Künstliche Intelligenz

Bayes'sche Netze

Dr.-Ing. Stefan Lüdtke

Universität Leipzig

Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI)

Motivation

- Die Full Joint erlaubt es uns, jede Query zu beantworten
- Aber: Größe der Full Joint wächst exponentiell mit Anzahl der Variablen
- Können (bedingte) Unabhängigkeit ausnutzen, um Verteilung schlauer zu repräsentieren
- Bayes'sches Netz: Repräsentation einer Wahrscheinlichkeitsverteilung, in der die (Un)abhängigkeiten zwischen Variablen explizit gemacht werden

Bayesian networks

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

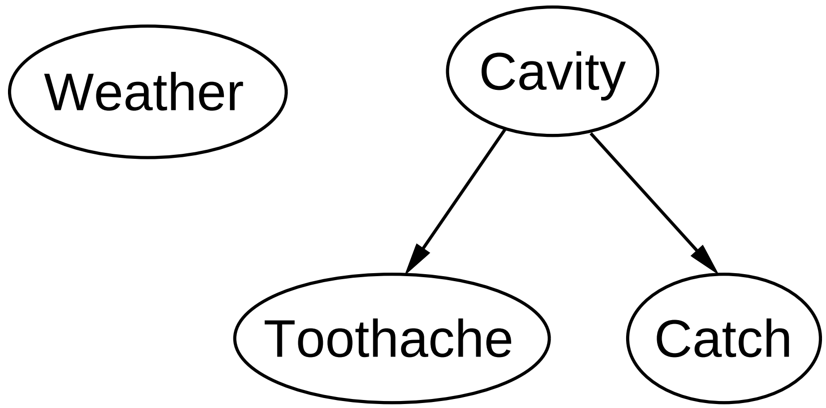
Syntax:

a set of nodes, one per variable
a directed, acyclic graph (link $\approx$ "directly influences")
a conditional distribution for each node given its parents:
$\mathbf{P}(X_i|Parents(X_i))$

In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

Example

Topology of network encodes conditional independence assertions:



