Last update: December 4, 2012

Bayesian networks

CMSC 421: Chapter 14, Sections 1–5

Outline

◇ Syntax

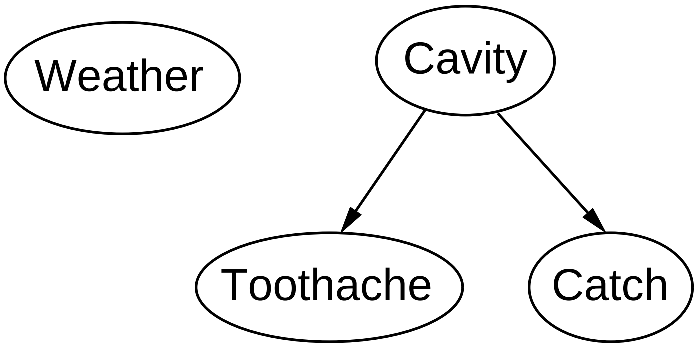
◇ Semantics

◇ Parameterized distributions

Bayesian networks

◇ Graphical network that encodes conditional independence assertions:

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



◇ $Weather$ is independent of the other variables

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

◇ For each node X_i, $\mathbf{P}(X_i \mid Parents(X_i))$ is represented as a *conditional probability table* (CPT); we'll have examples later

Example

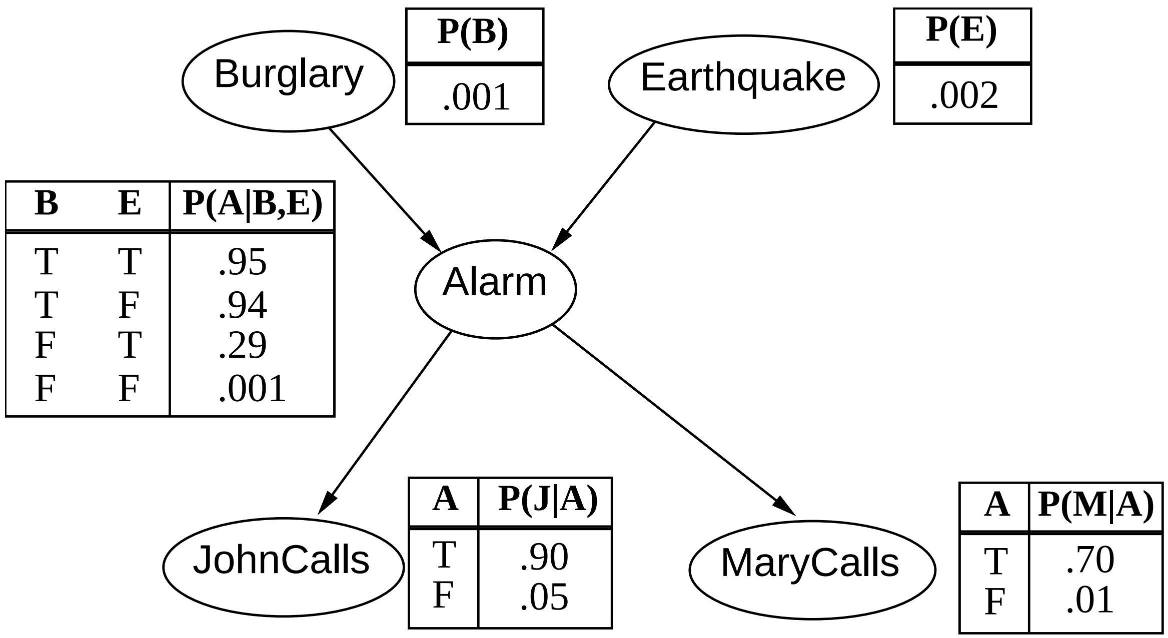
◇ Example from Judea Pearl at UCLA:

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
- So can an earthquake
- The alarm can cause Mary to call
- It can also cause John to call



