Fall 2022 | Lecture 13

Bayesian Networks

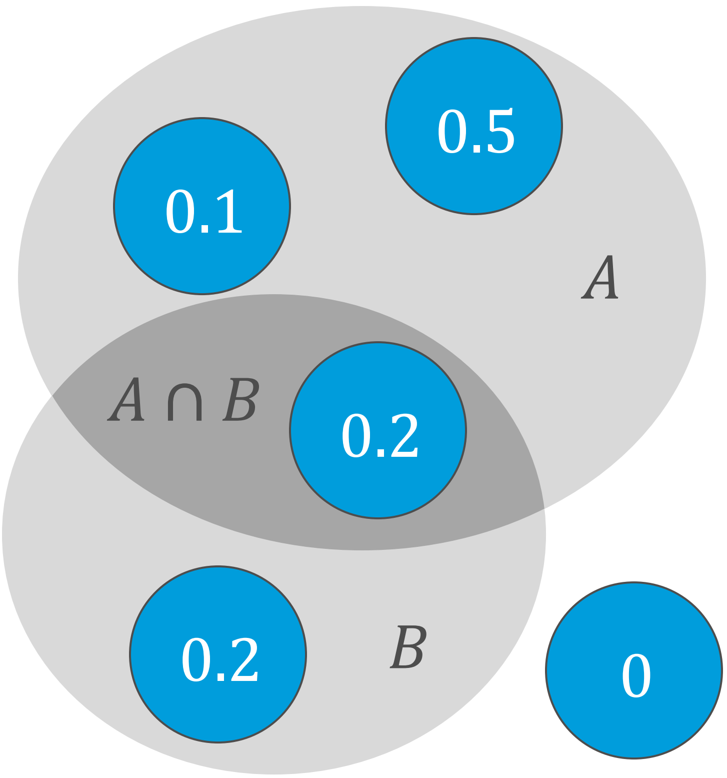
Ariel Procaccia | Harvard University

CONDITIONAL PROBABILITY

- The probability of event A given event B is defined as

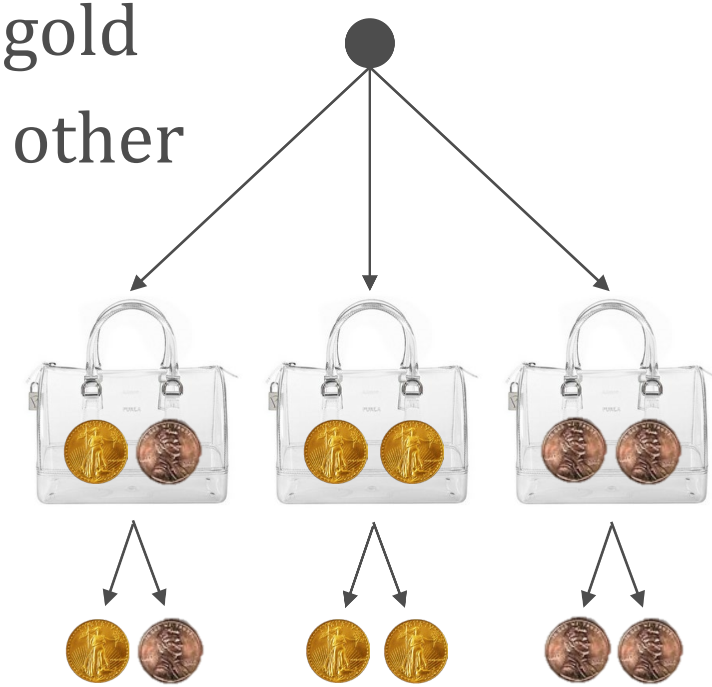
  $\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}$

- Think of it as the proportion of $A \cap B$ to B



CONDITIONAL PROBABILITY

- Three bags contain two gold coins, two pennies, and one of each
- Bag is chosen at random, and one coin from it is selected at random; the coin is gold
- What's the probability that the other coin is gold?
- G_i: coin $i \in \{1,2\}$ is gold
- $\Pr[G_1] = \frac{1}{2}, \Pr[G_1 \cap G_2] = \frac{1}{3}$
- $\Pr[G_2 | G_1] = \frac{1/3}{1/2} = \frac{2}{3}$



