15-381: Artificial Intelligence
Fall 2009

Uncertainty and Probabilistic Reasoning

Manuela Veloso

Thanks to past 15-381 instructors

Uncertainty

Let action A_t = leave for airport t minutes before flight
Will A_t get me there on time?

Problems:

1. partial observability (road state, other drivers' plans)
2. noisy sensors (traffic reports)
3. uncertainty in action outcomes (flat tire)
4. immense complexity of modeling and predicting traffic

A "purely logical" approach is too rigid. You have to enumerate a ton of things before you can actually assert truth of a statement (lest you be "flaky"). Also you don't know everything. You don't know what the traffic will be like, if there will be an accident, etc.

Logical Approach to Representing Uncertainty

A_{25} will get me there on time if there's no accident on the bridge *and* it doesn't rain *and* my tires remain intact etc etc."

A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport

Hence a purely logical approach either

1. risks falsehood: "A_{25} will get me there on time", or
2. leads to conclusions that are too weak for decision making

Methods for Handling Uncertainty

- Default or nonmonotonic logic:
  - Assume my car does not have a flat tire
  - Assume A_{25} works unless contradicted by evidence
- Issues: What assumptions are reasonable? How to handle contradiction?

- Rules with fudge factors:
  - $A_{25} \mid\!\rightarrow_{0.3}$ get there on time
  - $Sprinkler \mid\!\rightarrow_{0.99} WetGrass$
  - $WetGrass \mid\!\rightarrow_{0.7} Rain$
- Issues: Problems with combination, e.g., *Sprinkler* causes *Rain*??

With monotonic logic, if having certain premises necessitates a conclusion, then adding premises doesn't change the conclusion. For example: If A then B; A; Therefore B; If C is also true, B is still true.

With nonmonotonic logic, adding premises changes your conclusion. For example: Bob is an eagle. Can he fly? Yes. But Bob has a broken wing. Can he fly? No. But Bob has a jetpack. Can he fly? Yes (etc.)

Modeling Uncertainty

- Probability
  - Model agent's degree of belief
  - Given the available evidence,
  - A_{25} will get me there on time with probability 0.04

Imagine you are doing machine vision and tracking an orange soccer ball, which is currently at rest. Suddenly the ball appears across the field. Maybe it was just kicked, maybe this is a noisy observation, or maybe the ball teleported. You can use probability to model “degree of belief” given the available evidence and certain prior knowledge (like P(ball gets kicked)=0.5, P(ball teleports)=0.1, etc).

Advantages of Probabilistic Reasoning

- Appropriate for complex, uncertain, environments
  - Will it rain tomorrow?
- Applies naturally to many domains
  - Robot predicting the direction of road, biology, Word paper clip
- Allows to generalize acquired knowledge and incorporate prior belief
  - Medical diagnosis
- Easy to integrate different information sources
  - Robot's sensors

Probability

Probabilistic assertions summarize effects of

- laziness: failure to enumerate exceptions, qualifications
- ignorance: lack of relevant facts, initial conditions

Probabilities relate propositions to agent's own state of knowledge

e.g., $P(A_{25} \mid \text{no reported accidents}) = 0.06$

Probabilities of propositions change with new evidence:

e.g., $P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$

Here laziness and ignorance are not bad things. Using a probabilistic approach does not show personal weakness.

$P(A \mid B)$ is read "probability of A given B" (meaning the probability that A is true assuming that B is true).

Making Decisions Under Uncertainty

Suppose, agent's knowledge includes these facts/data:

$P(A_{25}$ gets me there on time | …) = 0.04
$P(A_{90}$ gets me there on time | …) = 0.70
$P(A_{120}$ gets me there on time | …) = 0.95
$P(A_{1440}$ gets me there on time | …) = 0.9999

- Which action to choose?
  Depends on my preferences, e.g., missing flight vs. time spent waiting
  - Utility theory is used to represent and infer preferences
  - **Decision theory: probability theory + utility theory**

Utility function combines information into a scoring system. All you might care about is making your flight, or maybe you care about not wasting your time.

Random Variables

- Basic element: random variable
  - possible worlds defined by assignment of values to random variables.
- Boolean random variables
  - e.g., *Cavity* (do I have a cavity?)
- Discrete random variables
  - e.g., *Weather* is one of <*sunny,rainy,cloudy,snow*>

Cavity can be either true or false

Weather is either sunny, rainy, cloudy, or snowy

Propositions

- Elementary proposition constructed by assignment of a value to a random variable:
  - e.g., *Weather* = *sunny*, *Cavity* = *false*
  - (abbreviated as $\neg$*cavity*)
- Complex propositions formed from elementary propositions and standard logical connectives
  - e.g., *Weather* = *sunny* $\vee$ *Cavity* = *false*
- Preconditions and effects of planning actions

$\neg$ means "not"

$\vee$ means "or"

Syntax

- Atomic event: A complete specification of the state of the world about which the agent is uncertain

  E.g., if the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:
  $Cavity = false \wedge Toothache = false$
  $Cavity = false \wedge Toothache = true$
  $Cavity = true \wedge Toothache = false$
  $Cavity = true \wedge Toothache = true$