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

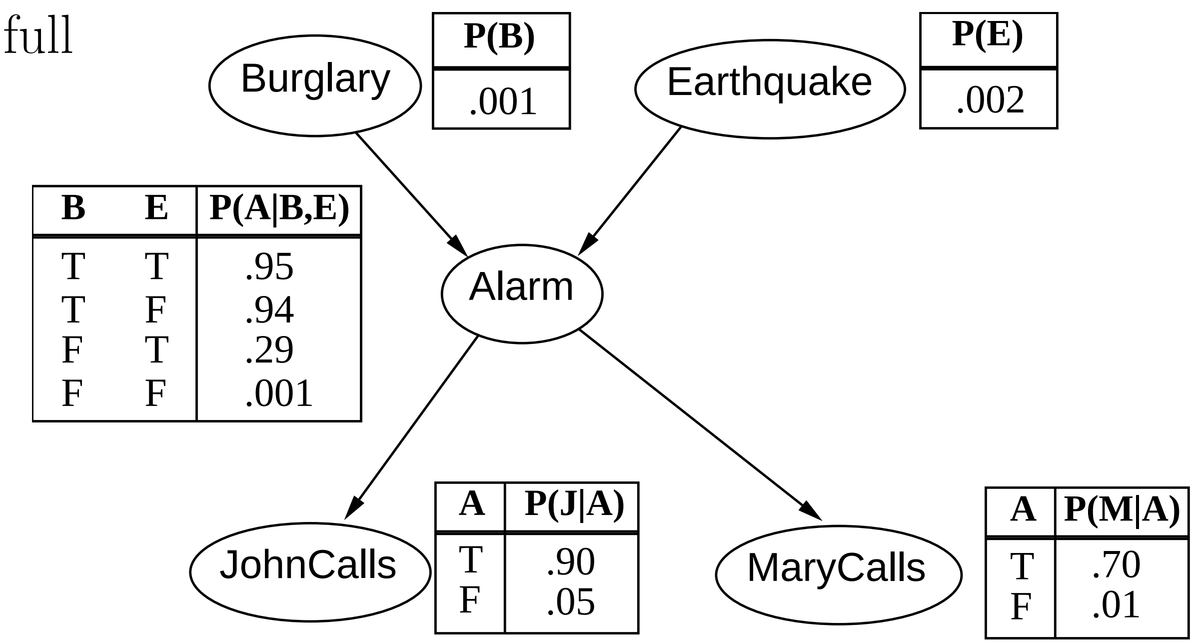
◇ For a Boolean node X_i with k Boolean parents, the CPT has 2^k rows, one for each combination of parent values

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

◇ If there are n variables and if each variable has no more than k parents, the complete network requires no more than $n \cdot 2^k$ numbers

- Grows linearly with n, vs. $O(2^n)$ for the full joint distribution

◇ How many numbers for the burglary net?



P(B)
.001

P(E)
.002

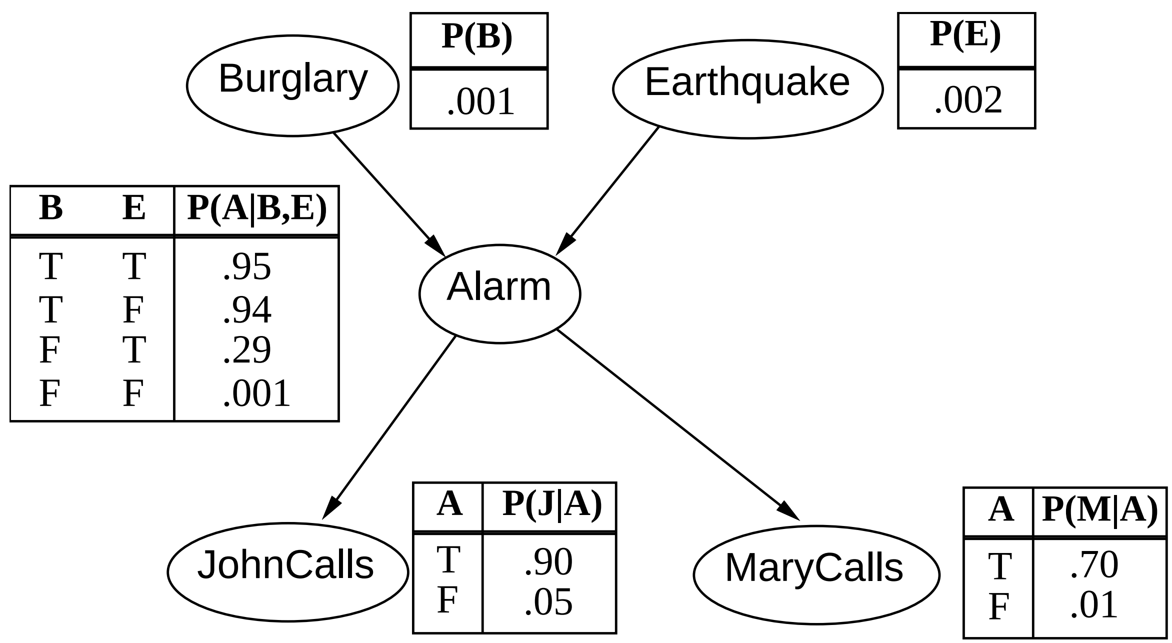
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

