Information Fusion

Information:
Information is data in context. It can be a description of a mapping toward human or technical knowledge.

Signal:
A signal is the carrier of data (information) and can be interpreted as a physical description of information. A signal can be one-dimensional or more-dimensional.

Sensor Fusion:
Sensor Fusion is information fusion from multiple sensors of different or the same type.

Information Fusion:
Information Fusion is combined information of data to estimate or predict the state of some real-world situations.

Information and Measurement:
Basically, we get information from the sensors. There are optical sensors, pressure, temperature, distance, angle, position, etc sensors.
But there are some problems with the information we get from the sensors like **random errors** (noise, further stochastic effects etc), and **systematic errors** (alignment error).
Measurement results from the sensors provide data. They are information if they are in the context. They might contain knowledge.

Creating reliable knowledge about machine processes is a challenge because it is a known fact that Data != Information != Knowledge.
In so far fusion process must create a low amount of data which creates reliable knowledge.

Data:
Data is a collection of facts in a raw or unorganized form such as numbers or characters. Data can range from abstract ideas to concrete measurements, even statistics. Without context, data can mean nothing.
Ex. 1 packet of bread, price of bread packet.

Information:
Information is the data that has been cleared of errors and further processed in a way that makes it easier to measure, visualize, analyze, etc.
Ex. 1 packet of bread contains 15 pieces of bread, and the raw materials used to make it.

Knowledge:
Knowledge is what we know. Knowledge comes up from the information that has been accumulated over time.
Ex. Purchase of bread based on past experience.

The main problem in information fusion is described below:
Too much data, poor models, bad features, or too many features and improperly analyzed applications.

Uncertainty:
Uncertainty is defined as the intrinsic absence of necessary knowledge. In many cases, as, e. g. in complex systems, it is usually impossible to acquire all the necessary information to create certain knowledge. Therefore, major attention should be given to this type of ignorance.

Aleatoric (statistical) uncertainty:
If data is complete and intrinsically non-deterministic in nature, it can be assumed as random (stochastic). The uncertainty is attributed to real-world phenomena and it can not be reduced or even eliminated by expanding an underlying knowledge base. Probabilistic approaches, such as classical Probability Theory (frequentist) and Bayesian Probability Theory are an effective way to model stochastic uncertainties, like measurement noise, etc. This type of uncertainty is referred to as aleatoric uncertainty. **Ex. Coin flip, chances of a win in the lottery, etc. are examples of aleatoric uncertainty**.

Epistemic (subjective, human-oriented) uncertainty:
In many situations we lack information, that is, not all intrinsically necessary knowledge is available at state. In this case, the uncertainty range should be reduced by expanding the underlying knowledge base. When data is scarce (short) the probabilistic approach may not be appropriate to reduce the system's uncertainty. Major types of this uncertainty are inconsistent and incomplete data, information, or knowledge. In many cases, this uncertainty can be reduced by multi-sensory fusion and expert knowledge. This type of uncertainty is referred to as epistemic uncertainty.

	Aleatoric Uncertainty	*Epistemic Uncertainty*
Type	irreducible	reducible
Data	random, stochastic	scarce
Origin	intrinsic variations in data	inconsistent & incomplete data, lack of knowledge
Model	Probability Theory	Evidence and Fuzzy Theories

Patterns:
Patterns are objects, which have pre-defined descriptive information signature One-dimensional patterns are defined by vectors (discrete signals). Two-dimensional patterns are defined by matrices (e.g. digital images).

Classes:
Patterns are mapped into different classes, independently regarding their appearance. Two discrete signals (patterns) are equivalent if they are allocated by a pattern recognition system to one specific class. All equivalent patterns are allocated to one so called equivalent class.

Classification:
Classification defines the assignment (or allocation) of objects in groups that belong to one class. The classification criteria are heavily dependent on the application. Ex. speech recognition (speaker-dependent, speaker-independent, content dependent, content-independent, etc.)

Automatic pattern recognition:
Systems that have the ability to allocate new objects (patterns) to known classes

Class allocation:
Semantical classes:

Semantical classes conclude all objects (technical or human-centric information) which are content-orientated similar or identical. Their generation is usually human expert based. The class allocation is knowhow or standpoint based.

