Introduction to Bayesian Network

Jiafei Han
Yanyue Wang

Outlines

- Introduction
- Bayesian Networks
- Probabilistic Inference
- Structure and Parameter learning

Introduction

The Alarm Example

- A burglar alarm
- Burglary or earthquakes
- Two neighbors (John, Mary)
- Given evidence about who has and hasn’t called, estimate the probability of a burglary

The Alarm Example

- Represent problem using 5 binary variables:
  - B = a burglary occurs at your house
  - E = an earthquake occurs at your house
  - A = the alarm goes off
  - J = John calls to report the alarm
  - M = Mary calls to report the alarm

- What is P(B | M) ?

  - We can use the full joint distribution to answer this question
    - Requires $2^5 = 32$ probabilities

  - Can we use prior domain knowledge to come up with a Bayesian network that requires fewer probabilities?

Bayesian Networks

- Definition: **BN = (DAG, CPD)**
  - **DAG**: directed acyclic graph (BN's **structure**)
    - **Nodes**: random variables $(X_1, X_2, \ldots, X_n)$
    - **Arcs**: indicate probabilistic dependencies between nodes
  - **CPD**: conditional probability distribution (BN's **parameters**)

$P(X_i \mid \pi(X_i))$, where $\pi(X_i)$ is the set of all parent nodes of X_i

- $P(X_1, X_2, \ldots, X_n) = P(X_1)\, P(X_2|X_1) \ldots P(X_n|X_1, X_2, \ldots, X_{n-1})$
$= \prod_i P(X_i \mid \pi(X_i))$

- Root nodes are a special case – no parents, so just use priors in CPD:

$$\pi(X_i) = \varnothing, \text{ so } P(X_i \mid \pi(X_i)) = P(X_i)$$

- Why Bayesian Networks are effective?
  - before, requires 2^N
  - after, requires $N \cdot 2^K$

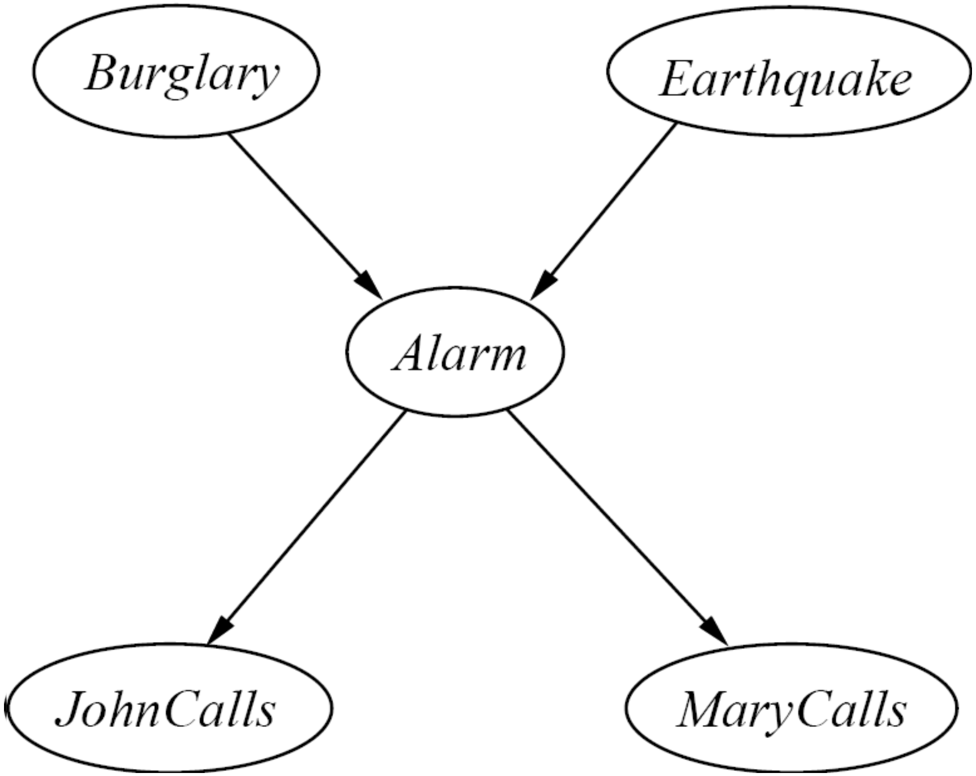
Constructing a BN: Step 1

- Order the variables in terms of causality (may be a partial order).
  - e.g., {E, B} -> {A} -> {J, M}

- Use these assumptions to create the graph structure of the Bayesian network.

