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

Independence

- Two variables are independent if:
  \[ P(x, y) = P(x)P(y) \]
  - This says that their joint distribution factors into a product two simpler distributions
  - Another form:
    \[ P(x|y) = P(x) \]
  - We write: \[ X \perp Y \]
- 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

- Conditional Independence
  - If I have a cavity, the probability that the probe catches in it doesn’t depend on whether I have a toothache:
    - \( P(\text{Catch} \mid +\text{toothache}, +\text{cavity}) = P(\text{Catch} \mid +\text{cavity}) \)
  - The same independence holds if I don’t have a cavity:
    - \( P(\text{Catch} \mid +\text{toothache}, -\text{cavity}) = P(\text{Catch} \mid -\text{cavity}) \)
  - Catch is conditionally independent of Toothache given Cavity:
    - \( P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity}) \)
  - Equivalent statements:
    - \( P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \)
    - \( P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) \)
    - One can be derived from the other easily.

Unconditional (absolute) independence very rare (why?)

- Conditional independence is our most basic and robust form of knowledge about uncertain environments.
- \( X \) is conditionally independent of \( Y \) given \( Z \) if and only if:
  - \( \forall x, y, z : P(x, y \mid z) = P(x \mid z)P(y \mid z) \)
  - or, equivalently, if and only if
  - \( \forall x, y, z : P(x, z, y) = P(x \mid z) \)

Conditional Independence and the Chain Rule

- Chain rule:
  - \( P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_2, X_1) \ldots \)
- Trivial decomposition:
  - \( P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic} \mid \text{Rain})P(\text{Umbrella} \mid \text{Rain}, \text{Traffic}) \)
- With assumption of conditional independence:
  - \( P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic} \mid \text{Rain})P(\text{Umbrella} \mid \text{Rain}) \)
- Bayes’ nets / graphical models help us express conditional independence assumptions

Ghostbusters Chain Rule

- Each sensor depends only on where the ghost is
  - \( P(T, B, G) = P(G)P(T \mid G)P(B \mid G) \)
- That means, the two sensors are conditionally independent, given the ghost position
  - \( P(T, B, G) = P(G)P(T \mid G)P(B \mid G) \)
- \( P(T, B, G) \)
  - \( T \): Top square is red
  - \( B \): Bottom square is red
  - \( G \): Ghost is in the top
- Given\: \( P(\text{tg}) = 0.5 \)
  - \( P(\text{tg}) = 0.5 \)
  - \( P(\text{tg}) = 0.4 \)
  - \( P(\text{tg}) = 0.8 \)
- \( P(\text{tg}) = 0.5 \)
  - \( P(\text{tg}) = 0.5 \)
  - \( P(\text{tg}) = 0.8 \)
- \( P(\text{tg}) = 0.5 \)
  - \( P(\text{tg}) = 0.5 \)
  - \( P(\text{tg}) = 0.8 \)
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

Graphical Model Notation

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

- Arcs: 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 Bayes’ Net: Insurance

Example Bayes’ Net: Car

Example: Coin Flips

- N independent coin flips

  \[ X_1, X_2, \ldots, X_N \]

- 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

- 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
    \[ P(X|\alpha_1 \ldots \alpha_n) \]
  - 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:

\[ P(\text{+cavity, +catch, -toothache}) \]
Why are we guaranteed that setting results in a proper joint distribution?

Chain rule (valid for all distributions):

\[
P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i))
\]

Assume conditional independences:

\[
P(x_1 | x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = P(x_1 | \text{parents}(X_i))
\]

Consequence:

\[
P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i))
\]

Not every BN can represent every joint distribution

- The topology enforces certain conditional independencies

Example: Coin Flips

\[
P(X_1, X_2, \ldots, X_n) = P(X_1) P(X_2) \cdots P(X_n)
\]

Example: Traffic

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

Example: Alarm Network

Example: Reverse Traffic

Causal direction

Reverse causality?
When Bayes’ nets reflect the true causal patterns:
- Often simpler (nodes have fewer parents)
- Often easier to think about
- Often easier to elicit from experts

Bayes’ nets need not actually be causal
- Sometimes no causal net exists over the domain (especially if variables are missing)
- E.g., consider the variables Traffic and Drips
- End up with arrows that reflect correlation, not causation

What do the arrows really mean?
- Topology may happen to encode causal structure
- Topology really encodes conditional independence

So far: how a Bayes’ net encodes a joint distribution

Next: how to answer queries about that distribution
- Today:
  - First assembled BNs using an intuitive notion of conditional independence as causality
  - Then saw that key property is conditional independence
- Main goal: answer queries about conditional independence and influence

After that: how to answer numerical queries (inference)