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

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

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

Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic
- Model 1: Independence
- Model 2: rain causes traffic

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
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
  - 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:

Example: Coin Flips

- Why are we guaranteed that setting results in a proper joint distribution?
  \[ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(x_i)) \]
- 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:
  \[ P(x_i | x_1, \ldots, x_{i-1}) = P(x_i | \text{parents}(x_i)) \]
  - 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: Traffic

- Chain rule (valid for all distributions):
  \[ P(T | R) = \prod_{i=1}^{n} P(T_i | R_i) \]

Example: Alarm Network

- 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:
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

- BNs 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
    \[ P(x_i | x_1, \ldots, x_{i-1}) = P(x_i | \text{parents}(x_i)) \]