**Probabilistic Logic and Reasoning**

**Perfect knowledge**

The KR languages we have studies are based on classic logic

Sentences are either true or false

They implicitly assume that knowledge is **perfect**

* Precise is clear the term
* certain if we say *raphael is a parent* we know he is a parent
* correct, should not have any inconsistency

Reality is not so gentle to us logic for uncertainty

**Representing uncertainty**

To represent and reason about uncertain knowledge we must first represent uncertainty

uncertainty may be **qualitative** (numerical representation 75% probability… number associated to it ) or **quantitative** (more abstract: don’t care about the number, saying that X is more likely than Y… *learn to ride a bike is more likely that learn to ride a car* or *since is winter is more likely to be cold outside*)

Quantitative uncertainty can be handled through

* probability theory
* possibility theory variant of probability, measure are easier to compute and have different properties
* log-linear computation logistic regression operator
* …

we will focus on probability theory, it has less definition

**Probability theory in a nutshell**

Consider a set of outcomes and a class ε of sets of outcomes (events)

throwing a dice, 6 possible outcomes (1,2,3,4,5,6)

an event is a set of outcomes, can be *fall in one, fall in an even number…*

A probability measure is a function P : ε → [0,1] such that P(Ω) = 1 (the probability of the universe, probability of *somethings* happens) is 1, it is certain that something will happen

if two events are disjoint (otherwise some events may be counted twice) the union of elements is the sum of the probability of the two events

Roll an even number on the first red dice and roll 2 on the second blue dice.

if E ∩ F = ø (no intersection) then P(E ∪ F) = P (E) + P(F)

**Some useful probabilities**

P(ø) = 0 the probability of *nothing* appen is always 0.Is also 1-P(Ω) = 0 (total probability 1 minus the probability of the universe (something will happen))

P (Ω \ E) = 1-P(E) the probability of a complement of an even is 1 - the probability of the event

P (E ∪ F) = P(E) + P(F) - P(E ∩ F) probability of the union of two events is the sum of the probabilities minus their intersection (if it exist) (otherwise some events may be counted twice)

**Probabilistic logic**

In its most basic form, probabilistic logic adapts the set-theoretic notions to formulas

probability based on set to probability based on formulas

Recall that propositional formulas are equivalent to sets of valuations (that make them true)

speaking about a formula is the set of speaking of sets

A logic probability distribution is a function P :F → [0,1]

to each formula assign a valuation between 0 and 1

* P(T) = 1 is tautology, all valuations makes it true, have probability 1, always true
* if a ^ b is unsatisfiable P(a^b) = 0 (meaning they are disjoint, the conjunction is unsatisfiable) then P(a v b) = P(a) + P(b) the union of the valuation that makes it true is the sum of the probability.

**Properties**

This probabilistic logic satisfies some desired probabilities

* P (⊥) = 0 probability of an empty event is 0
* P( ¬a) = 1- p(a)
* P(a v b) = P(a) + P(b) - P(a ^ b) if they are not disjoint, otherwise P(a^b)=0

**Truth functionality**

There is a problem

One issue that makes probabilistic logics difficult is that they are not truth functional (by knowing the value of the pieces we can not know the value of a complex expression) this is not possible

knowding P(a) and P(b) is **not enough** to find out P(a ^ b) or P(a v b)

(on some specific circumstances I can know it, but in general no)

* P(a v b) = P(a) + P(b) - P(a ^ b) we still need to know P(a ^ b)

this is specially bad for KR

need to find a way around

**Probabilistic scope**

For expressive logics (with interpretation-based semantics) we must make an additional choice

how each unary and binary predicate is interpreted

we have a lot of individuals, if we start with such an expressive logics and we speak about probably we need to make a choice on what is the scope of the probability

What is the scope of the uncertainty?

