Bayesian Network

Jinghao Zhao

11/12/2021

Bayesian Network

- Representation
  - Model conditional dependency (causation)
  - Represent <u>joint probability</u> over a set of random variables
- Inference
  - Typical Tasks
    - Conditional Probability Query P(X1,X2,...|E1,E2,...)
    - Marginalize one or a set variables P(X1, X2,...)

Bayesian Network

Goal: Represent joint probability over a set of random variables

- Facilitate probability computation

Component:

(1) Graph Structure: a Directed Acyclic Graph (DAG)

- Nodes: random variables (events)
- Edges: $y \rightarrow x$ means y causes/influences x

(2) Local Probability Model

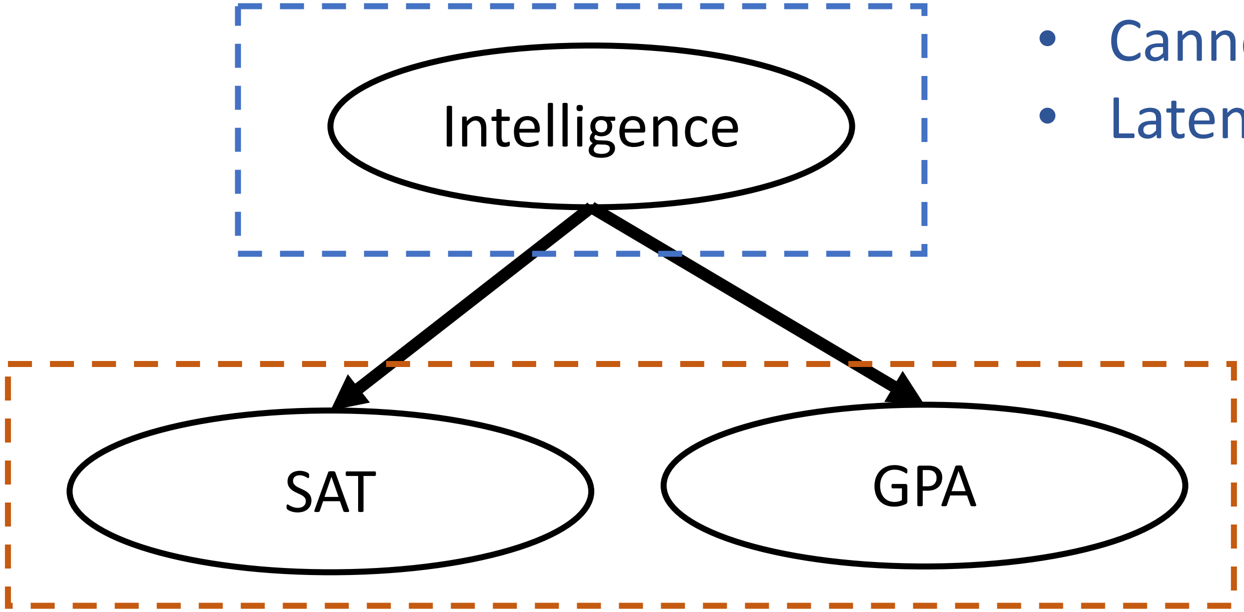
- Represent the dependence of each variable on its parents
- $y_1, y_2, \dots, y_k \rightarrow x$: conditional probability $p(x|y_1, y_2, \dots, y_k)$
- Root variables: marginal probability

Student Example – Graph Structure

- A company wants to hire an intelligent student.
  - But intelligence cannot be directly measured.
  - But the company may have access to the student's SAT and GPA score.

- Based on the observable evidence (SAT and GPA), company can try to infer whether this student is intelligent or not.

