Tufts UNIVERSITY

BAYESIAN NETWORKS 2

ARTIFICIAL INTELLIGENCE | COMP 131

TODAY ON AI

- Causal Networks
- Exact inference with Bayes networks
- Approximate inference with Bayes networks
- Questions?

- **Conditional probability**: $P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$

- **Product rule**: $P(X,Y) = P(X \mid Y)\,P(Y) = P(Y \mid X)\,P(X)$

- **Chain rule**: $P(X_1, \dots, X_n) = \prod_i P(X_i \mid X_1, \dots, X_{i-1})$

- **X and Y are independent**: iff $P(X,Y) = P(X)\,P(Y)$

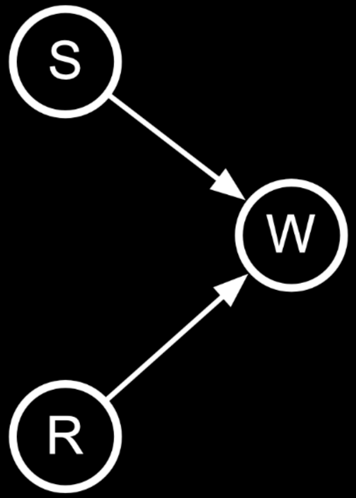
- **Bayes rule**: $P(Cause \mid Observation) = \frac{P(Observation \mid Cause) P(Cause)}{P(Observation)}$

- **Conditional independence**: $P(X,Y \mid Z) = P(X \mid Z)\,P(Y \mid Z)$

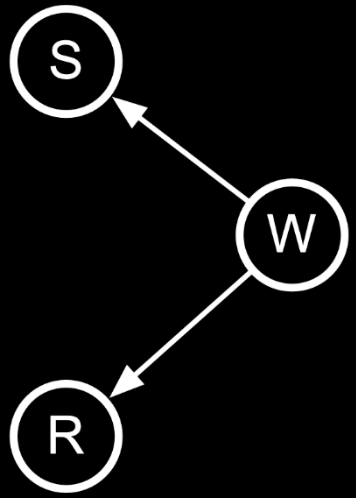
SECTION 01

Causal Bayesian networks

The word **causal** is usually contentious in Bayesian networks when the model of the data contains no explicit temporal information:



$$P(R,S,W) = P(S)\,P(R)\,P(W|R,S)$$



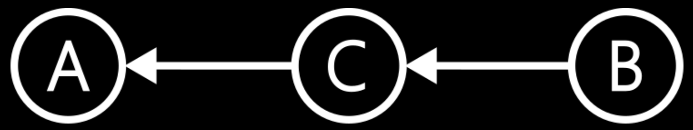
$$P(R,S,W) = P(W)\,P(R|W)\,P(S|W)$$

Bayes Networks **do not have to be** causal. The arcs simple **reflect some correlation**.

Causal networks are a special class of Bayes networks that **forbids** all but **causally compatible** relationships or orderings.

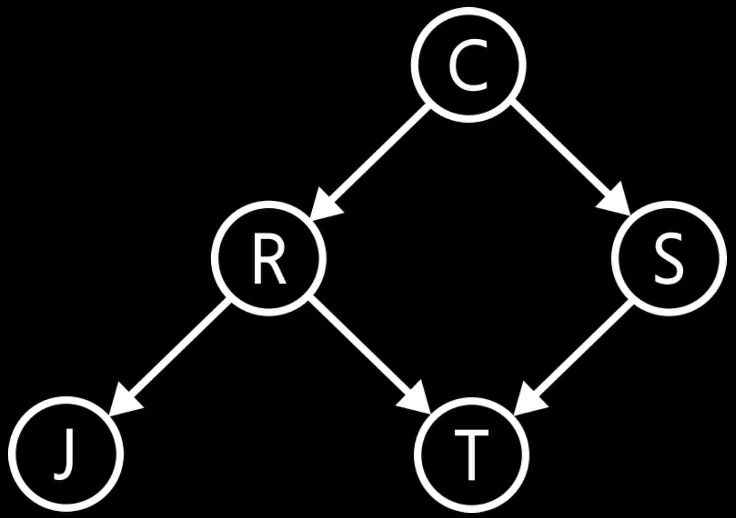
- When Bayes Networks reflect **a true causal** relationship:
  - They are more intuitive
  - They are simpler to represent from expert knowledge
  - They are topologically simpler

- In causal Bayesian networks the question is usually what variables comes **first** or what **variables causes other variables.**

