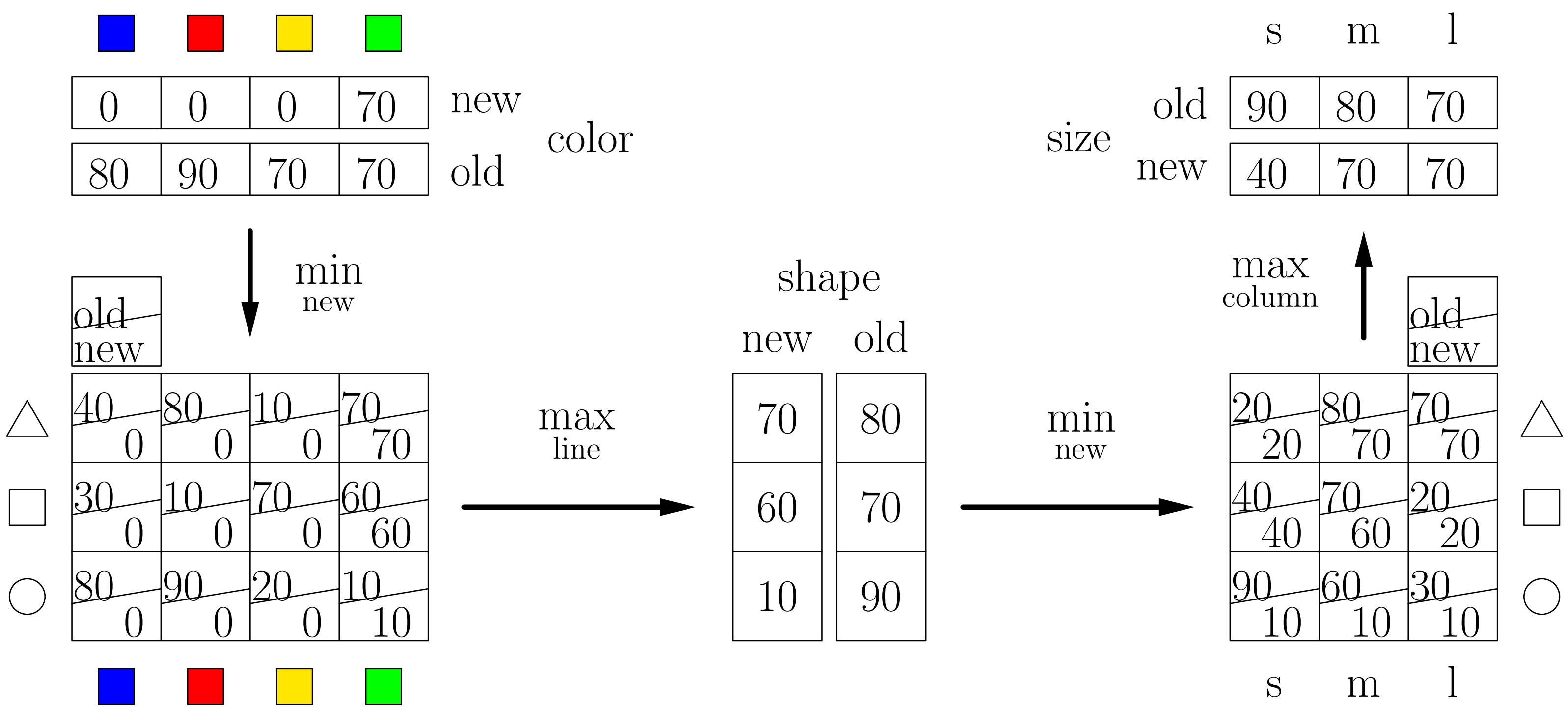
Possibilistic Decomposition

- As for relational and probabilistic networks, the three-dimensional possibility distribution can be decomposed into projections to subspaces, namely:
  – the maximum projection to the subspace color × shape and
  – the maximum projection to the subspace shape × size.
- It can be reconstructed using the following formula:

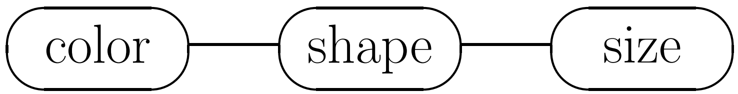
$$\begin{aligned}
\forall i,j,k:\ \pi\left(a_i^{(\text{color})}, a_j^{(\text{shape})}, a_k^{(\text{size})}\right) &= \min\left\{\pi\left(a_i^{(\text{color})}, a_j^{(\text{shape})}\right), \pi\left(a_j^{(\text{shape})}, a_k^{(\text{size})}\right)\right\} \\
&= \min\left\{ \max_k \pi\left(a_i^{(\text{color})}, a_j^{(\text{shape})}, a_k^{(\text{size})}\right), \right. \\
&\qquad \left. \max_i \pi\left(a_i^{(\text{color})}, a_j^{(\text{shape})}, a_k^{(\text{size})}\right) \right\}
\end{aligned}$$

- Note the analogy to the probabilistic reconstruction formulas.