$Weather$ is independent of the other variables

$Toothache$ and $Catch$ are conditionally independent given $Cavity$

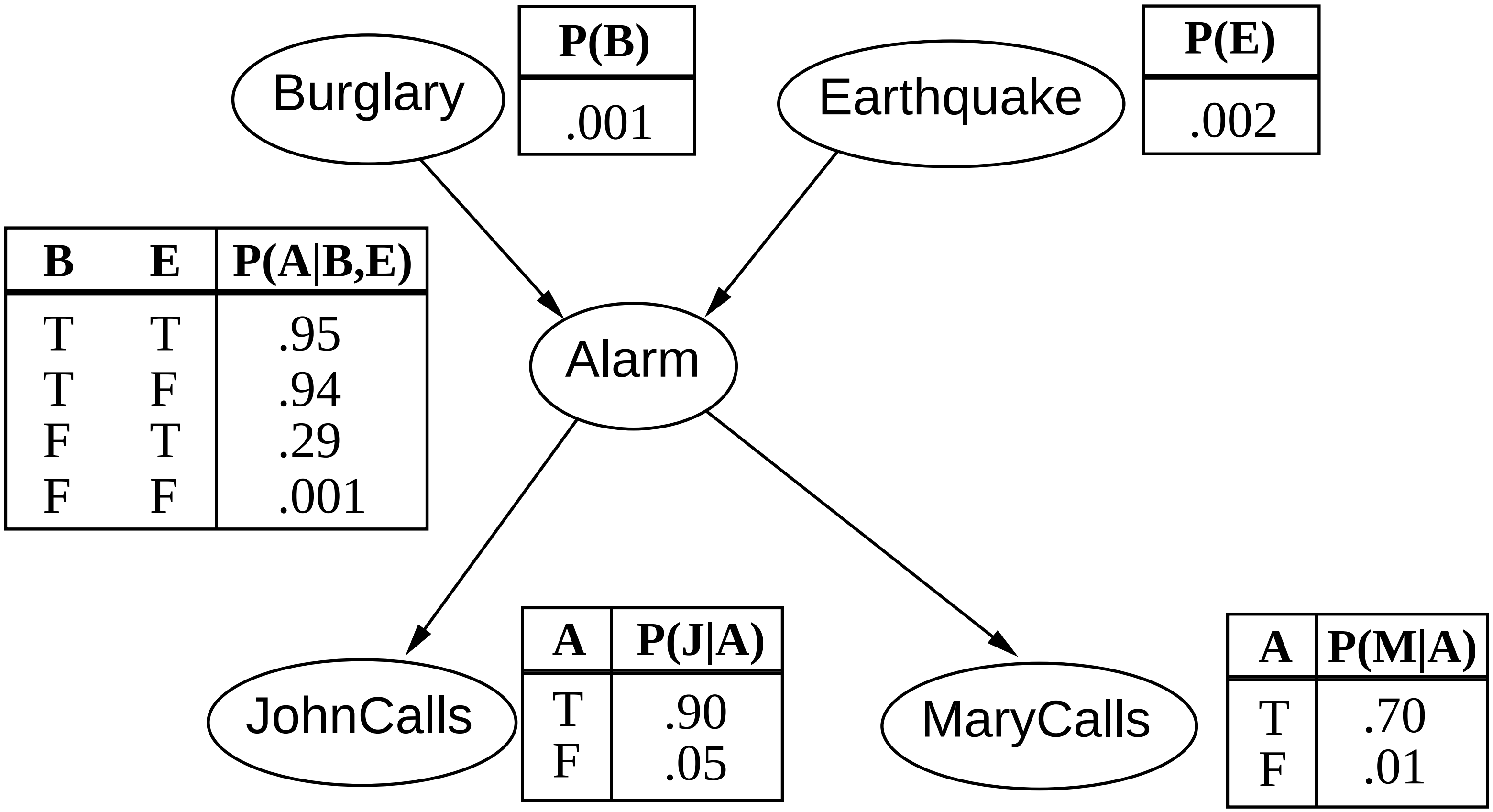
Example

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
Network topology reflects "causal" knowledge:

- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

Example contd.



P(B)
.001

P(E)
.002

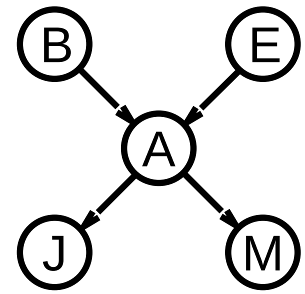
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactness

A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values



Each row requires one number p for $X_i = true$
(the number for $X_i = false$ is just $1 - p$)

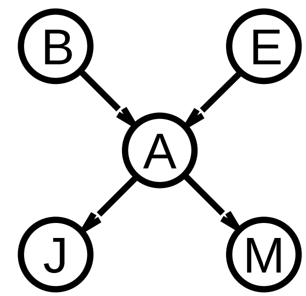
If each variable has no more than k parents,
the complete network requires $O(n \cdot 2^k)$ numbers

I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution

For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)

Global semantics

Global semantics defines the full joint distribution
as the product of the local conditional distributions:



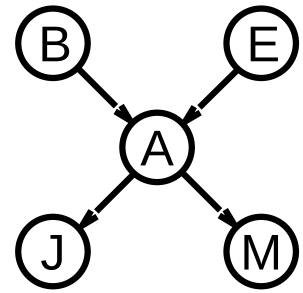
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

e.g., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$=$$

Global semantics

"Global" semantics defines the full joint distribution as the product of the local conditional distributions:



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

e.g., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

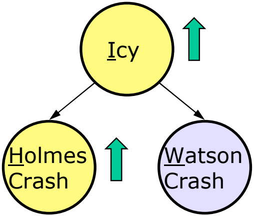
$$\begin{aligned} &= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) \\ &= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\ &\approx 0.00063 \end{aligned}$$