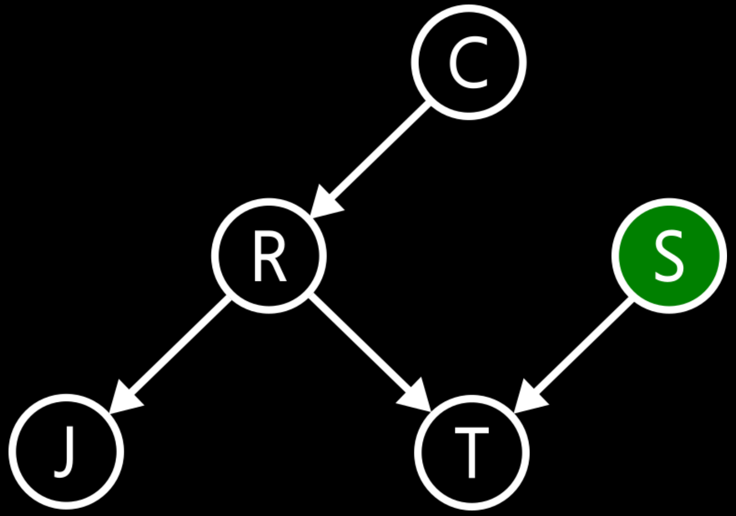


Causal networks are important also because they allow us to **predict** how **interventions** will affect the model.

We can **observe** the result of those interventions and determine the correct course of actions if we are planning something.



- The **do-operator** is an operation that **imposes a specific outcome** on a variable and **remove all conditional** dependencies to that variable:



$$P(C,R,S,J,T) = P(C)\,P(R \mid C)\,P(S \mid C)\,P(J \mid R)\,P(T \mid R,S)$$

$$P(C,R,J,T) = P(C)\,P(R \mid C)\,P(s)\,P(J \mid R)\,P(T \mid R,s)$$

$$P(C,R,J,T) = P(C)\,P(R \mid C)\,P(J \mid R)\,P(T \mid R,s)$$

SECTION 02

Exact inference with Bayes network

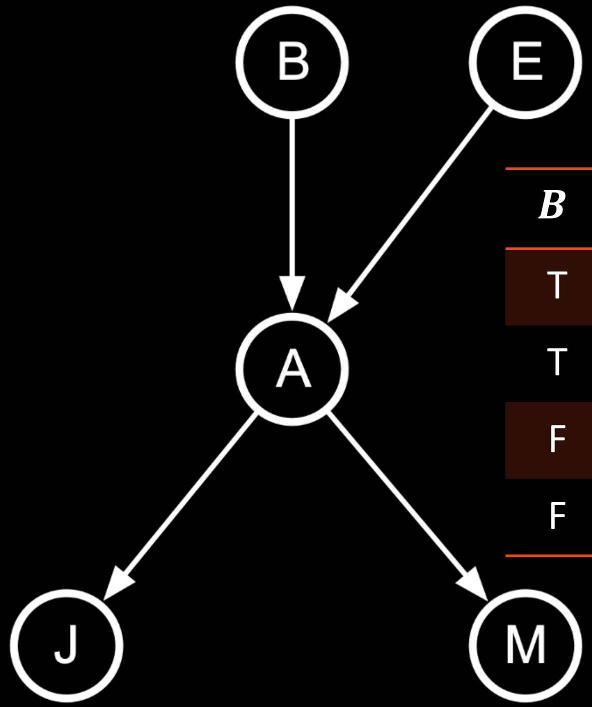
We have seen that answering a query with Bayesian networks is equivalent to **computing sums of products** of conditional probabilities from the networks. This is called **exact inference**.

A **variable elimination algorithm** can significatively reduce the overall number of calculations **caching** intermediate results that can used again later.

You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$P(b)$	$P(\neg b)$
0.001	0.999

$P(e)$	$P(\neg e)$
0.002	0.998



B	E	$P(a\|B,E)$	$P(\neg a\|B,E)$
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

$$P(B,E,A,J,M) = P(B)\,P(E)\,P(A|B,E)\,P(J|A)\,P(M|A)$$

J	John calls	**{T, F}**
M	Mary calls	**{T, F}**
A	Alarm	**{T, F}**
B	Burglary	**{T, F}**
E	Earthquake	**{T, F}**

