Uncertainty and information

1 - Ignorance, uncertainty, distortion, inaccuracy, vagueness, etc.

- Practical AI systems are constrained to deal with imperfect knowledge and are thus said to reason approximately under conditions of ingnorance
- The concept of uncertainty can be understood in the context of ignorance.
- A taxonomy of ignorance was created by Smithson [a].

[a] M. Smithson, Ignorance and uncertainty: Emerging Paradigms. Springer-Verlag. New York 1989.

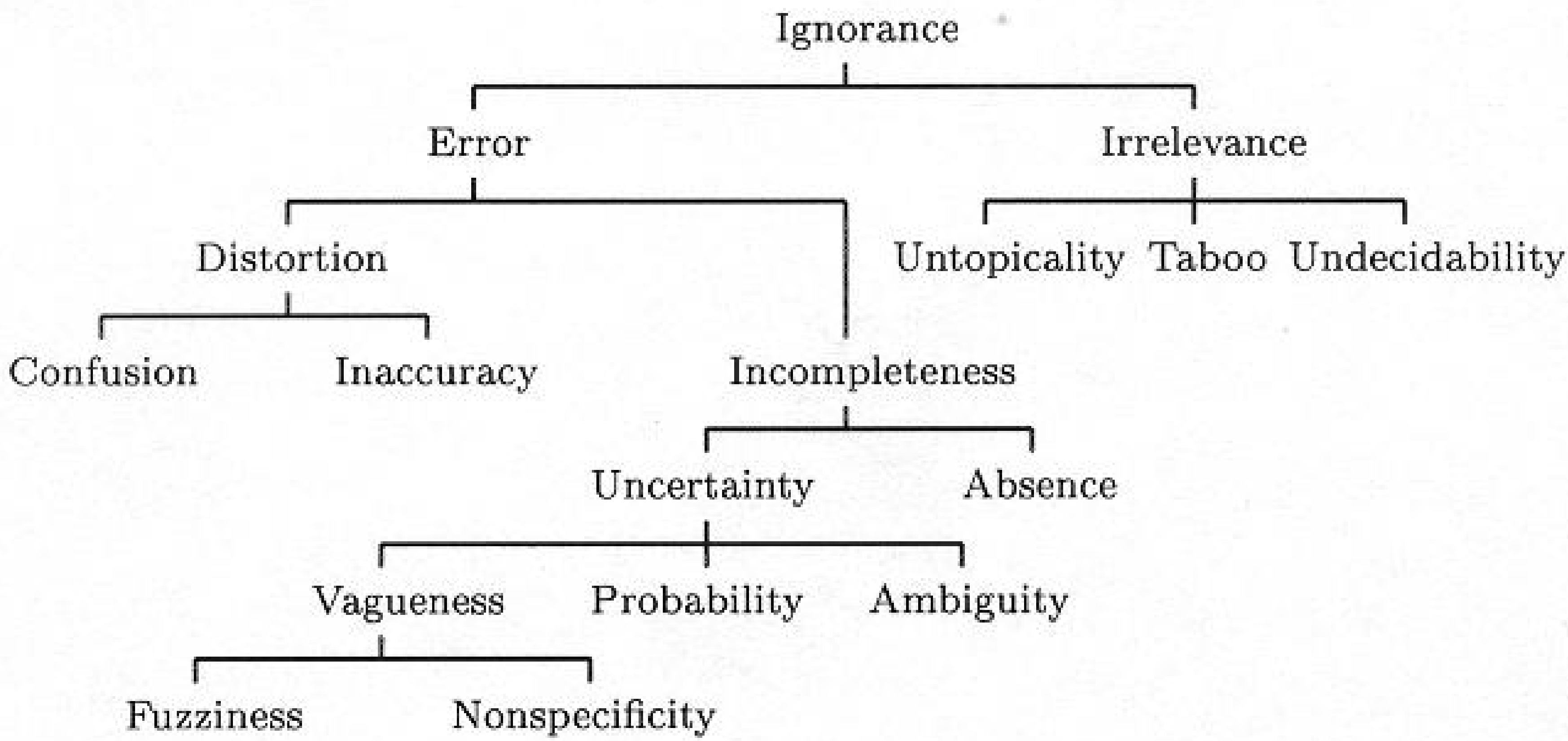


Fig. 1. Smithson's taxonomy of different types of ignorance.

2 - The taxonomy are not well established

- According to Philippe Smets imprecision, inconsistency and uncertainty are imperfect information types.
- Imprecision and inconsistency are properties related to the content of of the statement, i.e., they are properties of the information.
- Uncertainty is a property that results from a lack of information about the world for deciding if the statement is true or false, i.e., it is a property between the information and our knowledge about the world.

3 - Imprecision x Uncertainty

- John has at least two children am I am sure about it. Number of children is imprecise but certain.
- John has three children but I am not sure about it. Number of children is precise but uncertain.

4 - Uncertainty x Information

According to George Klir:

- Uncertainty: result of some information deficiency
- Information: uncertainty reduction.

A fully operation theory requires:

- First level: appropriate mathematical representation of the conceived type of uncertainty.
- Second level: a calculus for manipulation.
- Third level: a meaningful way of measuring relevant uncertainty.
- Fourth level: development of a methodology for making uncertainty principles operational within the theory.

5 - Precise and Imprecise probabilities

- Precise Probability:
  - events are required to be disjoint,
  - probability o each event is required to be expressed by a real number in the unit interval [0,1].
- Imprecise Probability:
  - In practice, precise probability is dificult to meet.
  - Unavoidable measurement errors, insufficient statistical information, missing data, conflicting evidence might arise.
  - In particular, in many situations we are dependent on subjective human judgments.

6 - Imprecise probabilities: are they real ?

- The first throughout study of imprecise probability was taken by Peter Walley [a] (1991).
- The main result: reasoning and decision making based on imprecise probabilities satisfy the principle of coherence and avoidance of sure loss, which are generally viewed as principles of rationality.
- Hence, the requirement of precision (or equivalently, the additivity axiom) cannot be justified as inevitable for rationality.

---

[a] Peter Walley, Statistical reasoning with imprecise probabilities, Chapman-Hall, 1991.

7 - The Quantification of Imprecision: Fuzzy Sets

- In the classical theory, sets are *crisp* in the sense that one element belongs to it or is excluded from it.
- Zadeh (1965) introduced the idea of fuzzy sets.
- Fuzziness is a property related to the use of vague predicates like:
  - *John is Tall*.

  In this case, we want to express the degree of membership of the element $x =$ *height of John* in the fuzzy set $A = Tall$.
- The function $\mu_A(x)$ specifies the degree of membership of the element x in the fuzzy set A.

- In this framework, new concepts like: fuzzy numbers (several, few), fuzzy quantifiers (most), fuzzy predicates (tall), linguistic hedges (very), can be formalised.

- Classical set operators like union, intersection and negation have been generalised:
  - Union: $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$.
  - Intersection: $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$.
  - Negation: $\mu_{\overline{A}}(x) = 1 - \mu_A(x)$.

8 - The Quantification of Uncertainty: Sugeno's Fuzzy Measures

- Another concept developed by Sugeno (1977) received the label fuzzy.

- Sugeno studied functions that express uncertainty with a statement
  x *belongs to* S,
  where S is a crisp set and x is a particular arbitrary element of the universe Ω.

- The fuzzy measure is a function

$$g : 2^{\Omega} \rightarrow [0, 1]. \tag{1}$$

  Where 2^{Ω} is the power set of crisp subsets of the universe Ω.