CONDITIONAL PROBABILITY

- $\Pr[A \cap B] = \Pr[B] \times \Pr[A|B]$
- Interpretation: For A and B to occur, B must occur, and A must occur given that B occurred
- Applying iteratively, we get the Chain Rule:

$$\Pr[A_1 \cap \cdots \cap A_n] \\ = \Pr[A_1] \times \Pr[A_2|A_1] \times \cdots \Pr[A_n|A_1, \cdots, A_{n-1}]$$

- We can also directly derive Bayes' Rule:

$$\Pr[A|B] = \frac{\Pr[A]\Pr[B|A]}{\Pr[B]}$$

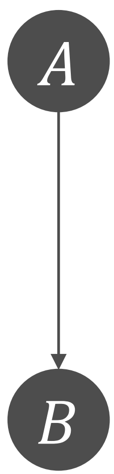
Thomas Bayes

1701–1761

English statistician, philosopher and minister. Also remembered for his probabilistic approach to miracles.

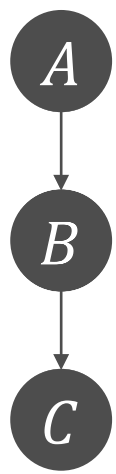
MODELING CAUSE AND EFFECT

Our goal is to graphically and concisely capture the dependencies among random variables



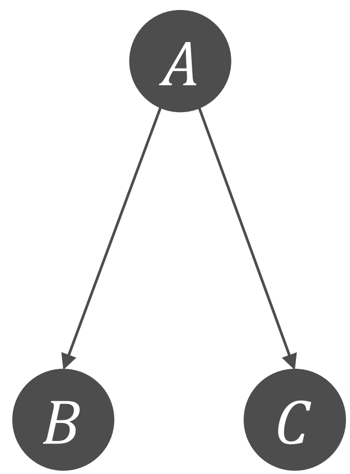
Direct cause

$\boldsymbol{P}[B|A]$



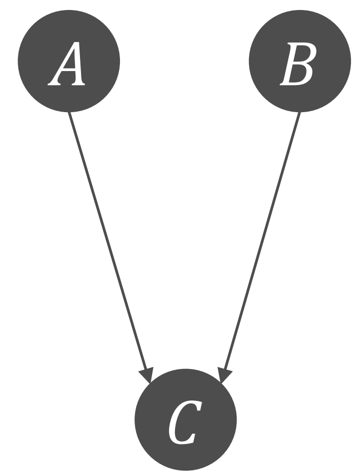
Indirect cause

$\boldsymbol{P}[B|A]\,,\boldsymbol{P}[C|B]$



Common cause

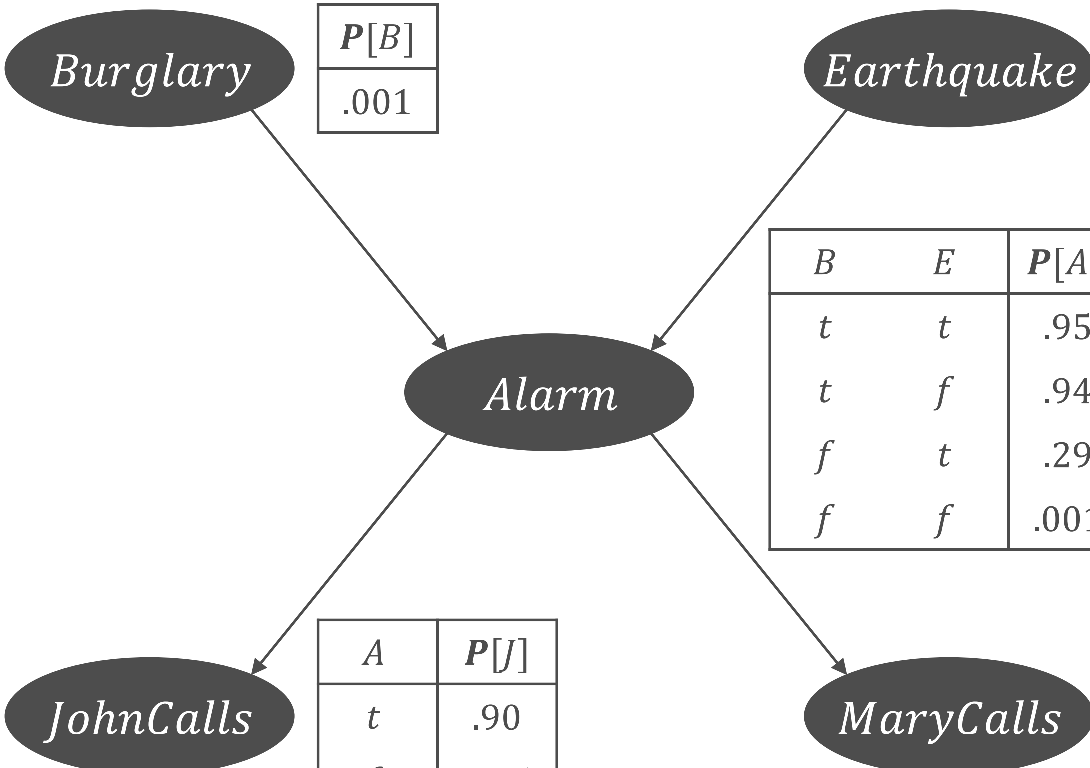
$\boldsymbol{P}[B|A],\boldsymbol{P}[C|A]$



