Representing Uncertainty

Material used

- Halpern: Reasoning about Uncertainty. Chapter 2

1 Motivating examples
2 Possible worlds
3 Probability measures
4 Lower and upper probabilities
5 Inner and outer measures
6 Possibility measures
7 Ranking functions
8 Choosing a formalism

0 Introduction

Standard way to represent uncertainty:

- Bayesian probability theory

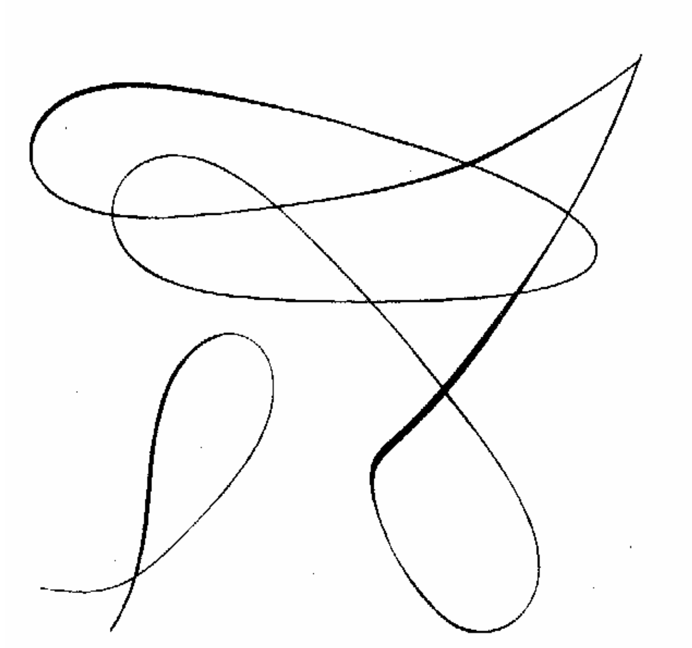
But probability is not the only way to represent uncertainty.

- Belief functions
- Possibility measures
- Ranking functions
- Relative likelihood
- Defaults

Considering many different approaches makes it easier to illustrate the relative advantages and disadvantages of each approach handling likelihood.

1 Motivating examples

Some puzzles and problems should convince you that reasoning about uncertainty can be subtle and that it requires a careful mathematical analysis



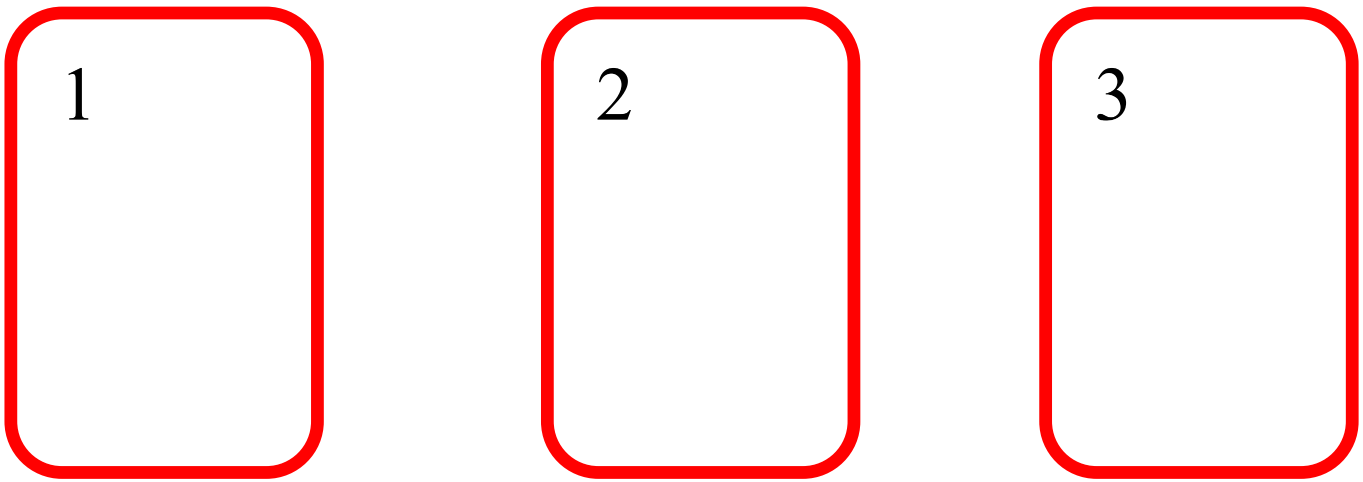
Example 1: Classical urn examples

- There is an urn containing three white and two black balls. You grab two at random. What is the probability that you grab two white ones?