Semantics of Bayesian nets

- ♢ In general, *semantics* = "what things mean"
  - Here, we're interested in what a Bayesian net means
- ♢ We'll look at *global* and *local* semantics



Global semantics

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

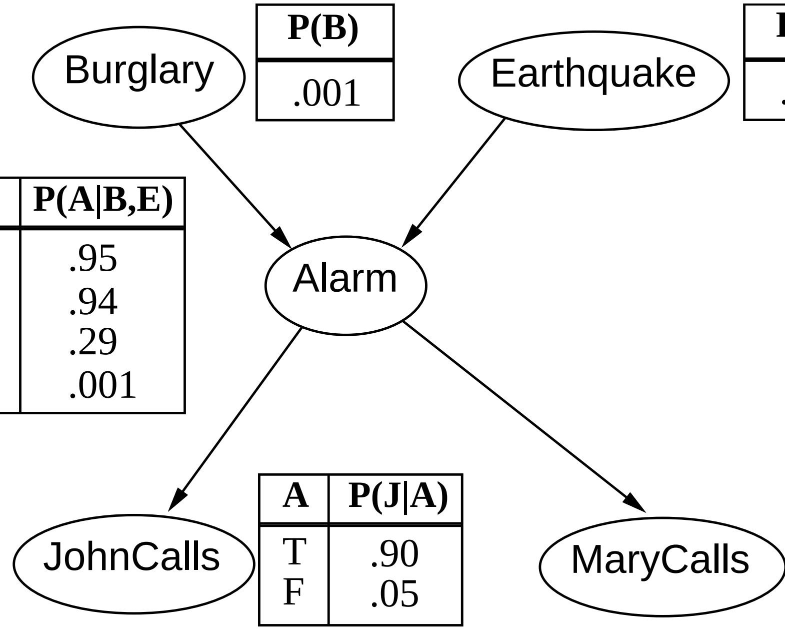
- If $X_1, \ldots, X_n$ are the random variables, the chain rule and conditional independence give us $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$

