2. CONDITIONALS AND PROBABILITY

- Representing uncertainty: basic settings
- Conditionals without probability
- Inference with conditionals : a three-valued logic.
- Bayesian probability: strengths and limitation

UNCERTAINTY : representing graded belief.

- AN AGENT IS UNCERTAIN ABOUT A PROPOSITION IF (S)HE DOES NOT KNOW ITS TRUTH VALUE
  - **Examples**
    - The **probability** that the trip is more than one hour long is 0.7.
    - It is quite **possible** it snows to-morrow.
    - The agent has no **certainty** that Jean comes to the meeting
- HOW TO EVALUATE THE PROBABILITY, THE POSSIBILITY, THE CERTAINTY, THAT A PROPOSITION IS TRUE OR FALSE

UNCERTAINTY THEORIES

- *Set-based representations:* Reasoning about belief in terms of possibility and certainty
  - *Propositional logic:* Believing = proving from a belief base.
  - *Interval analysis :* Propagation of incomplete information.
- *Probability theory*: statistical, subjective
- *Possibility Theory* ordinal or numerical:
  - Tells plausible states from less plausible ones
  - use fuzzy sets of mutually exclusive values
- *Disjunctive random sets* (Dempster-Shafer-Smets): probability on set-representations
- *Imprecise Probabilities* : the most general setting, with probability intervals.

SINGULAR vs. GENERIC INFORMATION

- PIECES OF EVIDENCE refer to a particular situation (measurement data, testimonies) and are singular.
  - E.g. results of medical tests on a patient
  - Observations about the current state of facts
  - *May be imprecise, incomplete, unreliable, irrelevant, wrong, etc.*
- BACKGROUND KNOWLEDGE refers to a class of situations and summarizes a set of trends
  - Laws of physics, commonsense knowledge (birds fly)
  - Professional knowledge (of medical doctor), Statistical knowledge
  - *Not absolutely true knowledge in the mathematical sense*: *tainted with exceptions, incompleteness, variability*

Warning about the word « **knowledge** »

- The term may refer to two notions
  - True belief: Knowing p => p is true in the real world (while beliefs need not be true)
  - Generic information: *background knowledge*, that summarizes the agent 's experience across a collection of situations (while belief refers to a singular situation)

- In this tutorial we adopt the second point of view

GENERIC KNOWLEDGE, EVIDENCE, BELIEFS

- An agent usually possesses three kinds of information on the world

  1. **Generic information *(background knowledge)*** : it pertains to a range of situations the agent is aware of.
     - **Examples** : statistics on a well-defined population commonsense knowledge (often ill-defined population)
  2. **Singular information on the current situation *(evidence)***
     - *Observed facts (results of tests, sensor measurement, testimonies)*
  3. **Beliefs about the current situation**
     – Derived from observed facts and singular observations

- **Generic knowledge may be tainted with exceptions**
  - It all comes down to considering some propositions are generally more often the case than other ones.
  - *Generic knowledge induces a normality or plausibility relation on the states of the world*.
  - *numerical* (frequentist) or *ordinal* (plausibility ranking):

- **Observed evidence** is often made of propositions known as true about the current world.
  - Can be encoded as disjunctive sets, or wff in propositional logic.
  - It delimits a reference class of situations for the case under study.
  - Can be uncertain (subjective probability, Shafer)

A first problem : PLAUSIBLE REASONING

- Inferring **beliefs** (plausible conclusions) on the current situation from observed evidence, using generic knowledge
  - **Example : medical diagnosis** Medical knowledge + test results $\Rightarrow$ believed disease of the patient.
- *This mode of inference makes sense regardless of the representation*, but
    - in a purely propositional setting, one cannot tell generic knowledge from contingent evidence
    - in the first order logic setting there is no exception.
  - **Need more expressive settings for representing background knowledge**

A second problem : MERGING UNCERTAIN EVIDENCE

- Observations about the current world may be unreliable, uncertain, inconsistent:
  - *Sensor failures, dubious testimonies*
  - Propositional logic cannot account for **unreliable** evidence
  - Probability theory alone cannot account for **incomplete** evidence
- A proper account of uncertain evidence needs to cope with uncertainty and the necessity for **merging unreliable evidence** in a flexible way, before even inferring beliefs