- There are three urns labeled one, two, three. These urns contain, respectively, three white and three black balls, four white and two black balls, and one white and two black balls. An experiment consists of selecting an urn at random, then drawing a ball from it. Find the probability of drawing a black ball.

Example 2: Monty Hall puzzle

Suppose you're in a game show and given a choice of three doors. Behind one is a car. Behind the others are goats. You pick door 1. Before opening door 1, host Monty Hall (who knows what is behind each door) opens door 3, which has a goat. He then asks you if you still want to take what's behind door 1, or to take instead what's behind door 2. Should you switch?



Example 3: a medical decision problem

Consider a doctor who is examining a patient Eric. The doctor can see that Eric has a cough, no temperature, and red hair. According to his medical textbook, 60 % of people with a cough have a flue and 80% of people with flue have a temperature. This is all the information he has that is relevant to the problem. Should he proceed under the assumption that Eric has a flue?

Example 4: Unknown probabilities

Compare: Tossing a coin which is known to be fair
Tossing a coin which is not known to be fair

In both cases, we assign a chance of 50% to the proposition that the result is heads. In the first case this assignment is based on probabilistic knowledge, in the second case it is based on the absence of such knowledge.

► Generalizations of probability theory which do allow the representation of ignorance. Partial variants of probability theory (probabilities are partially specified)

Example 5: 100 Marbles

Suppose that a bag contains 100 marbles; 30 are known to be red, and the remainder are known to be blue or yellow, although the exact proportion of blue and yellow is not known. What is the likelihood that a marble taken out of the bag is yellow?

Three bets:

B_r: pays \$1 if the marble is red and 0 otherwise

B_b: pays \$1 if the marble is blue and 0 otherwise

B_y: pays \$1 if the marble is yellow and 0 otherwise

People invariantly prefer B_r to B_b and B_y. Why?

Example 6: Three-prisoners puzzle

Three prisoners, A, B, C, are locked in their sells. It is common knowledge that two of them will be executed the next day and the other pardoned. Only the guard knows who will be executed. Prisoner A ask the guard a favour: “Since either B or C is certainly going to be executed, you will give me no information about my own chances if you give me the name of one man, either B or C, who is going to be executed.” Accepting this argument, the guard truthfully replies, “B will be executed”. Prisoner A feels happier because before the guard replied, his own chance of execution was 2/3, but afterwards there are only two people, himself and C, who could be the one not executed, and so his chance of execution is 1/2. Is this argumentation correct?

2 Possible worlds

Most representations of uncertainty start with a set of *possible worlds* (also called *states* or *elementary outcomes*). In probability the set of all possible worlds is called *sample space* W.

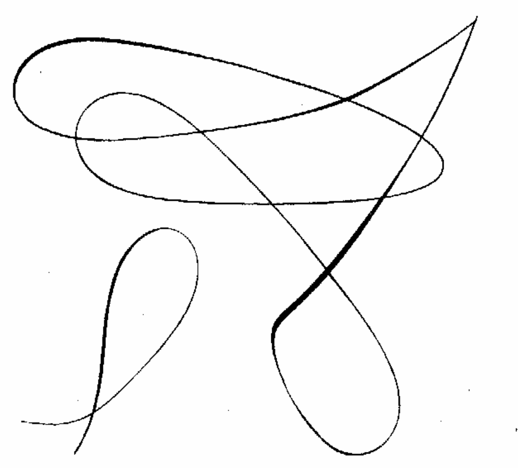
Example: Tossing a die.
Six possible worlds $W = \{w_1, w_2, w_3, w_4, w_5, w_6\}$

The objects that are known (or considered possible or probable) are called *propositions* (or *events*). Propositions are modelled as subsets of W.

Representing the three prisoner puzzle

Are three worlds enough?

a (representing a world where prisoner *a* is pardoned)
b (representing a world where prisoner *b* is pardoned)
c (representing a world where prisoner *c* is pardoned)



Representing the three prisoner puzzle

Are three worlds enough?

a (representing a world where prisoner *a* is pardoned)
b (representing a world where prisoner *b* is pardoned)
c (representing a world where prisoner *c* is pardoned)

No!!

