Bayesian Network

Important Rules

- Sum rule: $P(X = a) = \sum_{b \in dom(Y)} P(X = a|Y = b)P(Y = b)$
- Bayes' rule: $P(Y|X) = P(X|Y)\ P(Y)\ /\ P(X)$
- Chain rule: $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|X_{i-1}, \ldots, X_1)$

X is independent of Y if $P(X|Y) = P(X)$.

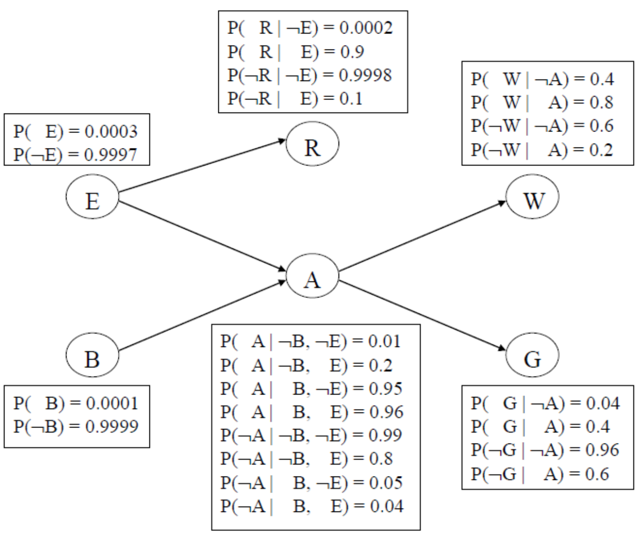
X is conditionally independent of Y given Z if $P(X|Y, Z) = P(X|Z)$.

Conditional independence assertions allow chain rule to be simplified.

Bayesian Network

A Bayesian network is a directed acyclic graph (DAG) where:

- nodes are random variables
- directed arcs connect pairs of nodes (if X → Y, then X has a direct influence on Y)
- each node has a conditional probability table specifying the parents' effects on the node



A Bayesian network is:

1. a representation of the joint probability distribution
2. an encoding of conditional independence assumptions

Inference

Task: Determine posterior probability of a set of query variables given exact values for some evidence variables.

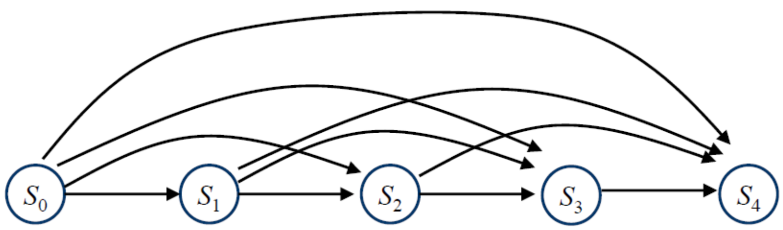
$$P(\text{Query}|\text{Evidence})$$

Dynamic System

So far, we assume the world does not change (i.e. static probability distribution).

Sometimes we need to reason over time:

- set of states
- set of time slices (snapshots of the world)
- different probability distribution over states at each time slice



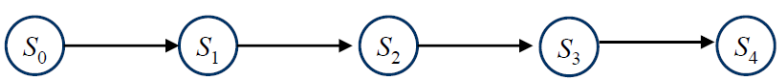
Conditional probability table: $\mathrm{P}(\mathrm{S}_{\mathrm{t}+1}|\mathrm{S}_{\mathrm{t}}, \dots, \mathrm{S}_0)$

This can be viewed as a Bayesian network with one random variable per time slice.

Markov Chain

Markov assumption: $\mathrm{P}(\mathrm{S}_{\mathrm{t}+1}|\mathrm{S}_{\mathrm{t}}, \dots, \mathrm{S}_0) = P(S_{t+1}|S_t)$

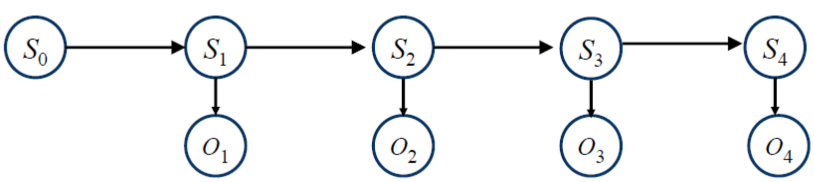
A Markov chain is stationary if $\mathrm{P}(\mathrm{S}_{\mathrm{t}+1}|\mathrm{S}_{\mathrm{t}})$ is the same for each $\mathrm{t} \geq 0$.



Hidden Markov Model

Set of states: S (hidden/latent random variables)

Set of observations: O (observable random variables)



Common tasks:

- Monitoring: $\mathrm{P}(\mathrm{S}_{\mathrm{t}}|\mathrm{O}_1, \dots, \mathrm{O}_{\mathrm{n}})$
- Prediction: $\mathrm{P}(\mathrm{S}_{\mathrm{t+k}}|\mathrm{O}_1, \dots, \mathrm{O}_{\mathrm{n}})$ where $\mathrm{k} \geq 1$
- Hindsight: $\mathrm{P}(\mathrm{S}_{\mathrm{t-k}}|\mathrm{O}_1, \dots, \mathrm{O}_{\mathrm{n}})$ where $1 \leq \mathrm{k} \leq \mathrm{t}$
- Most probable explanation: $\mathrm{argmax}_{\mathrm{S}_0,\dots,\mathrm{S}_{\mathrm{t}}} \mathrm{P}(S\_0, \dots, S\_t|O_1, \dots, O_t)$

HMMs are just DBNs with one state variable and one observation variable (???).