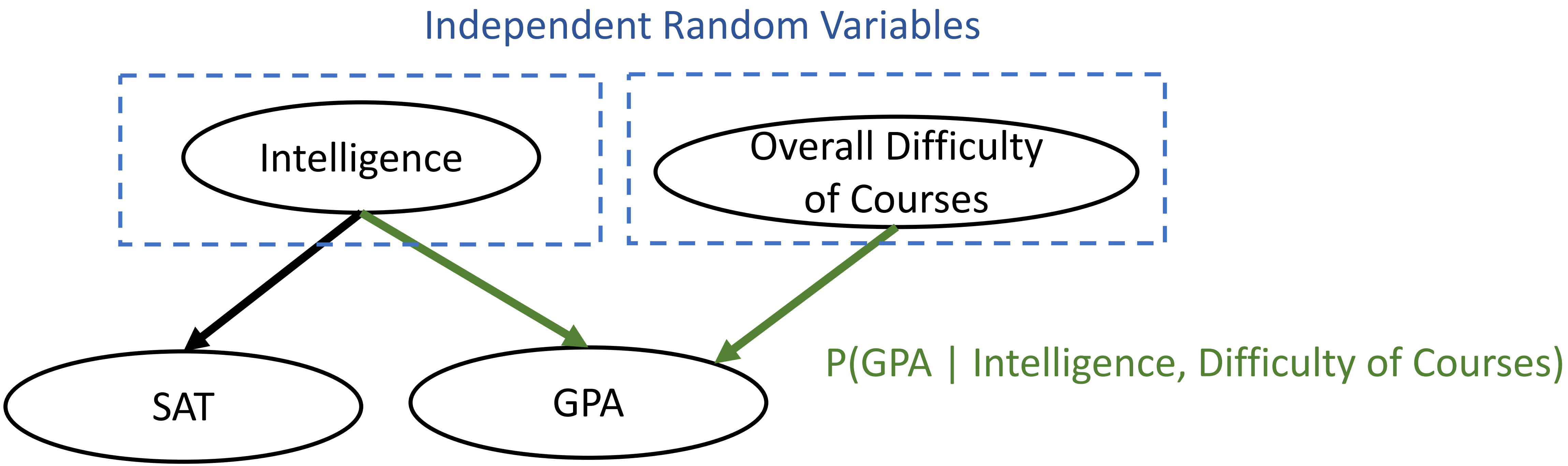
Student Example – Graph Structure



- Cannot be directly measured
- Latent factor of its children

Can be directly observed

Student Example – Graph Structure

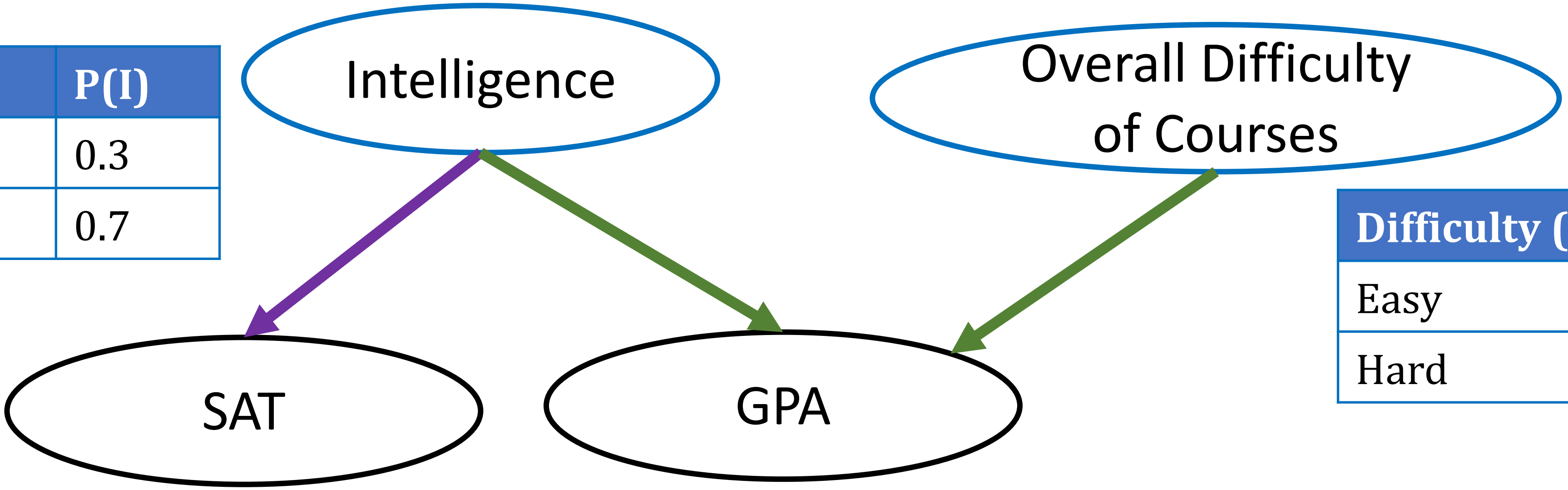


Student Example – A Full Bayesian Network

Component: (1) Graph Structure (DAG) (2) Local probability model (CPTs)

Conditional
Probability
Tabels

Intelligence (I)	P(I)