- In order to qualify as a fuzzy measure the function g must have certain properties. The following axioms must be verified:

  Axiom 1 (boundary conditions) $g(\emptyset) = 0$ *and* $g(\Omega) = 1$.

  Axiom 2 (monotonicity) *For every,* $A, B \in 2^{\Omega}$*, if* $A \subseteq B$ *then* $g(A) \leq g(B)$

  Axiom 3 (continuity) *For every sequence* $A_i \in 2^{\Omega} \mid i \in \mathbb{N}$ *of subsets of* X*, if either* $A_1 \subseteq A_2 \subseteq \ldots$ *or* $A_1 \supseteq A_2 \supseteq \ldots$ *(i.e. the sequence is monotonic), then*

$$\lim_{i \to \infty} g(A_i) = g(\lim_{i \to \infty} A_i) \tag{2}$$

  It is applicable only to an infinite universal set. The axiom requires that for every infinite sequence A_1, A_2, ..., of nested (monotonic) subsets of X that converge to the set

$$A = \lim_{i \to \infty} A_i. \tag{3}$$

- The concept of fuzzy measure fits probability measures, necessity measures, belief functions, etc.
- It has been called *fuzzy measure* but should not be confused with fuzzy sets.

9 - Belief and Plausibility Measures

- The theory of Belief functions provides one way to use mathematical probability in subjective judgement.
- It was pioneered by Dempster [a] (1967) and Shafer [b] (1976).
- It is a generalisation of the Bayesian theory of subjective probability.
- When we use Bayesian theory to quantify judgements about a question, we must assign probabilities to the possible answers to that question.
- The theory of Belief is more flexible, it allows us to derive degrees of belief for a question from probabilities for a related question.

---

[a] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Annals of mathematical Statistics, 38, 325–339, 1967.

[b] G. Shafer, A mathematical theory of evidence, Princeton University Press, 1976.

10 - Basic Probability Assignment

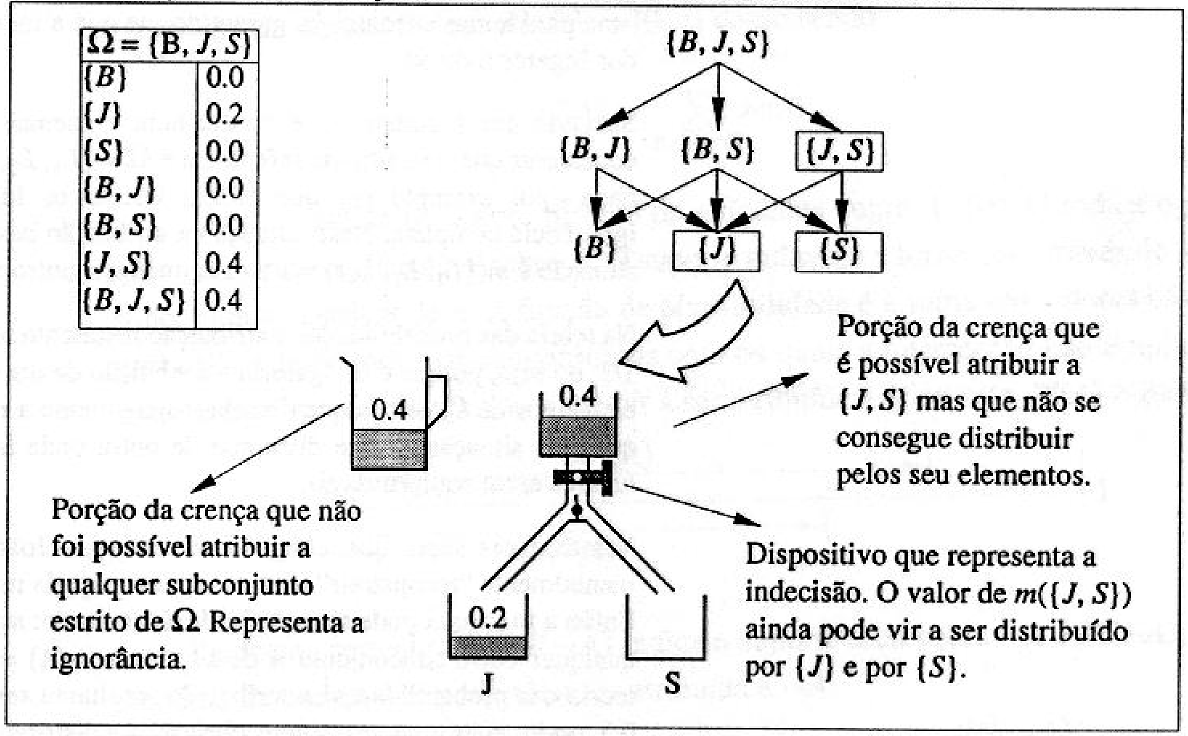
- Let q be a variable and Ω be the set of all possible values of q.
- The propositions can be represented as:
  - The value of q is in A, where $A \subseteq \Omega$.
- In the Dempster-Shafer theory, Ω is the frame of discernment.
- In Probability theory, it is necessary to designate the degree of belief for each one of the elements of Ω.
- In Belief theory, the probability assignment is not made over Ω, but on the space of all subsets of Ω, which is denoted by 2^Ω.
- Let $m : 2^\Omega \to [0, 1]$, be a basic assignment of probability function such that:
  1. $m(\emptyset) = 0$,

2.

$$\sum_{A \subseteq 2^{\Omega}} m(A) = 1. \tag{4}$$

- m is the degree of evidence that X belongs to the set A.
- (Ω, m) is defined as the body of evidence.
- The following figure illustrates these concepts.

DISTRIBUIÇÃO DA CRENÇA UNITÁRIA PELOS SUBCONJUNTOS DE Ω



- A hydraulic analogy is used illustrate the basic probability assignment.
- $m(\{J, S\}) \neq 0$ means that there is a belief that can be assigned to $\{J, S\}$, but there is no further evidence for a finer distribution.

11 - Where is John's wallet ?

- John has had a very hectic morning visiting three different places; L_1, L_2 and L_3. In the afternoon, while driving him to the train station, he has realised that he lost his wallet, but he is sure that he let the wallet in one of the three places where he has been, but does not know where exactly. He has left for his destiny but asked me to help him find his wallet.

- Knowing that the wallet is on one the three places, our frame of discernment is $\Omega = \{L_1, L_2, L_3\}$.

- As John has not informed any other information, like the order which he visited the places, we are in a complete ignorance state.

- In this context, the basic probability assignment must be $m(\{L_1, L_2, L_3\}) = 1$ and for any other subset $A \subset \Omega$ we must have $m(A) = 0$.

- In probability theory, the probability assignment would be $P(L_1) = 1/3,\ P(L_2) = 1/3,\ P(L_3) = 1/3.$

- There must be in this case, assignments to all elements of Ω, therefore our ignorance must be divided in equal parts. It would be the same to consider the three places equiprobable.

- After John arrived to his destiny, he phones me saying that he mentally reconstructed his steps during the morning and he has 40% belief that he let his wallet on L_3.

- In this case:
  - $m(\{L_1, L2\}) = 0.6$,
  - $m(L_3) = 0.4$,
  - $m(A) = 0$ for all other subset $A \subset \Omega$.
- In the case of probability theory, this would be:
  - $P(L_3) = 0.4$,
  - $P(L_1) = 0.3$,
  - $P(L_2) = 0.3$.