* type I (statical) vs
* type II (subjective)

**Statistical probabilities**

Statistical probabilities consider one interpretation but are uncertain about the **properties of its elements** (political idea, illness…)

probability about the population, we say 80% of the population is favorable on candidate A,

dealing in one possibile word

Check the proportion of people that are left handed, if you take a random person is probability of being left handed is 25% because 25% of the people are left handed

*taking a random person*

**Example**

75% of BAI students are italian the probability of a BAI student to be italian is 75%

**Subjective probabilities**

Subjective probabilities represent beliefs, uncertain about the validity of knowledge

Might be true or might be false, is either true or false, don’t know which one

represent through a **set of interpretations**

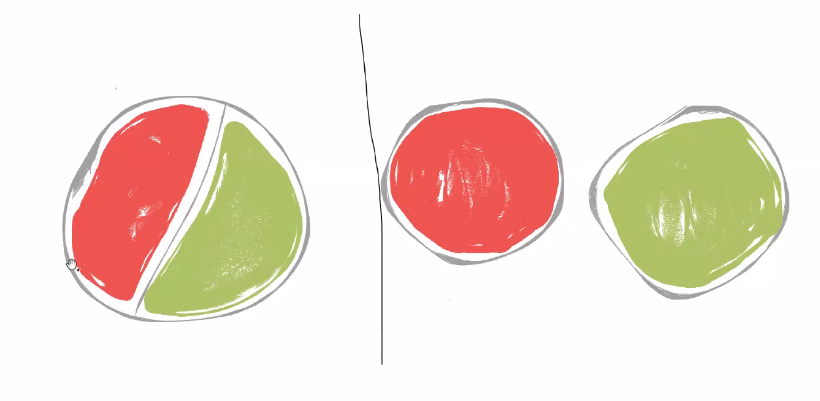
Since we have believes we have a set of interactions, we don’t divide the population but we divide the knowledge

**example**

the probability of Rafael having a daughter is 40%

speaking about one precise individual

2 options, or rafael has a daughter or he doesn't



**left**: staticial, or satisfy or not

**right**: subjective, two interpretations, in one the property is false in the other is true

**Probabilistic propositional rules**

We introduce a probabilistic KR language based on propositional rules

begin a propositional logic, type I probabilities make no sense (we don’t have a population)

In the language, we assign probabilities to clauses

**Syntax**

recall the main notions of clauses, rules, facts

clauses: disjunction of literals

horn clauses: exactly one positive literal

facts: only have a positive literal

rules: have some potentially negative literals

A probabilistic Horn clause is an expression p :: γ where

* p € [0,1]
* γ is a horn clause

In particular:

* p :: x ← is a probabilistic fact
* p :: x ← y1,...yn is a probabilistic rule

**probabilistic knowledge base**

a probabilistic knowledge base is a finite set K of

* classical rules and facts and knowledge that is correct
* probabilistic rules and facts +completare

formall, we can assume that **all** clauses are probabilistic but it is useful to differentiate for technical reason

dealing with probabilistic horn clauses is very expensive, the construction grows exponentially

**sematic intuition**

p :: γ expressed that the probability of γ is p (\* at least)

Under **subjective probabilities**, we have two scenations

* one where γ holds
* one where γ does not hold

we need to consider them both (we don’t know if γ is true or not but there can be consequences that do not depend on both)

**example**

0.4 :: x ← y ← x,u

u ← z ← u

three classical clauses and one probabilistic fact

two scenarios

define two options:

p = 0.4 scenario1, x is true

x ← y ← x,u y is true in this word

u ← z ← u

(x and u are facts)

p = 0.6 scenario2, x is not true

y ← x,u y is not true in this word

u ← z ← u

I have to take in account both queries

to answer a probabilistic query, we have to check both case

**Multiple probabilistic clauses**

What happens with multiple probabilistic causes?

more than two worlds may be necessary

this is known as the multiple-word semantics

semantic based on multiple interpretations

**multiple-word semantics**

given a KB K = KP U KC union of two knowledge bases

set of classical clauses and set of probabilistic causes

(the probabilistic and the classical parts)

each L ⊆ Kp defines a new KB

KL := L ∪ KC

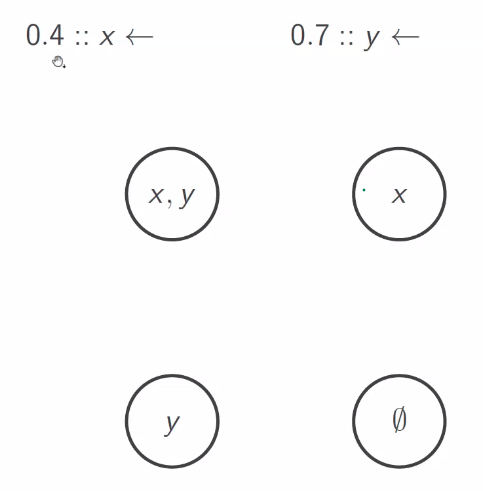
which of the clauses will be included and which not