True	0.3
False	0.7



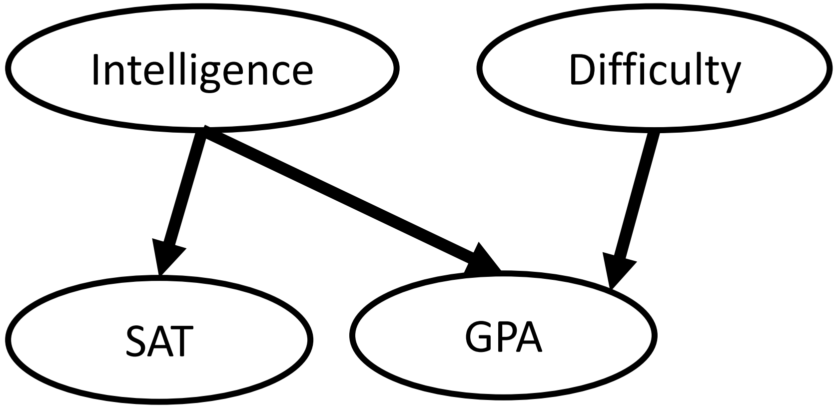
Difficulty (D)	P(D)
Easy	0.6
Hard	0.4

I	SAT	P(SAT|I)
True	High	0.7
True	Low	0.3
False	High	0.4
False	Low	0.6

I	D	GPA	P(GPA|I,D)
True	Easy	High	0.8
True	Easy	Low	0.2
True	Hard	High	0.6
…	…	…	…

Topological Semantics

- BN satisfies **local Markov property**:
  - A node is conditionally independent of its non-descendants given its parents.