◇ E.g., $P(j \land m \land a \land \neg b \land \neg e)$

$= P(j \mid a)\ P(m \mid a)\ P(a \mid \neg b, \neg e)\ P(\neg b)\ P(\neg e)$

$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$

≈ 0.00063



P(B)
.001

P(E)
.002

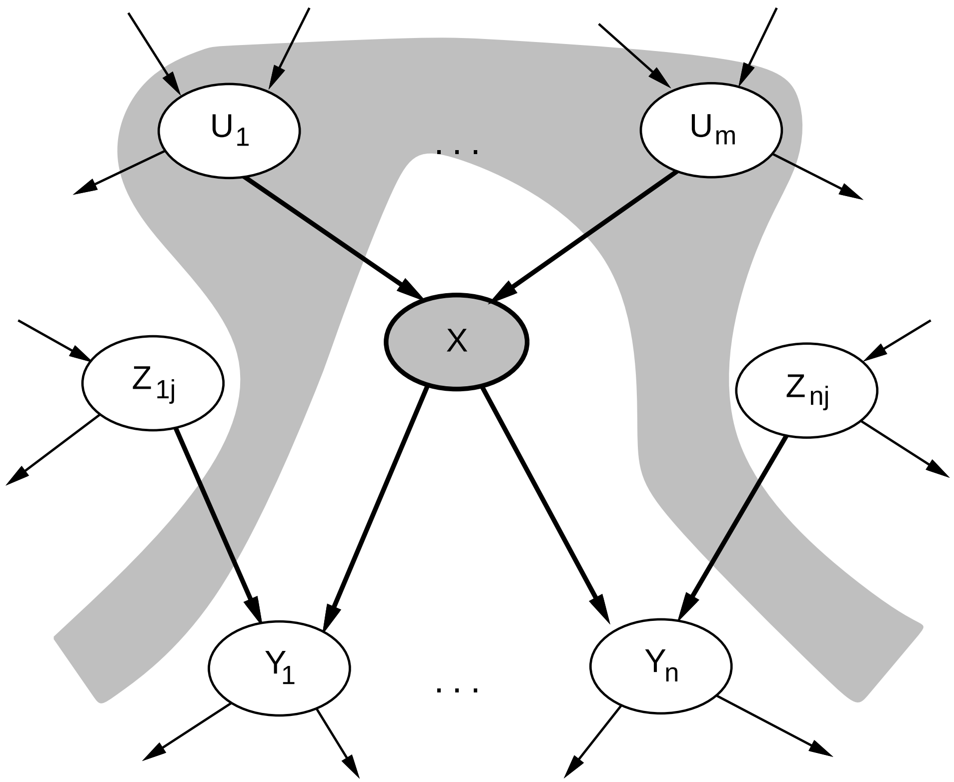
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

Local semantics

◇ *Local* semantics: each node is conditionally independent of its nondescendants given its parents

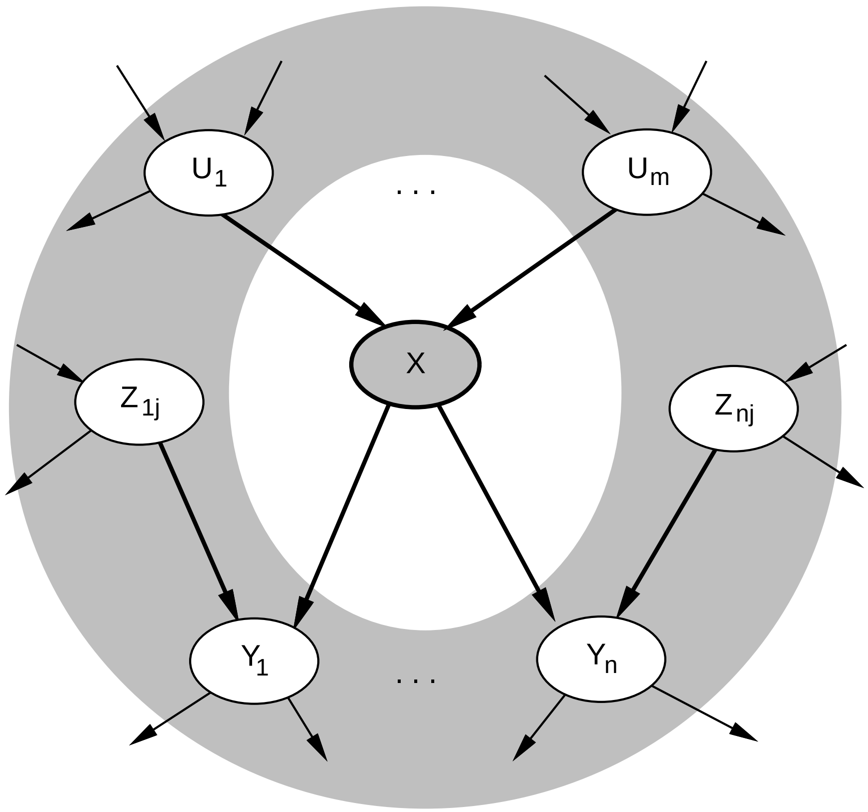


◇ Theorem: Local semantics $\Leftrightarrow$ global semantics

Markov blanket

◇ Each node is conditionally independent of all others given its *Markov blanket*:

- parents + children + children's parents



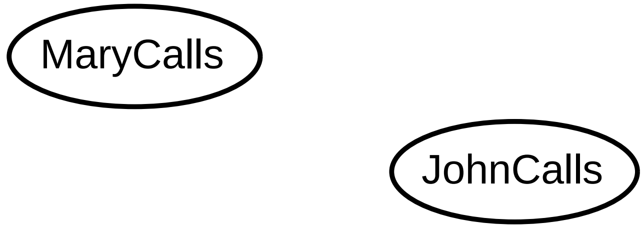
Constructing Bayesian networks

- Given a set of random variables
  1. Choose an ordering $X_1, \ldots, X_n$
     - In principle, *any* ordering will work
  2. For $i = 1$ to n, add X_i to the network as follows:
     - For $Parents(X_i)$, choose a subset of $\{X_1, \ldots, X_{i-1}\}$ such that X_i is conditionally independent of the other nodes in $\{X_1, \ldots, X_{i-1}\}$
     - i.e., $\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1,\ \ldots,\ X_{i-1})$
- This choice of parents guarantees the global semantics:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1,\ \ldots,\ X_{i-1}) \quad \text{(chain rule)} \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i)) \quad \text{(by construction)}\end{aligned}$$

Example

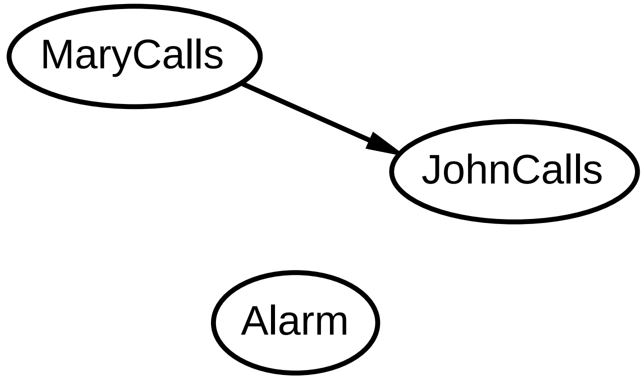
◇ Suppose we choose the ordering M, J, A, B, E



$P(J \mid M) = P(J)$?

Example

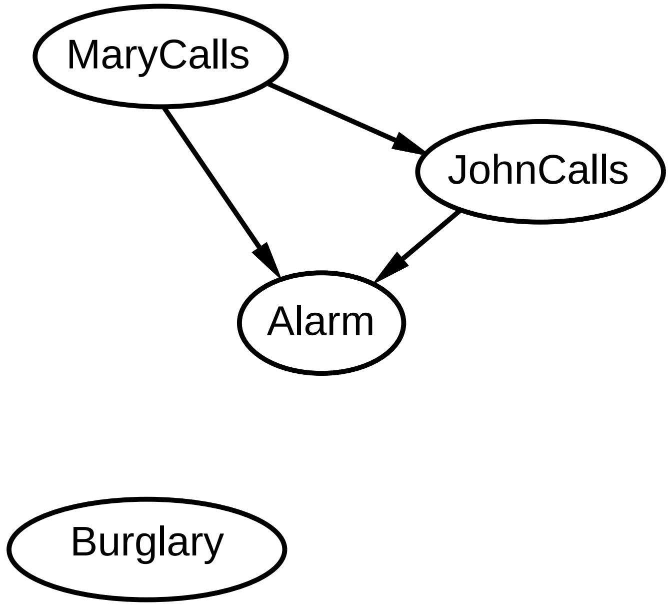
◇ Suppose we choose the ordering M, J, A, B, E



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$?