Belief construction

- *Beliefs of an agent about a situation are derived from generic knowledge and observed singular evidence about the case at hand*.
- Example :Statistical beliefs = **Hacking principle**
  - Generic knowledge = probability distribution P built from statistics
  - Singular observed fact = a set A
  - Computing the conditional probability $P(B|A)$ for the reference class A
  - $Bel_A(B) = P(B|A)$: equating degree of belief and frequency

Belief construction

- *Beliefs of an agent about a situation are inferred from generic knowledge AND observed singular evidence about the case at hand.*
- **Example :Commonsense inference**
  - Generic knowledge = birds fly, penguin are birds, penguins don't fly.
  - Singular observed fact = Tweety is a bird
  - Inferred belief = Tweety flies
  - Additional evidence = Tweety is a penguin
  - Inferred revised belief = Tweety does not fly

The Conditional: A naturally 3-valued proposition

- WHAT IS THE LOGICAL or MATHEMATICAL STATUS OF A "RULE" IN A RULE-BASED SYSTEM ?
  - Not a logical clause (material conditional $\neg p \vee q$) nor a classical inference rule « if p, always deduce q » BECAUSE THEY CANNOT HANDLE EXCEPTIONS
  - A RULE IS OFTEN A DEFAULT RULE: IF ALL I KNOW IS p THEN DEDUCE q »
- A CONDITIONAL PROBABILITY ATTACHES A "CERTAINTY FACTOR » to a rule and copes with exceptions (« Most p's are q's » quantified by $P(q|p)$).
- *But the probability of a material conditional is not a conditional probability! What is the entity whose probability is a conditional probability???*

A conditional event!!!!

Material implication : the raven paradox

- Testing the rule « all ravens are black » viewed as $\forall x, \neg Raven(x) \vee Black(x)$
- Confirming the rule by finding situations where the rule is true.
  - Seeing a black raven confirms the rule
  - Seeing a white swan also confirms the rule.
  - But only the former is an example of the rule.

3-Valued Semantics of conditionals

- A rule « if p then q » shares the world in 3
  - Examples : interpretations where $p \wedge q$ is true
  - Counterexamples: interpretations where $p \wedge \neg q$ is true
  - Irrelevant cases: interpretations where p is false
- Truth-table of $p \rightarrow q$
  - Truth($p \rightarrow q$) = T if truth(p)= truth(q) = T
  - Truth($p \rightarrow q$) = F if truth(p)=T and truth(q) = F
  - Truth($p \rightarrow q$) = I if truth(p)= F
- This truth-table is the solution X of $p \wedge q = X \wedge p$.
- Rules « all ravens are black » and « all non-black birds are not ravens » have the same exceptions (white ravens), but different examples (black ravens and white swans resp.)

A conditional event is a pair of nested sets

- The models of a conditional $p \rightarrow q$ can be represented by the pair $(A \cap B, A^c \cup B)$ if A and B are the sets of models of p and q respectively.
- The set of models $A^c \cup B$ of material implication $\neg p \vee q$ excludes exceptions to the rule $p \rightarrow q$.
- $(A \cap B, A^c \cup B)$ is an interval in the Boolean algebra of subsets of interpretations.
- It calls for a three-valued logic.

Inferring a rule from a rule

- A rule $p \rightarrow q$ implies another rule $r \rightarrow s$, if the latter has more examples and less exceptions than the former
- **Equivalent formulations**:
  - $p \rightarrow q \models r \rightarrow s$ iff $p \wedge q \models r \wedge s$ and $\neg p \vee q \models \neg r \vee s$

  (This is the canonical extenstion of the semantic inference relation $\models$ to intervals in the Boolean algebra)
  - Equipping the truth-set {T, F, I} with the logical ordering T > I > F:

    $p \rightarrow q \models r \rightarrow s$ iff $t(p \rightarrow q) \leq t(r \rightarrow s)$

Validity of a rule base

- Let Δ be a set of rules $p_i \rightarrow q_i$, $i = 1..N$.
  - Δ is verified by an interpretation if it verifies at least one rule and does not falsify any other.
  - The rule base Δ is falsified by an interpretation if it falsifies one rule.
- The validity of Δ is the one of the quasi-conjunction of its rules

$$\&_{i = 1..N} (p_i \rightarrow q_i) = (\vee_{i = 1..N} p_i) \rightarrow \wedge_{i = 1..N} \neg p_i \vee q_i$$