$W = \{(a,b), (a,c), (b,c), (c,b)\}$
where (x,y) represents a world where prisoner *x* is pardoned and the guard says that *y* will be executed.

lives-a =	{(a,b), (a,c)}	(a lives)
lives-b =	{(b,c)}	(b lives)
says-b =	{(a,b), (c,b)}	(the guard says b)

The proposition “the die landed on an even number” corresponds to a set $W^0 = \{w_2, w_4, w_6\}$. Similarly, the proposition "the guard says b" corresponds to a set {(a,b), (c,b)}.

If $W^0 \subseteq W$ represents the agent’s knowledge (and uncertainty) about the world, then we call (W, W^0) an *epistemic space*.

Definition 1: Let $\Sigma = (W, W^0)$ be a epistemic space.

- $\Sigma \Vdash \text{Possible}_x(U)$ iff $U \cap W^0 \neq \emptyset$ (x considers U possible)
- $\Sigma \Vdash \text{Know}_x(U)$ iff $W^0 \subseteq U$ (x knows U)

Fact 1: An agent knows U if and only if he doesn’t consider ¬U (the complement of U) possible:

- $\Sigma \Vdash \text{Know}_x(U) \Leftrightarrow \Sigma \nVdash \text{Possible}_x(\neg U)$

Fact 2: Let $\Sigma = (W, W^0)$ be an epistemic space, and $U, V \subseteq W$ propositions. Further define $U \to V =_{\text{def}} \neg U \cup V$: Then it holds:

- $\text{Know}_x(U)$ & $\text{Know}_x(V) \Leftrightarrow \text{Know}_x(U \cap V)$

  ($\Sigma \Vdash \ldots$ is omitted)

- $\text{Possible}_x(U)$ or $\text{Possible}_x(V) \Leftrightarrow \text{Possible}_x(U \cup V)$
- $\text{Know}_x(U \to V)$ & $\text{Know}_x(U) \Rightarrow \text{Know}_x(V)$
- $U \subseteq V$ & $\text{Know}_x(U) \Rightarrow \text{Know}_x(V)$
- $W^0 \subseteq V \Rightarrow \text{Know}_x(V)$

Proofs are exercises.

Remarks

- The choice of the set of possible worlds encodes many of the assumptions the modeler is making about the domain. There is not necessarily a single "right" set of possible worlds to use
- In the first part of the course we focus on the single-agent case. Multiple agents are not discussed here.
- I generally assume that the set W of possible worlds is finite. Most but not all results we discuss hold without change if W is infinite.

3 Probability measures

Most presentations of probability theory start with a set of events W and assigns to *all* subsets of W a probability μ. However, example 5 (100 marbles) shows that this is not always appropriate:

$W = \{R,B,Y\}$

$\mu(\{R\}) = 0.3$, $\mu(\{B,Y\}) = 0.7$, $\mu(\emptyset)=0$, $\mu(\{R,B,Y\})=1$

For all other subsets no probabilities are defined!

The system $\{\emptyset, \{R\}, \{B,Y\}, \{R,B,Y\}\}$ is called an algebra over W.

Definition 2:

An algebra over W is a set $\mathcal{F}$ of subsets of W that contains W and is closed under union and complementation.

That means, if $U \in \mathcal{F}$ and $V \in \mathcal{F}$, then $U \cup V \in \mathcal{F}$ and $\neg U \in \mathcal{F}$.

Note that an algebra is closed under intersection, since
$U \cap V = \neg(\neg U \cup \neg V)$

Example: Show that 2^W (the set of all subsets of W) is an algebra!

Definition 3:

A probability space is a tuple $(W, \mathcal{F}, \mu)$, where $\mathcal{F}$ is an algebra over W and $\mu: \mathcal{F} \to [0,1]$ satisfies the following two properties:

P1. $\mu(W)=1$

P2. $\mu(U \cup V) = \mu(U)+\mu(V)$ if U and V are disjoint elements of $\mathcal{F}$.

Simple consequence: $\mu(\emptyset) = 0$

Classical case: $\mathcal{F} = 2^W$, i.e. the algebra consist of *all* subsets of W. In this case we write (W, μ) instead of $(W, \mathcal{F}, \mu)$ for the probability space.

Fact 3: Let (W, μ) be a (classical) probability space, and write $\mu(\{u\}) = \mu(u)$. Then for $U \subseteq W$ we have: $\mu(U) = \sum_{u \in U} \mu(u)$

For the proof we need

Fact 4: $\mu(U_1 \cup \ldots \cup U_k) = \mu(U_1) + \ldots + \mu(U_k)$, for pairwise disjoint sets $U_1, \ldots, U_k$ (finite additivity).

Call a (classical) probability space uniform if $\mu(u) = \mu(u')$ for all $u, u' \in W$.