Markovian Assumptions

Intelligence $\perp$ Difficulty

SAT $\perp$ GPA | Intelligence

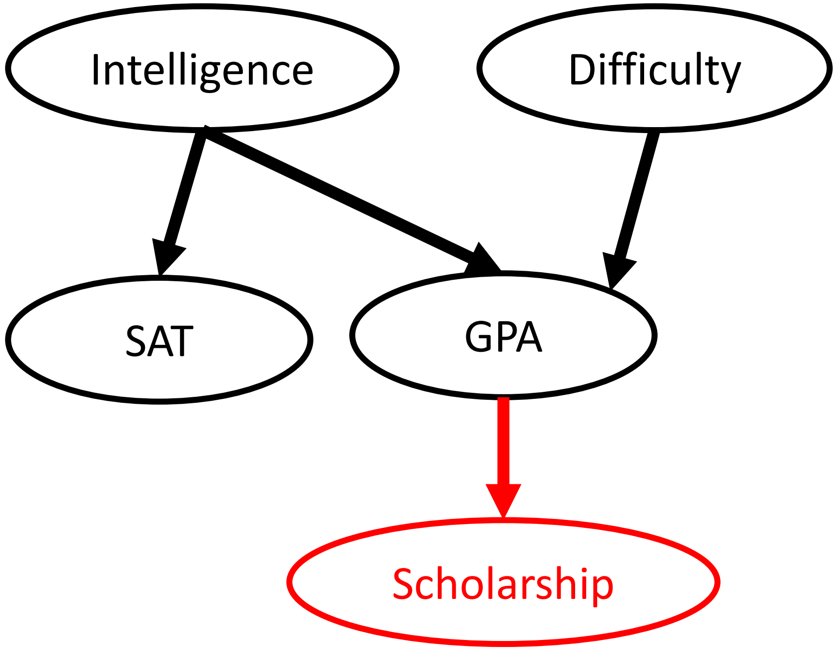
SAT $\perp$ Difficulty | Intelligence

GPA $\perp$ SAT | Intelligence, Difficulty

- A Bayesian network G encodes a set of (conditional) independence assumptions (Markovian assumptions)

Topological Semantics

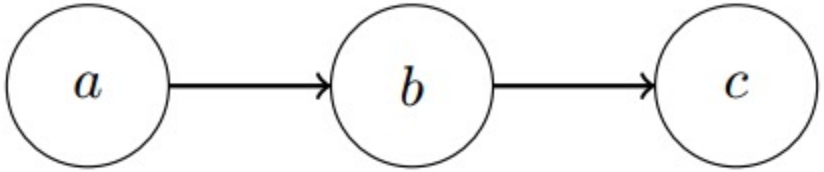
- BN satisfies **local Markov property:**
  - A node is conditionally independent of its non-descendants given its parents.



SAT $\perp$ Scholarship | Intelligence

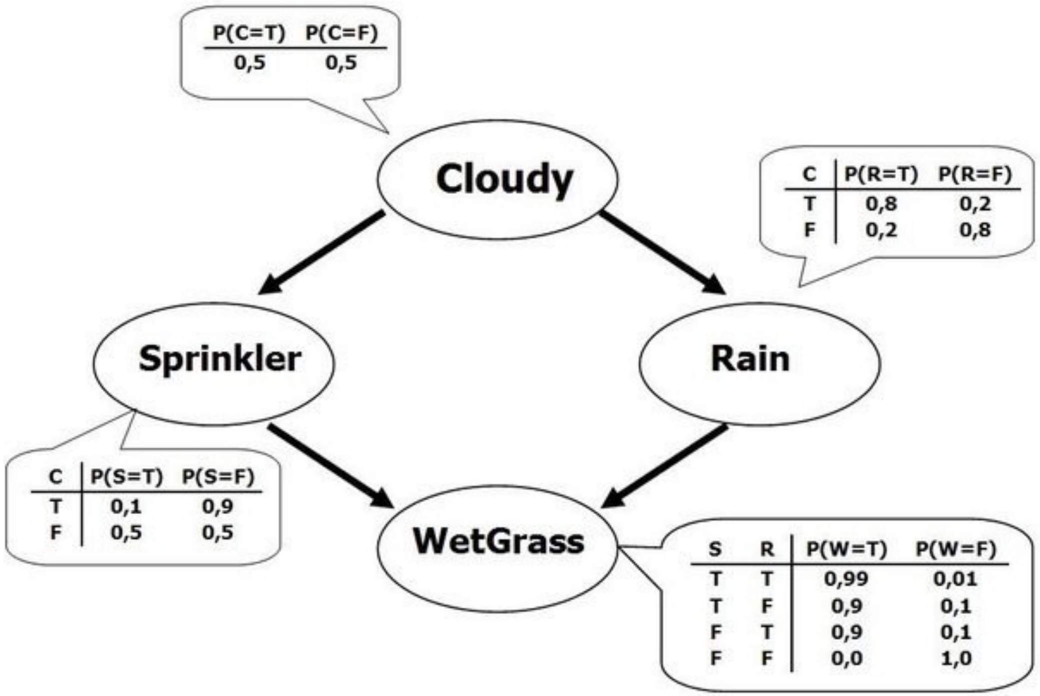
~~GPA $\perp$ Scholarship | Intelligence, Difficulty~~