Natural classes:
Natural classes are formed by information that is equivalent in a formal sense. They are based on mathematical formalisms which are based on distances between objects or different object clusters. Natural classes are the base for numerical classification.

Features:
A feature m, or a feature vector m is generated from sensory information. Features are the base of pattern recognition. They define different pattern signatures which are detectable and allocatable. Whether they are able to distinguish between all equivalent classes can not be deducted.

Redundant information fusion:
By means of redundant information the uncertainty of information, the signal-to-noise ratio (SNR) and the evidence plausibility can be increased. The reliability of data can be improved.

Complementary information:
is created by means of heterogeneous sensors. Incomplete information from a single sensor can be supplemented by another type of sensor.

Integration concepts of single sensors in a multi-sensor system strongly depend on the place of installation and the types of data.

Competitive integration – Fusion of similar sensor data with equivalent effective information. The goal is the reduction of signal uncertainty.

Sensors are configured competitive if each sensor delivers independent measurements of the same property. It can be the fusion of data from different sensors or the fusion of measurements from a single sensor taken at different instants. A competitive sensor configuration is also called a redundant configuration. A special case of competitive sensor fusion is fault tolerance. Fault tolerance requires an exact specification of the service and the failure modes of the system. Competitive configurations can also provide robustness to a system. Robust systems provide a degraded level of service in the presence of faults.

An example of a competitive integration is the temperature sensors which measures temperature at different time. It measures the same property of an object in the environment space. It can be also used in the context that if one breaks down another can measure the property.

Complementary integration – Fusion of different sensor data with the goal to close information gaps. In this case signal averaging is destructive.

A sensor configuration is called complementary if the sensors do not directly depend on each other, but can be combined in order to give a more complete image of the phenomenon under observation. This resolves the incompleteness of sensor data. Generally, fusing complementary data is easy, since the data from independent sensors can be appended to each other.

An example of a complementary integration is the employment of multiple cameras each observing disjunct parts of a room as applied in.

Cooperative integration – In this case, the information is distributed. Only after interpretation of all sensory data, the fusion result can be established.

A cooperative sensor network uses the information provided by two independent sensors to derive information that would not be available from the single sensor.

An example of a cooperative sensor configuration is stereoscopic vision – by combining two-dimensional images from two cameras at slightly different viewpoints a three-dimensional image of the observed scene is derived. Another example supposes we have 4 objects, 2 square, and 2 bolls, each one has a different color blue and red. In this example color sensor is needed to classify red or blue and a camera is needed to classify the shape.

Crisp Variable

A variable may only belong or not belong to the set into which it is defined. The function describing the membership of such a crisp variable to its appertaining set can therefore take only the value 1 if the variable belongs to the set, or 0 if the variable does not belong to the set.

Fuzzy Variable

For fuzzy variables and fuzzy sets, the function describing the membership of a variable to its appertaining set is allowed to take all the values between 0-1 interval. This means that given the referential set R of the real numbers, a fuzzy variable X is defined by its membership function $\mu_x(x)$ where x belongs to R. The membership function of the fuzzy variables satisfies the following property.

- $0 \leq \mu_X(x) \leq 1$;
- $\mu_X(x)$ is convex;
- $\mu_X(x)$ is normal (which means that at least one element x always exists for which $\mu_X(x) = 1$).

