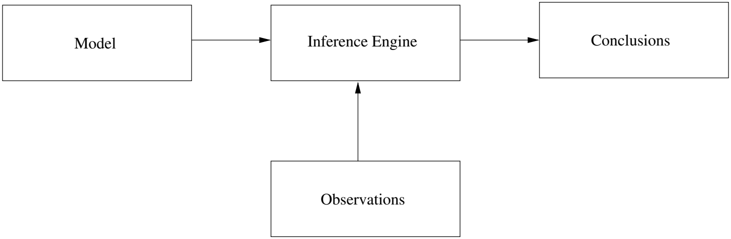
Introduction to Probabilistic Models

Brandon Malone

Much of this material is adapted from Chapter 1 of Darwiche's book
Many of the images were taken from the Internet

January 14, 2014

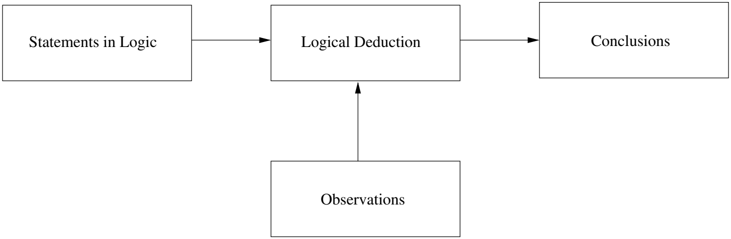
McCarthy's model-based systems



- Model: how the world works
- Inference Engine: how we reason about the world
- Observations: what we know to be true
- Conclusions: what we can say about the world, given the model and observations

Deductive logic

Specifically, McCarthy's system was based on **deductive logic**.



Flying birds

"If a bird is normal, then it will fly."

Do most birds fly?

Flying birds

"If a bird is normal, then it will fly."

Do *these* birds fly?

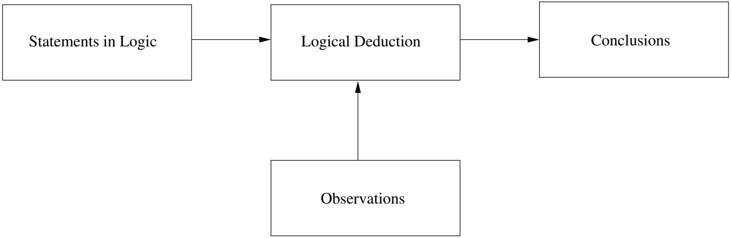
Flying birds

"If a bird is normal, then it will fly."

How can we assume most birds fly, but change our minds later?

Deductive logic

Specifically, McCarthy's system was based on **deductive logic**.



Deductive logic is **monotonic**: if $\delta \Rightarrow \alpha$, then$(\delta \wedge \gamma) \Rightarrow \alpha$

Once δ is `true`, no new information can invalidate α.

Non-monotonic logic to the rescue...?

Non-monotonic logics give a formalism for managing **assumptions**.

Assumptions can be dynamically **asserted** and **retracted**.

Initially, we assume birds can fly.

Do most birds fly? Yes.

Non-monotonic logic to the rescue...?

Non-monotonic logics give a formalism for managing **assumptions**.

Assumptions can be dynamically **asserted** and **retracted**.

As we observe more evidence, we can retract our assumptions.

Do *these* birds fly? They are penguins, so no.

Non-monotonic logic to the rescue...?

Non-monotonic logics give a formalism for managing **assumptions**.

Assumptions can be dynamically **asserted** and **retracted**.

As we observe more evidence, we can retract our assumptions.

Do *these* birds fly? Apparently so.

Non-monotonic logic to the rescue...?

Non-monotonic logics give a formalism for managing **assumptions**.

Assumptions can be dynamically **asserted** and **retracted**.

"A typical Quaker is a pacifist."
"A typical Republican is not a pacifist."

Non-monotonic logic to the rescue...?

Non-monotonic logics give a formalism for managing **assumptions**.

