Probabilistic Graphical Models in Intelligent Systems

Bertinoro International Spring School 2017

Luigi Portinale

University of Piemonte Orientale, Italy

March 7-10, 2017

Who I am

- Full professor of Computer Science at UPO (*University of Piemonte Orientale*); Head of the *Computer Science Institute* (Dept. of Science and Technological Innovation)
- My background: *Artificial Intelligence* in general (Knowledge-Based Systems)
- Main research interests: *Probabilistic Graphical Models* (this course); *Case-Based Reasoning, Machine Learning and IDA*; *Applications of AI* (system dependability, medicine, agri-food, finance, etc . . .)

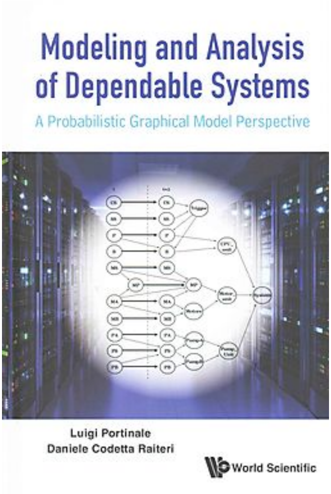
Who I am

- Full professor of Computer Science at UPO (*University of Piemonte Orientale*); Head of the *Computer Science Institute* (Dept. of Science and Technological Innovation)
- My background: *Artificial Intelligence* in general (Knowledge-Based Systems)
- Main research interests: *Probabilistic Graphical Models* (this course); *Case-Based Reasoning, Machine Learning and IDA*; *Applications of AI* (system dependability, medicine, agri-food, finance, etc ...)

Who I am

- Full professor of Computer Science at UPO (*University of Piemonte Orientale*); Head of the *Computer Science Institute* (Dept. of Science and Technological Innovation)
- My background: *Artificial Intelligence* in general (Knowledge-Based Systems)
- Main research interests: *Probabilistic Graphical Models* (this course); *Case-Based Reasoning, Machine Learning and IDA*; *Applications of AI* (system dependability, medicine, agri-food, finance, etc . . .)

A little ad


Modeling and Analysis of Dependable Systems
A Probabilistic Graphical Model Perspective
Luigi Portinale
Daniele Codetta Raiteri
World Scientific

Course Outline

Course Outline

- Uncertainty and Probability in AI
- Random Variables
- Probabilistic Graphical Models: Modeling
- Probabilistic Graphical Models: Inference
- Probabilistic Graphical Models: Learning
- Decision Theory and PGMs
- Suggested topics for in-depth study and references (exam)

Course Outline

- Uncertainty and Probability in AI
- Random Variables
- Probabilistic Graphical Models: Modeling
- Probabilistic Graphical Models: Inference
- Probabilistic Graphical Models: Learning
- Decision Theory and PGMs
- Suggested topics for in-depth study and references (exam)

Course Outline

- Uncertainty and Probability in AI
- Random Variables
- Probabilistic Graphical Models: Modeling
- Probabilistic Graphical Models: Inference
- Probabilistic Graphical Models: Learning
- Decision Theory and PGMs
- Suggested topics for in-depth study and references (exam)

Course Outline

- Uncertainty and Probability in AI
- Random Variables
- Probabilistic Graphical Models: Modeling
- Probabilistic Graphical Models: Inference
- Probabilistic Graphical Models: Learning
- Decision Theory and PGMs
- Suggested topics for in-depth study and references (exam)

Course Outline

- Uncertainty and Probability in AI
- Random Variables
- Probabilistic Graphical Models: Modeling
- Probabilistic Graphical Models: Inference
- Probabilistic Graphical Models: Learning
- Decision Theory and PGMs
- Suggested topics for in-depth study and references (exam)

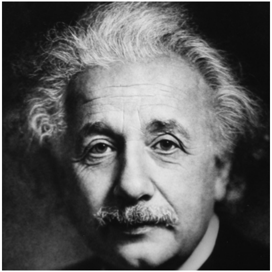
Course Outline

- Uncertainty and Probability in AI
- Random Variables
- Probabilistic Graphical Models: Modeling
- Probabilistic Graphical Models: Inference
- Probabilistic Graphical Models: Learning
- Decision Theory and PGMs
- Suggested topics for in-depth study and references (exam)

Course Outline

- Uncertainty and Probability in AI
- Random Variables
- Probabilistic Graphical Models: Modeling
- Probabilistic Graphical Models: Inference
- Probabilistic Graphical Models: Learning
- Decision Theory and PGMs
- Suggested topics for in-depth study and references (exam)

Uncertainty and Probability in AI

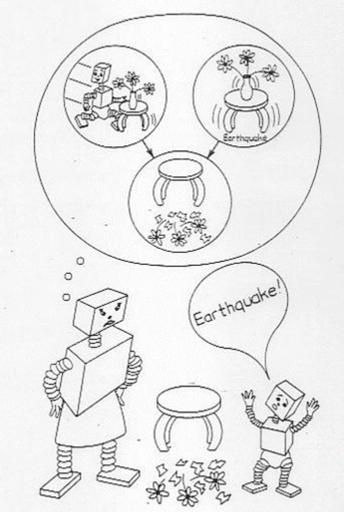


Albert Einstein

As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality.