Example

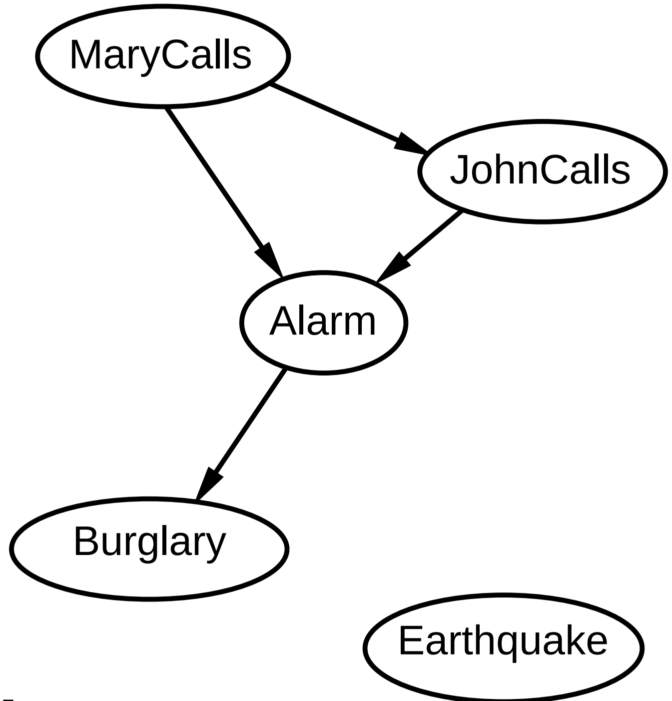
◊ Suppose we choose the ordering M, J, A, B, E



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$?
$P(B \mid A, J, M) = P(B)$?

Example

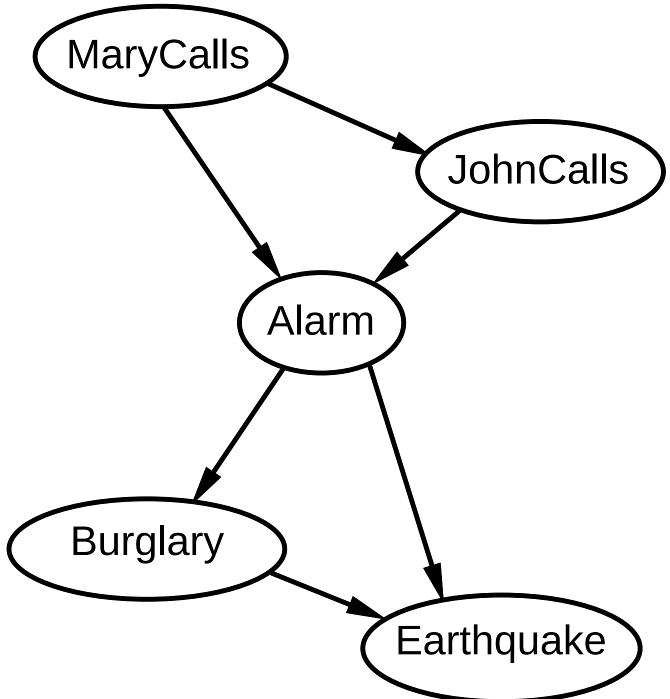
◇ Suppose we choose the ordering M, J, A, B, E



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$? Yes
$P(B \mid A, J, M) = P(B)$? No
$P(E \mid B, A, J, M) = P(E \mid A)$?
$P(E \mid B, A, J, M) = P(E \mid A, B)$?

