Uncertainty

George Konidaris
gdk@cs.duke.edu

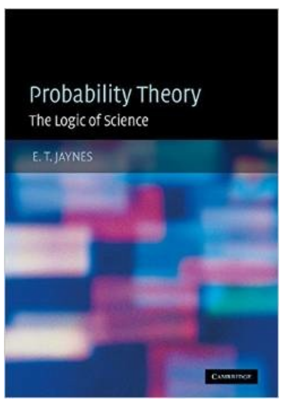
DUKE COMPUTER SCIENCE

Spring 2015

Logic is Insufficient

The world is not deterministic.
There is such thing as a fact.
Generalization is hard.
Sensors and actuators are noisy.
Plans fail.
Models are not perfect.
Learned models are *especially* imperfect.

$$\forall x, Fruit(x) \implies Tasty(x)$$



Probabilities

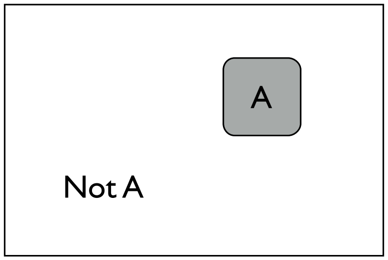
Powerful tool for reasoning about uncertainty.

But, they're tricky:
- Intuition often wrong or inconsistent.
- Difficult to *get*.

What do probabilities **really mean**?

Relative Frequencies

Defined over *events*.



P(A): probability random event falls in *A*, rather than *Not A*.
Works well for dice and coin flips!

Relative Frequencies

But this feels limiting.

What is the probability that the Blue Devils will beat Syracuse on Saturday?

- Meaningful question to ask.
- Can't count frequencies (except naively).
- Only really happens once.

In general, *all events only happen once*.

Probabilities and Beliefs

Suppose I flip a coin and hide outcome.

- What is P(Heads)?

This is a statement about *a belief*, not *the world*.
(the world is in exactly one state, with prob. 1)

Assigning truth values to probabilities is tricky - must reference speaker's *state of knowledge*.

Frequentists: probabilities come from relative frequencies.
Subjectivists: probabilities are degrees of belief.

For Our Purposes

No two events are identical, or completely unique.

Use probabilities as beliefs, but allow data (relative frequencies) to influence these beliefs.

We use *Bayes' Rule* to combine prior beliefs with new data.

Can prove that a person who holds a system of beliefs inconsistent with probability theory can be fooled.

To The Math

Probabilities talk about *random variables:*

- X, Y, Z, with domains d(X), d(Y), d(Z).
- Domains may be *discrete* or *continuous*.
- X = x: RV X has taken value x.
- P(x) is short for P(X = x).

Examples

X: RV indicating winner of Duke vs. Syracuse game.

d(X) = {Duke, Syracuse, tie}.

A probability is associated with each *event* in the domain:

- P(X = Duke) = 0.8
- P(X = Syracuse) = 0.19
- P(X = tie) = 0.01

Note: probabilities over *the entire event space* must sum to 1.

Expectation

Common use of probabilities: each event has *numerical value*.

Example: 6 sided die.
What is the *average die value?*

$$\frac{(1+2+3+4+5+6)}{6} = 3.5$$

In general, given RV X and function *f(x)*:

$$E[f(x)] = \sum_x P(x) f(x)$$

Expectation

For example, in min-max search, we assumed the opposing player took the min valued action (for us).

But that assumes perfect play. If we have *a probability distribution over the player's actions*, we can calculate their *expected value* (as opposed to min value) for each action.

Result: *expecti-max* algorithm.

Kolmogorov's Axioms of Probability

- 0 <= P(x) <= 1
- P(true) = 1, P(false) = 0
- P(a or b) = P(a) + P(b) - P(a and b)

Sufficient to completely specify probability theory for discrete variables.

Multiple Events

When several variables are involve, think about *atomic events*.

- Complete assignment of all variables.
- All possible events.
- Mutually exclusive.

RVs: Raining, Cold (both binary):

Raining	Cold	Prob.
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

joint distribution

Note: *still adds up to 1.*

Joint Probability Distribution

Probabilities to all possible atomic events *(grows fast)*

Raining	Cold	Prob.
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

Can define individual probabilities in terms of JPD:
P(Raining) = P(Raining, Cold) + P(Raining, not Cold) = 0.4.

$$P(a) = \sum_{e_i \in e(a)} P(e_i)$$

Independence

Critical property! But rare.

If A and B are independent:

- P(A and B) = P(A)P(B)
- P(A or B) = P(A) + P(B) - P(A)P(B)

Can break joint prob. table into separate tables.

Independence

Are *Raining* and *Cold* independent?

Raining	Cold	Prob.
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

P(Raining) = 0.4
P(Cold) = 0.7

Independence: Examples

Independence: two events don't effect each other.

- Duke winning NCAA, Dem winning presidency.
- Two successive, fair, coin flips.
- It raining, and winning the lottery.
- Poker hand and date.

Often we have an intuition about independence, but *always verify*. **Dependence does not mean causation!**

Mutual Exclusion

Two events are mutually exclusive when:

- P(A or B) = P(A) + P(B).
- P(A and B) = 0.

This is *different from* independence.

Independence is Critical

To compute P(A and B) we need a joint probability.

- This grows very fast.
- Need to sum out the other variables.
- Might require lots of data.
- NOT a function of P(A) and P(B).

If A and B are independent, then you can use separate, smaller tables.

Much of machine learning and statistics is concerned with identifying and leveraging independence and mutual exclusivity.

Conditional Probabilities

What if you have a joint probability, and you *acquire new data?*

My iPhone tells me that its cold. What is the probability that it is raining?

Raining	Cold	Prob.
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

Write this as:
- P(Raining | Cold)

Conditional Probabilities

We can write:

$$P(a|b) = \frac{P(a \text{ and } b)}{P(b)}$$

This tells us the probability of *a* **given only knowledge b**.

This is a probability w.r.t a **state of knowledge.**
- P(Disease | Symptom)
- P(Raining | Cold)
- P(Duke win | injury)

Conditional Probabilities

P(Raining | Cold)
 = P(Raining and Cold)
 / P(Cold)

… P(Cold) = 0.7
… P(Raining and Cold) = 0.3

Raining	Cold	Prob.
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

P(Raining | Cold) ~= 0.43.

Note!
P(Raining | Cold) + P(not Raining | Cold) = 1!

Bayes's Rule

Special piece of conditioning magic.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

If we have conditional P(B | A) and we **receive new data** for B, we can compute **new distribution for A.** (Don't need joint.)

As evidence comes in, revise belief.

Bayes Example

Suppose P(cold) = 0.7, P(headache) = 0.6.
P(headache | cold) = 0.57

What is P(cold | headache)?

$$P(c|h) = \frac{P(h|c)P(c)}{P(h)}$$

$$P(c|h) = \frac{0.57 \times 0.7}{0.6} = 0.66$$

Not always symmetric!
Not always intuitive!