

581287

Three Concepts: Probability  
(6 cr, 3 cu)

Administrivia & Introduction  
by  
Tomi Silander  
(Assistant Instructor Mika Urtela)

# Format of the course

- Period III
  - 6 (3 hour) lectures + weekly assignments
  - 30% of the grade
  - part of the lectures used for analyzing the results of the latest assignment
- Period IV
  - Assignments, projects, term papers, posters etc.
  - 70% of the grade
- Programming skills needed
  - <http://www.cs.helsinki.fi/group/cosco/Teaching/Probability/2008/index.html>
  - <https://moodle.cs.helsinki.fi/course/view.php?id=59>

# Yet another probability course?

- Computer science point of view
  - Artificial Intelligence Point of View
    - Agent point of view
    - Knowledge representation (KR)
    - Reasoning
    - Rationality
    - Decision making
    - Grounding
  - Machine learning point of view
    - Computational methods for data analysis
    - Large data bases

# Syllabus

- Part I (first 3 – 4 lectures)
  - Probability in knowledge representation
  - Probability calculus
  - Why probability?
  - Loosely follow Russel & Norvig Book (chapters 13-21)
- Part II (rest of the lectures)
  - Where do the probabilities come from?
  - Learning from observations / data
  - Generative models

# Syllabus with keywords

- Part I
  - KR, random variables, Bayesianism, qualification problem, decision making, betting, extensional vs. intensional systems, Cox's theorem, axioms, (joint) distribution, marginalization, relevance, independence, Bayesian networks, inference, prediction, calibration, proper score, causality, frame problem, temporal models, MDPs.
- Part II
  - Bayes, frequentism, propensity, likelihood, prior, data, posterior, generative models, HMMs, finite mixture models, Bayesian regression, MCMC, model selection, reinforcement learning.

# Andrew Moore, slide two

- The world is a very uncertain place
- Thirty years of Artificial Intelligence and Database research danced around this fact
- And then a few AI researchers decided to use some ideas from the eighteenth century

# But first there was logic

- Historically too it was first - syllogisms
  - a model of rationality
- Certainty, correctness, modularity, monotonicity
- BUT limited applicability since

**Agents almost never have access  
to the whole truth about their  
environment!**

# Acting by certain knowledge only?

- Is it enough to leave home 90 minutes before the flight departure?
- Anything can happen.
- How about X minutes before departure?
- Are you bound to stay home?

## **Qualification problem:**

What are the things that have to be taken into account?



# Knowledge representation in FOPL

- Let us try to use FOPL for dental diagnosis

$\forall p \text{Symptom}(p, \text{Toothache}) \Rightarrow \text{Disease}(p, \text{Cavity})$

Wrong!

$\forall p \text{Symptom}(p, \text{Toothache}) \Rightarrow$   
 $\text{Disease}(p, \text{Cavity})$   
 $\vee \text{Disease}(p, \text{GumDisease})$   
 $\vee \text{Disease}(p, \text{Abscess}) \dots$

Incomplete!

$\forall p \text{Disease}(p, \text{Cavity}) \Rightarrow \text{Symptom}(p, \text{Toothache})$

Wrong again, need to add qualifications!

# FOPL representation fails because

- Laziness
  - Its is too much work to list all the factors to ensure exceptionless rules
- Theoretical ignorance
  - We do not know all the factors that play role in the phenomenon
- Practical Ignorance
  - Even if we know all the factors in general, we do not know them for each particular case

# Probability to rescue

**Probability provides a way to  
summarize the uncertainty  
that comes from our  
laziness and ignorance**

# Acting by probability

- Probabilities are not enough to define rational behavior
- Utilities needed to define the desirability of world states
- Decision Theory = Probability + Utility
  - Choose an action A1, A2, or A3, when
    - A1: €100 with probability 0.1 and €50 with probability 0.9
    - A2: €200 with probability 0.2 and €5 with probability 0.8
    - A3: €80 with probability 0.5 and €50 with probability 0.5
- Maximum Expected Utility:  $\underset{a \in A}{argmax} \sum_{s \in S} P_a(s) U_a(s)$