Exercise – conditional independency



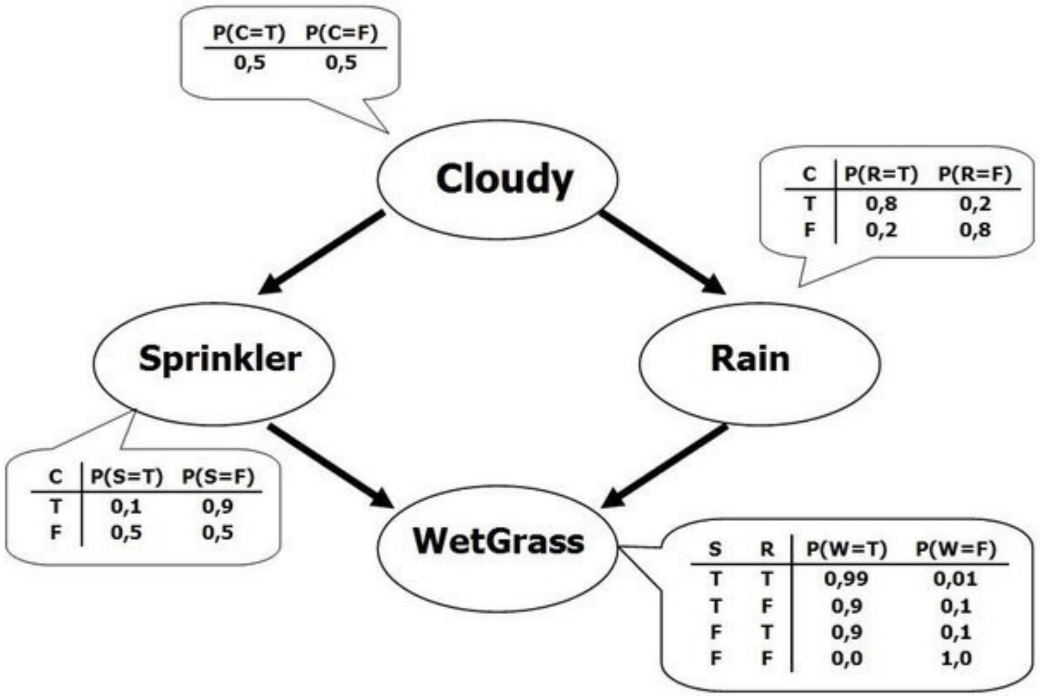
Give the topological semantics encoded in the BN.

Exercise – conditional independency



- Given Cloudy, what variables is Sprinkler conditional independent of?

Exercise – conditional independency



- Given Cloudy, what variables is Sprinkler conditional independent of?

Rain

Joint Probability – Chain Rule for BN

The joint probability modeled by BN:

$P(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} P(X_i | \text{parents of } X_i)$,

Why?

Joint Probability – Chain Rule for BN

Bayesian network models the following **joint probability**

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^{N} P(X_i | \text{parents of } X_i)$$

Why?

- Without loss of generality, assume $X_1, X_2, \dots, X_N$ is a topological ordering
- Chain rule

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^{N} P(X_i | X_1, X_2, \dots, X_{i-1})$$

- $P(X_i | X_1, X_2, \dots, X_n) = P(X_i | \text{parents of } X_i)$
  - Topological ordering => parents are in $X_1, X_2, \dots, X_{i-1}$
  - Local Markov property => given parents, independent of other variables in $X_1, X_2, \dots, X_{i-1}$

D-separation

- A graphical criterion used to identify independences (marginal or conditional) that hold in the BN graph
- Any complex example can be analyzed by considering relevant triples

Causal Chains

- “causal chain”

- Is it guaranteed that X is independent of Z given Y?

$$P(z|x,y) = \frac{P(x,y,z)}{P(x,y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

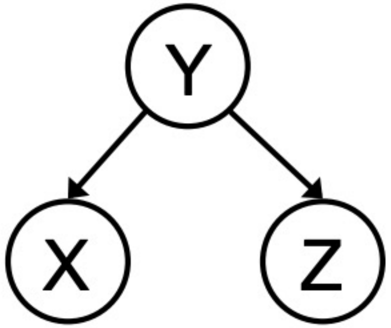
$$= P(z|y)$$

Yes!