Icy Roads: An Example

Universität Rostock Traditio et Innovatio

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

"Causal" Component

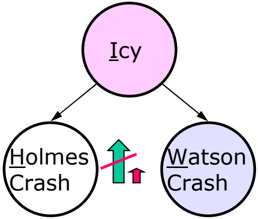


H and W are dependent,

Icy Roads: An Example

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

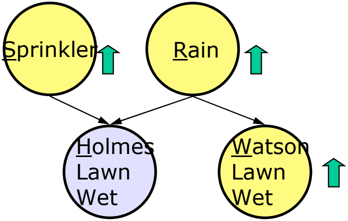
"Causal" Component



H and W are dependent, but conditionally independent given I

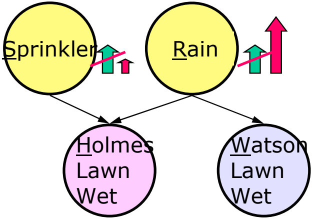
Holmes and Watson in LA: An Example

Holmes and Watson have moved to LA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.



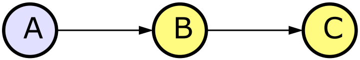
Holmes and Watson in LA: An Example

Holmes and Watson have moved to LA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.

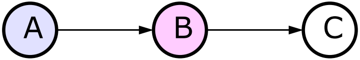


Given W, P(R) goes up and P(S) goes down – "explaining away"

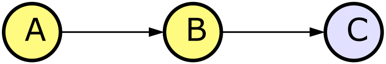
Forward Serial Connection



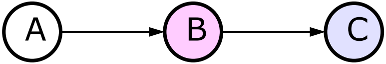
- transmit evidence from A to C through B, unless B is instantiated (its truth value is known)
  - A = battery dead
  - B = car won't start
  - C = car won't move
- knowing about A will tell you something about C
- but if we know B, then knowing about A will not tell us anything new about C



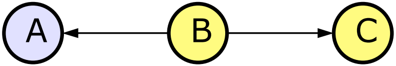
Backward Serial Connection



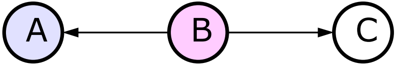
- transmit evidence from C to A through B, unless B is instantiated (its truth value is known)
  - A = battery dead
  - B = car won't start
  - C = car won't move
- knowing about C will tell you something about A
- but if we know B, then knowing about C will not tell us anything new about A



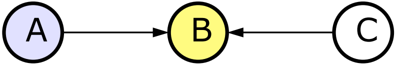
Diverging Connection



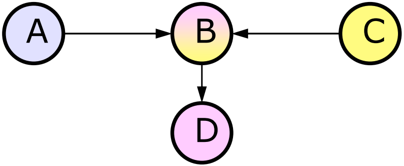
- transmit evidence through B, unless B is instantiated
  - A = Watson crash
  - B = icy
  - C = Holmes crash
- knowing about A will tell you something about C
- knowing about C will tell you something about A
- but if we know B, then knowing about A will not tell us anything new about C, or vice versa



Converging Connection

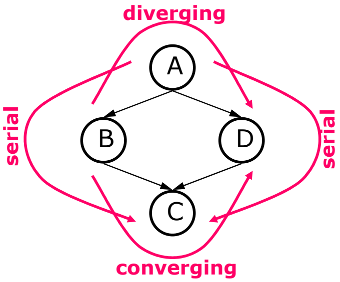


- transmit evidence from A to C only if B or a descendent of B is initiated
  - A = bacterial infection
  - B = sore throat
  - C = viral infection
- without knowing B, finding A does not tell us anything about C
- if we see evidence for B, then A and C become dependent (potential for “explaining away”). E.g. if we find bacteria in a patient with a sore throat, then viral infection is less likely.



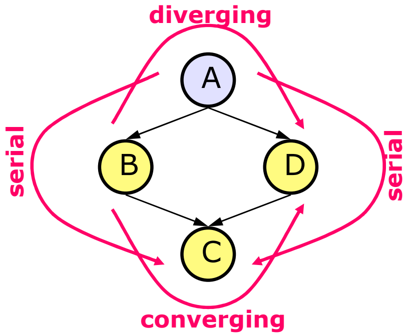
D-separation

- two variables are **d-separated** iff for every path between them, there is an intermediate node V such that either
  - the connection is serial or diverging and V is known
  - the connection is converging and neither or any descendent is instantiated
- two variables are **d-connected** if they are not d-separated



