Probabilistic Models

- Models describe how (a portion of) the world works
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - “All models are wrong; but some are useful.”
    – George E. P. Box

What do we do with probabilistic models?
- We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information

Model for Ghostbusters

- Reminder: ghost is hidden, sensors are noisy
- T: Top sensor is red
  B: Bottom sensor is red
  G: Ghost is in the top
- Queries:
P(+g) = ??
P(+g | +t) = ??
P(+g | +t, -b) = ??
- Problem: joint distribution too large / complex

Independence

- Two variables are independent if:
  \( \forall x, y : P(x,y) = P(x)P(y) \)
  - This says that their joint distribution factors into a product two simpler distributions
  - Another form:
    \( \forall x, y : P(x|y) = P(x) \)
  - We write: \( X \independent Y \)
- Independence is a simplifying modeling assumption
  - Empirical joint distributions: at best “close” to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?
Conditional Independence

- \( P(\text{Toothache}, \text{Cavity}, \text{Catch}) \)

- If I have a cavity, the probability that the probe catches in it doesn’t depend on whether I have a toothache:
  \( P(+\text{catch} | +\text{toothache}, +\text{cavity}) = P(+\text{catch} | +\text{cavity}) \)

- The same independence holds if I don’t have a cavity:
  \( P(+\text{catch} | +\text{toothache}, -\text{cavity}) = P(+\text{catch} | -\text{cavity}) \)

- Catch is conditionally independent of Toothache given Cavity:
  \( P(\text{Catch} | \text{Toothache}, \text{Cavity}) = P(\text{Catch} | \text{Cavity}) \)

- Equivalent statements:
  - \( P(\text{Toothache} | \text{Catch}, \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) \)
  - \( P(\text{Toothache, Catch} | \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) P(\text{Catch} | \text{Cavity}) \)
  - One can be derived from the other easily

The Chain Rule

- \( P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_2,X_1)\ldots \)

- Trivial decomposition:
  \( P(\text{Traffic, Rain, Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain, Traffic}) \)

- With assumption of conditional independence:
  \( P(\text{Traffic, Rain, Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}) \)

- Bayes’ nets / graphical models help us express conditional independence assumptions

Bayes’ Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time

- Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions
  - For about 10 min, we’ll be vague about how these interactions are specified

Ghostbusters Chain Rule

- Each sensor depends only on where the ghost is:
  - Traffic
    - Umbrella
    - Raining
  - What about fire, smoke, alarm?

Example Bayes’ Net: Insurance
Example Bayes’ Net: Car

Graphical Model Notation

- **Nodes**: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

- **Arrows**: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)
  - For now: imagine that arrows mean direct causation (in general, they don’t!)

Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

Example: Traffic

- **Variables**
  - R: It rains
  - T: There is traffic

- **Model 1**: independence

- **Model 2**: rain causes traffic

- Why is an agent using model 2 better?

Example: Traffic II

- Let’s build a causal graphical model

- **Variables**
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Example: Alarm Network

- **Variables**
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!
Bayes’ Net Semantics

- Let’s formalize the semantics of a Bayes' net
- A set of nodes, one per variable $X$
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
  \[ P(X|A_1 \ldots A_n) \]
- CPT: conditional probability table
- Description of a noisy ‘causal’ process

A Bayes net = Topology (graph) + Local Conditional Probabilities

Probabilities in BNs

- Bayes’ nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
  \[ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(x_i)) \]
  - Example:
  \[ P(+\text{cavity}, +\text{catch}, -\text{toothache}) \]

- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Example: Coin Flips

\[
\begin{align*}
X_1 & \quad X_2 & \quad \cdots & \quad X_n \\
P(X_1) & \quad P(X_2) & \quad \cdots & \quad P(X_n) \\
h & 0.5 & h & 0.5 & \cdots & h & 0.5 \\
t & 0.5 & t & 0.5 & \cdots & t & 0.5 \\

P(h, h, t, h) =
\end{align*}
\]

Example: Traffic

\[
P(R) = P(+r, -t) =
\]

Example: Alarm Network

\[
B P(B)
+ b 0.001
- b 0.999

Earthquake
B P(B)
+ e 0.002
- e 0.998

Alarm
B E A P(A|B,E)
+ b + e + a 0.95
+ b + e - a 0.05
+ b - e + a 0.94
+ b - e - a 0.06
- b + e + a 0.29
- b + e - a 0.71
- b - e + a 0.001
- b - e - a 0.999

John calls
M P(M|A)
+ a + m 0.7
- a - m 0.3

Mary calls
M P(M|A)
+ a + m 0.01
- a - m 0.99

Bayes’ Nets

- So far: how a Bayes’ net encodes a joint distribution
- Next: how to answer queries about that distribution
  - Key idea: conditional independence
  - Last class: assembled BNs using an intuitive notion of conditional independence as causality
  - Today: formalize these ideas
  - Main goal: answer queries about conditional independence and influence
- After that: how to answer numerical queries (inference)