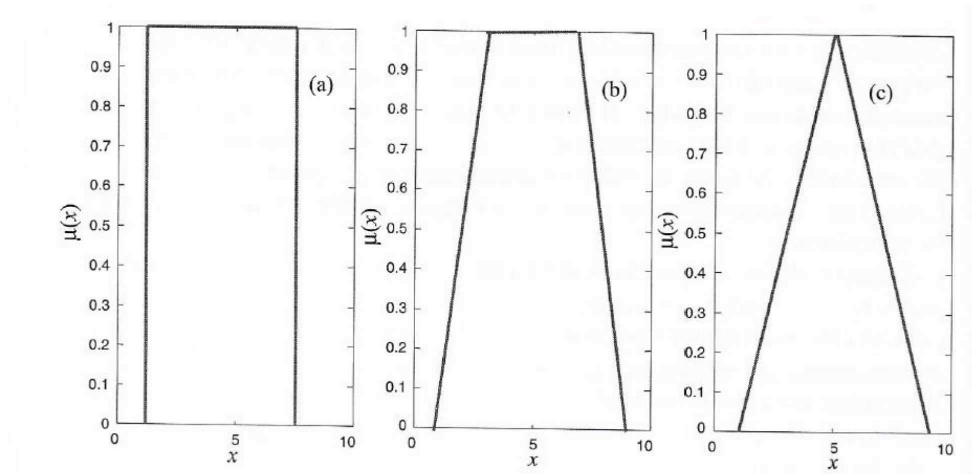
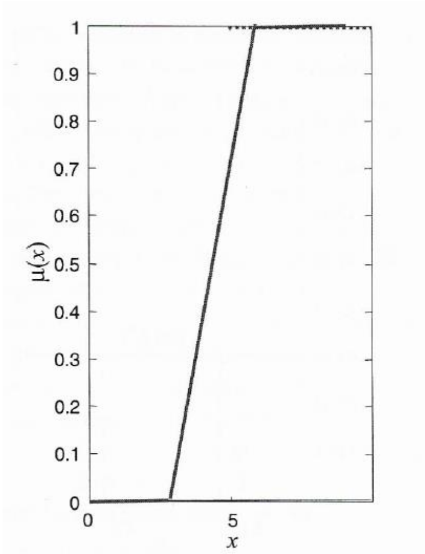


Fig. 2.1. Example of membership functions: (a) rectangular, (b) trapezoidal, and (c) triangular.

Membership function that only increases or only decreases qualify as a fuzzy variable and are used to represent the concept of a large number or small number in the context of each particular application.



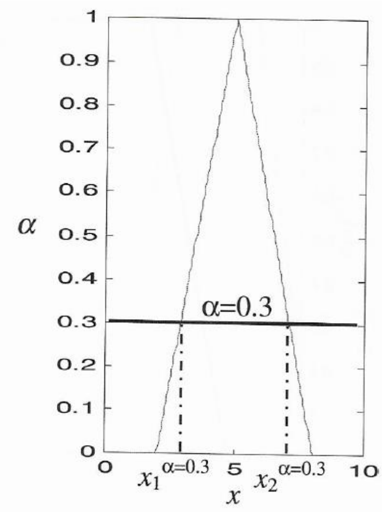


Fig. 2.3. α-cut of a triangular membership function for level $\alpha = 0.3$.

The membership function $\mu_X(x)$ of a fuzzy variable can be also described in terms of α-cuts at different vertical levels α. As the membership function of a fuzzy variable ranges, by definition, between 0 and 1, its α-cuts are defined for values of α between 0 and 1.

Each α-cut, at the generic level α, is defined as

$$X_\alpha = \{x \mid \mu_X(x) \geq \alpha\} \tag{2.1}$$

According to Eq. (2.1), each α-cut defines an interval $[x_1^\alpha, x_2^\alpha]$, where it is always $x_1^\alpha \leq x_2^\alpha$. The equality of x_1^α and x_2^α can be reached only for $\alpha = 1$ and only if the membership function has a single peak value, for instance, in the case of the triangular membership function reported in Fig. 2.1c. Generally, x_1^α and x_2^α take finite values, but in some cases, as the one in Fig. 2.2, it could be $x_1^\alpha = -\infty$ and/or $x_2^\alpha = +\infty$. If, for any value α, it is $x_1^\alpha = x_2^\alpha$, then the fuzzy variable degenerates into a crisp variable. An example of α-cut, for level $\alpha = 0.3$, is given in Fig. 2.3.

The importance of representing a fuzzy variable in terms of its α-cuts is that the α-cuts of a fuzzy variable and the corresponding levels α can be considered as a set of intervals of confidence and associated levels of certitude. The level of certitude contains information about how certain a person is about its knowledge. If a person, for instance, remembers exactly the birthday of a friend, his knowledge is certain; but if he only remembers the month, but not the day, the certainty of his knowledge is lower. In the first case, the available knowledge can be represented by a crisp value, that is, the exact day of the year (April, 15$^{\text{th}}$); in the second case, the available knowledge can be represented by a set, which contains the 30 days of April. Hence, as the level of certitude increases, the width of the corresponding confidence interval decreases.

As simply shown by the previous example, the link between the level α of certitude and the confidence interval at the same level corresponds to the natural, often implicit, mechanism of human thinking in the subjective estimation of a value for a measurement.