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

Alarm network

AN EXAMPLE

$P(B|j,m) = ?$ $\quad P(B|j,m) = \dfrac{P(B,E,A,J,M)}{P(J,M)}$ $\quad P(B|j,m) = \dfrac{P(B,E,A,j,m)}{P(J,M)}$ $\quad P(B|j,m) = \alpha\, P(B,E,A,j,m)$

$$P(B,E,A,J,M) = P(B)\,P(E)\,P(A|B,E)\,P(J|A)\,P(M|A)$$
$$P(B,E,A,j,m) = P(B)\,P(E)\,P(A|B,E)\,P(j|A)\,P(m|A)$$

Order of elimination: **A**, **E**, **B**

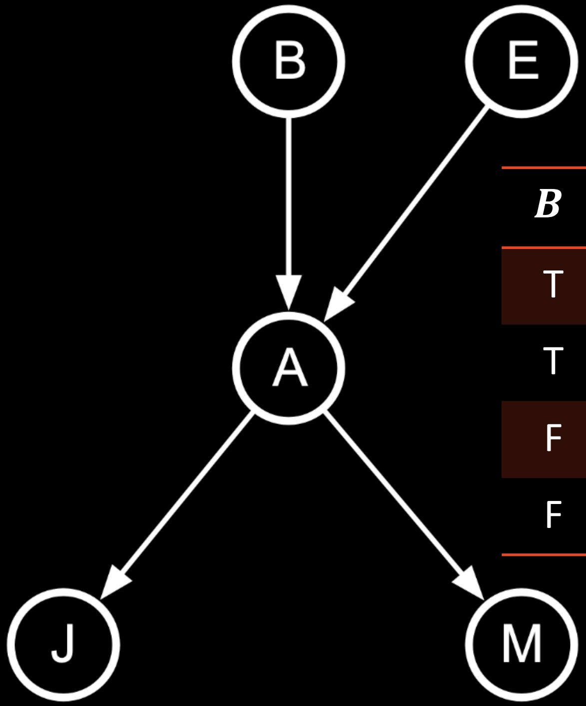
$$P(B|j,m) = \alpha \sum_{A,E,B} P(B)\,P(E)\,P(A|B,E)\,P(j|A)\,P(m|A) \qquad P(B|j,m) = \alpha\, P(B) \sum_{E} P(E) \sum_{A} P(A|B,E)\,P(j|A)\,P(m|A)$$

$$f_5(A) = P(m|A) \quad f_4(A) = P(j|A) \quad f_3(A,B,E) = P(A|B,E) \quad P(B|j,m) = \alpha\, P(B) \sum_{E} P(E) \underbrace{\sum_{A} f_3(A,B,E) \star f_4(A) \star f_5(A)}_{f_2(B,E)}$$

$$P(B|j,m) = \alpha\, P(B) \underbrace{\sum_{E} P(E) \star f_3(B,E)}_{f_1(B)} \qquad P(B|j,m) = \alpha\, P(B) \star f_1(B)$$

$P(b)$	$P(\neg b)$
0.001	0.999

$P(e)$	$P(\neg e)$
0.002	0.998



$$f_5(A) = P(m|A) \qquad f_5(A) =$$

A	$f_5(A)$
T	0.70
F	0.01

$$f_4(A) = P(j|A) \qquad f_4(A) =$$

A	$f_4(A)$
T	0.90
F	0.05

B	E	$P(a\|B,E)$	$P(\neg a\|B,E)$
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

Alarm network and earthquakes

AN EXAMPLE

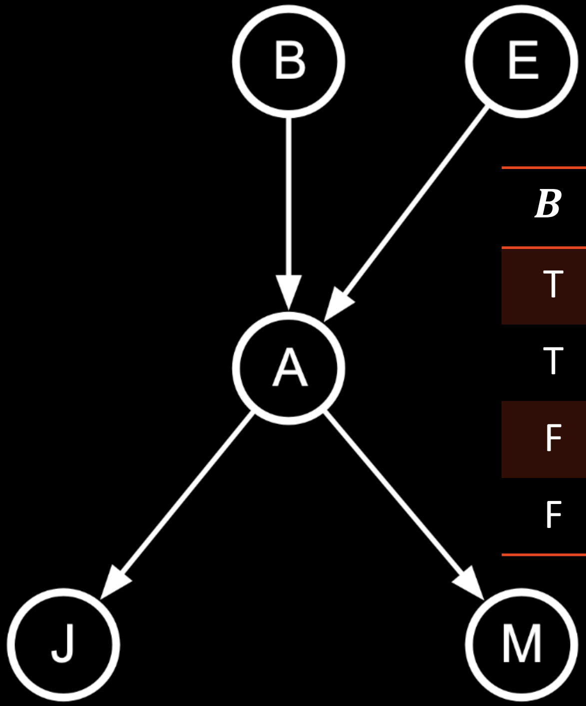
$$f_2(B,E) = \sum_A f_3(A,B,E) \star f_4(A) \star f_5\,(A)$$