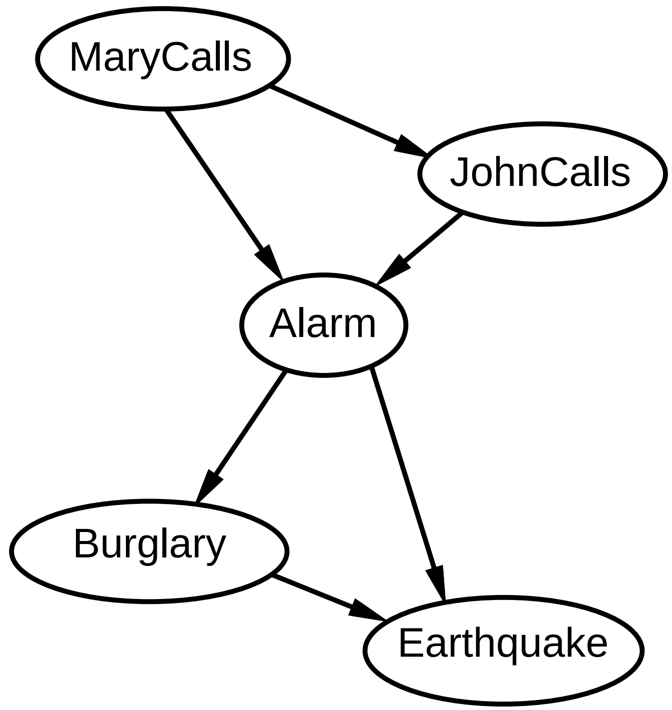
Example

◇ Suppose we choose the ordering M, J, A, B, E



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$? Yes
$P(B \mid A, J, M) = P(B)$? No
$P(E \mid B, A, J, M) = P(E \mid A)$? No
$P(E \mid B, A, J, M) = P(E \mid A, B)$? Yes