- Evidence along the chain “blocks” the influence

Common Cause

- Two effects of the same cause
- Is it guaranteed that X and Z are independent?
  - No!
- Is it guaranteed that X and Z are independent given Y?



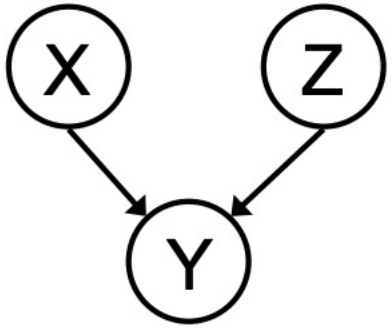
$$P(z|x,y) = \frac{P(x,y,z)}{P(x,y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y) \quad \textit{Yes!}$$

- Observing the cause blocks influence between effects.

D-separation Examples

- two causes of one effect (v-structures)
- Are X and Z independent?
  - Yes!
- Are X and Z independent given Y?
  - No!
- Observing an effect activates influence between possible causes.

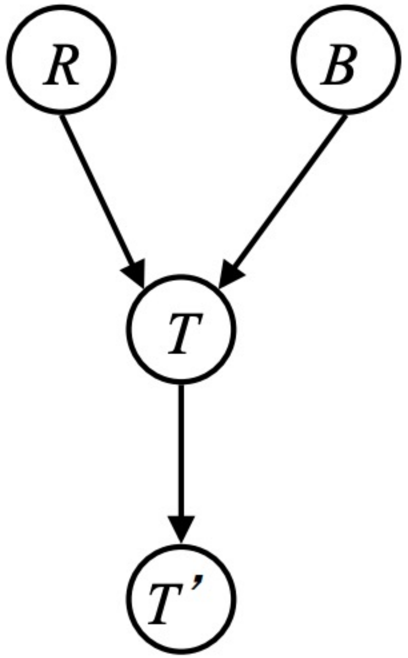


Example

$$R \perp\!\!\!\perp B$$

$$R \perp\!\!\!\perp B | T$$

$$R \perp\!\!\!\perp B | T'$$

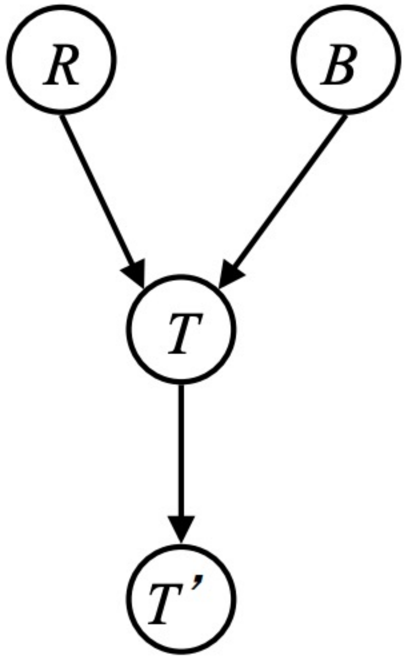


Example

$R \perp\!\!\!\perp B$ Yes

$R \perp\!\!\!\perp B | T$

$R \perp\!\!\!\perp B | T'$



Example

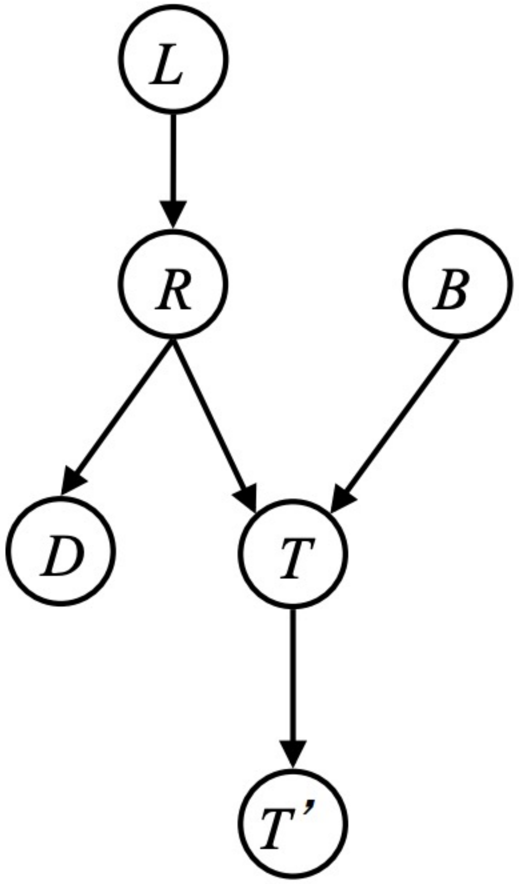
$$L \perp\!\!\!\perp T' | T$$

$$L \perp\!\!\!\perp B$$

$$L \perp\!\!\!\perp B | T$$

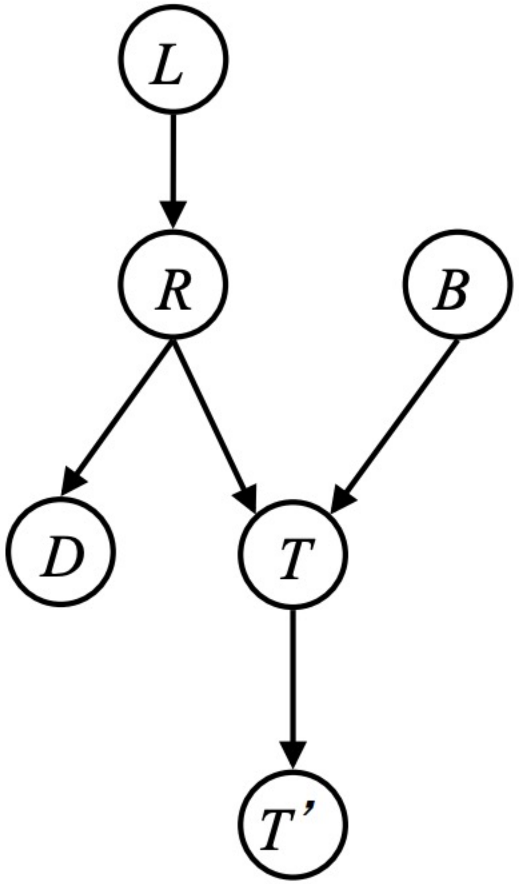
$$L \perp\!\!\!\perp B | T'$$

$$L \perp\!\!\!\perp B | T, R$$