"Rapid pulse, sweating, shallow breathing ... According to the computer, you've got gallstones."

Uncertainty = lack of information



From *R. Neapolitan. "Probabilistic reasoning in expert systems: theory and algorithms", J. Wiley & Sons, 1992*


Uncertainty
Just Ahead

Probabilities are epistemologically inadequate.

J. McCarthy, P. Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence, Machine Intelligence, 1969

Probabilities are computationally inadequate.

Lots and lots of people in AI, since (maybe) 1956

$\Longrightarrow$ Logic, more logic and only logic !

Probabilities are epistemologically inadequate.

J. McCarthy, P. Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence, Machine Intelligence, 1969

Probabilities are computationally inadequate.

Lots and lots of people in AI, since (maybe) 1956

$\Longrightarrow$ Logic, more logic and only logic !

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

Probabilistic Graphical Models in Intelligent Systems

Luigi Portinale

Course Outline

Uncertainty and Probability in AI

Interpretation of Probability

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

Probabilistic Graphical Models in Intelligent Systems

Luigi Portinale

Course Outline

Uncertainty and Probability in AI

Interpretation of Probability

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

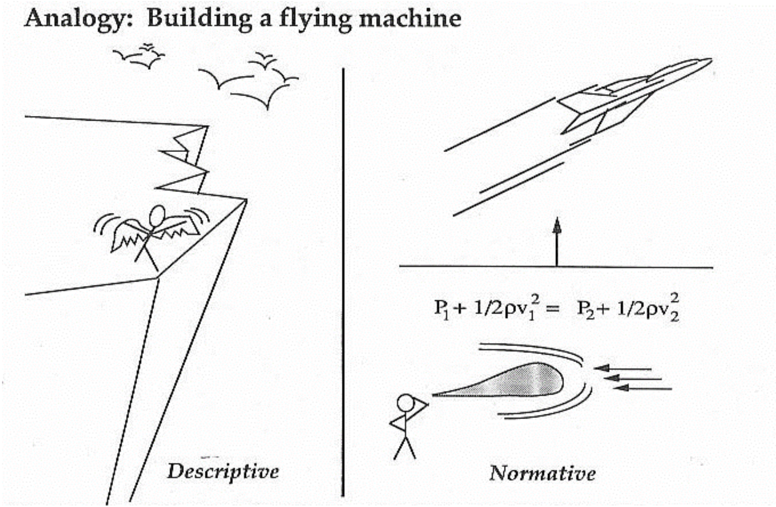
Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

Cox's Justification of Probability (Cox's Theorem)

1. **Clarity.** Propositions must be well defined (either `true` or `false`)
2. **Scalar representation.** A single real number is sufficient to represent *belief* or *plausibility*
3. **Completeness.** A belief $Bel(X)$ can be assigned to any proposition X
4. **Context dependency.** Beliefs depend on other beliefs: $Bel(X|Y)$
5. **Hypothetical conditioning.** $Bel(X \wedge Y) = f(Bel(X|Y), Bel(Y))$
6. **Complementarity.** $Bel(\neg X) = f(Bel(X))$
7. **Consistency.** $X \equiv Y \rightarrow Bel(X) = Bel(Y)$

Only axioms of probability calculus satisfy such principles (R. Cox, 1948)

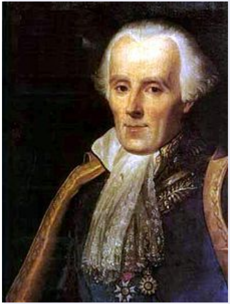
Normative vs Descriptive



Interpretation of Probability

Classical Interpretation (Laplace)

Pierre Simone Laplace, *A Philosophical Essay on Probabilities (1814)*



The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible

Frequentist Interpretation (Richard von Mises)

Let s_E be the number of occurences of event E in n trials, the probability of the event is defined as

$$P(E) = \lim_{n \to \infty} \frac{s_E}{n}$$

- What about events related to trials that are not repeatable? (e.g., *probability of Lakers winning NBA this year*)
- If only a finite number of trials are perfomed, we can estimate the limiting frequency; but if we acknowledge the fact that we have an error in measuring probability, then we still get into problems as the error of measurement can only be expressed as a probability, the very concept we are trying to define (*circularity*)

Bruno de Finetti (1906, 1985)

(Objective) Probability does not exist.