Common effect

$\boldsymbol{P}[C|A,B]$

BAYES NETS BY EXAMPLE



$\mathbf{P}[B]$
.001

$\mathbf{P}[E]$
.002

B	E	$\mathbf{P}[A]$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$\mathbf{P}[J]$
t	.90
f	.005

A	$\mathbf{P}[M]$
t	.70
f	.01

THE JOINT DISTRIBUTION

- Using x_i as a shorthand for the event $X_i = x_i$, it holds by the chain rule that

$$\Pr[x_1, \dots, x_n] = \prod_{i=1}^{n} \Pr[x_i \mid x_{i-1}, \dots, x_1]$$

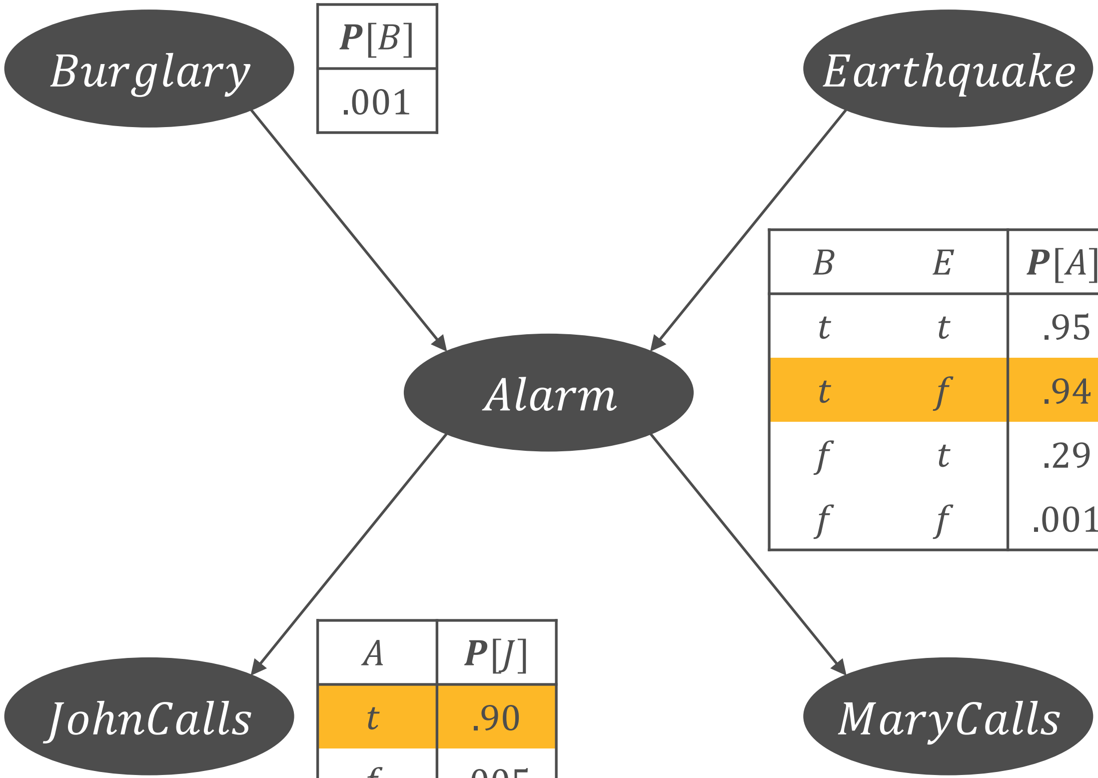
- In a Bayes net each random variable is conditionally independent of its predecessors given its parents:

$$\Pr[x_i \mid x_{i-1}, \dots, x_1] = \Pr[x_i \mid parents(X_i)]$$

- It follows that

$$\Pr[x_1, \dots, x_n] = \prod_{i=1}^{n} \Pr[x_i \mid parents(X_i)]$$

THE JOIN DISTRIBUTION



$\boldsymbol{P}[B]$
.001

$\boldsymbol{P}[E]$
.002

B	E	$\boldsymbol{P}[A]$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

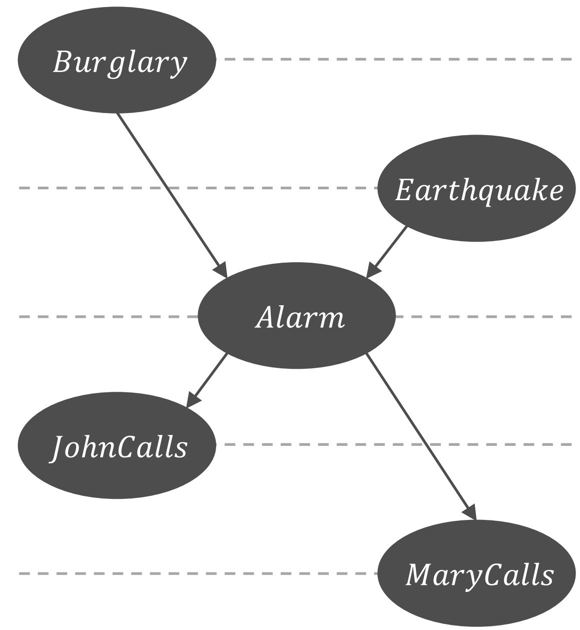
A	$\boldsymbol{P}[J]$
t	.90
f	.005

A	$\boldsymbol{P}[M]$
t	.70
f	.01

$$\Pr[B = t, E = f, A = t, J = t, M = f] = 0.01 \times 0.998 \times 0.94 \times 0.9 \times 0.3$$