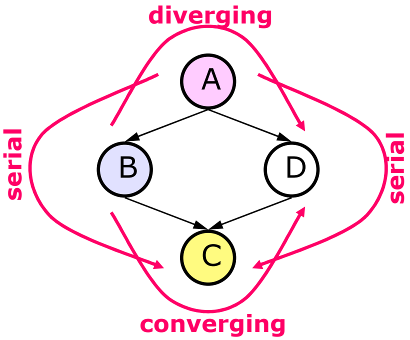
- A-B-C: serial, blocked when B is known, connected otherwise
- A-D-C: serial, blocked when D is known, connected otherwise
- B-A-D: diverging, blocked when A is known, connected otherwise
- B-C-D: converging, blocked when C has no evidence, connected otherwise

D-separation: Example



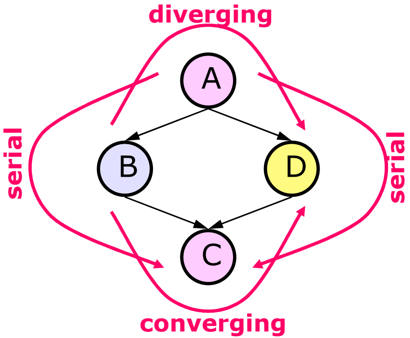
- no instantiation
  - A, C are d-connected (A-B-C connected, A-D-C connected)
  - B, D are d-connected (B-A-D connected, B-C-D blocked)

D-separation: Example



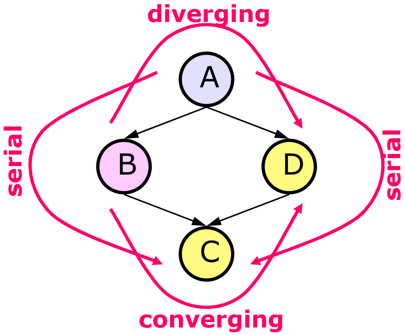
- A is instantiated
  - B, D are d-separated (B-A-D blocked, B-C-D blocked)

D-separation: Example



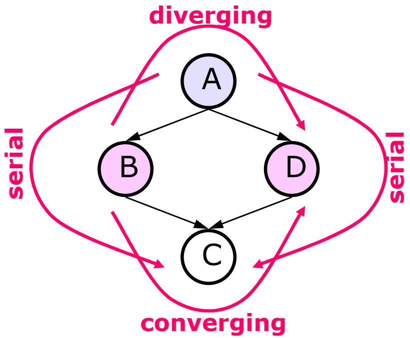
- A and C are instantiated
  - B, D are d-connected (B-A-D blocked, B-C-D connected)

D-separation: Example



- B is instantiated
  - A, C are d-connected (A-B-C blocked, A-D-C connected)

D-separation: Example



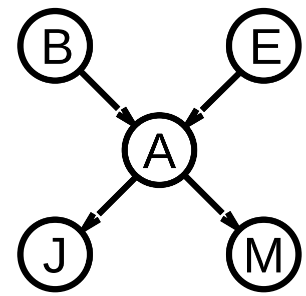
- B and D are instantiated
  - A, C are d-separated (A-B-C blocked, A-D-C blocked)

Inference by enumeration

Slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation

Simple query on the burglary network:

$\mathbf{P}(B|j,m)$
$= \mathbf{P}(B,j,m)/P(j,m)$
$= \alpha\mathbf{P}(B,j,m)$
$= \alpha \sum_e \sum_a \mathbf{P}(B,e,a,j,m)$