- Equipping the truth-set {T, F, I} with the q-c ordering $I \gg T \gg F$:
  - $t(\Delta) = \min\{t(p_i \rightarrow q_i), i = 1..N\}$

Inferring a rule from a rule base: semantic entailment

- Let Δ be a set of rules, and $QC(\Delta)$ be its quasi-conjunction.
- The rule base Δ is consistent iff

  $\forall \Sigma \subseteq \Delta$, Σ has one example
- Definition:

  $\Delta \models p \rightarrow q$ iff $\exists \Sigma \subseteq \Delta, QC(\Sigma) \models p \rightarrow q$

Valid patterns of inference for conditionals

- Left logical equivalence:

  if $q \dashv\vDash r$ then $q \to p \dashv\vDash r \to p$
- Right weakening: if $q \vDash r$ then $p \to q \vDash p \to r$
- Cautious monotony: $\{p \to q, p \to r\} \vDash p \wedge q \to r$
- Cut: $\{p \to q, p \wedge q \to r\} \vDash p \to r$
- AND: $\{p \to q, p \to r\} \vDash p \to q \wedge r$
- OR : $\{p \to q, r \to p\} \vDash p \vee r \to q$
- Half deduction theorem: $p \wedge q \to r \vDash p \to \neg q \vee r$

Invalid patterns of inference for conditionals

- Monotony: $p \to q \not\models p \wedge q \to r$
  - Indeed $p \wedge q \to r$ has less examples than $p \to q$
- Transitivity $\{p \to q, q \to r\} \not\models p \to r$
  - An example to $q \to r$ that falsifies p verifies the quasi-conjunction of the two premises.
- Half deduction theorem $p \to \neg q \vee r \not\models p \wedge q \to r$
  - $p \wedge \neg q$ verifies the premise, not the conclusion

Syntactic inference with conditional knowledge

- **Definition** : $\Delta \mid\!- p \to q$ iff $p \to q$ can be produced from $\{r \to r, \forall r \neq \bot\} \cup \Delta$ using Left logical equivalence, Right weakening, Cautious monotony, AND, OR
- This is basically « *system P* » of Kraus, Lehmann and Magidor.
- **Soundness + Completeness** (Dubois &Prade 1994):

  $\Delta \mid\!- p \to q$ iff $\Delta \models p \to q$

  So we can reason in system P in a 3-valued logic with truth set {T, F, I} equipped with 2 orderings.

Belief construction in the logic of conditionals

- *Observed singular evidence on a situation = propositional formula p*
- *Generic knowledge = a conditional knowledge base Δ*
- *Proposition q is believed about the situation after observing p and under generic knowledge Δ iff* $p \to q$ can be inferred from Δ.
- **Example : Commonsense inference**
  - Knowledge : $\Delta = \{b(x) \to f(x), p(x) \to b(x), p(x) \to \neg f(x)\}$
  - Singular observed fact = b(Tweety)
  - Inferred belief = f(Tweety)
  - Additional evidence = p(Tweety)
  - Inferred belief = ¬f(Tweety) ($p(x) \to f(x)$ fails as no transitivity)
- But this system is notoriously too weak
  - from b(Tweety) red(Tweety) , f(Tweety) is unknown

GRADUAL REPRESENTATIONS OF UNCERTAINTY

Belief is a matter of degree !

- **Family of propositions or events $\mathcal{E}$ forming a Boolean Algebra**
  – S, Ø are events that are certain and ever impossible respectively.
- **A confidence measure** g : a function from $\mathcal{E}$ in [0,1] such that
  – $g(\emptyset) = 0$; $g(S) = 1$
  – if A implies (= included in) B then g(A) **≤** g(B) (monotony)
- g(A) quantifies the confidence of an agent in proposition A.

