

## Bayesian Networks

- Nodes: set of random variables. Try to order them in a way s.t. causes precede effects. Let the ordering be  $\{X_1, X_2, \dots, X_n\}$
- Links: For each node  $X_i$ 
  - Choose a minimal set of parents for  $X_i$  from  $\{X_1, \dots, X_{i-1}\}$
  - $P(X_i | X_{i-1}, \dots, X_1) = P(X_i | \text{Parents}(X_i))$
  - For each parent, insert a link from parent to  $X_i$
  - Write down the CPTs or Conditional Probability Tables,  $P(X_i | \text{Parents}(X_i))$
- $\text{Parents}(X_i)$  directly influences  $X_i$
- No cycles and no redundancy by construction
- Causal mode: parents  $\leftrightarrow$  cause and child  $\leftrightarrow$  effect

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(x_i))$$

## Conditional Independence in Bayesian Networks

- Each variable is conditionally independent of its non-descendants, given its parents.
  - Instead of full joint distribution, define a set of conditional independence properties.
  - The full joint distribution can be derived from those properties
- A variable is conditionally independent of all other nodes in the network, given its parents, children, and children's parents.

## Efficient Representation of Conditional Distributions

- **Deterministic node:**  $X_i = f_i(\text{Parents}(X_i))$
- **Noisy-OR node:**  $P(X_i = 1 | \text{Parents}(X_i)) = 1 - \prod_{j: X_j = \text{true}} (q_j)$ , where  $q_j$  is the probability that  $X_j$  is the cause of  $X_i$ 
  - All causes should be listed (use misc if some are unknown)
  - inhibition of each parent is independent of inhibition of any other parent
- Noisy logical relationships in which a variable depends on  $k$  parents can be described using  $\mathcal{O}(k)$  parameters instead of  $\mathcal{O}(2^k)$  for the full conditional probability table

## Hybrid Bayesian Networks

- Need to specify 2 things
  - The conditional distribution for a continuous variable given discrete or continuous parents
  - The conditional distribution for a discrete variable given continuous parents

## Inference in Bayesian Networks

- Inference: given some evidence, compute the posterior distribution over the unobserved variables
- **Exact inference:** compute the posterior distribution exactly
- Inference by enumeration: Sum up all the paths that are consistent with the evidence.  $P(X|e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y)$ 
  - Variable elimination: basically memoization
  - pointwise product:  $f(X_1, \dots, X_j, Y_1, Y_k) \times g(Y_1, \dots, Y_k, Z_1, \dots, Z_l) = h(X_1, \dots, X_j, Y_1, \dots, Y_k, Z_1, \dots, Z_l)$
- **Approximation Inference:** They work by generating random events based on the probabilities in the Bayes net and counting up the different answers found in those random events. With enough samples, we can get arbitrarily close to recovering the true probability distribution—provided the Bayes net has no deterministic conditional distributions.