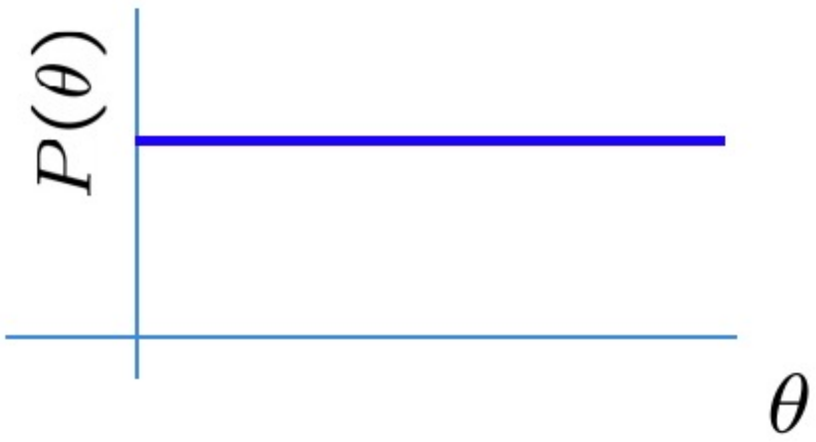
Example

$L \perp\!\!\!\perp T' \mid T$	Yes
$L \perp\!\!\!\perp B$	Yes
$L \perp\!\!\!\perp B \mid T$	
$L \perp\!\!\!\perp B \mid T'$	
$L \perp\!\!\!\perp B \mid T, R$	Yes



Prior distribution

- What prior? What distribution do we want for a prior?
  - Represents expert knowledge (philosophical approach)
  - Simple posterior form (engineer's approach)
- For example, uninformative priors is a uniform distribution:



Chain Rule & Bayes Rule

- Chain rule:

$$
\begin{aligned}
P(A,B,C,D,E) &= P(A|B,C,D,E) \cdot P(B,C,D,E) = P(A|B,C,D,E)\ P(B|C,D,E)\ P(C,D,E) \\
&= P(A|B,C,D,E)\ P(B|C,D,E)\ P(C,D,E) \\
&= \ldots \\
&= P(A|B,C,D,E)\ P(B|C,D,E)\ P(C|D,E)\ P(D|E)\ P(E)
\end{aligned}
$$

- Bayes rule: important for reverse conditioning

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Bayesian Learning

- Use Bayes rule:

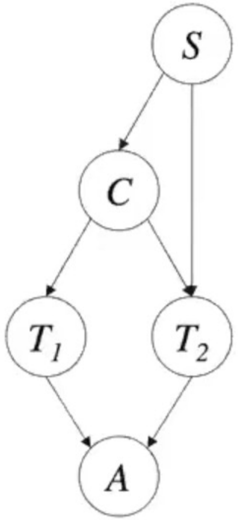
$$P(\theta \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \theta) P(\theta)}{P(\mathcal{D})}$$

- Or equivalently:

$$\underbrace{P(\theta \mid \mathcal{D})}_{\text{posterior}} \propto \underbrace{P(\mathcal{D} \mid \theta)}_{\text{likelihood}} \underbrace{P(\theta)}_{\text{prior}}$$