- Atomic events are mutually exclusive and exhaustive

Basic Notations

- Random variable
  - referring to an element / event whose status is unknown:
    A = “it will rain tomorrow”
- Domain (usually denoted by Ω)
  - The set of values a random variable can take:
    - “A = The stock market will go up this year”: Binary
    - “A = Number of Steelers wins in 2007”: Discrete
    - “A = % change in Google stock in 2007”: Continuous

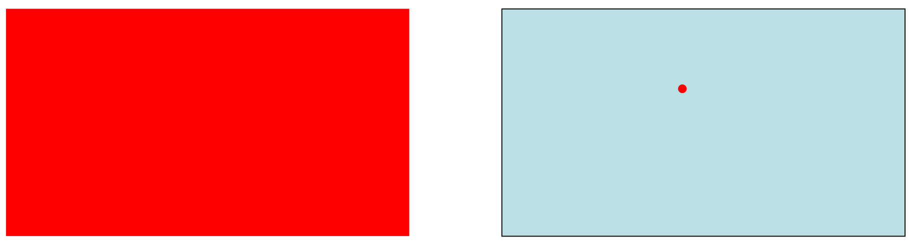
What is the domain of a die? If you had a random variable D, which is the value you get when you roll a die, what are the possible values D can have?

Axioms of Probability (Kolmogorov's Axioms)

Andrey Kolmogorov 1903-1987

- A variety of useful facts can be derived from just three axioms:

1. $0 \leq P(A) \leq 1$
2. $P(\mathit{true}) = 1$, $P(\mathit{false}) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Axioms of Probability (Kolmogorov's axioms)

- A variety of useful facts can be derived from just three axioms:

1. $0 \leq P(A) \leq 1$
2. $P(true) = 1$, $P(false) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

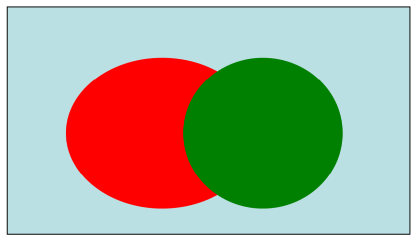
P(Today is September 23rd, 2009) = 1

Well, I suppose now P(Today is September 23rd, 2009) = 0

Axioms of Probability (Kolmogorov's Axioms)

- A variety of useful facts can be derived from just three axioms:

1. $0 \leq P(A) \leq 1$
2. $P(true) = 1$, $P(false) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

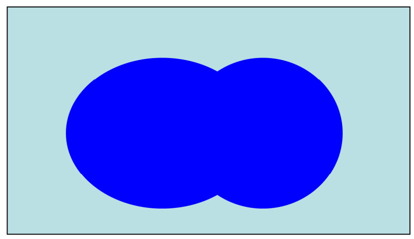


P(A) + P(B) double counts the intersecting area, hence you subtract P(A,B)

Axioms of Probability (Kolmogorov's Axioms)

- A variety of useful facts can be derived from just three axioms:

1. $0 \leq P(A) \leq 1$
2. $P(\mathit{true}) = 1$, $P(\mathit{false}) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Axioms of Probability (Kolmogorov's Axioms)

- A variety of useful facts can be derived from just three axioms:

1. $0 \le P(A) \le 1$
2. $P(\mathit{true}) = 1$, $P(\mathit{false}) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

There have been several other attempts to provide a foundation for probability theory. Kolmogorov's axioms are the most widely used.

Examples - Head / tail

Example - dice

Using the Axioms

- How can we use the axioms to prove that:
  $P(\neg A) = 1 - P(A)$
  ?

1 - P(A)

$= 1 - (-P(\neg A) + P(A \vee \neg A) + P(A \wedge \neg A))$ [Using rule 3]

$= 1 - (-P(\neg A) + 1 + 0)$ [Using rule 2, and the knowledge "A or not A is true", "A and not A is false"]

$= P(\neg A)$ [Cleaning up]

Priors

Degree of belief in an event in the absence of any other information

P(rain tomorrow) = 0.2

P(no rain tomorrow) = 0.8

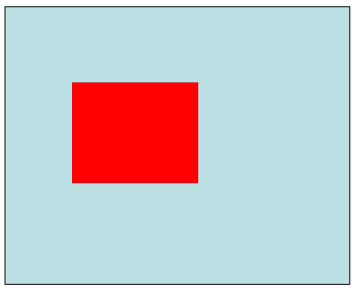
Conditional Probablity

- What is the probability of an event given knowledge of another event
- Example:
  - P(raining | sunny)
  - P(raining | cloudy)
  - P(raining | cloudy, cold)

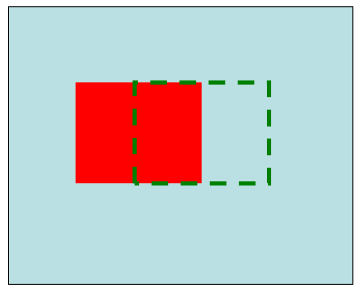
Conditional Probability

- P(A = 1 | B = 1): The fraction of cases where A is true if B is true

P(A = 0.2)