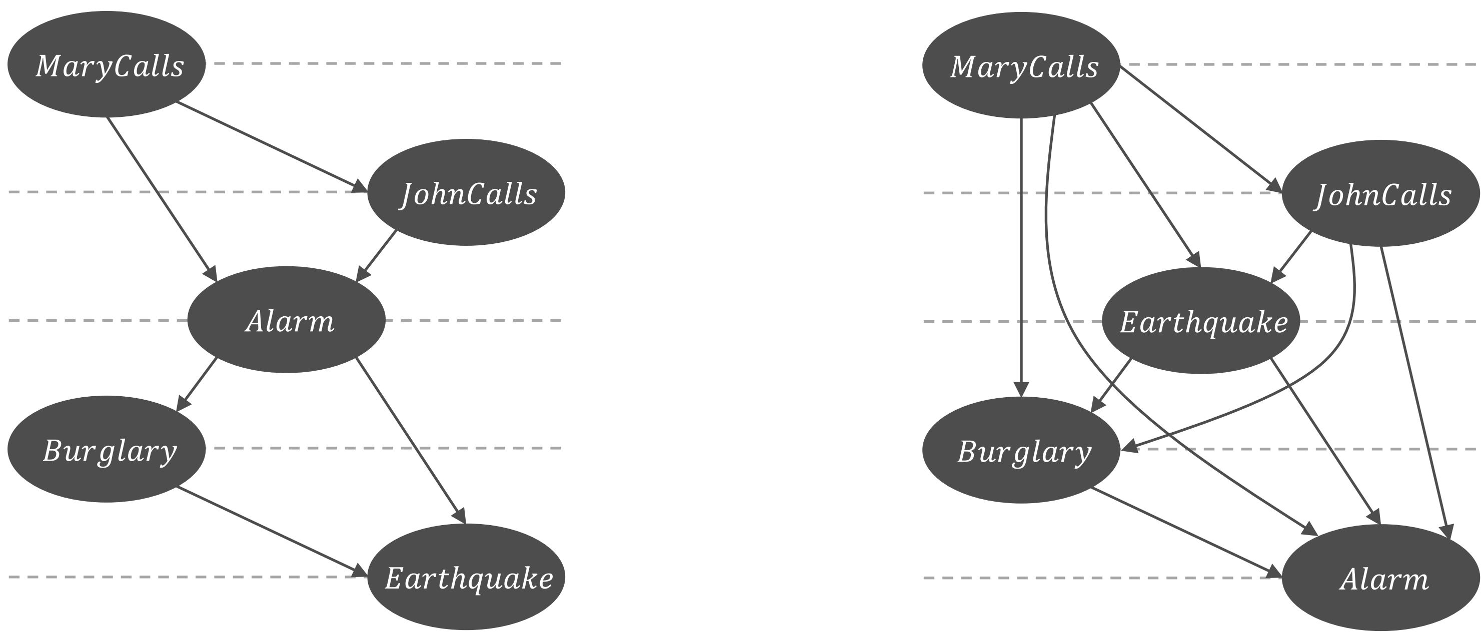
CONSTRUCTING BAYES NETS

- We can choose any ordering of the nodes $X_1, \dots, X_n$
- We must insert links so that the conditional independence condition holds



COMPACTNESS AND NODE ORDERING

We will get a compact representation only if we choose the node ordering well



INFERENCE IN BAYES NETS

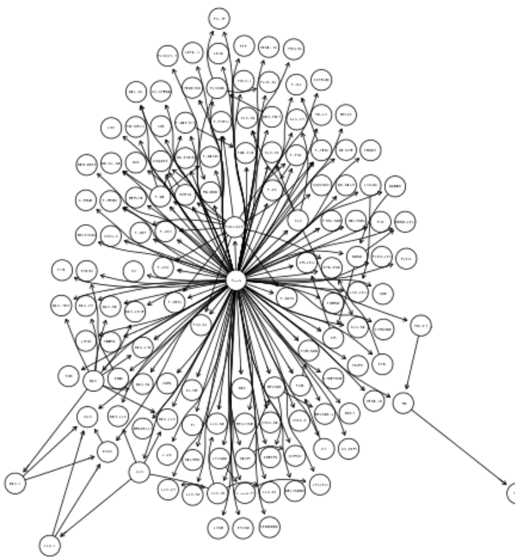
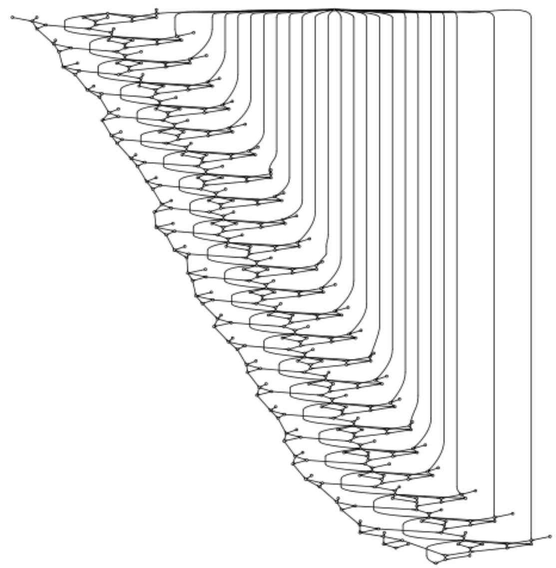
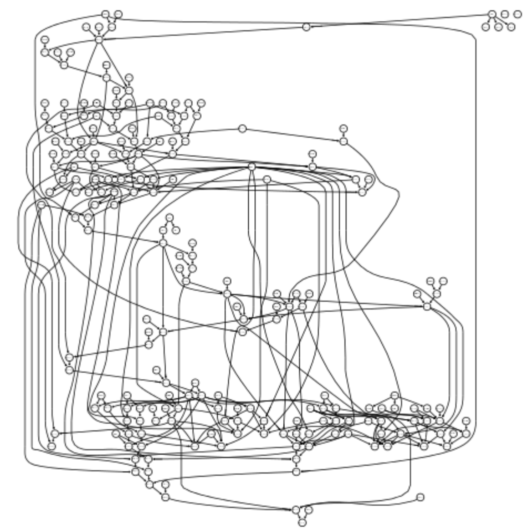
- Our goal is to calculate a useful quantity from a joint probability distribution
- Posterior probability:

$$\Pr[Q = q \mid E_1 = e_1, \ldots, E_k = e_k]$$

- Most likely explanation:

$$\operatorname{argmax}_q \Pr[Q = q \mid E_1 = e_1, \ldots, E_k = e_k]$$

INFERENCE BY ENUMERATION?