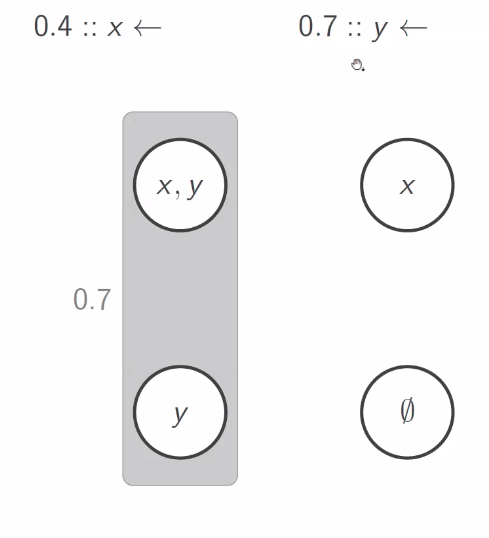
each such KB is called a world

probabilistic reasoning reduces to reasoning over worlds

but we need to assign probabilities to each of them

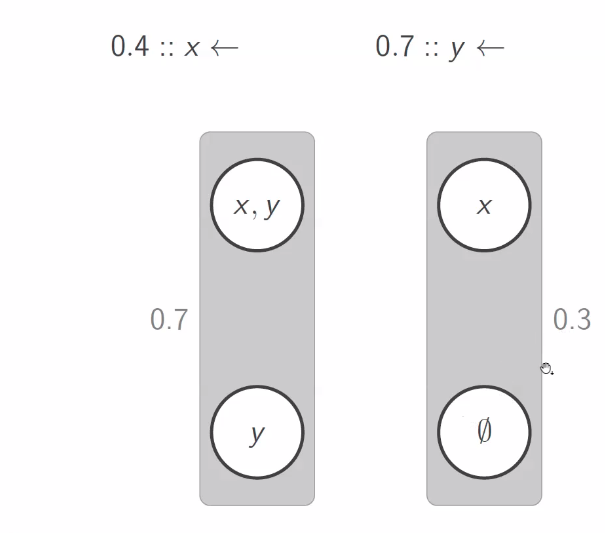
**example**



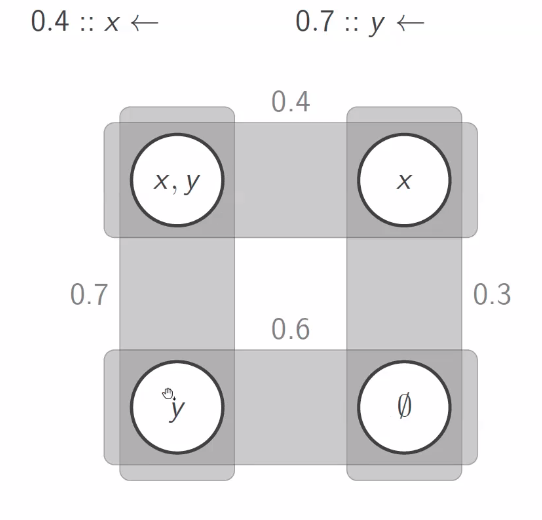


y is true with probability 0.7

together they give me probability 0.7



1-0.7=0.3



this is underspecified, need others informations

**The issue of truth-functionality**

remember that probabilities are not truth function

the probability of a singular clauses does not yield their joint probability

there are different approaches to handle this issue

**probabilistic independence**

the easiest and most commonly used approach is to assume independence

if two events E, F are independent then

P( E ∩ F) = P(E) P(F) product of their probability

knowing that one holds, does not change the probability of the other (don’t affect the likelihood of the other)

it two clauses a, b are independent then

P(a ^ b) = P(a) P(b)