Fact 5: In a classical and uniform probability space we have $\mu(U) = |U| \,/\, |W|$ (where $|.|$ designates the cardinality of the corresponding set).

W = {hh, ht, th, tt}
Uniformity: all worlds are equally probable
$\mu(hh) = \mu(ht) = \mu(th) = \mu(tt) = 1/4.$

$H^1 = \{hh, ht\}, H^2 = \{hh, th\}$
$T^1 = \{tt, th\}, T^2 = \{tt, ht\}$
$\mu(H^1) = \mu(H^2) = \mu(T^1) = \mu(T^2) = 1/2$

Independence (this will be introduced later more systematically)

$$\mu(H^2 | H^1) =_{def} \mu(H^2 \cap H^1) \;/\; \mu(H^1) \;=\; \mu(H^2)$$
$$1/4 \;/\; 1/2 \;=\; 1/2$$

$W = \{(a,b), (a,c), (b,c), (c,b)\}$ where (x,y) represents a world where prisoner x is pardoned and the guard says that y will be executed.

Principle of Indifference

lives-a =	{(a,b), (a,c)}	1/3
lives-b =	{(b,c)}	1/3
lives-c =	{(c,b)}	1/3

But what is the probability of (a,b)? That depends on the guard's strategy in the one case where he has a choice, namely when a lives: $\mu(\mathit{says\text{-}b} | \mathit{lives\text{-}a}) = \alpha$.

Make a plausible choice of α and calculate $\mu(\{(a,b)\})$!

Let (W, W') be the epistemic space of an agent x. Then we can take $(W, 2^{W'}, \mu)$ with a uniform μ as the *associated probability space* of the agent (who considers all possible states as equally probable)

Fact 6: Let $(W, 2^{W'}, \mu)$ be the associated probability space of the epistemic space $\Sigma = (W, W')$ of agent x. Then for each $U \subseteq W$:

- $\Sigma \Vdash \text{Possible}_x(U)$ iff $\mu_x(U) \neq 0$,
- $\Sigma \Vdash \text{Know}_x(U)$ iff $\mu_x(U) = 1$

4 Lower and upper probabilities

Suppose Alice has a coin and she knows that it has either bias 2/3 (head is preferred) or bias 1/3 (tail preferred). But she doesn't know which bias is more likely. How to represent this situation?

$W = \{h, t\}$
Because of bias uniformity doesn't longer hold.
We have to assume two probability measures:
$\mu_{2/3}(h) = 2/3$, $\mu_{2/3}(t) = 1/3$;
$\mu_{1/3}(h) = 1/3$, $\mu_{1/3}(t) = 2/3$

Definition of lower and upper probability

Because we do not know how likely it is that the coin is head-biased (tail-biased) we can give only an interval for the probability that head/tail results after tossing:

$\mu(h) \in [1/3, 2/3]$, $\mu(t) \in [1/3, 2/3]$.

Definition 4: Given a set $\mathcal{P}$ of probability measures defined on a algebra over a (finite) set W, and $U \in \mathcal{F}$, define

- $\mathcal{P}_*(U) = \mathbf{min}\{\mu(U) \colon \mu \in \mathcal{P}\}$ (lower probability)
- $\mathcal{P}^*(U) = \mathbf{max}\{\mu(U) \colon \mu \in \mathcal{P}\}$ (upper probability)

$\mathcal{P} = \{\mu_{2/3}, \mu_{1/3}\}$; $\mathcal{P}_*(h) = \mathbf{min}\ \{2/3, 1/3\} = 1/3$

$\mathcal{P}^*(h) = \mathbf{max}\ \{2/3, 1/3\} = 2/3$

5 Inner and outer measures

100 marble example again: Suppose that a bag contains 100 marbles; 30 are known to be red, and the remainder are known to be blue or yellow, although the exact proportion of blue and yellow is not known. What is the likelihood that a marble taken out of the bag is yellow?

$W = \{R,B,Y\}$

$\mu(\{R\}) = 0.3$, $\mu(\{B,Y\}) = 0.7$, $\mu(\emptyset)=0$, $\mu(\{R,B,Y\})=1$

Intuitively, we feel that $\mu(\{Y\}) \in [0, 0.7]$. How to calculate the interval in this case?