The following example shows how a fuzzy variable can be built, starting from the information, or personal idea, of one, or more human beings. Let us consider again the birthday of a friend, whose name is Sam. Tom is sure that Sam's birthday is April 15^{th}; John is not so sure about the date, even if he remembers well that the birthday falls in April. From a mathematical point of view, the interval of confidence estimated by Tom is [April 15, April 15], whereas the interval of confidence estimated by John is [April 1, April 30]. The levels of certitude associated with the two intervals are 1 and 0, respectively. If now a fuzzy variable must be associated with the available information, the following applies.

Let us suppose the only available information is from Tom. In this case, the available information can be represented by a crisp value. In fact, in this case, full certainty is given.

Let us now suppose the only available information is from John. In this case, the available information can be represented by a rectangular fuzzy variable, like the one in Fig. 2.1a.

However, it is also possible to combine the information given by Tom and John. As the level of certitude associated with John's interval is zero and those associated with Tom's interval is one, these two intervals can be considered as the α-cuts of the fuzzy variable, which represents Sam's birthday, at levels of α zero and one, respectively. Since, as the level of certitude increases, the interval of confidence becomes narrower, the fuzzy variable could be, for instance, triangular, like the one in Fig. 2.1c.

This simple example also shows, in an intuitive way, a relationship between the level of certitude and the level of confidence, which, according to the Theory of Uncertainty, should always be associated with an interval of confidence. The level of certitude indicates how much a person is sure about a certain event. If, for instance, the example of Sam's birthday is considered again, this means that the surer a person is about the birthday's date, the smaller is the extimated range of days. On the other hand, the level of confidence indicates the probability of a certain event. Therefore, considering the same example of Sam's birthday, the smaller the given range of days, the smaller the probability that Sam's birthday falls within those dates. If only one date is given, the probability that this is exactly Sam's birthday is zero, as also shown by Eq. (1.4). In other words, although the level of certitude increases as the width of the confidence interval decreases, the opposite applies to the levels of confidence. Hence, the levels of confidence equal to one and zero are assigned to intervals [April 1, April 30] and [April 15, April 15], respectively. Moreover, intuitively, it can also be assessed that, given the confidence interval at level α, the associated level of confidence is $1-\alpha$[1] [FS03, FGS04, FS04, FS05a, FS05b, FS05c].

Possibility Theory

Possibility theory is an uncertainty theory devoted to the handling of incomplete information. It uses a pair of dual set functions (possibility and necessity). Imprecision and vagueness are rather possibilistic than probabilistic. It is reasonable to describe the meaning of information, especially the meaning of incomplete information within a possibilistic framework.

Possibility Theory can be interpreted as a framework for handling incomplete information and aggregate information coming from multiple sources (sensors, experts, databases, etc.).

3.4 Possibility Theory – Axioms

- Finite frame of discernment $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$
- Possibility measure is a mapping $\Pi : 2^\Theta \to [0,1]$
  - Axiom 1 (empty set)

    $$\Pi(\varnothing) = 0.$$

  - Axiom 2 (universal set)

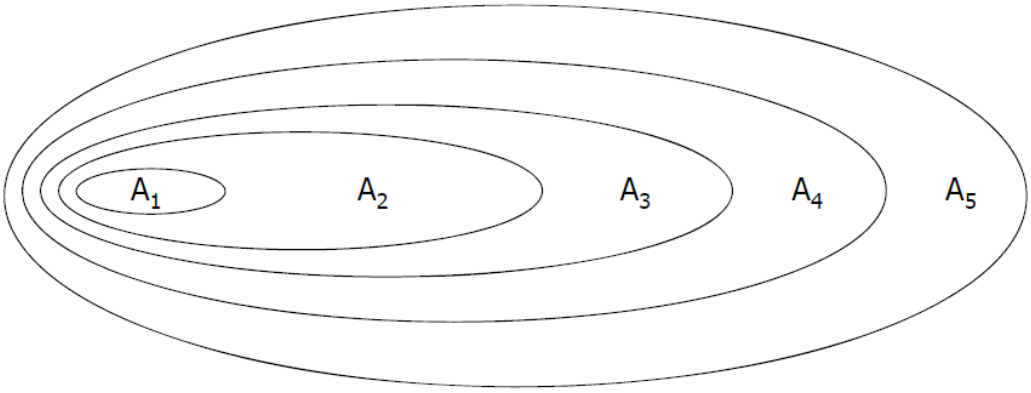
    $$\Pi(\Theta) = 1.$$

  - Axiom 3 (maxitivity)