P(A|B = 0.5)



Conditional Probability

- In some cases, given knowledge of one or more random variables we can improve upon our prior belief of another random variable
- For example:

Liked movie	Slept	P
1	1	0.2
1	0	0.4
0	0	0.1
0	1	0.3

p(slept in movie) =

p(slept in movie | liked movie) =

p(didn't sleep in movie | liked movie) =

p(slept in movie) = 0.5

p(slept in movie | liked movie) = 1/3

p(didn't sleep in movie | liked movie) = 2/3

Joint Distributions

- The probability that a set of random variables will take a specific value is their joint distribution.
- Notation: $P(A \wedge B)$ or $P(A,B)$
- Example: P(liked movie, slept)

Liked movie	Slept	P
1	1	0.2
1	0	0.4
0	0	0.1
0	1	0.3

Joint Distribution (cont)

P(class size > 20) = 0.5

P(summer) =

P(class size > 20, summer) = ?

Evaluation of classes

Time (regular =2, summer =1)	Class size	Evaluation (1-3)
1	10	2
2	34	3
1	12	2
2	65	1
2	15	3
2	43	1
1	13	3
2	51	2

Joint Distribution (cont)

P(class size > 20) = 0.5

P(summer) = 3/8

P(class size > 20, summer) = 0

Evaluation of classes

Time (regular =2, summer =1)	Class size	Evaluation (1-3)
1	10	2
2	34	3
1	12	2
2	65	1
2	15	3
2	43	1
1	13	3
2	51	2

Joint Distribution (cont)

P(class size > 20) = 0.5

P(eval = 1) = ?

P(class size > 20, eval = 1) = ?

Evaluation of classes

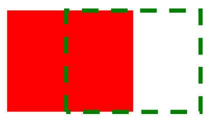
Time (regular =2, summer =1)	Class size	Evaluation (1-3)
1	10	2
2	34	3
1	12	2
2	65	1
2	15	3
2	43	1
1	13	3
2	51	2

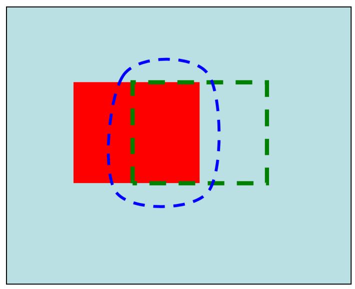
Chain Rule

- The joint distribution can be specified in terms of conditional probability:

$$P(A,B) = P(A|B)*P(B)$$

- Together with Bayes rule (which is actually derived from it) this is one of the most powerful rules in probabilistic reasoning





Bayes Rule

- One of the most important rules for AI usage.
- Derived from the chain rule:

  $P(A,B) = P(A \mid B)P(B) = P(B \mid A)P(A)$
- Thus,

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

Thomas Bayes was an English clergyman who set out his theory of probability in 1764.

Very simple rule, but very important.

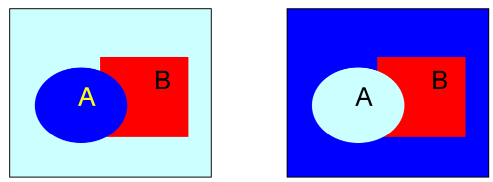
Bayes Rule (cont)

Often it would be useful to derive the rule a bit further:

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

This results from:
$P(B) = \sum_A P(B,A)$

P(B,A=1) P(B,A=0)



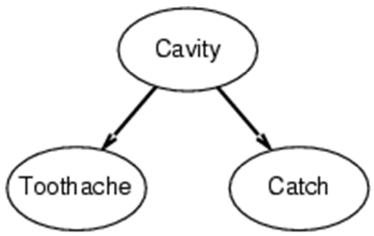
Conditional Independence

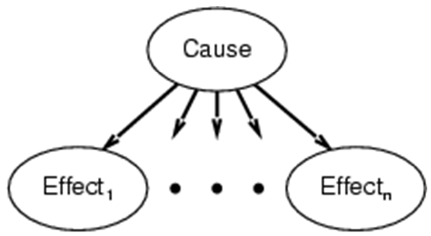
- *A* and *B* are independent
  iff
  $\mathbf{P}(A|B) = \mathbf{P}(A)$
  or $\mathbf{P}(B|A) = \mathbf{P}(B)$
  or $\mathbf{P}(\mathrm{A}, \mathrm{B}) = \mathbf{P}(A)\, \mathbf{P}(B)$

3 different ways of rephrasing the same idea.

Bayes' Rule and Conditional Independence

- Naïve Bayes model:
- $\mathbf{P}(\text{Cause},\text{Effect}_1, \ldots ,\text{Effect}_n) = \mathbf{P}(\text{Cause})\ \pi_i \mathbf{P}(\text{Effect}_i|\text{Cause})$





- Total number of parameters is linear in *n*

Naïve Bayes assumes conditional independence. It makes life a LOT easier.

Important Points

- Uncertainty
- Handling uncertainty
- Random variables
- Chain rule
- Bayes rule
- Joint distribution, independence, conditional independence