Assumptions can be dynamically **asserted** and **retracted**.

"A typical Quaker is a pacifist."
"A typical Republican is not a pacifist."

Non-monotonic logic to the rescue...?

Non-monotonic logics give a formalism for managing **assumptions**.

Assumptions can be dynamically **asserted** and **retracted**.

"A typical Quaker is a pacifist."
"A typical Republican is not a pacifist."

Which assumptions do we retract? When?

Shortcomings of symbolic logic

Monotonic logics do not allow us to change our minds.

Non-monotonic logics require non-obvious conflict resolution.

Can we do something better?

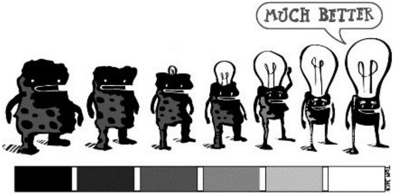
Degrees of belief to the rescue!

Both of these require a hard commitment to a proposition.

Degrees of belief to the rescue!

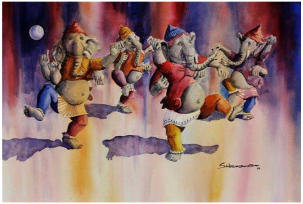
Both of these require a hard commitment to a proposition.

A **degree of belief** is a probability assigned to a proposition.

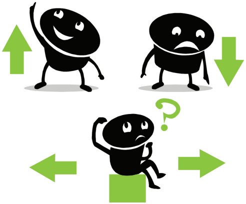


Retracting assumptions vs. revising beliefs

Assumptions are asserted or retracted.



Degrees of beliefs are adjusted up or down.



Why use numbers instead of logic?

Probabilities model non-monotonicity without special machinery.

Other questions...

- Do people think like this?
- Where do the numbers come from?

How can we represent the probabilities?

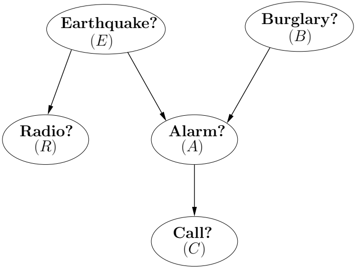
Probability distributions are naively exponential.

Bayesian networks give a compact representation.

Explicit distribution: 31 parameters

E	B	A	R	C	$\Pr(\cdot)$
T	T	T	T	T	0.075
T	T	T	T	F	0.01
T	T	T	F	T	0.001
...					
F	F	F	F	F	0.1

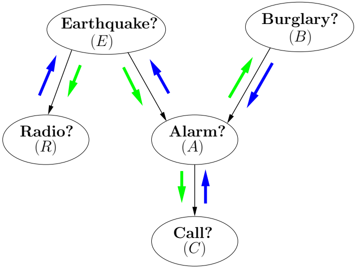
Bayesian network: 10 parameters



How can we efficiently compute desired quantities?

Pearl [1986] proposed the **polytree algorithm** for performing efficient inference in Bayesian networks with polytree structures.

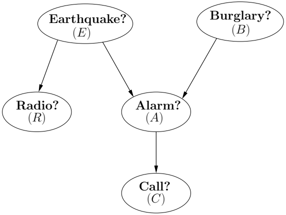
Lauritzen and Speigelhalter [1988] followed with the **jointree algorithm** for arbitrary structures.



The Bayesian network representation

A Bayesian network consists of two components.

A DAG in which vertices correspond to random variables and edges encode dependencies.	Conditional probability distributions which quantify the relations among the variables.