$f_3(A,B,E) =$

A	B	E	$f_3(A,B,E)$
T	T	T	0.95
T	T	F	0.94
T	F	T	0.29
T	F	F	0.001
F	T	T	0.05
F	T	F	0.06
F	F	T	0.71
F	F	F	0.999

$P(b)$	$P(\neg b)$
0.001	0.999

$P(e)$	$P(\neg e)$
0.002	0.998



B	E	$P(a\|B,E)$	$P(\neg a\|B,E)$
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

Alarm network and earthquakes

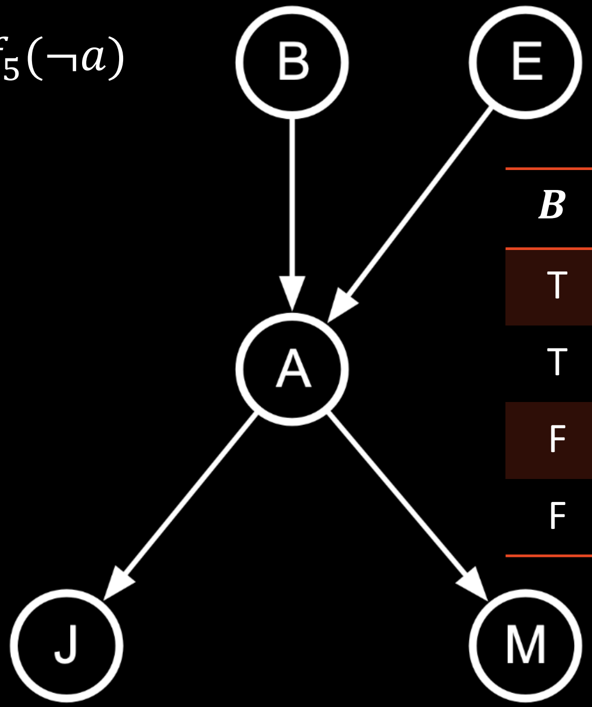
AN EXAMPLE

$$f_2(B,E) = \sum_A f_3(A,B,E) \star f_4(A) \star f_5\,(A)$$

$P(b)$	$P(\neg b)$
0.001	0.999

$P(e)$	$P(\neg e)$
0.002	0.998

$$f_2(B,E) = f_3(a,B,E) \star f_4(a) \star f_5(a) + f_3(\neg a,B,E) \star f_4(\neg a) \star f_5(\neg a)$$



$f_2(B,E) =$

B	E	$f_2(B,E)$
T	T	0.5985
T	F	0.5922
F	T	0.1831
F	F	0.0011

B	E	$P(a\|B,E)$	$P(\neg a\|B,E)$
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

Alarm network and earthquakes

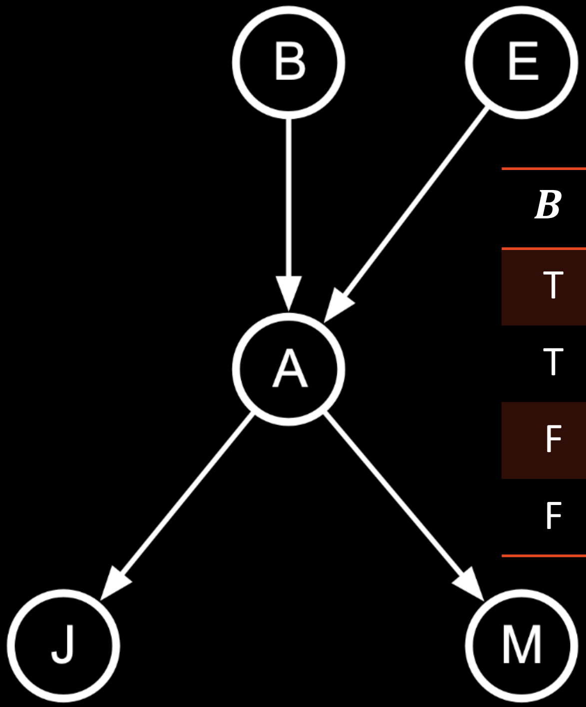
AN EXAMPLE

$$f_E(E) = P(E) \qquad f_E(E) =$$

E	$f_E(E)$
T	0.002
F	0.998

$P(b)$	$P(\neg b)$
0.001	0.999

$P(e)$	$P(\neg e)$
0.002	0.998



B	E	$P(a\|B,E)$	$P(\neg a\|B,E)$
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