- A belief measure is a function:

$$Bel : 2^{\Omega} \rightarrow [0, 1], \tag{5}$$

which can be defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B). \tag{6}$$

- It satisfies axioms 1-3 of fuzzy measures and the following axiom:

$$Bel(A_1 \cup A_2 \cup \ldots \cup A_n) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \ldots + (-1)^{n+1} Bel(A_1 \cap A_2 \cap \ldots \cap A_n) \tag{7}$$

for every $n \in \mathbb{N}$ and every collection of subsets of X.

- $m(A)$ gives the direct evidence support of A, $Bel(A)$ is total support, direct and indirect, of A.

- A plausibility measure is a function

$$Pl : 2^{\Omega} \rightarrow [0, 1], \tag{8}$$

which can be defined as:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \tag{9}$$

- It satisfies the axioms 1-3 and the following axiom:

$$Pl(A_1 \cap A_2 \cap \ldots \cap A_n) \geq \sum_i Pl(A_i) - \sum_{i<j} Pl(A_i \cup A_j) + \ldots + (-1)^{n+1} Pl(A_1 \cup A_2 \cup \ldots \cup A_n) \tag{10}$$

for every $n \in \mathbb{N}$ and every collection of subsets of X.

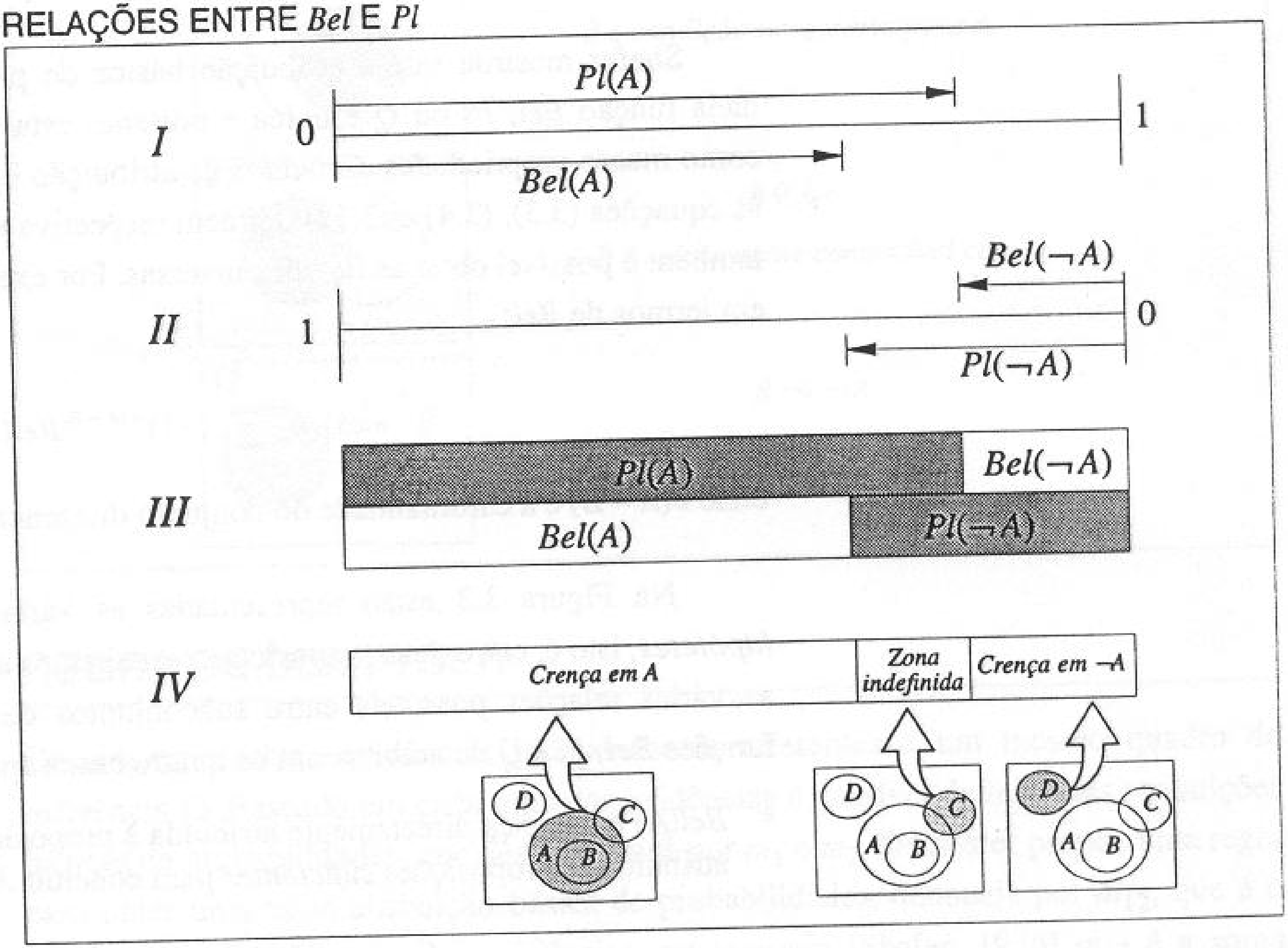
- The plausibility measure is the sum of all basic probability assignment of all subsets with non null intersection with A.

- Therefore, it measures the maximum belief that can be credited to A.

12 - Some properties

- It is easy to note that $Pl(A) \geq Bel(A)$.
- Similarly, $Bel(\neg A)$ indicates the total support to $\neg A$ and $Pl(\neg A)$ measures the plasubility of $\neg A$.
- As Belief and Plausibility functions are dual measures then:
  - $Pl(A) = 1 - Bel(\neg A)$,
  - $Bel(A) = 1 - Pl(\neg A)$.
- This can be represented by the following figure.

RELAÇÕES ENTRE *Bel* E *Pl*



13 - Combining Evidences

- Evidences from two different independent sources inside the same frame of discernment Ω can be combined

- Dempster's Rule of Combination:

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} \tag{11}$$

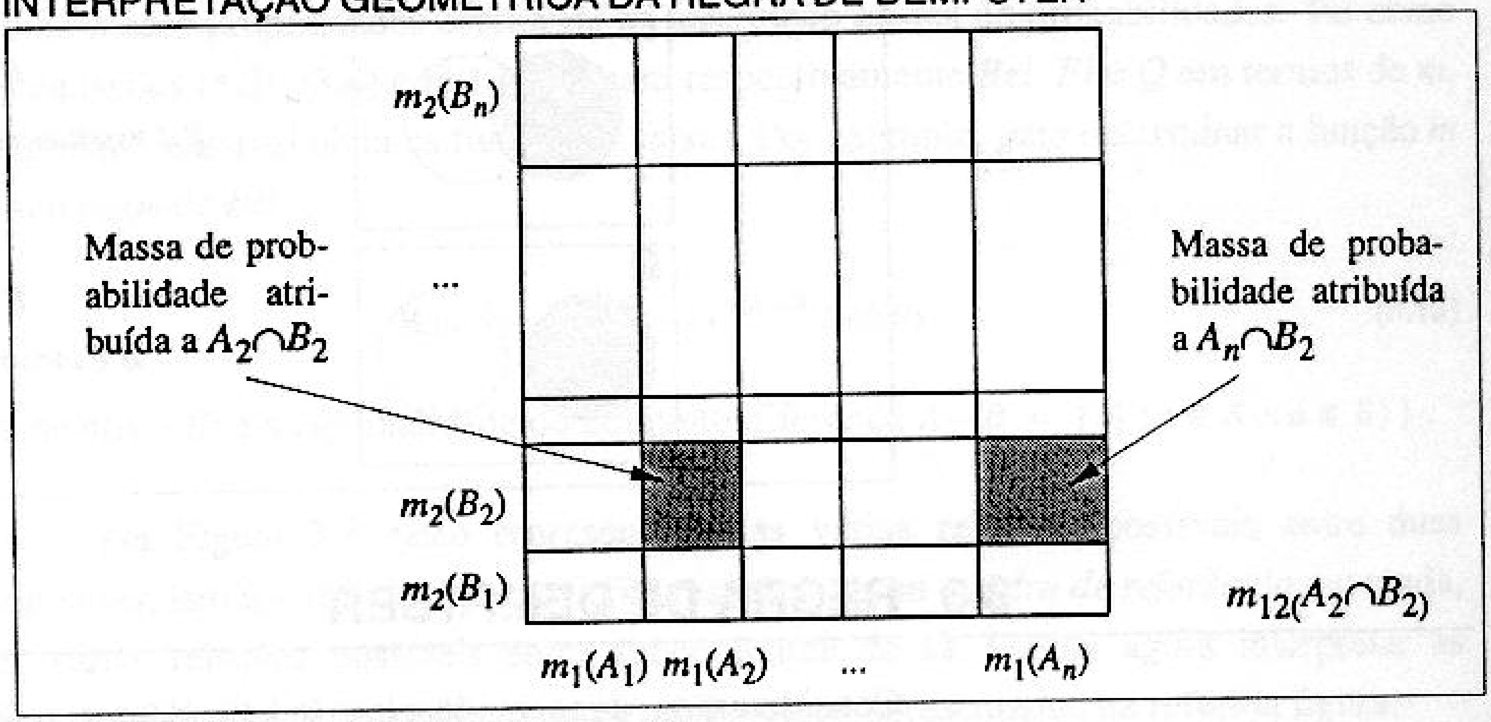
for $A \neq \emptyset$, where

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C). \tag{12}$$

- This can be illustrated by the following figure.

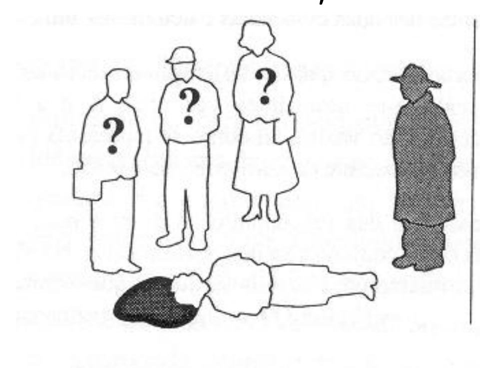
–

INTERPRETAÇÃO GEOMÉTRICA DA REGRA DE DEMPSTER



14 - Who murdered Karen ?

- Four persons are locked in a room: Bob, Jim, Sally and Karen. Suddenly, there is a electricity shortage supply and the room become dark. When the electricity supply is normalised Bob, Jim and Sally realises that Karen is dead. A police inspector arrives and concludes that Karen was murdered, and the



murder is one of the three persons.

- After detailed investigation and based on evidences found in the room, the inspector produces a basic probability assignment which reflects his beliefs and doubts.

$\Omega = \{B, J, S\}$	
$\{B\}$	0.1
$\{J\}$	0.2
$\{S\}$	0.1
$\{B, J\}$	0.1
$\{B, S\}$	0.1
$\{J, S\}$	0.3
$\{B, J, S\}$	0.1

- The functions Bel e Pl derived from these basic probability assignment are summarised in the following table:

	m_1	Bel_1	Pl_1
$\{B\}$	0.1	0.1	0.4
$\{J\}$	0.2	0.2	0.7
$\{S\}$	0.1	0.1	0.6
$\{B, J\}$	0.1	0.4	0.9
$\{B, S\}$	0.1	0.3	0.8
$\{J, S\}$	0.3	0.6	0.9
$\{B, J, S\}$	0.1	1.0	1.0

- It is easy to verify that Jim is the main suspect, but the hypothesis of Sally being the murderer is very plausible $Pl(\{S\}) = 0.6$.

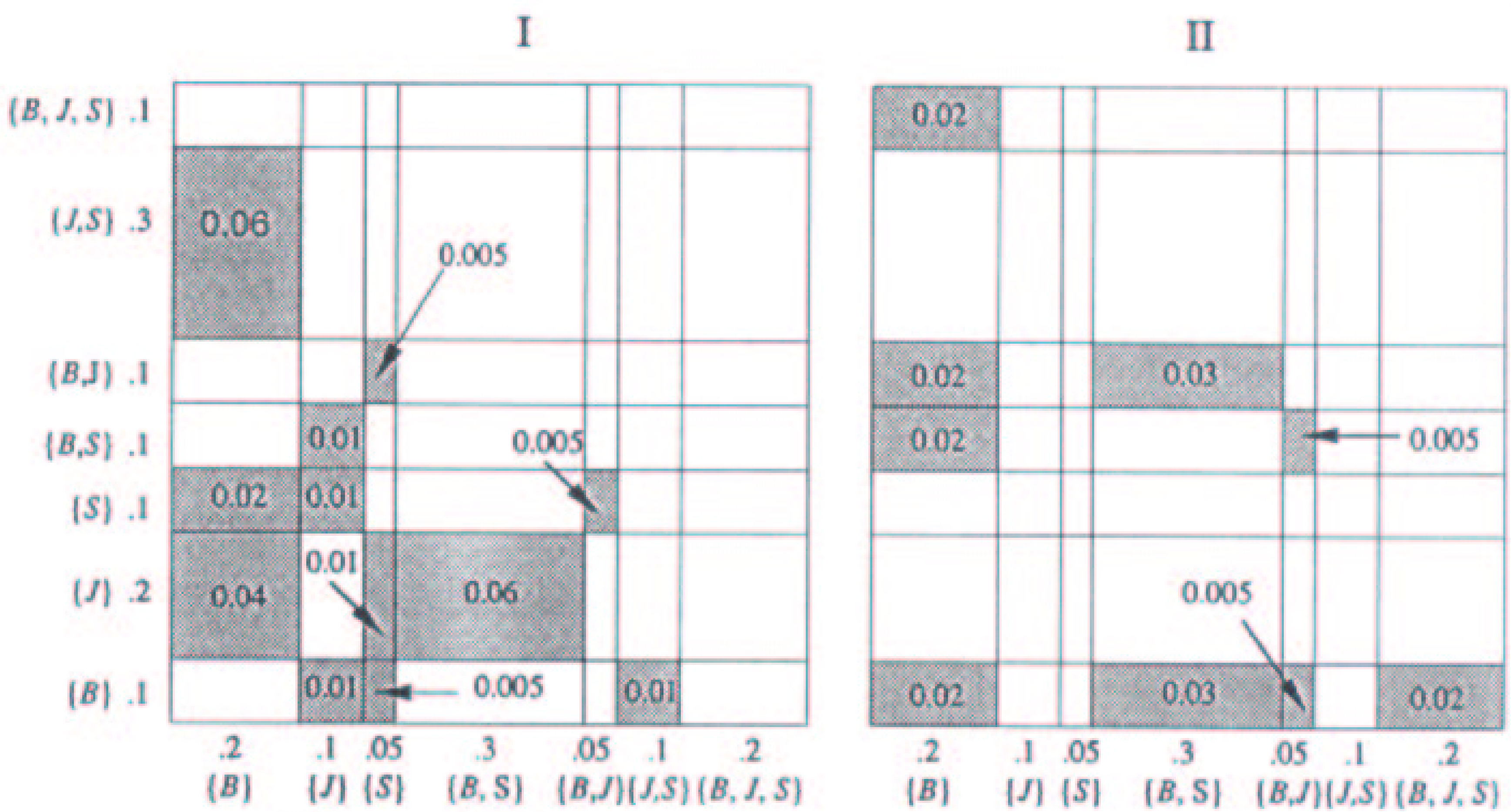
- Later, based on further evidence, information about the dead and her personal relationships with the suspects, the inspector builds up another basic probability assignment m_2.