BASIC PROPERTIES OF CONFIDENCE MEASURES

- **$g(A \cup B) \geq \max(g(A), g(B))$;**
- **$g(A \cap B) \leq \min(g(A), g(B))$**
- It includes :
  - probability measures : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - possibility measures $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$
  - necessity measures $N(A \cap B) = \min(N(A), N(B))$
- ***The two latter functions do not require a numerical setting***

Probability Representations (on finite sets)

- A finite set S with n elements: A probability measure is characterized by a set of non negative weights $p_1, \ldots, p_n$, such that $\sum_{i=1,n} p_i = 1$.
  - p_i = probability of state s_i
- **Possible meanings of a degree of probability** :
  - Counting *favourable cases* for s_i over the number of possible cases assuming uniform distribution (coins, dice, cards,…)
  - *Frequencies from statistical information*: p_i = limit frequency of occurrence of s_i (**Objective probabilities**)
  - *Money involved in a betting scheme* (**Subjective probabilities**)

SUBJECTIVE PROBABILITIES
(Bruno de Finetti, 1935)

- p_i = *belief degree* of an agent on the occurrence of s_i
- measured as the price of a lottery ticket with reward 1 € if state is s_i in a betting game
- **Rules of the game**:
  - gambler proposes a price p_i
  - banker and gambler exchange roles if price p_i is too low
- **Why a belief state is a single distribution**:
  - Assume player buys all lottery tickets i = 1, m.
  - If state s_j is observed, the gambler gain is $1 - \sum_j p_j$
  - and $\sum_j p_j - 1$ for the gambler
  - *if* $\sum p_j > 1$ gambler *always loses money ;*
  - *if* $\sum p_j < 1$ *banker exchanges roles with* gambler

Probabilistic belief from statistical probabilities

- Subjective probability of the particular occurrence of an event may derive from its statistical probability.
- Probabilistic beliefs: **Hacking principle**
  - Generic knowledge = probability distribution P
  - BetP(A) = FreqP(A): equating belief and frequency
- Beliefs can be directly elicited as subjective probabilities with no frequentist flavor if frequencies are not available or for non repeatable events.

Remarks on the representation of belief by a single probability distribution

- Computationally simple : $P(A) = \sum_{s \in A} p(s)$
- $P(A) = 0$ iff A impossible; $P(A) = 1$ iff A is certain; usually $P(A) = 1/2$ for ignorance
- **Meaning :**
  - Objective probability is generic knowledge (statistics from a population)
  - Subjective probability is contingent (degrees of belief)
- The counterpart of a conditional knowledge base is a Bayesian network: a set of conditional probability assessments that represent a unique distribution

Conditional Probability

- Two definitions:
  - derived (Kolmogorov): $P(A \mid C) = \dfrac{P(A \cap C)}{P(C)}$

    requires $P(C) \neq 0$
  - primitive: $P(A|C)$ is assigned a value and P is derived such that $P(A \cap C) = P(A|C) \cdot P(C)$.

    Makes sense even is $P(C)= 0$

The probability of A if C represents all that is known on the situation.

THE MEANING OF CONDITIONAL PROBABILITY

- P(A|C) : probability of a conditional event « A in epistemic context C » (when C is all that is known about the situation).
- *It is NOT the probability of A, if B is true*.
- **Counter-example** :
  - Uniform Probability on $\{1, 2, 3, 4, 5\}$
  - P(Even $|\{1, 2, 3\}$) = P(Even $|\{3, 4, 5\}$) = 1/3
  - Under a classical logic interpretation :
    - From « if result $\in \{1, 2, 3\}$ then P(Even) = 1/3 »
    - And« if result $\in \{3, 4, 5\}$ then P(Even) = 1/3 »
    - Then (classical inference) : P(Even) = 1/3 unconditionally!!!!!
  - **But of course : P(Even) = 2/5.**

Probability of conditionals

- Let $[q] = A$, $[p] = C$, $P(A|C) = P(p \rightarrow q)$ where $p \rightarrow q$ is a 3-valued conditional.
- Indeed $P(A|C)$ is totally determined by
  - $P(A \cap C)$ (proportion of examples)
  - $P(A^c \cap C) = 1 - P(A \cup C^c)$ (proportion of examples)

$$P(A|C) = \frac{P(A \cap C)}{P(A \cap C) + 1 - P(A \cup C^c)}$$

- $P(A|C)$ is increasing with $P(A \cap C)$ and decreasing with $P(A^c \cap C)$
- If $p \rightarrow q \models r \rightarrow s$ then $P([q]|[p]) \leq P([s]|[r])$