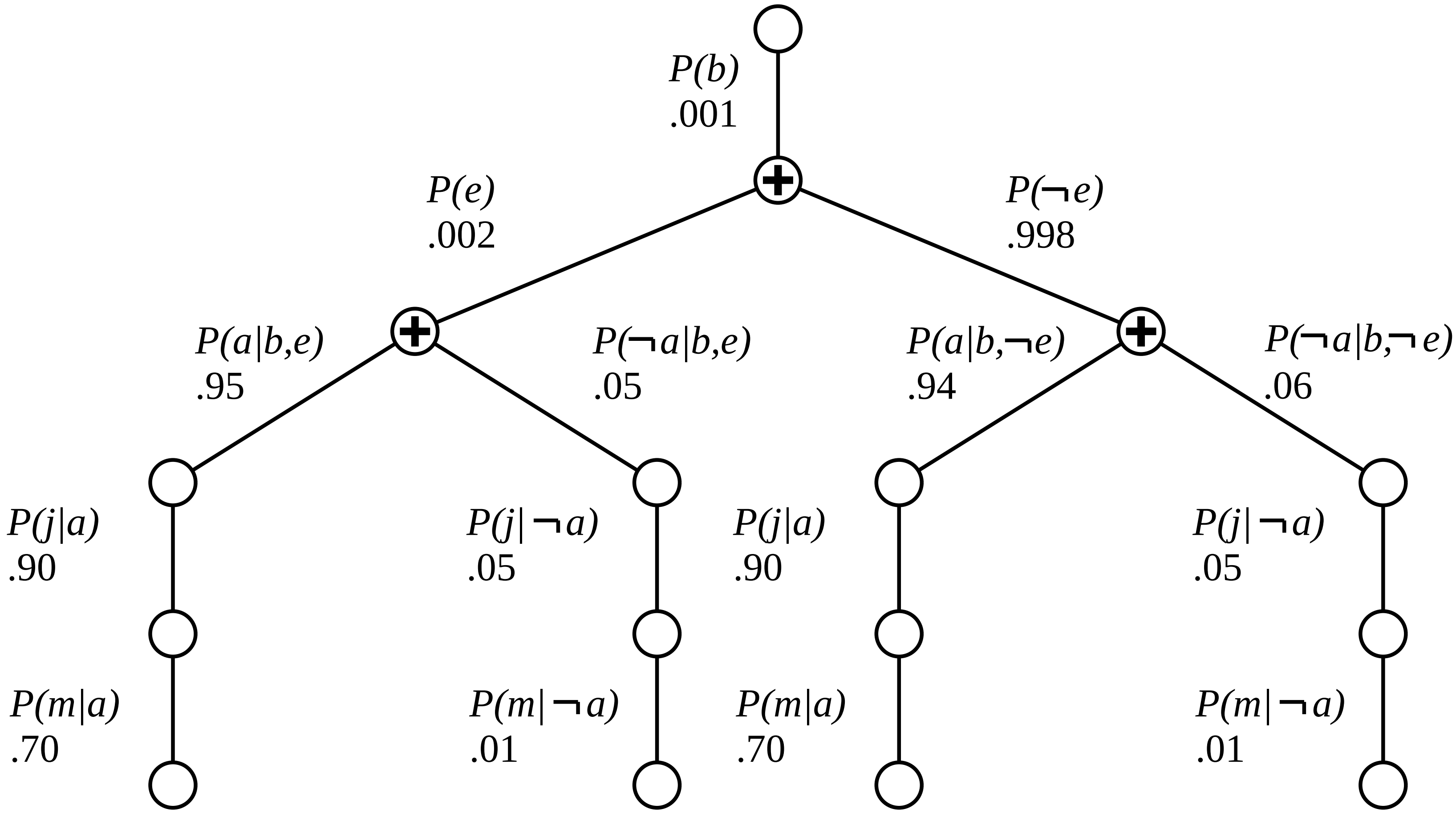


Rewrite full joint entries using product of CPT entries:

$\mathbf{P}(B|j,m)$
$= \alpha \sum_e \sum_a \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)$
$= \alpha\mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a|B,e)P(j|a)P(m|a)$

Recursive depth-first enumeration: $O(n)$ space, $O(d^n)$ time

Evaluation tree



Enumeration is inefficient: repeated computation
e.g., computes $P(j|a)P(m|a)$ for each value of e

Zusammenfassung

- Bayes'sche Netze sind natürliche Repräsentation für Verteilungen mit bedingten Unabhängigkeiten
- Topologie + CPTs= Kompakte Repräsentation der Full Joint
- Einfach zu spezifizieren, auch für Nichtexperten
- Exakte Inferenz durch Aufzählen nur in einfachen Fällen möglich