$m_1 : 2^\Omega \rightarrow [0,1]$	
$\Omega = \{B, J, S\}$	
$\{B\}$	0.1
$\{J\}$	0.2
$\{S\}$	0.1
$\{B, J\}$	0.1
$\{B, S\}$	0.1
$\{J, S\}$	0.3
$\{B, J, S\}$	0.1

$m_2 : 2^\Omega \rightarrow [0,1]$	
$\Omega = \{B, J, S\}$	
$\{B\}$	0.2
$\{J\}$	0.1
$\{S\}$	0.05
$\{B, J\}$	0.05
$\{B, S\}$	0.3
$\{J, S\}$	0.1
$\{B, J, S\}$	0.2

- Now, using the Dempster's rule the beliefs of both evidences can be combined in a new basic probability assignment m_{12}.

- The
  following figure illustrates the combination of the two evidences.



Massas atribuídas aos conjuntos contraditórios

Intersecções que resultam em {B}

- The results for m_{12}, Bel_{12}, Pl_{12} are summarised in the following table.

	m_{12}	Bel_{12}	Pl_{12}
$\{B\}$	0.225	0.225	0.4
$\{J\}$	0.219	0.219	0.485
$\{S\}$	0.25	0.25	0.451
$\{B, J\}$	0.04	0.515	0.781
$\{B, S\}$	0.105	0.549	0.751
$\{J, S\}$	0.131	0.6	0.775
$\{B, J, S\}$	0.03	1.0	1.0

15 - The Classical Probability Measure

- When the axiom for belief measures is replaced with a stronger axiom:

$$Bel(A \cup B) = Bel(A) + Bel(B), \tag{13}$$

  whenever $A \cap B = \emptyset$ we obtain a special type of belief measures the classical probability measures.

16 - There are many possible interpretations for probabilities

- **The classical theory**: as defended by Laplace, assumes the existence of a fundamental set of equipossible events.

- **The relative frequency theory**: in this case, probability is essentially the convergence limit of relative frequencies under repeated independent trials.

- **The subjective (or Bayesian) probability**: in this case the probability measure quantifies the credibility that an event will occur. It is a subjective, personal measure.

17 - The Kolmogorov's axioms

Definition 1 *Let Ω be a set of mutually exclusive events.*

Definition 2 *let F be a set of subsets of Ω such that:*

1. $\Omega \in F$,
2. $E_1, E_2 \in F \Rightarrow E_1 \cup E_2 \in F$,
3. $E \in F \Rightarrow \neg E \in F$.

then F is defined by event algebra related Ω.

Definition 3 *Let Ω, F and a function P which associates a real number to each $E \in F$ If P satisfies the following axioms:*

1. $P(E) \geq 0 \quad \forall E \in F$,

2. $P(\Omega) = 1$,

3. *If E_1 e E_2 are two subsets with $E_1 \cap E_2 = \emptyset$ then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$. The triple (Ω, F, P) are defined as probability space, and P is the probability measure of F.*

18 - Some consequences

- According to the third axiom, the probability associated to any set of events its the sum of probabilities of its disjoint components.
- Any event event A can be written as $(A \cup B) \cup (A \cap \neg B))$, therefore we have:

$$P(A) = P(A \cup B) + P(A \cap \neg B) \tag{14}$$

.

- This can be extended to n mutually exclusive events $B_1, B_2, \ldots, B_n$,

$$P(A) = \sum_{i=1}^{n} P(A \cup B_i). \tag{15}$$

- A direct consequence of this argument is:

$$P(A) + P(\neg A) = 1. \tag{16}$$

19 - The Bayesian Theory

- The classical theory, pioneered by J. Neyman, E.S. Pearson e R.A. Fisher, considers only information obtained by sampling tohgether with the frequentist concept of probability.

- On the contrary, the subjectivist approach considers that probabilities are conditioned on prior information.

- Probabilities which are dependent on additional information are defined as conditional probabilities:

$$P(E|K) = \frac{P(E,K)}{P(K)}. \tag{17}$$

- If $P(A|B) = P(A)$ then A and B are independent events,

- If $P(A|B,C) = P(A|C)$ then A and B are conditionally independent on C.

20 - Thinking axiomatically

- Let (Ω, F, P) be probability space and $E_1 \in F$ such that $P(E_1) > 0$,

- For $\forall E_2 \in F$ the conditional probability $P(E_2|E_1)$ can be defined as:

$$P(E_2|E_1) = \frac{P(E_2 \cap E_1)}{P(E_1}. \tag{18}$$

- Therefore the joint probability can be written as:

$$P(E_2 \cap E_1) = P(E_1)P(E_2|E_1). \tag{19}$$

- The probability of an event A can be calculated conditioned on a mutually exclusive set of events $\{B_1, B_2, \ldots, B_n\}$:

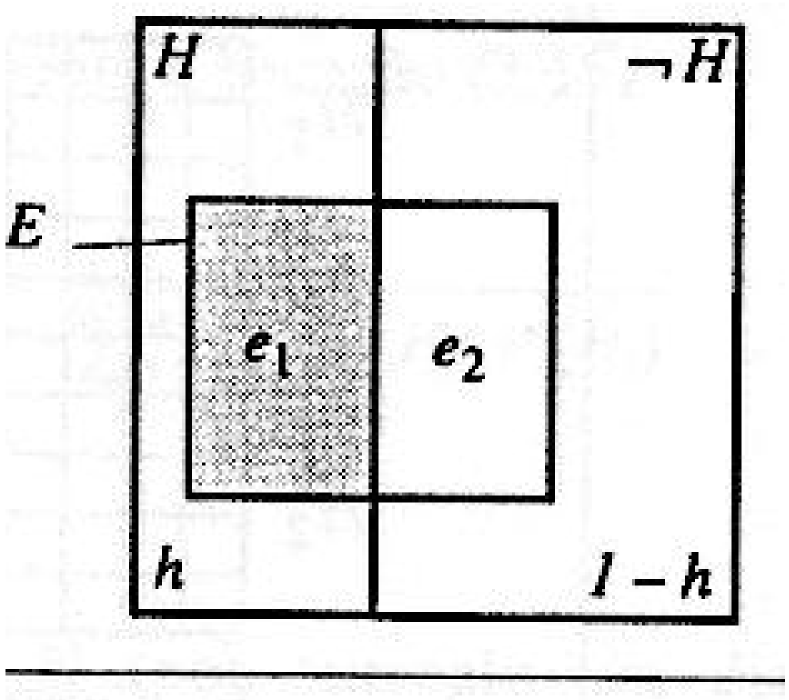
$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{20}$$

- It is possible to write the a posteriori probability as:

$$P(E_2|E_1) = \frac{P(E_1|E_2)P(E_2)}{P(E_1)}. \tag{21}$$

21 - Understanding the Bayesian concept in probabilistic diagrams

- Starting from:
  - $P(H) = h$, $P(\neg H) = 1 - h$,
  - $P(E|H) = e_1/h$, $P(E|\neg H) = e_2/(1 - h)$.



- What does it mean the shadowed area in the figure ?

- from Ω, it represents $P(H \cup E) = e_1/1 = e_1$.
- from H, i.e., if H is certain then it represents $P(E|H) = e_1/h$.
- from E, i.e., if E is certain then $P(H|E) = e_1/(e_1 + e_2)$.

