Bayesian Network: A directed graph in which each node is annotated with quantitative probability info.

- It is a representation of joint probability distribution.
- It is encoding of a collection of conditional independent statements.
- A generic entry in JPD is the probability of a conjuction of particular assignments to each variable, such as $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$ abbreviated as $P(x_1, \ldots x_n)$. The value of this entry is given by formula -

$$P(x_1, \ldots x_n) = \prod_{i=1}^{n} \theta(x_i \mid parents(X_i))$$ where

$parent(X_i)$ denotes the value of $Parents(X_i)$.

- Baysеian Network can too answer any query, by summing all relevant joint entries.
- Locally structured system: A component in LSS interacts directly with only a bounded no. of components.
- In fully connected ~~Bayesian~~ Bayesian Nw,
- Topological Sematics of BN: Specifies the conditional independence relⁿ encoded by graph structures.
  + Specifies that each ~~node~~ node ~~variable~~ is conditionally - independent of its non-descendants.
  + Markov blanket: A node is conditionally independent of all other nodes, given its parents, children & children's parents.
- The basic task of any probabilistic inference system is to compute the posterior probability distribution for a set of query variables, given some observed event - that is, some assignment of values to evidence variables

- A query can be answered using

$$P(X|e) = \alpha P(X,e) = \alpha \sum_{y} P(X,e,y)$$

+ A query can be computed using Baysean Nw by computing some of products of conditional probabilities from the network. (Inference by enumeration)
+ Enumeration algo can be improved substantially by eliminating repeated calculations. one approach variable elimination algo:
  - Singly connected networks / Polytrees: Contains at the most on undirected path betw any 2 nodes in the network.
    + Time & Space complexity is linear in size of the network.
  - Multiply connected networks: contains more than one undirected paths betw any 2 nodes.

- Clustering Algos
  + Variable elimination is less efficient in computing posterior probabilities. Total time $O(n^2)$
  + Clustering Algo: Time total $O(n)$
  + Joins individual nodes to form clusters to create ~~singly~~ polytree.

Approximate Inference in Bayesian Network

Method 1: Direct Sampling Methods

- Generates events from a Network that has no evidence associated with it.
- The frequency of sampling a specific event converges, in the limit, to expected value according to the event's sampling probability
- An estimated probability becomes exact in large sample limit - such an estimate is called consistent.

Method 2 : Rejection Sampling.

- Rejects samples that are not consistent with the evidence
- Given $P(X|\underline{e})$, the algo first generate samples from prior distribution & reject those that are not consistent with the given evidence
- Produces a consistent estimate of the true probability.
- Rejects too many samples.

Method 3 : Likelihood weighting

- Avoids inefficiency with rejection sampling by generating only events that are consistent with $\underline{e}$
- Suffers a degradation in performance & no. of $\underline{e}$ increases. Because most samples will have very low weights & ∴ the weighted estimate will be dominated by tiny fraction of samples that accord more than an infinitesimal likelihood to evidence.

Method 4 : Markov Chain Monte Carlo (MCMC) Algo.

- Makes random change to preceeding event in previous event. ~~Ass~~ Network can be assumed to be in current state. specifying value for each variable.
- Next state is generated by randomly sampling one of the non-evidence variables. condition on current values of variables in the Markov blanket..

※ Markov blanket of a variable consists of its parents, children & other parents of its children.

Extending Probability to First-order representation (4)

- First order probabilistic KB should specify probabilities for all possible first order models.
- For any FO sentence ϕ, the probability $P\phi$ is given in the usual way by summing over possible worlds where ϕ is true;

$$P\phi = \sum_{M:\phi \text{ is true in } M} \mu(M)$$, where $\mu(M)$ is model prob.

- Relational Probability Model can be used to represent FOL Models.

- Ashish R. Gavande