JOINT PROBABILITY and GRAPHICAL REPRESENTATIONS

- If the finite domain is a Cartesian product $S = S_1 x S_2 x \ldots x S_n$ with **variables** : $x_1, \ldots, x_n$, a **joint probability** is a big table containing $p(s_1, \ldots, s_n)$, for all $(s_1, \ldots, s_n) \in S$
- **Claim** : Any *positive* joint probability can be represented by a set of conditional probabilities forming a **directed graph**:
  - rank variables in arbitrary order $x_1, \ldots, x_n$
  - express $p(x_1, \ldots, x_n)$ as

    $$p(x_1 | x_2 \ldots, x_n) \cdot p(x_2 | x_3 \ldots, x_n) \cdot \ldots \cdot p(x_{n-1} | x_n) \cdot p(x_n)$$
  - simplify the expression if conditional independence relations hold (e.g $p(x_1 | x_2 \ldots, x_n) = p(x_1 | x_2)$)

Examples

- $p(x_1, x_2, x_3) = p(x_1 | x_2, x_3) \cdot p(x_2 | x_3) \cdot p(x_3)$

1. If x_2 and x_3 are independent :
   $p(x_3 | x_2) = p(x_3)$, then
   $p(x_1, x_2, x_3) = p(x_1 | x_2, x_3) \cdot p(x_2) \cdot p(x_3)$

$$
\begin{tikzcd}
x_2 \arrow[dr] & \\
 & x_1 \\
x_3 \arrow[ur] &
\end{tikzcd}
$$

2. If x_1 is independent from x_3 given x_2 :

$P(x_1 | x_2, x_3) = P(x_1 | x_2)$, then:

$p(x_1, x_2, x_3) = p(x_1 | x_2) \cdot p(x_2 | x_3) \cdot p(x_3)$

$$x_3 \longrightarrow x_2 \longrightarrow x_3$$

PLAUSIBLE REASONING WITH BAYES NETS

- *A Bayes net represents* ***generic knowledge*** *(especially frequentist) in the form of a probability measure P.*
- *Querying a Bayes net comes down to* ***instantiating the values of some variables*** *and computing the* ***conditional probability of a proposition A of interest in the context C*** *described by all instantiated variables.*
  - *E* ***is contingent evidence on a case*** *(it is not true that P(C) = 1, generally)*
  - *P(A|C) is the probability (frequency) that in general A occurs in context C.*
  - *P(A|C) is then* ***interpreted as the degree of belief*** *$Bel_C(A)$ that A holds for the case at hand about which all that is known is that C is true.*
  - *This framework handles non-monotonicity: one may have P(A|C) high and $P(A | C \cap B)$ low.*

LIMITATIONS OF BAYESIAN PROBABILITY FOR THE REPRESENTATION OF BELIEF

- A single probability cannot represent ignorance
- Subjective specification of a Bayes net imposes unnatural conditions on conditional probabilities to be assessed: complete and consistent conditional probability assessments are requested

Why the unique distribution assumption?

- Laplace principle of insufficient reason : What is EQUIPOSSIBLE must be EQUIPROBABLE
  - *It postulates the identity between IGNORANCE and RANDOMNESS*
  - *like the principle of maximal entropy*
- The exchangeable betting framework enforces the elementary probability assessments to sum to 1.
  - It enforces uniform probability when there is no reason to believe one outcome is more likely than another
  - *Betting rates are induced by belief states, but are not in one-to-one correspondence with them* : ignorance and knowledge of randomness justify uniform betting rates.

THE PARADOX OF IGNORANCE

- Case 1: life outside earth/ no life
  - ignorant's response 1/2 1/2
- Case 2 : Animal life / vegetal only/ no life
  - ignorant's response 1/3 1/3 1/3
- They are inconsistent answers :
  - case 1 from case 2 : $P(\text{life}) = 2/3 > P(\text{no life})$
  - case 2 from case 1 : $P(\text{Animal life}) = 1/4 < P(\text{no life})$
- ***ignorance produces information***
- **Conclusion :** *a probability distribution cannot model incompleteness*