E	Θ_E
T	.6

B	Θ_B
T	.1

E	R	$\Theta_{R\|E}$
T	T	.2
F	T	.75

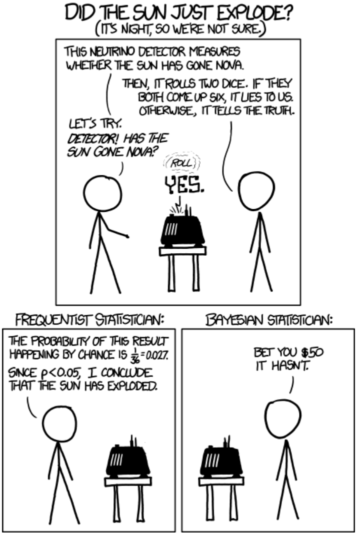
A	C	$\Theta_{C\|A}$
T	T	.3
F	T	.6

E	B	A	$\Theta_{A\|E,B}$
T	T	T	.95
T	F	T	.9
F	T	T	.8
F	F	T	0

Important characteristics of Bayesian networks

- **Consistent** and **complete** - define a unique probability distribution
- **Localized** - consider only variables and their direct causes
- **Compact** - require parameters only for direct dependencies

Frequentists and Bayesians



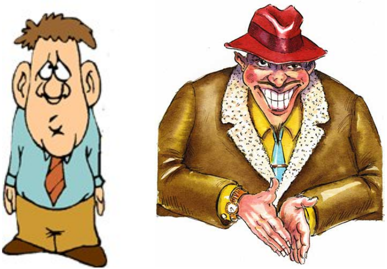
Frequentists and Bayesians

Frequentists. There is a "truth", and we have noisy samples of this truth. Probability is the long run frequency of the noisy samples.

Bayesians. There is no "truth", just data we can use as evidence. Probability is the plausibility of a hypothesis given this (incomplete) data.

"A frequentist can calculate probabilities precisely, but often not the probabilities we want. A Bayesian can calculate the probabilities we want, but often not precisely."

The Dutch Book Argument



A person with belief p in S will pay up to p€ to bet on S, where the payout is 100 €.
For example, if my belief that Barcelona will win the World Cup (S) is .55 (p), then I will bet 55 € that they will win.

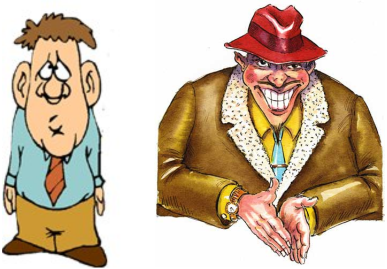
See `http://plato.stanford.edu/entries/dutch-book` for much more discussion.

The Dutch Book Argument

Suppose my belief that Barcelona will win is $p = .55$. Further, say my belief that Barcelona will not win the World Cup is $p = .51$. So I will bet 55 € that Barcelona will win and 51 € that Barcelona will not win.

See `http://plato.stanford.edu/entries/dutch-book` for much more discussion.

The Dutch Book Argument



Whether Barcelona wins or loses, I have bet 106 €, but I can only win 100 €. Therefore, I am guaranteed to lose money based on these beliefs. A rational agent will always attach consistent degrees of belief (*i.e.*, sum to 1) to outcomes of a random event.

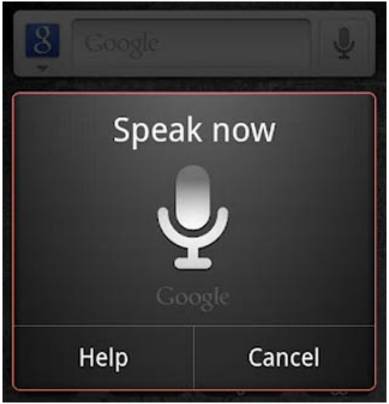
See `http://plato.stanford.edu/entries/dutch-book` for much more discussion.

Spam filtering

Many spam filters are based on the naive Bayes classifier.

Speech recognition

Hidden Markov models are the state of the art in speech recognition.



Natural language processing

Many modern NLP techniques are based on probabilistic graphical models.

Conclusions

The development of probabilistic models arose because symbolic logic often has trouble mimicking commonsense.

Pearl and other proposed solutions to many of the traditional arguments against numerical artificial intelligence.

Bayesian networks form a philosophically defensible cornerstone of many of these solutions.

In practice, probabilistic models have been used to solve many real-world problems.

Let's have a good semester!