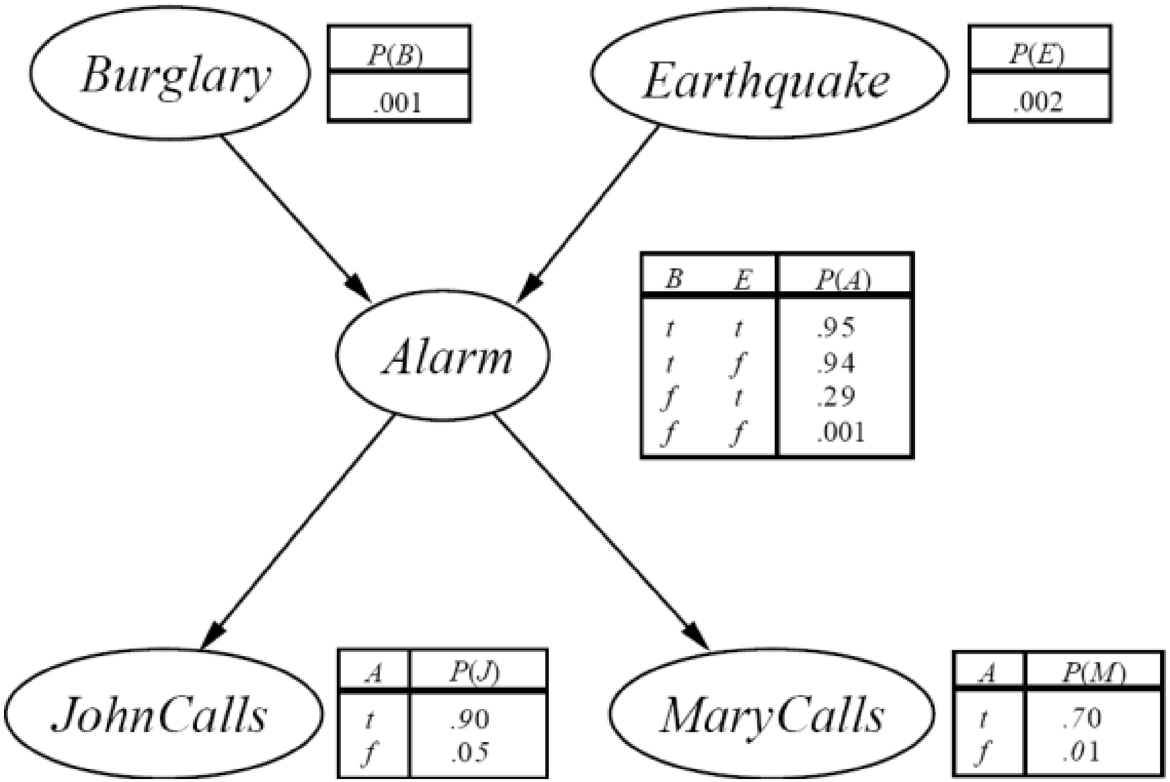
The Resulting Bayesian Network

- DAG



Constructing a BN: Step 2

- Fill in conditional probability tables (CPTs)
  - One for each node
  - 2^p entries, where p is the number of parents