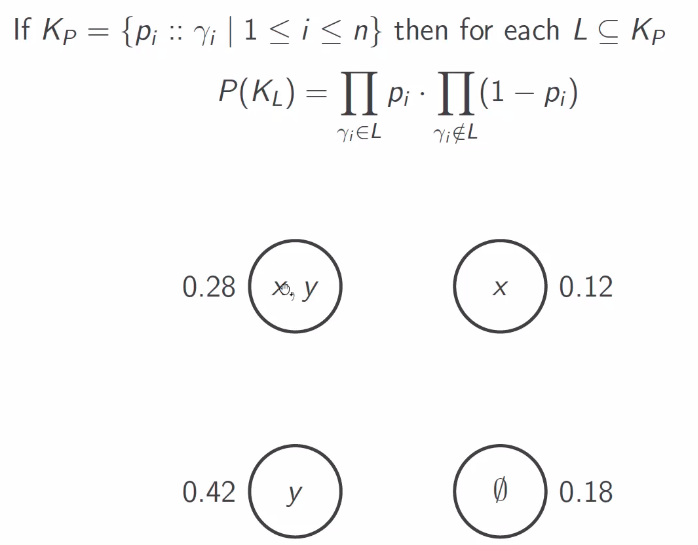
also if a b are independent, then also a and ¬a is independent

**world probabilities under independence**

if KP = {pi :: y1 | 1 ≤ i ≤ n } then for each L ⊆ KP



is the product of all the clauses in the world multiple by the complement of all the clauses that are not in the word



x =0.4

y = 0.7

x,y have both so 0.4x0.7=0.28

x and ¬ y is 0.4x(1-0.7) = 0.12

¬ x and y (1-0.4)x(0.7) = 0.42

¬ x and ¬ y (1-0.4)x(1-0.7)= 0.18

the sum of 0.28 and 0.42 (the two words where y is true) is 0.7 and so on

**maximum entropy**

a second idea is to consider the entropy of a distribution

briefly the entropy measures how **uninformative** a distribution is

more entropy means less informative

entropy is how random the values are

for each distribution we can compute an entropy value

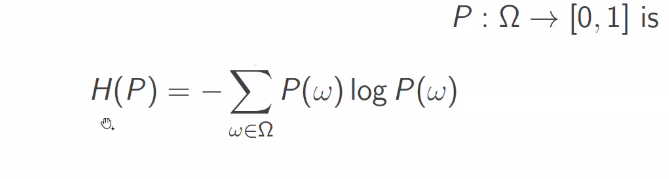
you should always try to use the weakest possible assumptions

the closer to a uniform distribution, the higher the entropy, put to 0.5

make the weakest assumption possible about the distribution

**Entropy**

the entropy of a discrete probability distribution



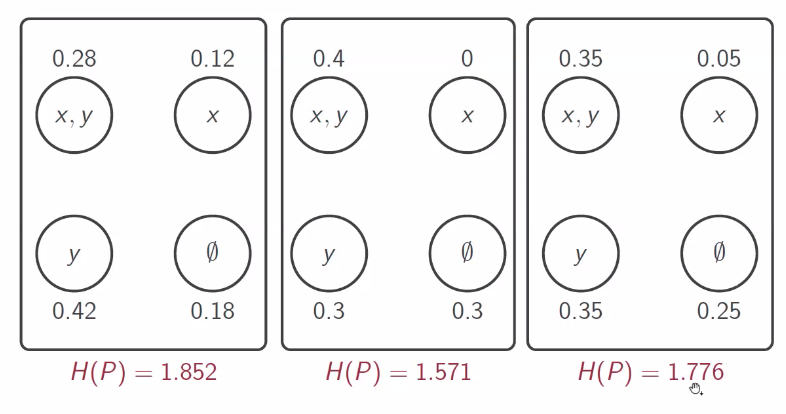
log**2**P(*ω*)

!base 2!

the bigger the entropy the closer it is to the uniform

use the distribution that maximizes this entropy

**Example**



| **0,28** | -0,514220355 |  | **0,4** | -0,528771238 |  | **0,35** | -0,5301006105 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0,12** | -0,3670672427 |  | **0,3** | -0,5210896782 |  | **0,05** | -0,2160964047 |
| **0,42** | -0,5256462821 |  | **0,3** | -0,5210896782 |  | **0,35** | -0,5301006105 |
| **0,18** | -0,4453076139 |  |  |  |  | **0,25** | -0,5 |
|  |  |  |  |  |  |  |  |
| H(P) | **1,852241494** |  | H(P) | **1,570950594** |  | H(P) | **1,776297626** |

entropy maximization is not trivial but can be done

**join probability expression**

The most robust approach is to request the joint distribution explicit

this amounts to specifying the probability of each world

(an exponential number of values)

not very used

often a partial specification together with one of the previous approaches

02 dic 2021

**Summary**

A probabilistic knowledge base K = KP ∪ KC ( union of probabilistic part (probabilistic close) and some classical close)) defines