[https://www.bnlearn.com/bnrepository/discrete-verylarge.html]

SAMPLING METHODS

- Warmup: Generate events from a network with no evidence
- For a large number k of samples,

$$\Pr[X_1 = x_1, \ldots, X_n = x_n] \approx \frac{\#(x_1, \ldots, x_n)}{k}$$

- How do we obtain samples?

DIRECT SAMPLING

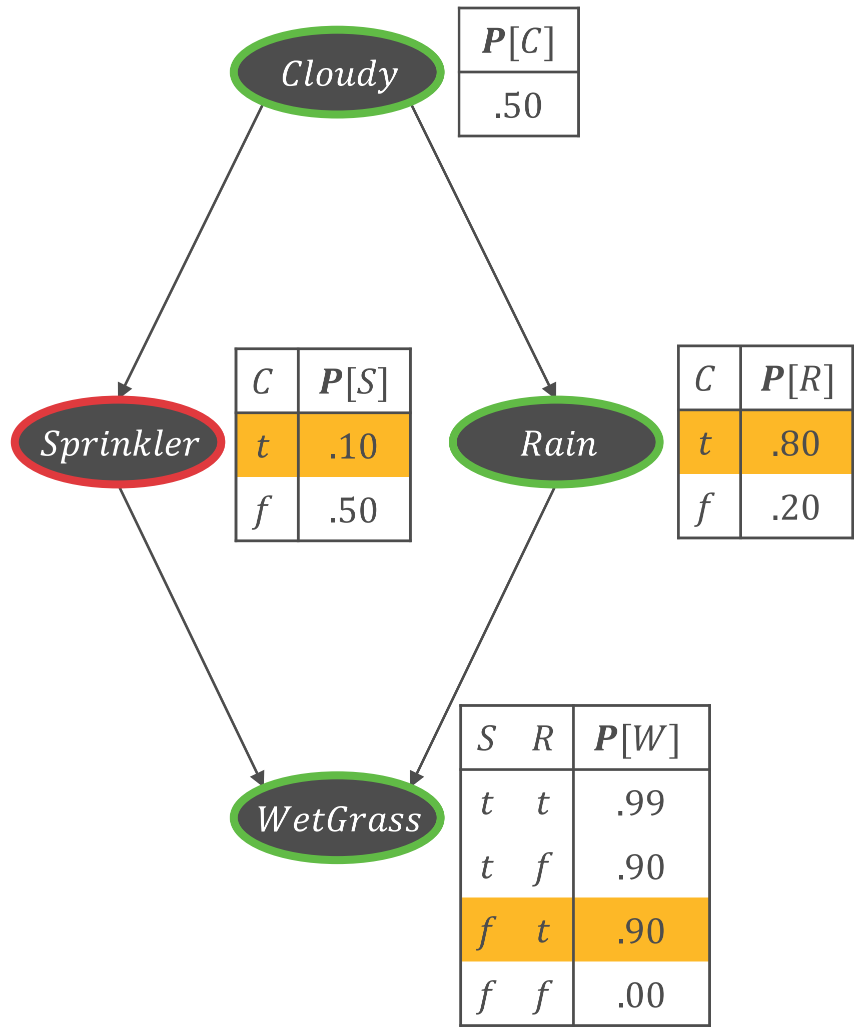
Sample $C \leftarrow t$

Sample $S \leftarrow f$

Sample $R \leftarrow t$

Sample $W \leftarrow t$

Out: (t, f, t, t)



REJECTION SAMPLING

- We want to estimate $\Pr[R = t|S = t]$ using 100 samples
- 73 have $S = f$ and are rejected
- 27 have $S = t$
- Of the 27, 8 have $R = t$
- Our estimate would be 8/27

IT'S RAINING FROGS!