- Then a posteriori probability can be written as:

$$P(H|E) = \alpha P(E|H)P(H). \tag{22}$$

Where $\alpha = P(E) = 1/(e_1 + e_2)$ is a normalisation factor.

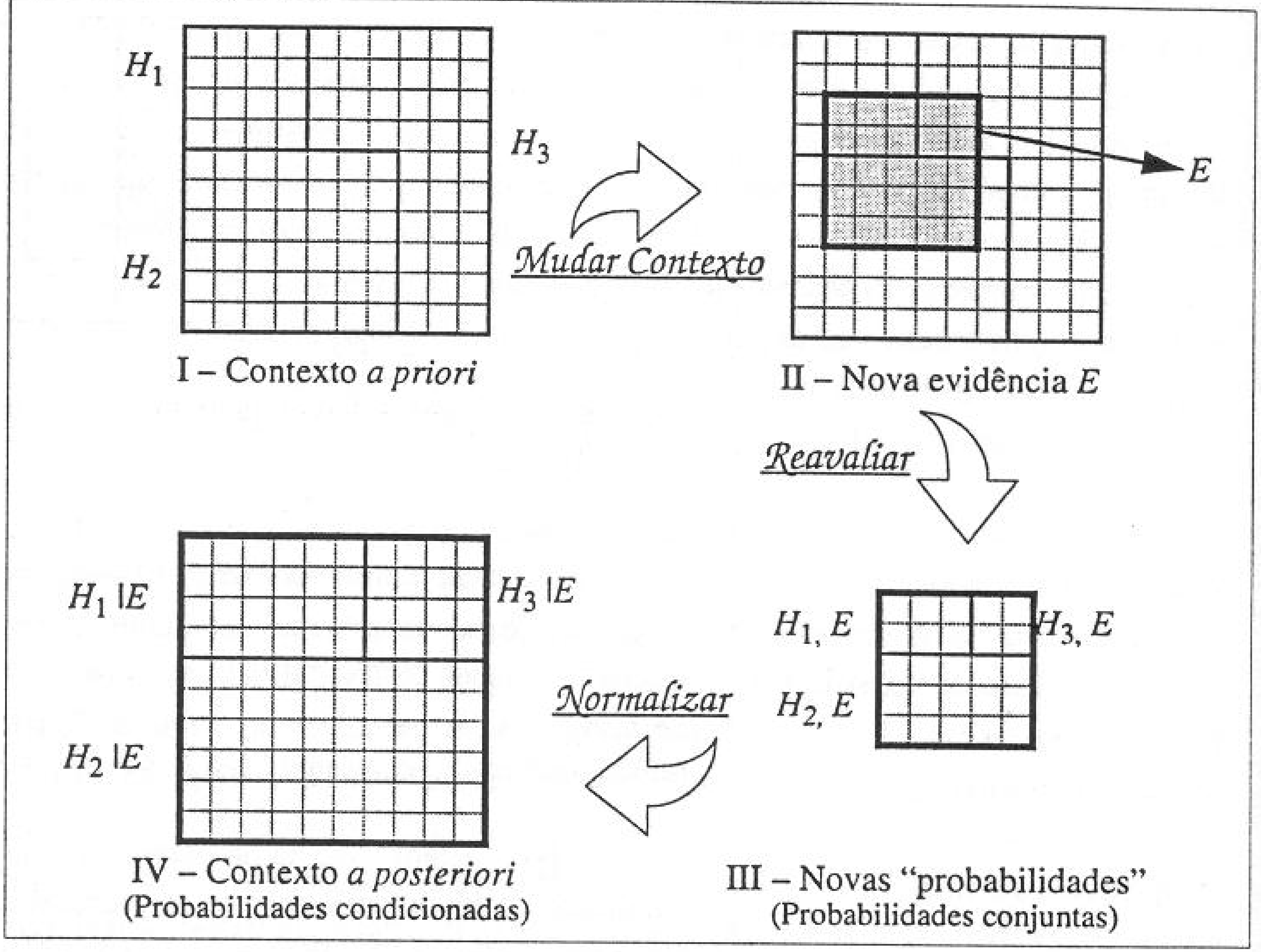
22 - The Bayes theorem

- Given the probabilistic space (Ω, F, P) and a set F of mutually exclusive events $\{H_1, H_2, \ldots, H_n\}$ with non null probabilities, then:

$$P(H_j|E) = \frac{P(E|H_j)P(H_j)}{\sum_{i=1}^{n} P(E|H_i)P(H_i)} \tag{23}$$

- This can be summarised in the following figure:

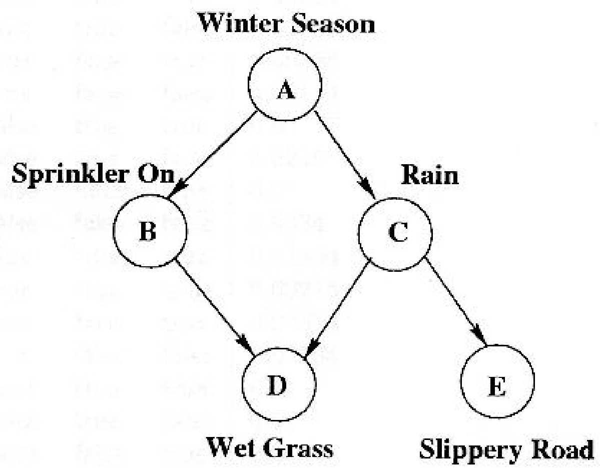
INTERPRETAÇÃO GRÁFICA DO TEOREMA DE BAYES



O surgir de uma nova informação relevante para o problema significa que alguma

23 - Bayesian networks

- A Bayesian network is a graphical representation of a probability distribution.
- Each Bayesian network is comprised of two distinct parts:
  1. A directed acyclic graph,
  2. a set of probability tables.
- Edges represent direct causal influences between variables, i.e., conditional independence.
- it is a parsimonious representation.
- It is possible to infer things like: *given that D is true What state is the most probable state for B or C ?*
- This is illustrated in the following example:



A	T_A
true	.6
false	.4

A	B	T_B
true	true	.2
true	false	.8
false	true	.75
false	false	.25

A	C	T_C
true	true	.8
true	false	.2
false	true	.1
false	false	.9

B	C	D	T_D
true	true	true	.95
true	true	false	.05
true	false	true	.9
true	false	false	.1
false	true	true	.8
false	true	false	.2
false	false	true	0
false	false	false	1

C	E	T_E
true	true	.7
true	false	.3
false	true	0
false	false	1

Figure 1.1: A Bayesian network consisting of two parts: a DAG and a set of conditional probability tables (CPTs). The DAG declares a set of conditional independence constraints. The tables declare conditional probability constraints. The combined set of constraints is satisfied by a unique probability distribution, shown in Figure 1.2.

A	B	C	D	E	
true	true	true	true	true	0.06384
true	true	true	true	false	0.02736
true	true	true	false	true	0.00336
true	true	true	false	false	0.00144
true	true	false	true	true	0.0
true	true	false	true	false	0.0216
true	true	false	false	true	0.0
true	true	false	false	false	0.0024
true	false	true	true	true	0.21504
true	false	true	true	false	0.09216
true	false	true	false	true	0.05376
true	false	true	false	false	0.02304
true	false	false	true	true	0.0
true	false	false	true	false	0.0
true	false	false	false	true	0.0
true	false	false	false	false	0.096
false	true	true	true	true	0.01995
false	true	true	true	false	0.00855
false	true	true	false	true	0.00105
false	true	true	false	false	0.00045
false	true	false	true	true	0.0
false	true	false	true	false	0.243
false	true	false	false	true	0.0
false	true	false	false	false	0.027
false	false	true	true	true	0.0056
false	false	true	true	false	0.0024
false	false	true	false	true	0.0014
false	false	true	false	false	0.0006
false	false	false	true	true	0.0
false	false	false	true	false	0.0
false	false	false	false	true	0.0
false	false	false	false	false	0.09

Figure 1.2: A probability distribution (table) specified by the Bayesian network in Figure 1.1. Each row in the table has two components: an instantiation and a number. The instantiation represents an assignment of values to all network variables and the number represents the probability of its corresponding instantiation. The set of probabilities appearing in the table sum to 1.