Definition 5: Let $(W, \mathcal{F}, \mu)$ be a finite probability space. Then the inner and outer measures μ_* and μ^* induced by μ are defined as follows:

- For every $U \subseteq W$, $\mu_*(U) = \mathbf{max}\{\mu(V): V \subseteq U, V \in \mathcal{F}\}$
- For every $U \subseteq W$, $\mu^*(U) = \mathbf{min}\{\mu(V): V \supseteq U, V \in \mathcal{F}\}$

100 marble example again:

$\mu(\{R\}) = 0.3$, $\mu(\{B,Y\}) = 0.7$, $\mu(\emptyset) = 0$, $\mu(\{R,B,Y\}) = 1$

$\mu_*(Y) = \mathbf{max}\{0\} = 0$

$\mu^*(Y) = \mathbf{min}\{0.7, 1\} = 0.7$

Remark: In the case of probability spaces with infinite W we have to replace **max** by **sup** and **min** by **inf**, respectively.

Connection to lower and upper probability

Given a probability space $(W, \mathcal{F}, \mu)$, we can try to extend μ by considering all algebras $(W, 2^W, \mu')$ such that $(W, \mathcal{F}, \mu)$ is a subalgebra of the classical algebra $(W, 2^W, \mu')$. That means: $\mu'(U) = \mu(U)$ for all $U \in \mathcal{F}$

Definition 6: Let μ be a probability measure on a subalgebra $(W, \mathcal{F}, \mu)$. Then the extension set $\mathcal{P}_\mu$ is defined as set of all extensions of μ to the classical algebra.

Remark: the lower and upper probabilities then become:

- For every $U \subseteq W$, $(\mathcal{P}_\mu)_*(U) = \mathbf{min}\{\mu'(U): \mu' \in \mathcal{P}_\mu\}$
- For every $U \subseteq W$, $(\mathcal{P}_\mu)^*(U) = \mathbf{max}\{\mu'(U): \mu' \in \mathcal{P}_\mu\}$

Fact 7: Let μ be a probability measure on a subalgebra $(W,\mathcal{F},\mu)$ and let $\mathcal{P}_\mu$ consist of all extensions of μ to the classical algebra. Then $\mu_*(U) = (\mathcal{P}_\mu)_*(U)$ and $\mu^*(U) = (\mathcal{P}_\mu)^*(U)$.

Fact 8: The inner and the outer measure are dual, i.e.
$\mu_*(U) = 1-\mu^*(\neg U)$.

Fact 9: The following inequalities hold for inner and outer measure:

- $\mu_*(U) \leq \mu^*(U)$
- If $U \subseteq V$ then $\mu_*(U) \leq \mu_*(V)$ and $\mu^*(U) \leq \mu^*(V)$ [monotonicity]
- $\mu_*(U \cup V) \geq \mu_*(U)+\mu_*(V)$ for disjoint U,V [superadditivity]
- $\mu^*(U \cup V) \leq \mu^*(U)+\mu^*(V)$ for disjoint U,V [subadditivity]

6 Possibility measures

Classical probability theory start with a set of events W and assigns to *all* subsets of W a probability μ. Possibility theory is just another approach to assign numbers to subsets of W.
Instead of the probabilistic axioms P we assume axioms Poss:
In the infinite case we have to replace max by sup(reme).

P1. $\mu(\emptyset)=0$
P2. $\mu(W)=1$
P3. $\mu(U \cup V) = \mu(U)+\mu(V)$ if U and V are disjoint

Poss1. $\mathrm{Poss}(\emptyset)=0$
Poss2. $\mathrm{Poss}(W)=1$
Poss3. $\mathrm{Poss}(U \cup V) = \max(\mathrm{Poss}(U),\mathrm{Poss}(V))$ if U and V are disjoint

Facts

Fact 10: Let W be a finite set of possible worlds and Poss a possibility measure satisfying Poss1-Poss3 (not necessarily defined for all subsets of W).

- Poss3 holds even if U and V are not disjoint!
- If the possibility measure is defined for all subset of (finite) W, then it can be characterized by its behaviour on singleton sets:

$$\mathrm{Poss}(U) = \max_{u \in U} \mathrm{Poss}(u)$$

- Under the same conditions at least one element in W must have maximum possibility: $\exists w \in W$ [Poss(w) = 1].

Dual

The dual of possibility is called *necessity* and is defined in the usual way: $\mathrm{Nec}(U) = 1 - \mathrm{Poss}(\neg U)$

Fact 11: Let W be a set of possible worlds. Then $\mathrm{Nec}(U \cap V) = \min(\mathrm{Nec}(U), \mathrm{Nec}(V))$

Example: Poss is defined on $W = \{1, \ldots, 10\}$ by taking $\mathrm{Poss}(U) = \max_{n \in U}(n/10)$ and stipulating $\mathrm{Poss}(\emptyset)=0$.