• a class of (classical) KBs KL = L ∪ KC for L ⊆ KP

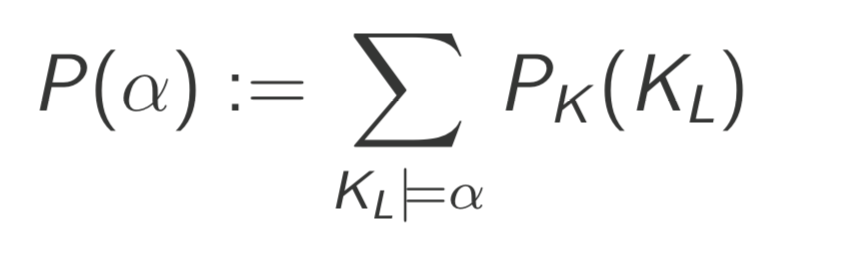
• a probability distribution PK over {KL | L ⊆ KP} (the exact distribution depends on the chosen approach). If they are independent do the product…

Depending on the approach we will get a different result

What do we do with these? We want to find out the probability of some consequences

**Query Answering**

The probability of a consequence α is



The sum of the probabilities of all worlds that entail α

all the words are disjoint, give a number between 0 and 1

**Black-box Approach**

As long as we know how to compute PK , probabilistic reasoning is technically very easy

Build all the KBs KL and check (with a classical reasoner) whether the consequence holds

if we know it does not follows from a world, it will not also follow from a subworld

This is independent of the underlying classical language (as long as a reasoner exists)

saving space does not depends on the propositional rules, depends on the reasoner that can do the classical reasoning

The reasoner is a “black-box:” we don’t care how it works

we don’t care how it works, we only care about the answer, we just use it

using classical knowledge base whether the consequence holds

**Open-world Distributions**

Recall that in KR we usually employ an open-world assumption (OWA)

Idea: when we are representing knowledge it is always incomplete, we don't want to express every possible term, just want to represent the things that are important for our application, let free what we don’t care about.

What I don’t mention could be true or false

Knowledge is always incomplete and what is not expressed could be true or false

*all father are male* if I speak with mothers (since I never talk about them) I can’t guarantee is a female or a male I never speak of male or female (as long I haven’t express the constraints about them)

Why not apply the same idea to the probability distribution?

giving a probability to each closure, it is like a constraint, why do we have to say that exist one probability. why not keep track of all the probability (same of classical logic)

Rather than choosing one (probabilistic distribution), reason over all coherent ones

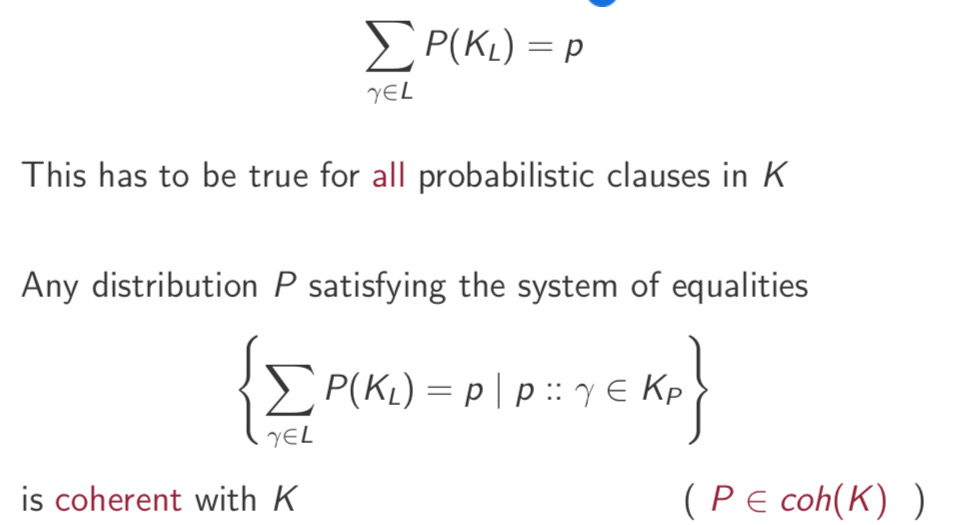
consider the set of all distribution and express something about all of them

