Chapter 13: Probability

- Properties:
    - Partially observable world.
    - Still deterministic, discrete.

Example: driving to the airport.

- Uncertainty: Road state, other drivers' plans, weather, flat tire, ...
- A purely logical approach either
    - Risks falsehood: Plan A_25 will get me there on time.
    - Gives answers that are too weak for decision making: A_25 will get me there on time if I don't get a flat tire, no accident, etc.
    - Is more certain than necessary: Get to the airport the day before.

Probabilistic analysis is present in almost all practical AI tasks.

- Health: diagnosis, testing, response to treatment
- Logistics: weather, traffic, demand
- Finance: weather, global forces

Probabilities, atomic events and events

Atomic event: Specific possible world.

$$\omega \in \Omega$$

Probability model: Assignment P(ω) to each atomic event ω.

- Atomic event: Specific possible world
  - Example: Roll a die
  - Example: Roll two dice. Atomic events are pairs of rolls = (1,1), (1,2), (2,1), ... (6,6)
  - Atomic event ω in Ω (set of all atomic events)

- Probability model: Assignment P(w) to each w in Omega.
  - 0 <= P(w) <= 1
  - sum_{w in Omega} P(w) = 1
  - One die: P(1) = P(2) = ... = P(6) = 1/6
  - Two dice: P((1,1)) = P((1,2)) = ... = P(6,6) = 1/36

Events

An *event* A is a subset of of atomic events.

$$A \subseteq \Omega$$

- An Event A is a subset of Omega
  - A = die roll < 4
  - Probability of an event is the sum of probabilities of atomic events
  - P(A) = P(1) + P(2) + P(3) = 1/2

Random variables

A *random variable* is a mapping from atomic events to values.

Domain of a random variable: Set of possible values it can take.

- Examples:
  - Odd = First die is odd.
  - Sum = Sum of dice rolls.
  - Doubles = Both dies have the same value

- We can define events according to random variable values.
  - P(Odd = true) = 1/2
  - P(Sum = 2) = 1/36
  - P(Sum = 3) = 2/36
  - P(Sum >= 4) = 33/36
  - P(Double = true) = 6/36

- Types of domains:
  - Boolean: Odd in {true, false}. Also called propositional random variables.
  - Discrete: (Weather in {sunny, rainy, cloudy})
  - Continuous: Temperature = 21.3

Notation and shorthand

We use capital letters for random variables: Odd*,* *Total*, ...

We use lowercase letters for atomic events: *true*, *false, heads, sunny...*

Shorthand:

- *P(sunny): P(Weather = sunny)*
- *P(odd): P(Odd = true)*
- *P(¬odd): P(Odd = false)*
- *P(Weather) = [0.6, 0.1, 0.2]*

Propositional logic using Boolean random variables

- We can define events based on logical expressions of random variables
  - P(sunny ^ odd)
  - P(sunny v odd)

Joint probability distribution

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

- Defines a probability for every atomic event.
- P(Cavity = true, Weather = sunny)
- Could also be written as P(Cavity = true ^ Weather = sunny)

Conditional probability

- $P(a \mid b) = P(a, b) / P(b)$
- Alternatively: $P(a, b) = P(a \mid b) P(b)$
- Chain rule: $P(X1, X2, ..., Xn) = P(X1) P(X2 \mid X1) P(X3 \mid X1, X2) ...$

Inference by enumeration

We want to compute P(y I e).

We have the joint distribution P(Y,E,H).

Each random variable is in one of three sets:
- Query Y
- Evidence E
- Hidden H

$$P(y|e) = \frac{P(y,e)}{P(e)} = \sum_{h \in H} \frac{P(y,e,h)}{P(e,h)} = \sum_{h \in H} \frac{P(y,e,h)}{\sum_{y' \in Y} P(y,e,h)}$$

- Variables divided into three sets:
  - Query Y
  - Evidence E
  - Hidden H
- We want to compute: P(Y I E) = P(Y,E) / P(E) = \sum_{h in H} P(Y,E,h) / P(E,h)

Normalization factor

$$P(Y|e) = \frac{P(Y,e)}{P(e)} = \alpha P(Y,e) = \alpha \sum_{h \in H} P(Y,e,h)$$

$$\alpha = \sum_{y \in Y} P(Y,e)$$

- We know P(y I e) sums to 1.
- Instead of calculating the denominator explicitly, we can calculate the whole distribution P(Y, e) and multiply it by whatever factor makes is sum to 1.
- alpha: "normalization factor", "partition function"

Inference by enumeration

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

- First task: inference.
- Given an event, sum the atomic events where it is true.
- P(toothache) = 0.2
- P(cavity v toothache) = 0.28
- P(cavity ^ toothache) = 0.12

Inference by enumeration

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

- Inference in general:

Inference by enumeration

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

Inference by enumeration

Rain	Traffic	Late	Probability
F	F	F	0.729
F	F	T	0.081
F	T	F	0.063
F	T	T	0.027
T	F	F	0.018
T	F	T	0.002
T	T	F	0.056
T	T	T	0.024

- P(no rain, traffic, not late)? 0.063
- P(late I rain, traffic). Two ways:
    - Normalization factor: [0.056, 0.024]. Alpha = 0.08. 0.024/0.08 = 0.3 .
    - Denominator: 0.024 / (0.056 + 0.024) = 0.3 .
- P(rain): 0.1
- P(late I rain):
    - P(Late I rain) = [0.018+0.056, 0.024+0.002] = 0.074, 0.026
    - Alpha = 0.1
    - P(late I rain) = 0.26

Independence

- W43-Wed:
  - Contest:
    - Focus on: choice of symbol to branch on.
    - Branch on symbol that appears in the most clauses.
      - Better: Favor variables in shorter clauses.
    - Start with the value that satisfies the most clauses.

- Consider variables: Cavity, Catch, Toothache, Weather
- Intuitively, Weather is unrelated.
- Two sets of variables: Cavity, Toothache, Catch. Weather.
- A and B are independent (A _I_ B) iff P(A,B) = P(A)*P(B)
  - Equivalently, P(A|B) = P(A)
- Makes inference much easier. For n independent biased coins:
  - Size of probability table: 2^n-1 -> n
  - Time for inference on a single variable: O(2^n) -> O(1)
- Full independence is rare.

Conditional independence

- Assuming I have a cavity, probability that probe catches doesn't depend on toothache.
    - Intuitively
    - Evaluate based on table
- toothache <- cavity -> catch
- Conditional independence:
    - A is conditionally independent of B given C iff P(A|B,C) = P(A|C)
    - Equivalent: P(A,B | C) = P(A | C) P(B | C)

P(Toothache, Catch, Cavity) has 8 independent entries
Chain rule:
P(Toothache, Catch, Cavity)
= P(Toothache|Catch, Cavity)P(Catch, Cavity)
= P(Toothache|Catch, Cavity)P(Catch|Cavity)P(Cavity)
= P(Toothache|Cavity)P(Catch|Cavity)P(Cavity)
-> 5 independent entries

Using just the conditional probability tables, we can reconstruct the full joint distribution.

We almost always represent probability distributions using conditional probability tables. Benefits
- Much smaller representation.
- Usually more intuitive.
- Usually more efficient for inference.

Another example: Rain, Traffic, Late

Bayes' rule

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

- We can derive from product rule.
- P(cause I effect) = P(effect I cause) P(cause) / P(effect)

Example: Meningitis causes stiff neck.
P(m) = 1/50,000
P(s) = 0.1
P(s I m) = 0.7
->
P(m I s) = 0.0008