Then Poss(n) = n/10; Nec(n)=0 if n<10, and Nec(10) = 1/10

Example: Poss is defined on $\mathbb{N}$ (set of natural numbers) by taking $\mathrm{Poss}(U) = \sup_{n\in U}(1\text{-}1/n)$ and stipulating $\mathrm{Poss}(\varnothing)=0$.
[that means that $\mathrm{Poss}(n) = 1\text{-}1/n$ for any $n\in\mathbb{N}$]

It can be shown that

- $\mathrm{Poss}(W) = 1$
- $\mathrm{Poss}(U\cup V) = \max(\mathrm{Poss}(U),\mathrm{Poss}(V))$

Hence, Poss is a possibility measure.

Possible exercise: Calculate Nec(n)!

Interpretation and importance

- Degree of surprise (low possibility indicates high degree of surprise). See next section about ranking functions!
- The most common interpretation given to possibility (and necessity) is not in terms of surprise/likelihood but as degree of uncertainty regarding the truth of a *vague* statement. Even if there is no uncertainty about John's actual height (say 1.78 meters) there might be uncertainty about the statement S: "John is tall". This can be described, by assuming Nec(S) = 0.3, Poss(S) = 0.7
- Possibility measures are compositional w.r.t. "$\cup$"
- Possibility measures can be used to define defaults.

7 Ranking functions

- Ranking functions are very similar in spirit to possibility measures.
- A ranking function κ assigns to the subsets of *W* a natural number or infinity; that is any $\kappa: 2^W \rightarrow \mathbb{N}^*$, where $\mathbb{N}^* = \mathbb{N}^* \cup \{\infty\}$
- The numbers can be thought of as denoting degrees of surprise; that is $\kappa(U)$ is the degree of surprise the agent would feel if the actual world were in *U*. 0 denotes "unsurprising", …, ∞ denotes "so surprising as to be impossible".

Axioms

Poss1. $\mathrm{Poss}(\varnothing) = 0$

Poss2. $\mathrm{Poss}(W) = 1$

Poss3. $\mathrm{Poss}(U \cup V) = \max(\mathrm{Poss}(U), \mathrm{Poss}(V))$ if U and V are disjoint

Rk1. $\kappa(\varnothing) = \infty$

Rk2. $\kappa(W) = 0$

Rk3. $\kappa(U \cup V) = \min(\kappa(U), \kappa(V))$ if U and V are disjoint

- Again, Rk3 holds eve if U and V are not disjoint
- As with probability and possibility, a ranking function is defined by its behaviour on singletons in finite spaces:

$$\kappa(U) = \min_{u \in U} \kappa(\mathrm{u})$$

Fact 12: Ranking functions can be viewed as possibility measures. Given a ranking function κ define the possibility measure Poss_κ by taking

$$\mathrm{Poss}_\kappa(U) = 1/(1+\kappa(U))$$

It is not difficult to prove that Poss_κ satisfies the three axioms

Poss1. $\mathrm{Poss}(\varnothing) = 0$

Poss2. $\mathrm{Poss}(W) = 1$

Poss3. $\mathrm{Poss}(U \cup V) = \max(\mathrm{Poss}(U), \mathrm{Poss}(V))$ if U and V are disjoint

Ranking functions can also be viewed as providing a way of doing order-of-magnitude probabilistic reasoning.
Let $\varepsilon << 1$ and take $\mu(U) = \varepsilon^{\kappa(U)}$. Then

- $\mu(W) = 1$
- $\mu(\varnothing) = 0$
- $\mu(U \cup V) = \varepsilon^{\kappa(U \cup \mathrm{V})} = \max(\varepsilon^{\kappa(U)}, \varepsilon^{\kappa(V)}) \approx \varepsilon^{\kappa(U)} + \varepsilon^{\kappa(\mathrm{V})} = \mu(U) + \mu(V)$ for disjoint U and V

With this interpretation the ranking function defines extremely small probabilities (in terms of infinitesimal ε). The laws for the ranking functions correspond to the laws of probability in this limit.

8 Choosing a formalism

Probability: Well understood, many technical results have been proved. Dutch book arguments that "probability" is the only rational way to represent uncertainty.

Sets of probability measures, inner & outer measures: Many advantages of probability, deals better in settings where there is uncertainty about the likelihood.

Possibility measure, ranking functions: Deals well with default reasoning and counterfactual reasoning.