$$f_1(B) = \sum_E f_E(E) \star f_2(B,E)$$

$$f_1(E) =$$

B	$f_1(B)$
T	0.5922
F	0.0015

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

Alarm network and earthquakes

AN EXAMPLE

$$f_B(B) = P(B)$$

$$f_B(B) =$$

B	$f_B(B)$
T	0.001
F	0.999

$$P(B|j,m) = \alpha\, f_B(B) \star f_1(B)$$

$$P(B|j,m) = \alpha$$

B	$\alpha P(B\|j,m)$
T	0.0006
F	0.0016

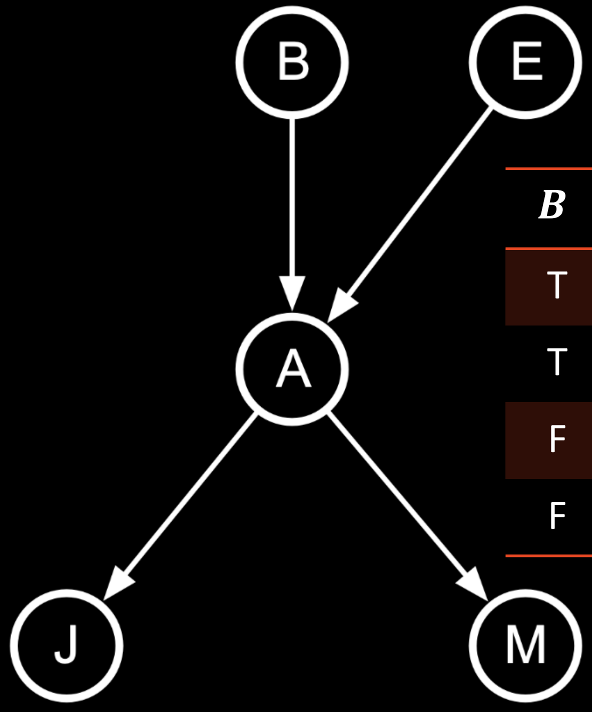
$$1 = \alpha \times P(b|j,m) + \alpha \times P(\neg b|j,m)$$

$$1 = \alpha[P(b|j,m) + P(\neg b|j,m)]$$

$$\alpha = \frac{1}{P(b|j,m)+P(\neg b|j,m)}$$

$P(b)$	$P(\neg b)$
0.001	0.999

$P(e)$	$P(\neg e)$
0.002	0.998



B	E	$P(a\|B,E)$	$P(\neg a\|B,E)$
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

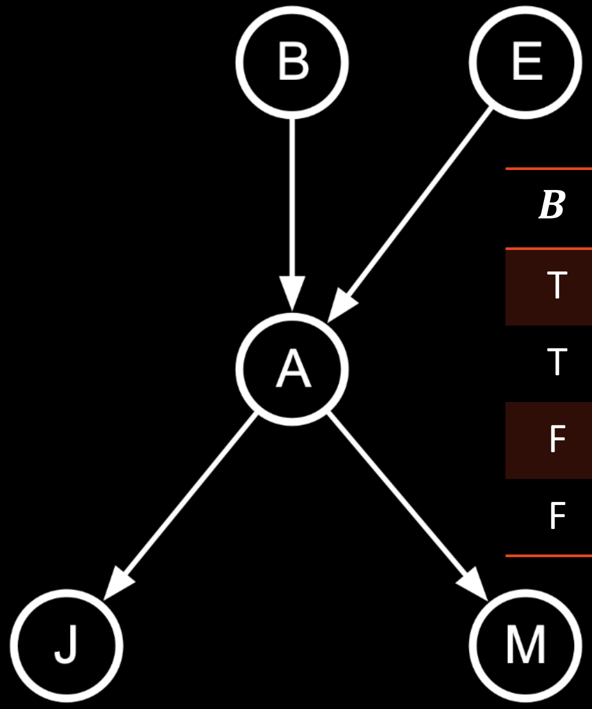
A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

$P(B|j,m) =$

B	$P(B|j,m)$
T	0.2727
F	0.7272

$P(b)$	$P(\neg b)$
0.001	0.999

$P(e)$	$P(\neg e)$
0.002	0.998



B	E	$P(a|B,E)$	$P(\neg a|B,E)$
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	$P(j|A)$	$P(\neg j|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m|A)$	$P(\neg m|A)$
T	0.70	0.30
F	0.01	0.99

Alarm network and earthquakes

AN EXAMPLE

- The complexity of **exact inference** in Bayes networks depends on the topology of the network:

  - For **singly connected networks** or **polytrees** the time and space complexity is **linear** in the size of the network.