    $$\Pi(A \cup B) = \max[\Pi(A), \Pi(B)]$$

3.4 Possibility Theory – Consonant Sets

- Definition: **Consonant sets** aka **Nested sets**



Nested Sets: $A_1 \subset A_2 \subset A_3 \subset A_4 \subset A_5$ (Consonant Evidence)

$$\boxed{\mathrm{Bel}(A_i \cap A_j) = \min\left[\mathrm{Bel}(A_i), \mathrm{Bel}(A_j)\right]}$$

$$\boxed{\mathrm{Pl}(A_i \cup A_j) = \max\left[\mathrm{Pl}(A_i), \mathrm{Pl}(A_j)\right]}$$

Necessity and possibility function

If Belief and Plausibility are based on nested sets, then the intersection and union of two nested sets are called Necessity* (credibility, confidence) and Possibility (functions).

$$\mathrm{Pl}(A_i \cup A_j) = \max\left[\mathrm{Pl}(A_i), \mathrm{Pl}(A_j)\right] \Leftrightarrow \Pi(A_i \cup A_j) = \max\left[\Pi(A_i), \Pi(A_j)\right]$$

$$\mathrm{Bel}(A_i \cap A_j) = \min\left[\mathrm{Bel}(A_i), \mathrm{Bel}(A_j)\right] \Leftrightarrow \mathrm{N}(A_i \cap A_j) = \min\left[\mathrm{N}(A_i), \mathrm{N}(A_j)\right]$$

$$\boxed{\mathrm{N}(A) + \mathrm{N}\left(A^c\right) \le 1}$$

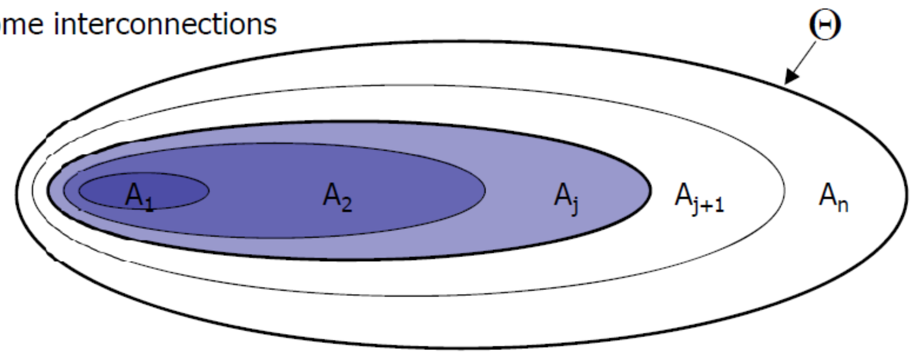
$$\boxed{\Pi(A) + \Pi\left(A^c\right) \ge 1}$$

$$\boxed{\Pi(A) = 1 - \mathrm{N}\left(A^c\right), \quad \mathrm{N}(A) = 1 - \Pi\left(A^c\right).}$$

Possibility and Necessity implications based on
$\max\left[\Pi(A), \Pi(A^c)\right] = 1, \min\left[\mathrm{N}(A), \mathrm{N}(A^c)\right] = 0$

- Let i) $\mathrm{N}(A) > 0 \Rightarrow \Pi(A) = 1,$
  ii) $\Pi(A) < 1 \Rightarrow \mathrm{N}(A) = 0.$

- Some interconnections



$$\mathrm{N}(A_j) = \sum_{k=1}^{j} m(A_k) \qquad \Pi(A_j) = \sum_{k=1}^{n} m(A_k) = 1$$