24 - Possibility and Necessity Measures

- Incomplete information such as:
  - John's height is above 1.70m,

  implies that any height above 1.70m is possible and any height equal or below 1.70m is impossible.

- This can be represented by a possibility measure $\Pi(x)$ defined on the height domain whose value is:

$$\Pi(x) = \begin{cases} 0 & \text{height} < 1.70\text{m} \\ 1 & \text{height} >= 1.70\text{m} \end{cases} \tag{24}$$

- When the predicate is vague like:
  - John is tall,

  possibility can admit degrees, the largest the degree the largest the possibility.

- Possibility is usually associated with fuzziness, however, non-fuzzy (crisp) events can admit different degrees of possibility. Suppose that there is a box where you try to squeeze soft balls. You can say:

  – It is possible to put 20 balls in it, impossible to put 30 balls in it, quite possible to put 24 balls, but not so possible to put 26 balls, . . .

  These degrees of possibility are degrees of realizability and totally unrelated to any underlying random process.

A salesman could also forecast about next year sales. He could say:

– It is possible to sell about 50K, impossible to sell more than 100K, quite possible to sell 70K, hardly possible to sell more than 90K, . . .

This might express the possible values of sales based on the sale capacity. On the contrary, he could express what he will actually sell next year, but this concerns another problem for which the theories of probability and belief functions are more adequate.

25 - Possibility and Necessity Measures: the definition

- Let Π and N denote a possibility and a necessity measure on 2^{Ω}, respectively. Then,

$$\Pi(A \cup B) = \max[\Pi(A), \Pi(B)] \tag{25}$$

and

$$N(A \cap B) = \min[N(A), N(B)] \tag{26}$$

for all $A, B \in 2\Omega$.

- The necessity and possibility measures are dual measures.

- Let $E \subseteq X$ a certain event, then we can define:

$$\Pi_E(A) = \begin{cases} 1 & A \cap E \neq \emptyset \\ 0 & A \cup E = \emptyset \end{cases} \tag{27}$$

$\Pi_E(A) = 1$ means that A is a possible event.

- If A and $\neg A$ are two complementary events then,

$$\max(\Pi(A), \Pi(\neg A)) = 1. \tag{28}$$

  If two events are complementary at least one of them is possible.

- Let $E \subseteq X$ a certain event, then we can define:

$$NE(A) = \begin{cases} 1 & A \subseteq E \\ 0 & A \supset E \end{cases} \tag{29}$$

  $N_E(A) = 1$ means that A is a certain event, i.e., necessarily true.

- They are related to each other by the equation

$$N(A) = 1 - \Pi(\neg A), \tag{30}$$

  for all $A \in 2^{\Omega}$.

- $\Pi(A)$ is the degree of possibility that A is true.
- The necessity of a proposition is the negation of the possibility of its negation.
- An event is necessary when the occurrence of its negation is impossible.
- An equivalent argument could be:

$$\min(N(A), N(\neg A)) = 0. \tag{31}$$

Which means that an event and its negation cannot be necessary simultaneously.

- Using intuition an event must be possible before being necessary, threfore:

$$\Pi(A) \geq N(A) \tag{32}$$

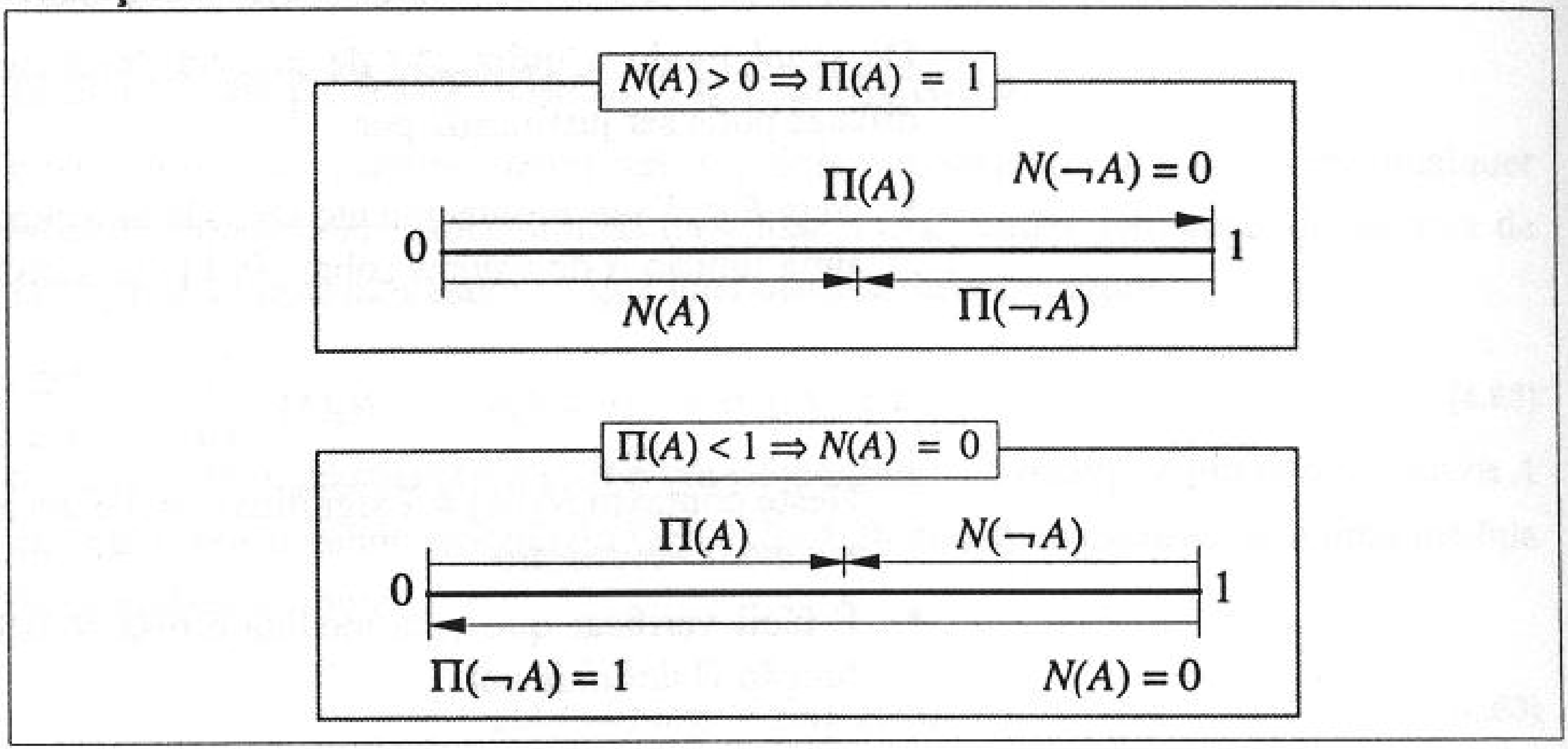
- It is easy to note that:

1. $N(A) > 0 \rightarrow \Pi(A) = 1$

2. $\Pi(A) < 1 \rightarrow N(A) = 0$

- The following figure illustrates the relation between the necessity and possibility measures:
- Some comments:
  - $N(A) + N(\neg A) \leq 1$,
  - $\Pi(A) + \Pi(A) \geq 1$.
  - An event to be considered as something necessary, $N(A) > 0$, must be completely possible i.e. $\Pi(A) = 1$. As a consequence, its complementary cannot be something necessary, i.e. $(N(\neg A) = 0)$.
  - Similarly, If an event is not completely possible $(\Pi(A) < 1)$ then the event is not necessary $(N(A) = 0)$.