P(B)
.001

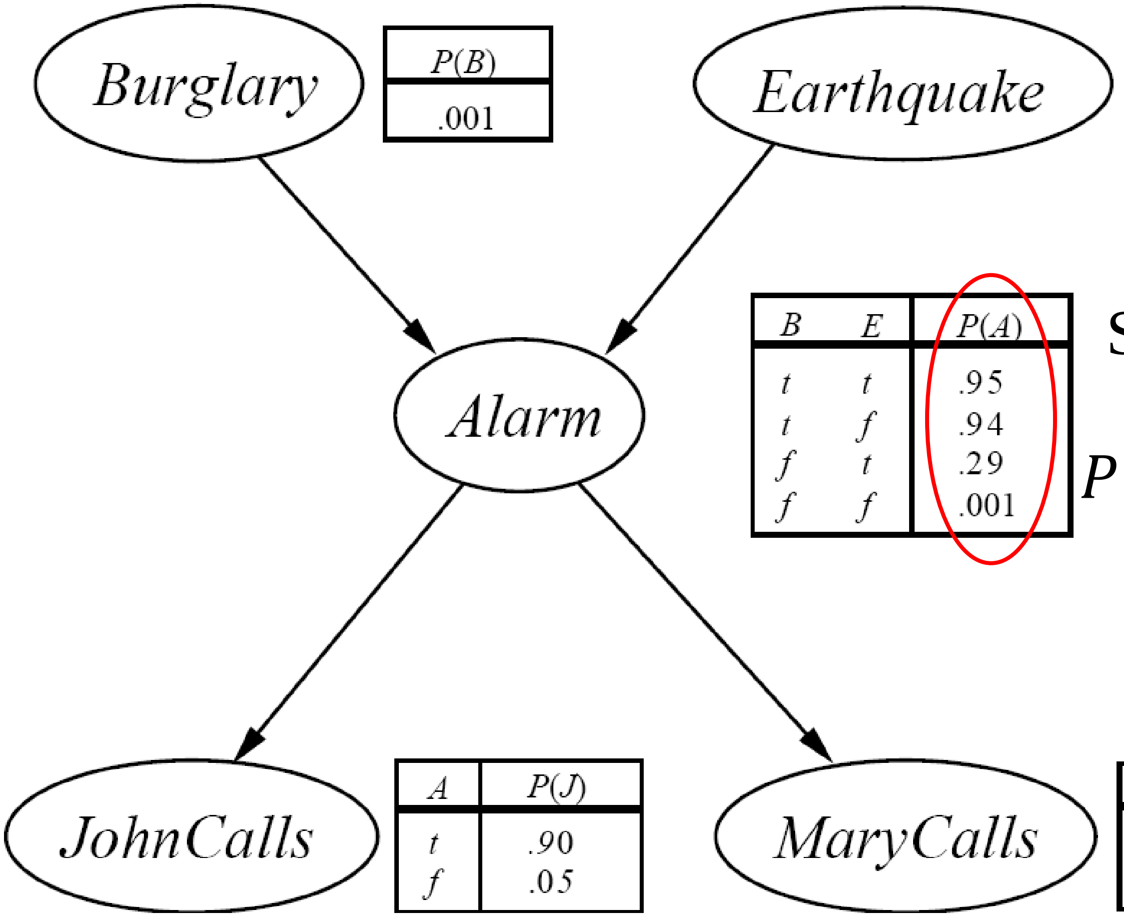
P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

The alarm example



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

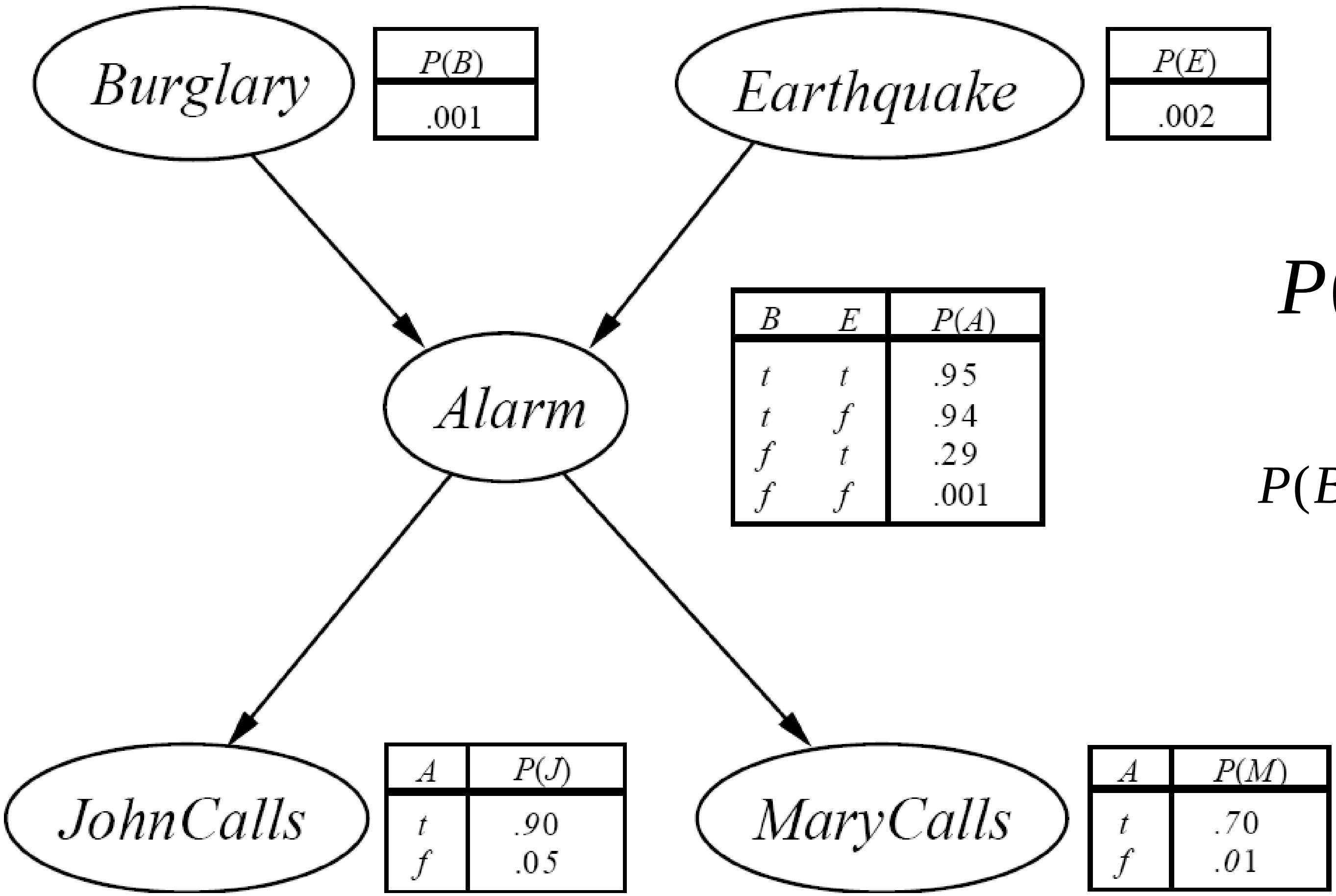
A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Shouldn't these add up to 1?