Coherent: make sense with respect to the constraint of our KB

**The Class of Coherent Distributions**

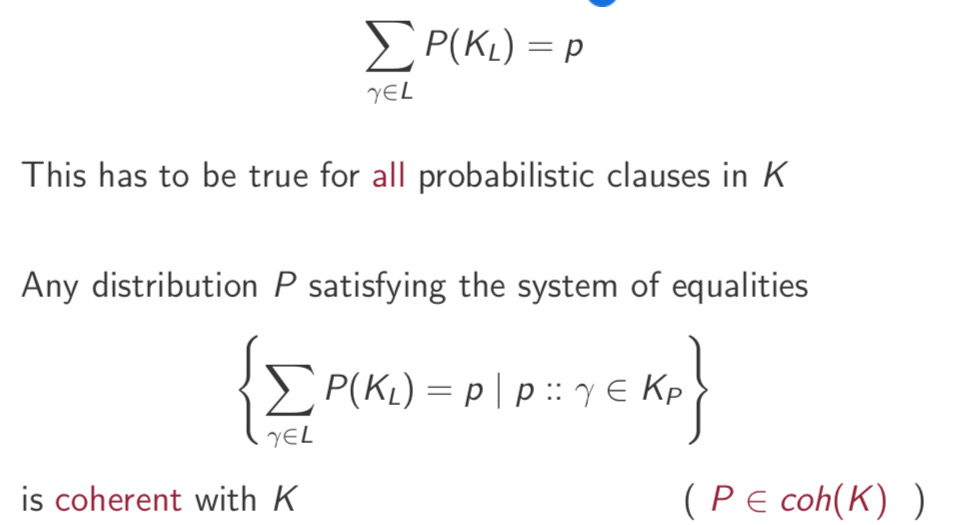
The probabilistic clause p :: γ requires that the worlds containing γ have in total probability p

don’t say anything about independence. Take all the words that contain γ, take all the probability , should be exactly p



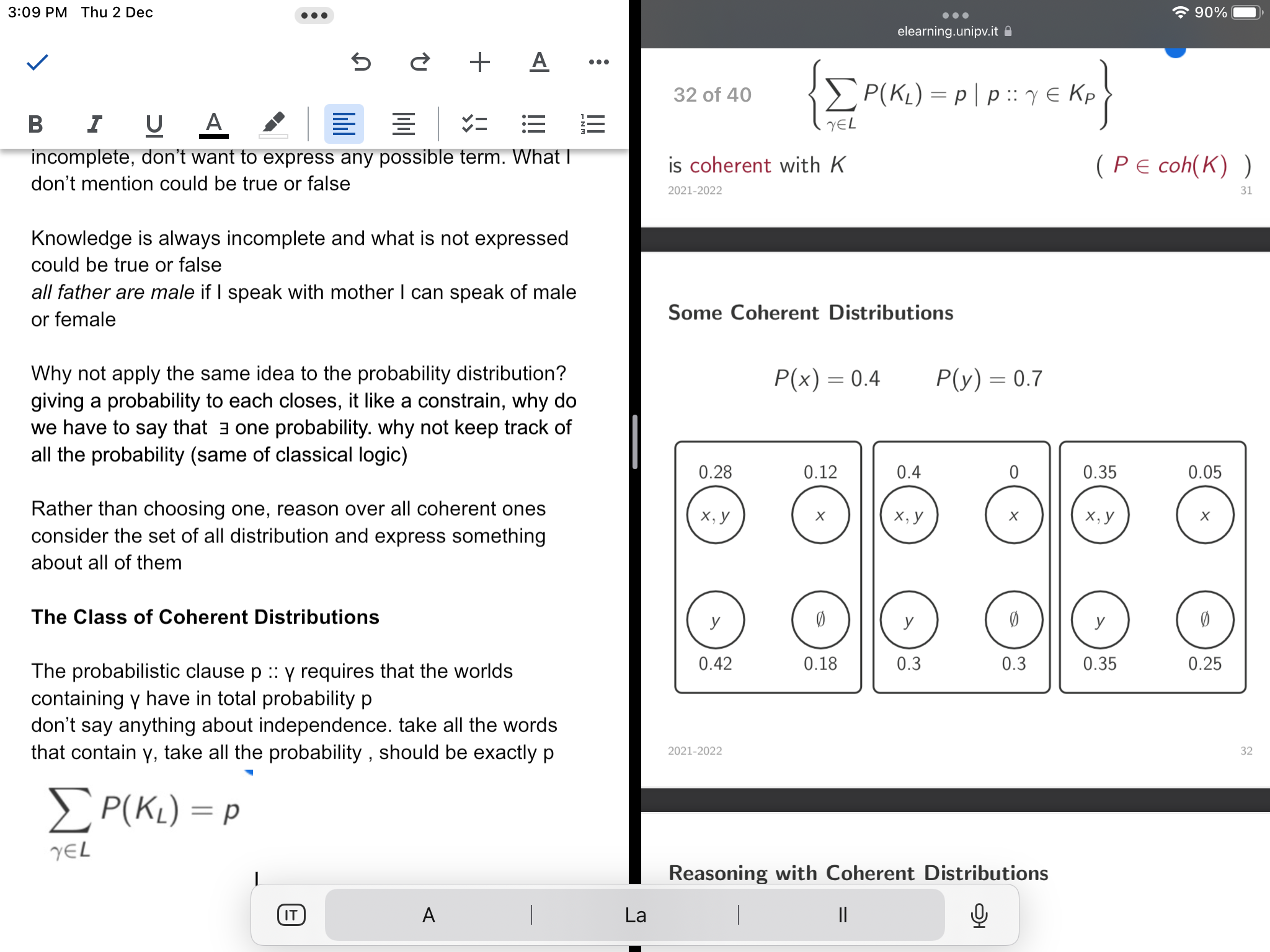
Take all the words that contains γ add the probability of all the words that contains γ and that’s is *p*