Subjectivist Interpretation (Bruno de Finetti)

The probability of an event E is the price $p = P(E)$ that one would bet in order to get a reward of 1 if E occurs and a reward of 0 if E does not occur.

- The definition applies both if we consider buying or selling a bet (p is the price at which one would either buy or sell a bet such as above)
- The bet should conform probability axioms (fair bet or coherent assignment)
- Dutch book argument

Dutch Book

- A Dutch book is a series of bets acceptable to an agent that guarantee his/her loss, regardless of the actual outcome
- A Dutch book may arise if bets are not coherent. Example: the agent violates 3^{rd} axiom and assigns $P(A \cup B) < P(A) + P(B)$ with A, B mutually exclusive. A bookmaker can buy a bet on $A \cup B$ at $P(A \cup B)$ and sell to the agent individual bets on A and B for $P(A)$ and $P(B)$ respectively, making a positive profit of $P(A) + P(B) - P(A \cup B)$ regardless of the actual outcomes of A and B.
- A Dutch book will never arise if probability assignments are coherent (*Kemeny 1955*)

Probabilistic Graphical Models in Intelligent Systems

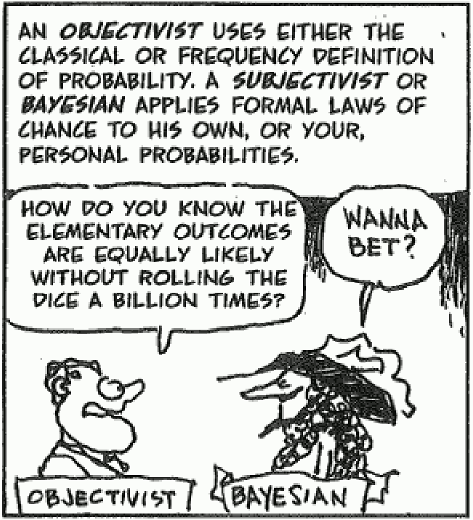
Luigi Portinale

Course Outline

Uncertainty and Probability in AI

Interpretation of Probability

Frequentists vs Bayesians



Frequentist and Bayesian Statistics

- **Frequentist:** unknown parameters are fixed, data may vary, sampling is in principle infinite, each repeated experiments starts from ignorance, decision rules can be sharp
- **Bayesian:** unknown parameters are uncertain (thus random variables), data are fixed (observations), prior information can be incorporated and probabilities can always be updated

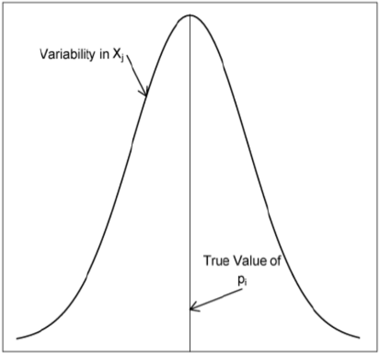
What to compute:

- **Frequentist:** probability of data given a hypothesis (*likelihood*);
- **Bayesian:** posterior probability of a hypothesis, given the data

Frequentist

- Describe variability in data X_j for a fixed hypothesis (parameter) p_i.

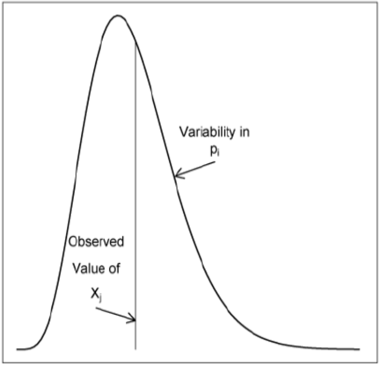
Distribution of Sample



Bayesian

- Describe variability in parameter p_i for fixed data X_j.

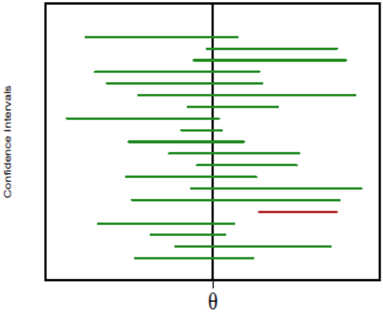
Distribution of Parameter



Interpretation of Confidence

Frequentist

- α% Confidence Intervals: a collection of intervals with α% of them containing the true parameter



Bayesian

- α% Credibility Interval: an interval having α% probability of containg the true parameter.