  - For **multiply connected** networks the **variable elimination algorithm too** can exponential complexity.

In general inference in Bayesian network is **NP-hard**.

SECTION 03

Approximate inference with Bayes network

Approximate inference uses a family of randomized sampling algorithms called **Monte Carlo algorithms** that approximate answers given several samples.

They **approximate the joint distribution** of the network drawing several samples from the network. The **accuracy** of these algorithms depends on the samples generated.

In the **Direct Sampling Algorithm** **every variable** for which we draw the sample is **already conditioned** to its parents.

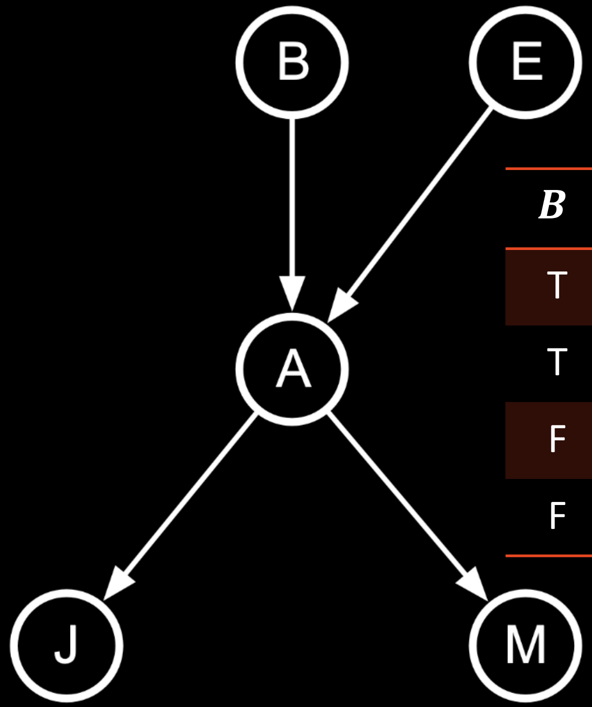
- Start from the variables that have **no evidence** associated with it
- Generate a sample $s_i = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$
- Equal samples are counted
- The **final distribution of the samples** will **approximate** the joint distribution of the network

```
function Prior-Sample(BE) return P(X_1, X_2, … X_n)
x = an event with n elements
for i = 1 to n
  x[i] = random sample from P(X_i | parents(X_i))
return x
```

You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$P(b)$	$P(\neg b)$
0.001	0.999

$P(e)$	$P(\neg e)$
0.002	0.998



B	E	$P(a\|B,E)$	$P(\neg a\|B,E)$
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

J	John calls	{T, F}
M	Mary calls	{T, F}
A	Alarm	{T, F}
B	Burglary	{T, F}
E	Earthquake	{T, F}

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

Alarm network

AN EXAMPLE

1. Sample from $P(B) = < 0.001, 0.999 >$ ➡ TRUE
2. Sample from $P(E) = < 0.002, 0.998 >$ ➡ FALSE
3. Sample from $P(A \mid b, \neg e) = < 0.94, 0.06 >$ ➡ FALSE
4. Sample from $P(J \mid \neg a) = < 0.05, 0.95 >$ ➡ TRUE
5. Sample from $P(M \mid \neg a) = < 0.01, 0.99 >$ ➡ TRUE

B E A J M

The atomic event [**true**, **false**, **false**, **true**, **true**] is stored and its frequency is calculated.

In the **Direct Sampling Algorithm** **every variable** for which we draw the sample is **already conditioned** to its parents.

- Start from the variables that have **no evidence** associated with it
- Generate a sample $s_i = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$
- Equal samples are counted
- The **final distribution of the samples** will **approximate** the joint distribution of the network

```
function Prior-Sample(BE) return P(X_1, X_2, … X_n)
x = an event with n elements
for i = 1 to n
  x[i] = random sample from P(X_i | parents(X_i))
return x
```

Direct sampling method

READINGS

Chapter 14

QUESTIONS ?

Tufts UNIVERSITY

ARTIFICIAL INTELLIGENCE

COMP 131

FABRIZIO SANTINI

VERSION 4.1