This has to be true for all probabilistic clauses in K Any distribution P satisfying the system of equalities



For each probabilistic closest we have that the sum of some probabilities should be exactly p

**Some coherent distributions**



infinitely many probabilistic distribution, coherent probabilistic distribution

**Reasoning with Coherent Distributions**

Since we have multiple distributions, we cannot assign one probability to a given consequence

don’t have a specific probability value, can make some guarantees, that are true for all distribution

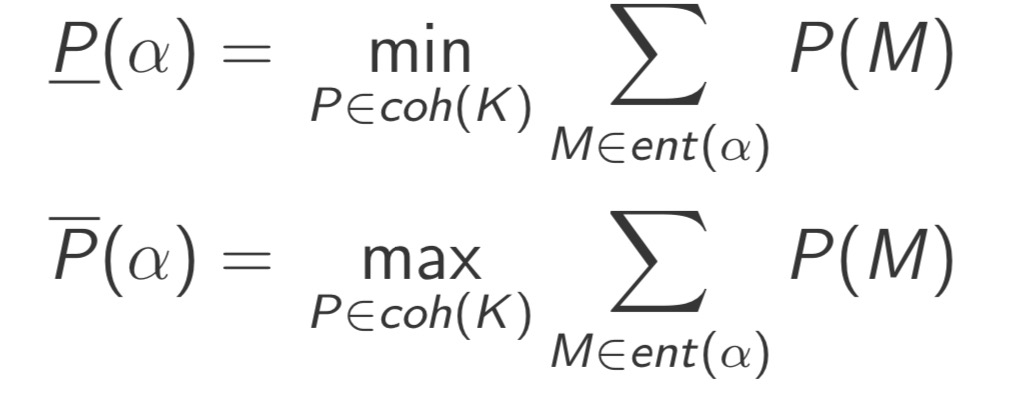
But we can make some guarantees about all distributions

**Extreme Semantics**

Given a consequence α, let ent(α) := {KL | KL ⊨ α}

(all the worlds that entail α)

The extreme probabilities of α are



lower is something true for all distributions, will be greater or equal to P(A)

formally infimum, min if there is one distribution that satisfy this, than is a min

Look at all possible distribution, compute all the entailment of that distribution, check which is the smallest one

identify the set that satisfies the entailment, once that we found the set find the possible probabilities that we can assign to them

can find a min and max, we are dealing with all possible probabilities distribution

if I have a occerent distribution ,is always true that the lower is smallest or equal to the biggest

**Properties**

Obviously, for every coherent distribution P,

P(α) ≤ P(α) ≤ (α)

More interestingly, for every P(α) ≤ p ≤(α) there exists a P ∈ *coh*(K) (continuous) such that P(α) = p