Again the same result can be obtained using only projections to subspaces (maximal degrees of possibility):



This justifies a graph representation:



Conditional Possibility and Independence

Definition: Let Ω be a (finite) sample space, Π a possibility measure on Ω, and $E_1, E_2 \subseteq \Omega$ events. Then

$$\Pi(E_1 \mid E_2) = \Pi(E_1 \cap E_2)$$

is called the **conditional possibility** of E_1 given E_2.

Definition: Let Ω be a (finite) sample space, Π a possibility measure on Ω, and A, B, and C attributes with respective domains $\mathrm{dom}(A)$, $\mathrm{dom}(B)$, and $\mathrm{dom}(C)$. A and B are called **conditionally possibilistically independent** given C, written $A \perp\!\!\!\perp_\Pi B \mid C$, iff

$$\begin{aligned}&\forall a \in \mathrm{dom}(A) : \forall b \in \mathrm{dom}(B) : \forall c \in \mathrm{dom}(C) :\\ &\qquad \Pi(A = a, B = b \mid C = c) = \min\{\Pi(A = a \mid C = c), \Pi(B = b \mid C = c)\}.\end{aligned}$$

- Similar to the corresponding notions of probability theory.

Possibilistic Evidence Propagation

$$
\begin{aligned}
&\pi(B = b \mid A = a_{\text{obs}}) \\
&= \pi\left(\bigvee_{a\in\text{dom}(A)} A = a, B = b, \bigvee_{c\in\text{dom}(C)} C = c \,\middle|\, A = a_{\text{obs}} \right) \\
&\overset{(1)}{=} \max_{a\in\text{dom}(A)}\{ \max_{c\in\text{dom}(C)} \{\pi(A = a, B = b, C = c \mid A = a_{\text{obs}})\}\} \\
&\overset{(2)}{=} \max_{a\in\text{dom}(A)}\{ \max_{c\in\text{dom}(C)} \{\min\{\pi(A = a, B = b, C = c), \pi(A = a \mid A = a_{\text{obs}})\}\}\} \\
&\overset{(3)}{=} \max_{a\in\text{dom}(A)}\{ \max_{c\in\text{dom}(C)} \{\min\{\pi(A = a, B = b), \pi(B = b, C = c), \\
&\qquad\qquad \pi(A = a \mid A = a_{\text{obs}})\}\}\} \\
&= \max_{a\in\text{dom}(A)} \{\min\{\pi(A = a, B = b), \pi(A = a \mid A = a_{\text{obs}}), \\
&\qquad \underbrace{\max_{c\in\text{dom}(C)} \{\pi(B = b, C = c)\}\}\}}_{=\pi(B=b)\geq\pi(A=a,B=b)} \\
&= \max_{a\in\text{dom}(A)} \{\min\{\pi(A = a, B = b), \pi(A = a \mid A = a_{\text{obs}})\}\}
\end{aligned}
$$

A:	color
B:	shape
C:	size

References

References

[Baldwin *et. al* 1995]
J.F. Baldwin, T.P. Martin, and B.W. Pilsworth.
FRIL — Fuzzy and Evidential Reasoning in Artificial Intelligence. Research Studies Press/J. Wiley & Sons, Taunton/Chichester, United Kingdom, 1995

[Borgelt *et al.* 2009]
Christian Borgelt, Matthias Steinbrecher and Rudolf Kruse.
Graphical Models — Representations for Learning, Reasoning and Data Mining, 2nd edition, John Wiley & Sons Ltd, United Kingdom, 2009

[Castillo *et al.* 1997]
Enrique Castillo, José Manuel Gutiérrez and Ali S. Hadi.
Expert Systems and Probabilistic Network Models, Springer Verlag, 1997

[Cooper and Herskovits 1992]
Gregory F. Cooper and Edward Herskovits.
A Bayesian Method for the Induction of Probabilistic Networks from Data, Machine Learning, Volume 9, pp. 309–347, 1992

References

[Dubois and Prade 1988]
D. Dubois and H. Prade.
Possibility Theory. Plenum Press, New York, NY, USA, 1988

[Gebhardt 1997]
J. Gebhardt.
Learning from Data: Possibilistic Graphical Models. Habilitation Thesis, University of Braunschweig, Germany, 1997

[Gebhardt and Kruse 1993]
J. Gebhardt and R. Kruse.
The Context Model — **An Integrating View of Vagueness and Uncertainty**. Int. Journal of Approximate Reasoning 9:283–314, North-Holland, Amsterdam, Netherlands, 1993

References

[Kruse *et. al* 1991]
Rudolf Kruse, Erhard Schwecke and Jochen Heinsohn.
Uncertainty and Vagueness in Knowledge Based Systems, Springer Verlag, 1991

[Laplace 1812]
P.S. Laplace
Théorie analytique des probabilités. Paris: Veuve Courcier, 1812

[Nguyen 1978]
H.T. Nguyen.
On Random Sets and Belief Functions. Journal of Mathematical Analysis and Applications 65:531–542. Academic Press, San Diego, CA, USA, 1978

Homepages

- Otto-von-Guericke-University of Magdeburg
  `http://www.uni-magdeburg.de/`

- School of Computer Science
  `http://www.cs.uni-magdeburg.de/`

- Computational Intelligence Group
  `http://fuzzy.cs.uni-magdeburg.de/`