Single distributions do not distinguish between incompleteness and variability

- VARIABILITY: Precisely observed random observations
- INCOMPLETENESS: Missing information
- **Example:** probability of facets of a die
  - *A fair die tested many times :* Values are known to be equiprobable
  - *A new die never tested:* No argument in favour of an hypothesis nor its contrary, but frequencies are unknown.
- *BOTH NOTIONS LEAD TO TOTAL INDETERMINACY BUT THEY DIFFER AS TO THE QUANTITY OF INFORMATION*

Example

- **Variability**: daily quantity of rain in Toulouse
  - May change every day
  - It is objective: can be estimated through statistical data
- **Incomplete information** : Birth date of Brazilian President
  - It is not a variable: it is a constant!
  - Information is subjective: Most may have a rough idea (an interval), a few know precisely, some have no idea.
  - Statistics on birth dates of other presidents do not help much.

Instability of prior probabilities

1. A uniform prior on x induces a non-uniform prior on $f(x)$ if f is non-affine : again Laplacean ignorance produces information
2. When information is missing, decision-makers do not always choose according to a single subjective probability (Ellsberg paradox).

Ellsberg Paradox

- Savage claims that rational decision-makers choose according to expected utility with respect to a subjective probability.
- Counterexample :An Urn containing
  - 1/3 red balls ($p_R = 1/3$)
  - 2/3 black or white balls ($p_W + p_B = 2/3$)
- For the ignorant subjectivist: $p_R = p_W = p_B = 1/3$
- But this is contrary to overwhelming empirical evidence

Ellsberg Paradox

1. Choose between two bets

   B1 : Win 1$ if red (1/3) and 0$ otherwise (2/3)

   B2 : Win 1$ if white (≤ 1/3) and 0$ otherwise

   Most people prefer B1 to B2

2. Choose between two bets (just add 1$ on Black)

   B3 : Win 1$ if red or black (≥ 1/3) and 0$ if white

   B4 : Win 1 $ if black or white (2/3) and 0$ if red (1/3)

   Most people prefer B4 to B3

Ellsberg Paradox

- Let $0 < u(0) < u(1)$ be the utilities of gain.
- If decision is made according to a subjective probability assessment for red black and white: $(1/3, p_B, p_W)$:
  - B1 > B2:

    $EU(B1) = u(1)/3 + 2u(0)/3 > EU(B2) = u(0)/3 + u(1)p_w + u(0)p_B$
  - B4 > B3:

    $EU(B4) = u(0)/3 + 2u(1)/3 > EU(G) = u(1)\ (1/3 + p_N) + u(0)p_W$

  $\Rightarrow$ (summing, as $p_B + p_N = 2/3$) $2(u(0) + u(1))/3 > 2(u(0) + u(1))/3$: CONTRADICTION!
- Such an agent cannot reason with a unique probability distribution: **Violation of the sure thing principle.**

Ellsberg Paradox

- **Plausible Explanation: In the face of ignorance, the decision maker is pessimistic:**
- In the first choice, agent supposes $p_w = 0$: no white ball

  $EU(B1) = u(1)/3 + 2u(0)/3 > EU(B2) = u(0)$
- In the second choice, agent supposes $p_B = 0$: no black ball

  $EU(B4) = u(0)/3 + 2u(1)/3 > EU(B3) = 2u(0)/3 + u(1)/3$
- **The agent does not use the same probability in both cases** (because of **pessimism): the subjective probability depends on the proposed game.**

Beyond classical logic and probability

- Classical logic
  - is not expressive enough to grasp the difference between singular and generic information
  - Does not express shades of belief
  - Cannot account for non-monotonic feature of plausible reasoning with incomplete knowledge
- Bayesian Probability
  - Cannot account for incomplete knowledge
  - Does not tell the difference between variability and ignorance
  - Is too information-demanding when only subjective sources are available
  - Handles exceptions and non-monotonicity of inference
- ***The way out:*** *ordinal uncertainty theories and imprecise probabilities (strengthening the logic of conditional events).*

Probability vs. Classical logic: a basic difference

- In classical logic,
  - all variables are supposed to be independent.
  - All pieces of knowledge express (logical) dependencies.
- In probability theory
  - variables are **not** supposed to be independent
  - Independence assumptions are pieces of knowledge

The two frameworks are at odds with each other!

Next question*: how to extend classical logic in an ordinal setting so as to account for the presence of exceptions*