$$P(\neg A|B,E) = 1 - P(A|B,E) = 1 - 0.95 = 0.05$$

The alarm example



$$P(B \mid M)?$$

$$P(B \mid M) = \frac{P(M \mid B)P(B)}{P(M)}$$

- What are they?
  - a network-based framework, uncertainty

- Where did BNs come from?
  - artificial intelligence, decision analysis, and statistic communities

- What are they used for?
  - Intelligent decision aids, data fusion, feature recognition, intelligent diagnostic aids, automated free text understanding, data mining

Bayesian Network Inference

- The process of inference:

Known Information → Unknown Probability

- The process of inference:

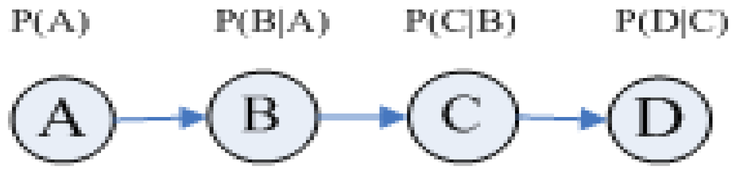
Joint Distribution (Complex) $\xrightarrow{\text{Algorithms}}$ Marginalized Distribution (Concise)

Bayesian Network Inference

(1)Variable Elimination (VE)

- **Purpose**: Finding the posterior distribution
- **Method:** Factorizing the probability distribution
- Simplify the inference