Any probability is reasonable, can’t tell the exact probability

**Beyond Propositional Rules**

The main ideas go beyond propositional logic base logic to extend with probability logic

In particular, the multiple-world semantics translate to any KR language

**Probabilistic EL**

A probabilistic EL-GCI is of the form p :: α where p∈[0,1] α aGCI

assign a probability to a TBox

GCI with a probability degree

A probabilistic EL TBox is a finite set of probabilistic GCIs A

TBox with n probabilistic GCIs defines 2n worlds

The probability of a consequence is the probability of the worlds that entail it

**Example**

Consider the following knowledge:

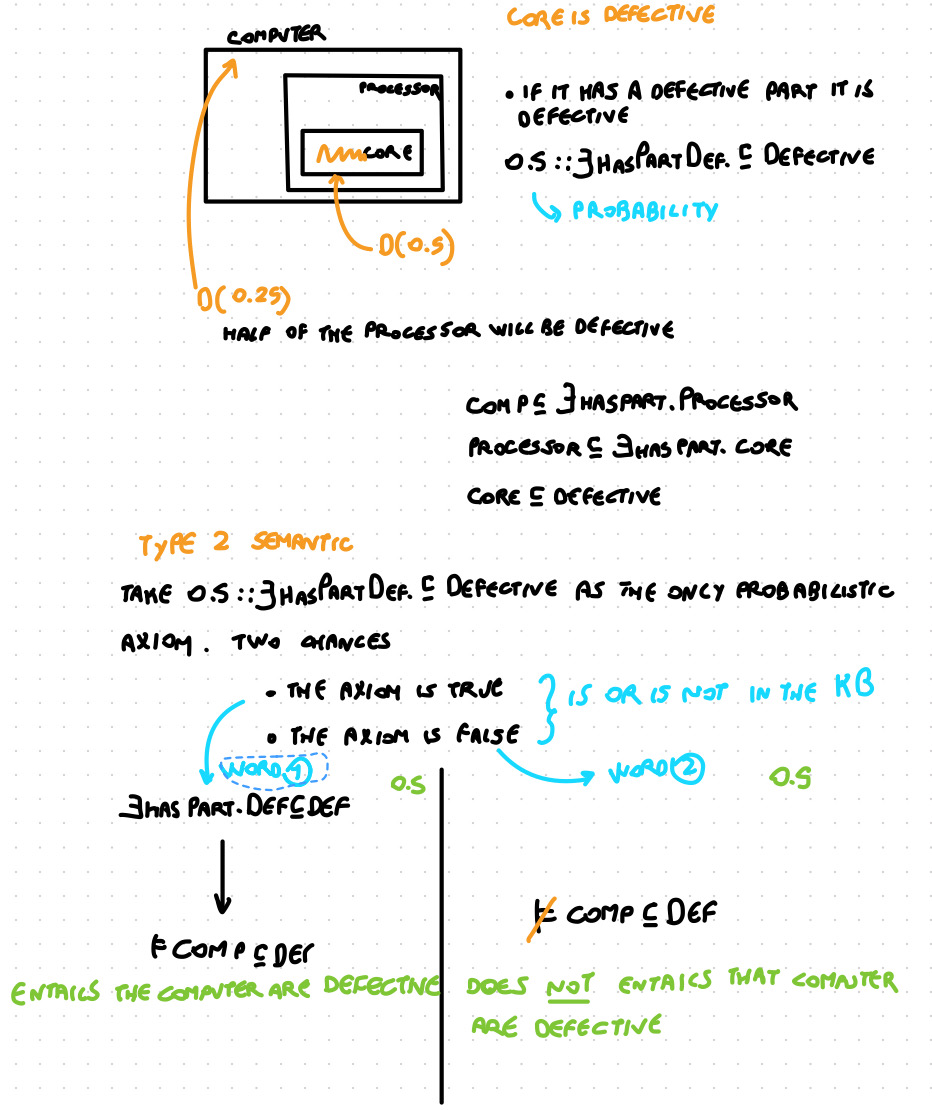
• A piece with a defective part is defective with probability 0.5

• Computers have processors as parts

• Processors have cores as parts

• All cores are defective

What is the probability of computers being defective?



**Bottom Line**

Be always careful about the meaning of the semantics

Be aware of the tradeoffs between expressivity, simplicity, and computational complexity

More expressivity: less simple, more compact to express, but the computational complexity is higher (might be problematic to be implemented)