Most probable explanation (MPE)

- Most probable explanation (MPE), also known as max propagation, computes the most probable configuration of variables that do not have evidence.
- The MPE given the evidence A=yes:
  - C=no
  - S=female
  - T1=-ve
  - T2=-ve



S	θ_s
male	.55
female	.45

S	C	$\theta_{c\|s}$
male	yes	.05
male	no	.95
female	yes	.01
female	no	.99

C	T_1	$\theta_{t_1\|c}$
yes	+ve	.80
yes	−ve	.20
no	+ve	.20
no	−ve	.80

S	C	T_2	$\theta_{t_2\|c,s}$
male	yes	+ve	.80
male	yes	−ve	.20
male	no	+ve	.20
male	no	−ve	.80
female	yes	+ve	.95
female	yes	−ve	.05
female	no	+ve	.05
female	no	−ve	.95

T_1	T_2	A	$\theta_{a\|t_1,t_2}$
+ve	+ve	yes	1
+ve	+ve	no	0
+ve	−ve	yes	0
+ve	−ve	no	1
−ve	+ve	yes	0
−ve	+ve	no	1
−ve	−ve	yes	1
−ve	−ve	no	0

Maximum a posteriori (MAP) estimation

- Similar with MPE, choose value that is most probable given observed data and prior belief

$$\begin{aligned}\widehat{\theta}_{MAP} &= \arg\max_{\theta} P(\theta|D) \\ &= \arg\max_{\theta} P(D|\theta)P(\theta)\end{aligned}$$

Complexity of Inference

- n: number of variables
- d: number of values
- w: treewidth

- The complexity for marginal is $O(nd^w)$