RELAÇÕES ENTRE AS MEDIDAS DE NECESSIDADE E POSSIBILIDADE



- Every possibility measure Π on $\mathcal{P}(X)$ can be uniquely determined by a possibilistic distribution function:

$$\pi : X \to [0, 1] \tag{33}$$

via the formula:

$$\Pi(A) = \max_{x \in A} \pi(x). \tag{34}$$

26 - Possibility x Probability

- Consider the following statement:
  - *Hans ate X eggs for breakfast*, with $X \in 1, \ldots, 8$
- We can associate both a possibility distribution (based on our view o ease which Hans can eat eggs) and a probability distribution (based on our observations of Hans at breakfast).
- We could have:

u	1	2	3	4	5	6	7	8
$\Pi_X(u)$	1	1	1	1	0.8	0.6	04	0.2
$Pr_X(u)$	0.1	0.8	0.1	0	0	0	0	0

- So, while is perfectly possible that Hans can eat three eggs for breakfast, he is unlikely to do so.

- There is a heuristic connection between possibility and probability, since if something is impossible it is likely to be improbable.

- A high degree of possibility does not imply a high degree of probability, nor a low degree of probability reflect a low degree of possibility.

- Let A be a nonfuzzy set of X and v be a variable on X. The statement ¨v takes its values in A¨ can be viewed as inducing a possibility distribution π over X associating with each element x the possibility that x is a value of v:

$$\Pi(v = x) = \pi(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

- Next, assume that A is a fuzzy set that acts as a fuzzy restriction on the possible value of v. An extension of our above representation is that A induces a possibility distribution that is equal to μ_A on the values of v:

$$\Pi(v = x) = \pi(x) = \mu_A(x). \tag{36}$$

- Since the expression of a possibility distribution can be viewed as a fuzzy set, possibility distributions may be manipulated by the combination rules of fuzzy sets, and more particular of fuzzy restrictions.

27 - Relation between fuzziness and possibility

(Fuzzy sets and Systems: Theory and Applications, Didier Dubois, Henri Prade, p.136-137)

- Note that a fuzzy set and a possibility distribution have a common mathematical expression, the underlying concepts are different.
- Zadeh introduced both the concept of fuzzy set (1965) and the concept of possibility measure (1978).
- The first allows one to describe the grade of membership of a well-known individual to an ill-defined set.
- The second allows one to describe what are the individuals that satisfy some ill-defined constraints or that belong to some ill-defined sets.
- A fuzzy set A can be viewed as a fuzzy value that we assign to a variable. Viewed as a possibilistic restriciton A is the fuzzy

set of nonfuzzy values that can be possibly assigned to v.

- For instance, $\mu_{Tall}(x)$ quantifies the membership of a person with height h to the set of *Tall* men and $\pi_{\mathsf{Tall}}(h)$ quantifies the possibility that the height of a person is h given the person belongs to the set of *Tall* men.

- Zadeh's possibilistic principle postulates the following equality:

$$\pi_{Tall}(h) = \mu_{Tall}(h), \quad \forall h \in H, \tag{37}$$

where H is the set of height.

- The writing is often confusing and would have been better written as:

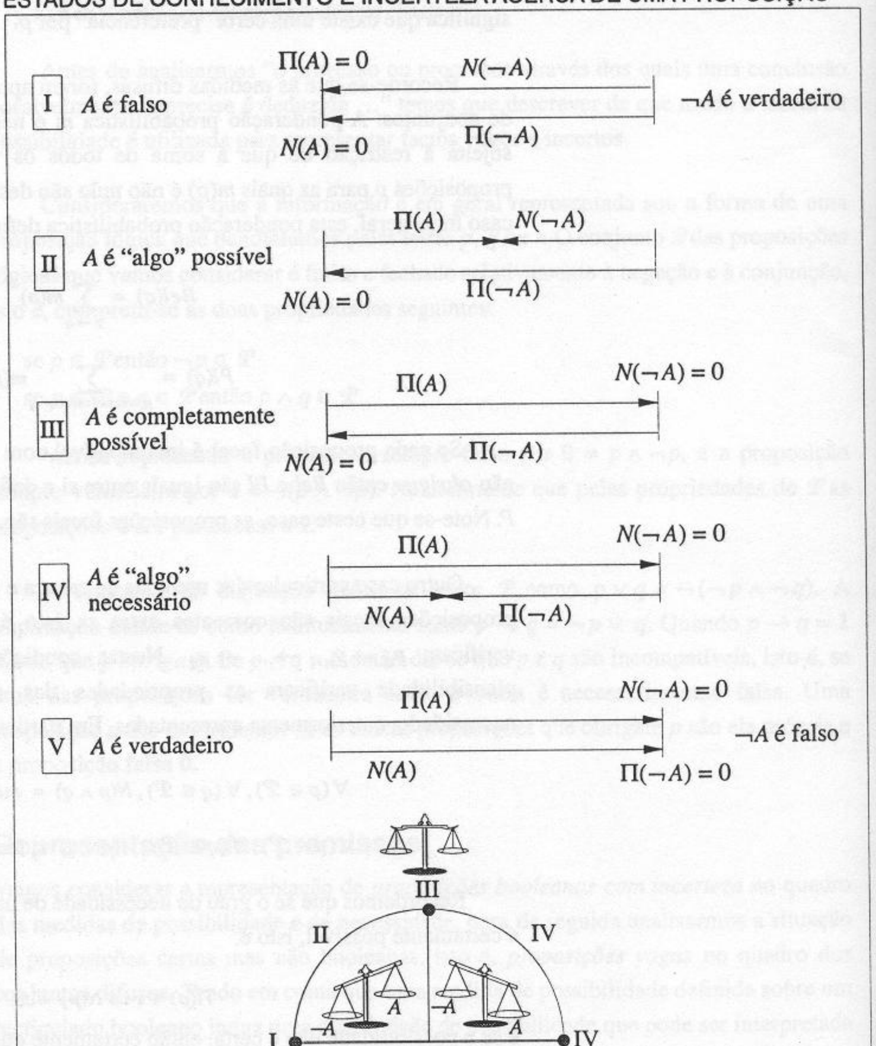
$$\pi(h|Tall) = \mu(Tall|h), \quad \forall h \in H, \tag{38}$$

or still better: If $\mu(Tall|h) = x$ then $\pi(h|Tall) = x$

- The difference is analogous to the difference between a probability distribution $p(x|\theta)$ (the probability of the observation x given the hypothesis θ) and a likelihood function $l(\theta|x$ (the likelihood of the hypothesis θ given the observation x).

28 - The following figure illustrates the states of knowledge and uncertainty of typical propositions

ESTADOS DE CONHECIMENTO E INCERTEZA ACERCA DE UMA PROPOSIÇÃO



Na Figura 4.30 estão representadas as situações qualitativamente diferentes que podemos considerar ao representar a incerteza acerca de uma proposição através das medidas de possibilidade e de necessidade.