□Example



A,B,C,D are binary variables

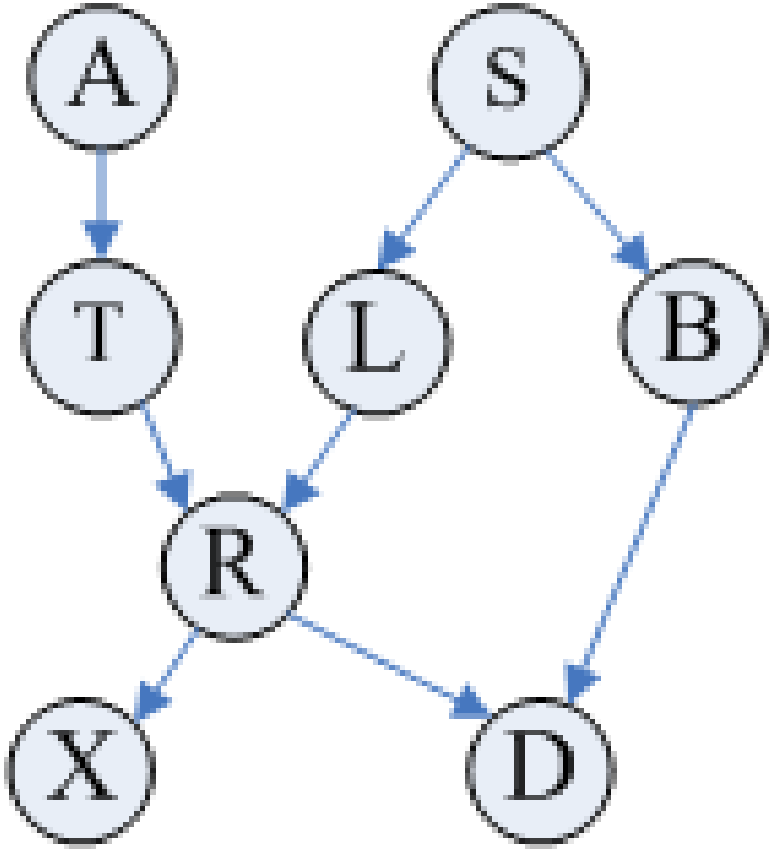
$$P(D) = \sum_{A,B,C} P(A,B,C,D) = \sum_{A,B,C} P(A)P(B|A)P(C|B)P(D|C) =$$

$$\sum_{C} P(D|C) \sum_{B} P(C|B) \sum_{A} P(A)P(B|A)$$

□Calculating the posterior

$$P(A|D=0) = \frac{h(A)}{\sum_{A} h(A)} \qquad h(A) = \sum_{B,C} P(B,C,D=0)$$

Bayesian Network Inference



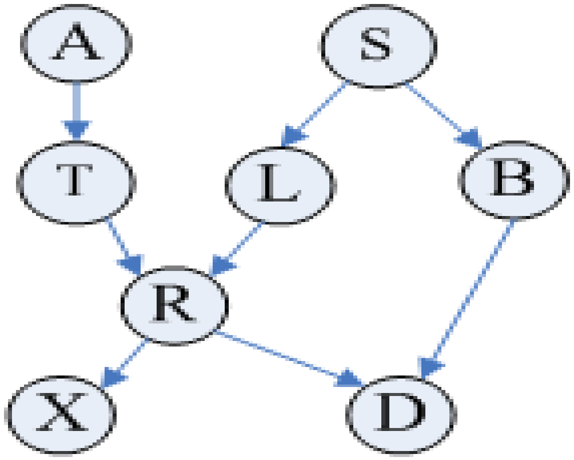
How to make inference?

How to calculate the posterior $P(A|X)$?

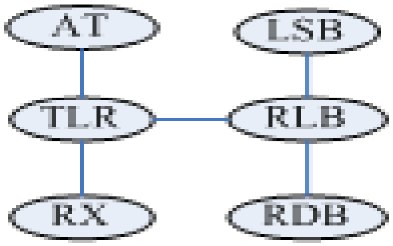
Bayesian Network Inference

(2) Clique Tree Propagation (CTP)

- Purpose: Calculating the posterior
- Method: Sharing the steps
- Simplify the inference