$$\mathrm{N}(A_j^c) = 0 \qquad \Pi(A_j^c) = \sum_{k=j+1}^{n} m(A_k)$$

3.4 Possibility Theory – Possibility Distribution Function

- Degree of Possibility $\theta \in A \subset \Theta$

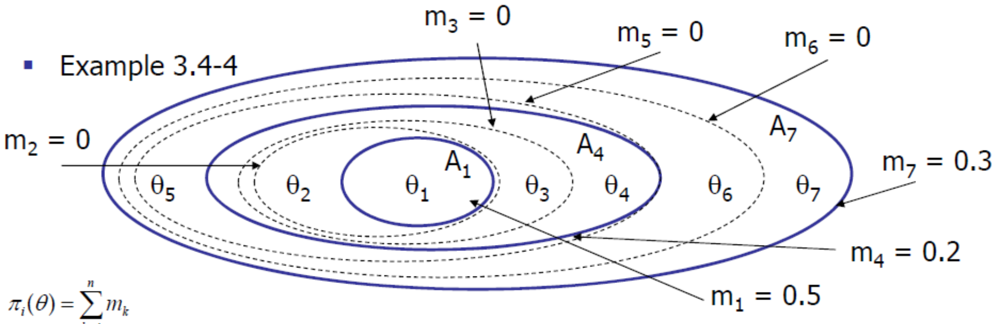
$$\Pi(A) = \max_{\theta \in A}\left(\pi_A(x) \mid x = \theta \in A\right)$$

- Degree of Necessity (confidence) $\theta \notin A \subset \Theta$

$$\mathrm{N}(A) = \min_{\theta \notin A}\left(1 - \pi_A(x) \mid x = \theta \notin A\right)$$

3.4 Possibility Theory – Possibility Distribution Function

- Example 3.4-4



$$\pi_i(\theta) = \sum_{k=i}^{n} m_k$$

$$\begin{aligned}
\pi_1 &= 0.5+0+0+0.2+0+0+0.3 = 1.0 & A_1 &: \pi_1(\theta) = \Pi(\{\theta_i\}) = \Pi(\{\theta_1\}), \\
\pi_2 &= 0+0+0.2+0+0+0.3 = 0.5 & A_2 &: \pi_2(\theta) = \Pi(\{\theta_i\}) = \Pi(\{\theta_1,\theta_2\}), \\
\pi_3 &= 0+0.2+0+0+0.3 = 0.5 & A_3 &: \pi_3(\theta) = \Pi(\{\theta_i\}) = \Pi(\{\theta_1,\theta_2,\theta_3\}), \\
\pi_4 &= 0.2+0+0+0.3 = 0.5 & A_4 &: \pi_4(\theta) = \Pi(\{\theta_i\}) = \Pi(\{\theta_1,\theta_2,\theta_3,\theta_4\}), \\
\pi_5 &= 0+0+0.3 = 0.3 & A_5 &: \pi_5(\theta) = \Pi(\{\theta_i\}) = \Pi(\{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5\}), \\
\pi_6 &= 0+0.3 = 0.3 & A_6 &: \pi_6(\theta) = \Pi(\{\theta_i\}) = \Pi(\{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6\}), \\
\pi_7 &= 0.3 = 0.3 & A_7 &: \pi_7(\theta) = \Pi(\{\theta_i\}) = \Pi(\{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7\}).
\end{aligned}$$

- The function $\pi_A(x)$ reflects the more or less plausible values of the unknown quantity roughly. We tend to another notation proposed by Dubois and Prade to circumvent any ambiguity:

$$\boxed{\pi_A(\theta \mid \theta \in A) = \mu_A(\theta \in A \mid \theta).}$$

The possibility that $x = \theta$, knowing that $\theta \in$ A, is $\pi(\theta \mid$ A). The level of possibility is described, knowing that θ is element of a crisp set A, $\theta \in$ A, with $\mu(\mathrm{A} \mid \theta)$.

- An interesting interrelationship exists between α-cuts, basic belief assignments m, possibility distributions, and membership functions ***iff the sets are nested***.

- Membership Function as graded possibility
  - It follows that $\alpha_i > \alpha_{i+1}$ and $A_{\alpha_i} \subset A_{\alpha_{i+1}}$.
  - Therefore, necessity and possibility can be rewritten in terms of alpha-cuts:

$$\mathrm{N}(A_{\alpha_i}) = \sum_{k=1}^{i} m(A_{\alpha_k})$$

$$\mathrm{N}(A_{\alpha_i}^c) = 0$$

$$\Pi(A_{\alpha_i}) = \sum_{k=1}^{n} m(A_{\alpha_k}) = 1$$

$$\Pi(A_{\alpha_i}^c) = \sum_{k=i+1}^{n} m(A_{\alpha_k})$$