Try to estimate

$$\Pr[WetGrass = f \mid RainOfFrogs = t]$$

LIKELIHOOD WEIGHTING

```
function WEIGHTED-SAMPLE(bn, e)
    w ← 1; x ← initialized from e
    foreach variable X_i in X_1, …, X_n do
        if X_i is an evidence variable with value x_i in e
            then w ← w · Pr[X_i = x_i | parents(X_i)]
            else x_i ← random sample conditioned on parents
    return x, w
```

Why is this the right thing to do? If $\boldsymbol{e}$ is evidence and $\boldsymbol{z}$ is sampled, then:

$$\underbrace{\prod_i \Pr[z_i \mid parents(Z_i)]}_{\text{Probability of sampling } \boldsymbol{z}} \cdot \underbrace{\prod_j \Pr[e_j \mid parents(E_j)]}_{\text{Weight of sample}} = \Pr[\boldsymbol{z}, \boldsymbol{e}]$$

LIKELIHOOD WEIGHTING: EXAMPLE

Evidence: $C = t, W = t$

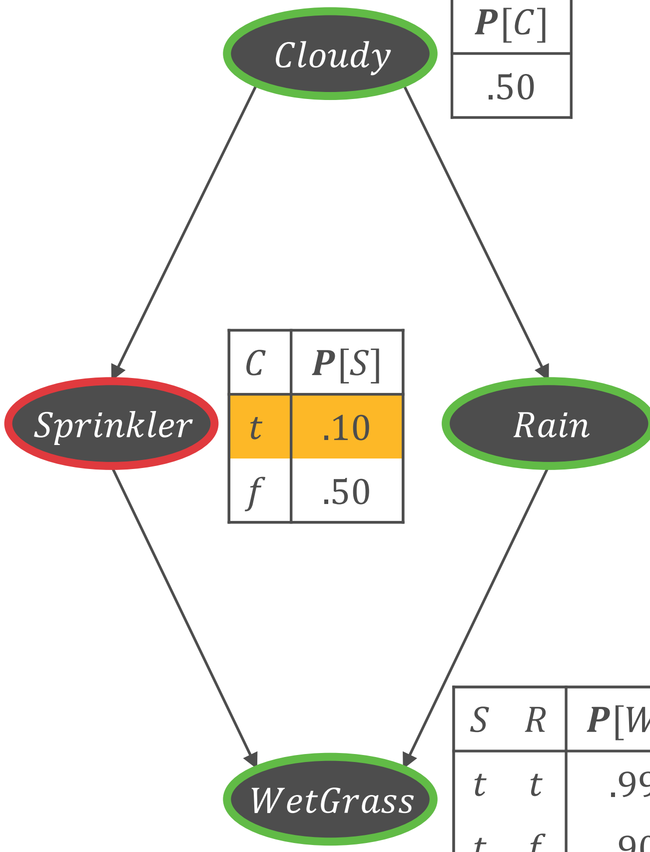
C is evidence:
$w = 1 \cdot 0.5 = 0.5$

Sample $S \leftarrow f$

Sample $R \leftarrow t$

W is evidence:
$w = 0.5 \cdot 0.9 = 0.45$

Out: $(t, f, t, t), w = .45$



$\mathbf{P}[C]$
.50

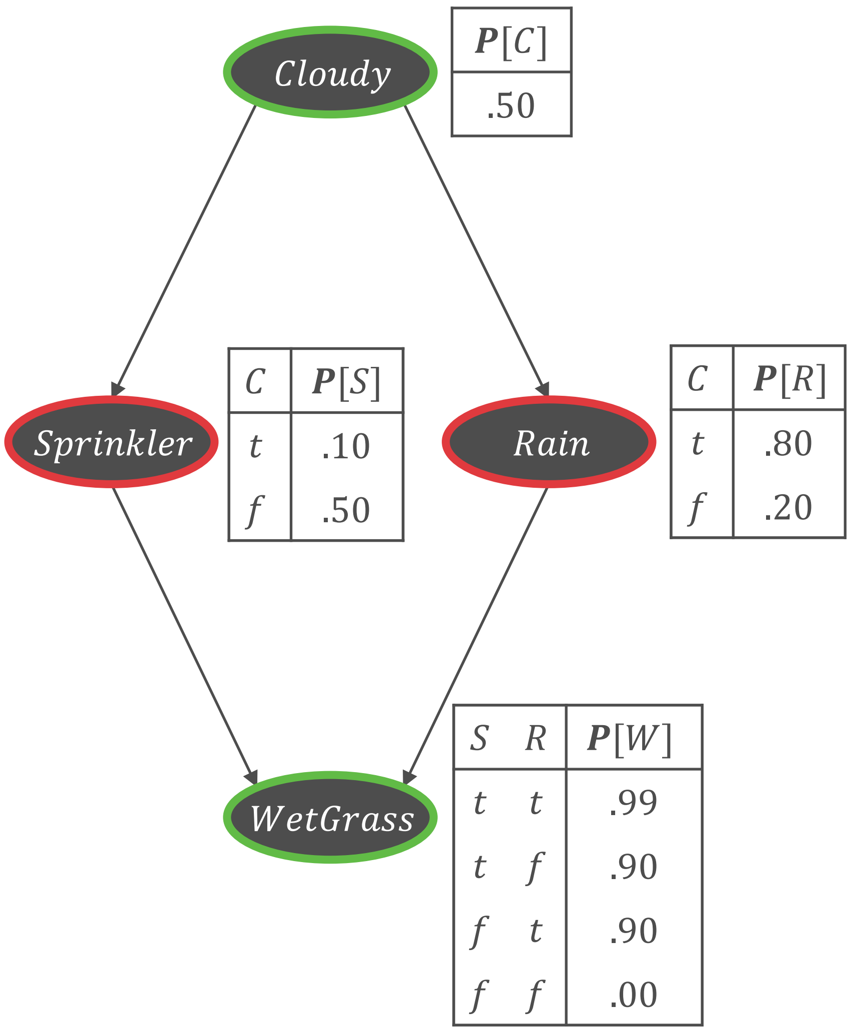
C	$\mathbf{P}[S]$
t	.10
f	.50

C	$\mathbf{P}[R]$
t	.80
f	.20

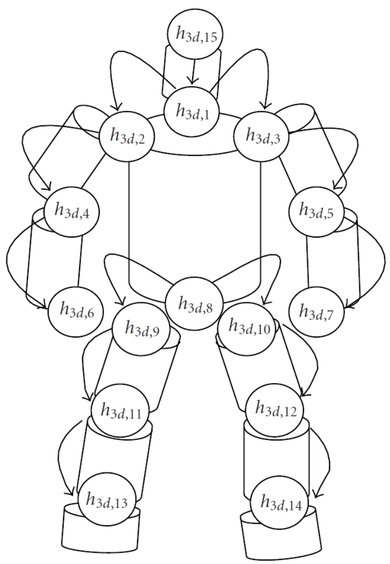
S	R	$\mathbf{P}[W]$
t	t	.99
t	f	.90
f	t	.90
f	f	.00

LIKELIHOOD WEIGHTING: EXAMPLE

- The evidence is $C = t$, $W = t$ and we sample $S = f,\ R = f$
- Poll: What is the weight of this sample?
  - 0.495
  - 0.45
  - 0.1
  - 0



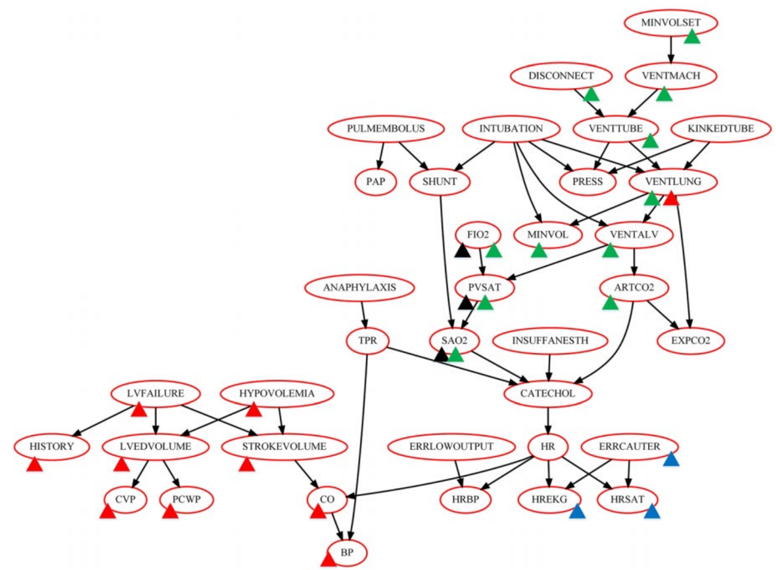
APPLICATIONS



Computer vision

Pose estimation

[Wang and Cheng, 2010]



Bioinformatics

Gene regulatory networks

[Xing et al., 2017]