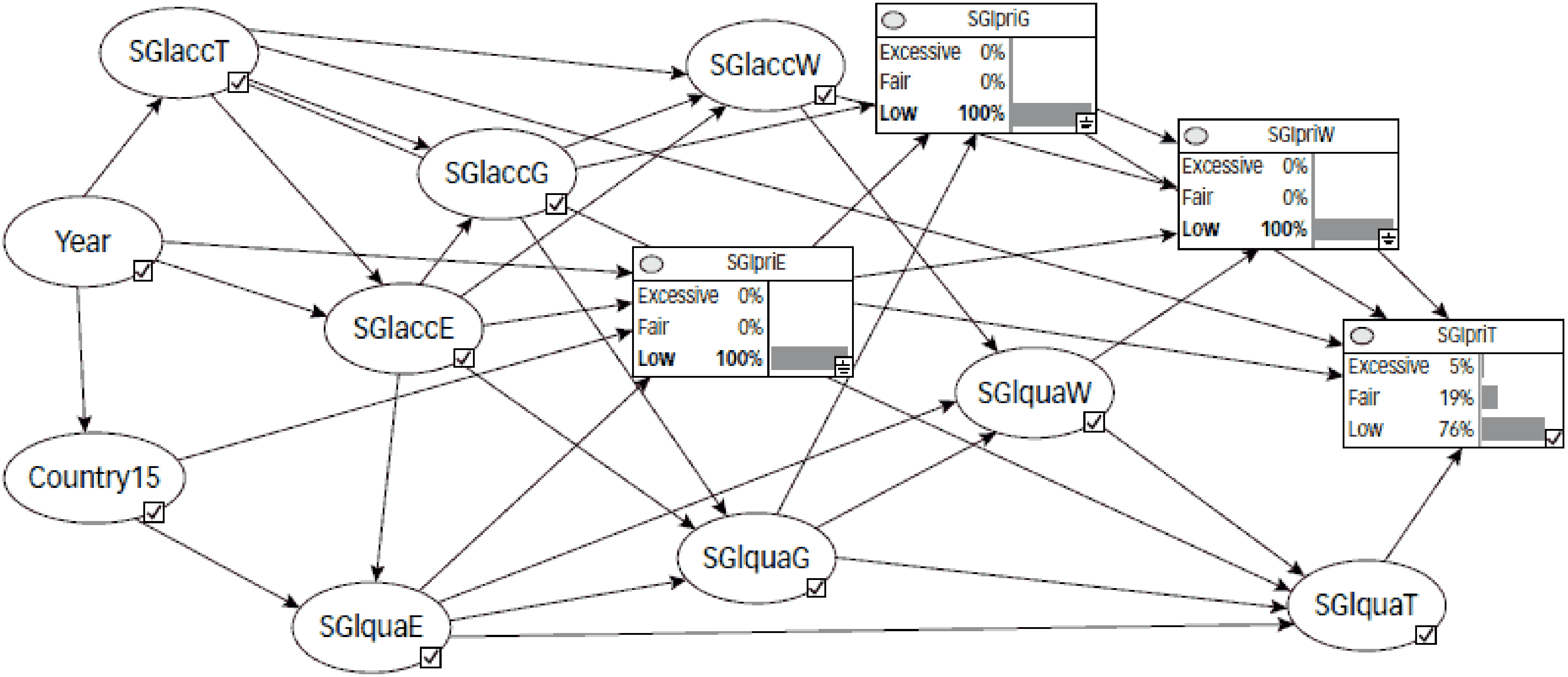
BN



Clique tree

Bayesian Network Inference

How about this network?



Bayesian Network Inference

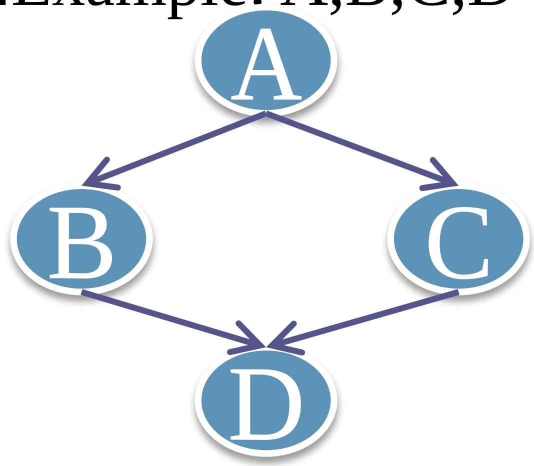
(3)Markov chain Monte Carlo (MCMC)

- Markov Chain: memoryless

  Monte Carlo algorithms: random sampling algorithms

- An approximate inference ⇨ An approximate estimation

□ Example: A,B,C,D are binary variables.



A,B,C,D are binary variables

Calculate the posterior: P(A=1|D=1)

Simulating the samples：

D_1={A=1,B=1,C=0,D=1}

D_2={A=0,B=0,C=0,D=1}

…

D_n={A=1,B=1,C=1,D=1}

- P(A=1|D=1)=frequency of A=1

Now, we have known

- What is BN?
- Why we use BN?
- How to compute the posterior that we interest in?

- Now there is a question : If we have a dataset, how to construct a Bayesian Network based on samples?

Bayesian Network Learning

- Structure is known:
- Structure is unknown:

Bayesian Network Learning

- Structure is known: Parameter Learning
- Structure is unknown: Structure Learning

Bayesian Network Learning

- Structure is known: Parameter Learning

(1)Maximum Likelihood Estimation(MLE)

(2)Bayesian Estimation

Bayesian Network Learning

- Structure is known: Parameter Learning
- Structure is unknown: Structure Learning

Step 1: Model selection (scoring function etc.)

Step 2: Model optimization

Thanks!