Example, continued



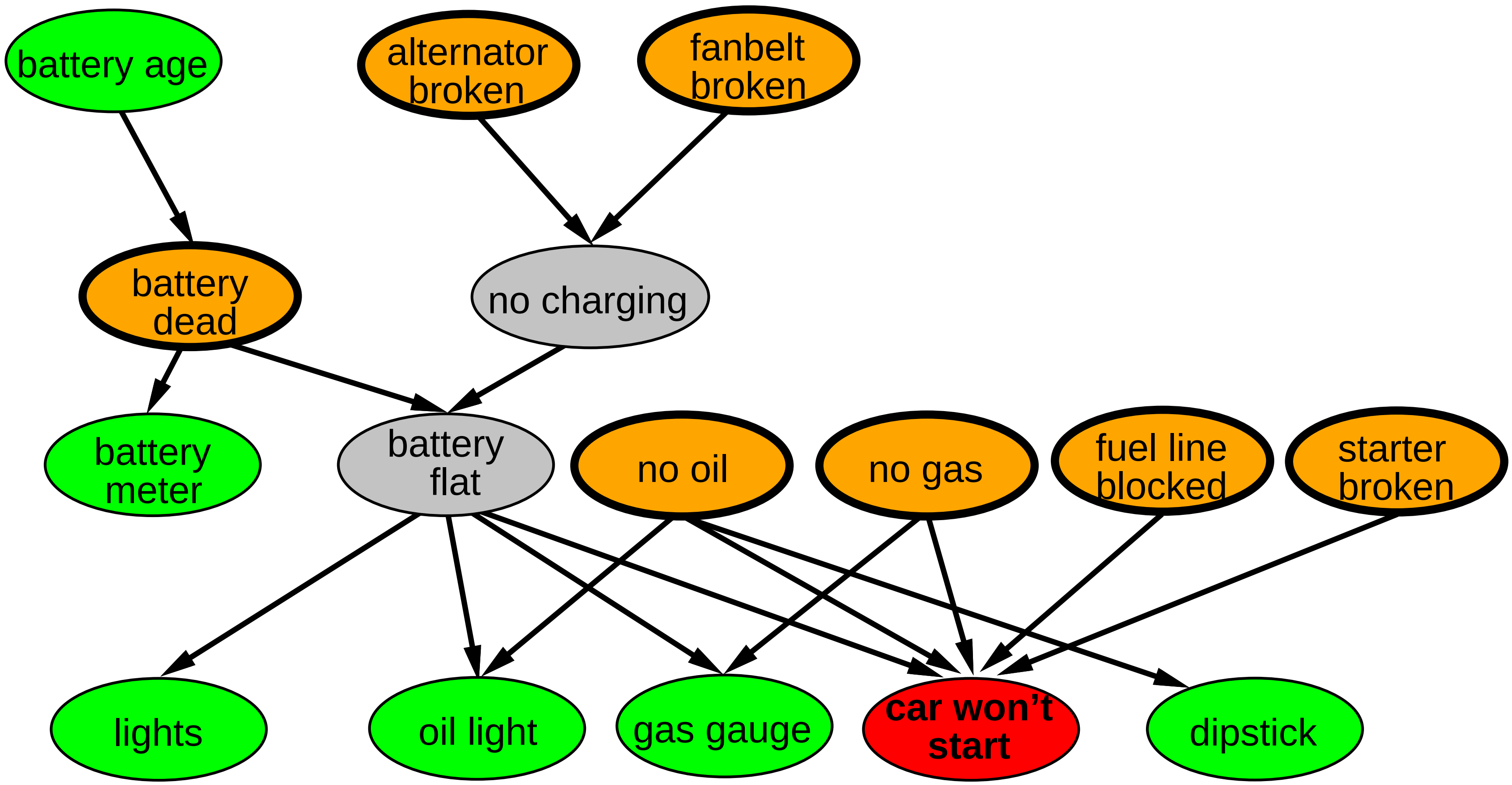
◇ In noncausal directions,

- Deciding conditional independence is hard
- Assessing conditional probabilities is hard
- Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers needed

Example: Car diagnosis

♢ Initial evidence: car won't start

- Testable variables (green), "broken, so fix it" variables (orange)
- Hidden variables (gray) ensure sparse structure, reduce parameters



Compact conditional distributions

◇ Problem: CPT grows exponentially with number of parents.

◇ Can overcome this if the causes don't interact: use *Noisy-OR* distribution

1) Parents $U_1 \ldots U_k$ include all causes (can add *leak node*)

2) Independent failure probability q_i for each cause U_i by itself

$$\Rightarrow \; P(\neg X \mid U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = \Pi_{i=1}^{j} q_i$$

◇ Number of parameters **linear** in number of parents

Cold	*Flu*	*Malaria*	$P(Fever)$	$P(\neg Fever)$
F	F	F	**0.0**	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Compact conditional distributions, continued

◇ Problem: CPT becomes infinite with continuous-valued parent or child

◇ Solution: *canonical* distributions that are defined compactly

- i.e., standard math formulas

◇ *Deterministic* nodes are the simplest case:
$X = f(Parents(X))$ for some function f

◇ Examples:

- Boolean functions

$$NorthAmerican \Leftrightarrow Canadian \vee US \vee Mexican$$

- Numerical relationships among continuous variables

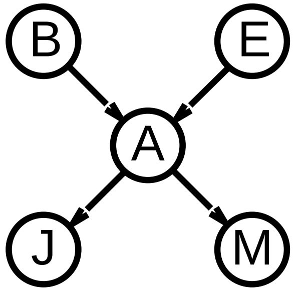
$$\frac{\partial Level}{\partial t} = \text{inflow + precipitation - outflow - evaporation}$$

◇ More details in the book

Inference tasks

- *Simple queries*: compute posterior marginal distribution $\mathbf{P}(X_i \mid \mathbf{E}=\mathbf{e})$
  - e.g., $P(NoGas \mid Gauge=empty, Lights=on, Starts=false)$
- *Conjunctive queries*:
  - $\mathbf{P}(X_i, X_j \mid \mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i \mid \mathbf{E}=\mathbf{e})\mathbf{P}(X_j \mid X_i, \mathbf{E}=\mathbf{e})$
- *Optimal decisions*: decision networks include utility information; probabilistic inference required for $P(outcome \mid action, evidence)$
- *Value of information*: which evidence to seek next?
- *Sensitivity analysis*: which probability values are most critical?
- *Explanation*: why do I need a new starter motor?

Inference by enumeration



◇ Simple query on the burglary network

- probability of burglary, given John and Mary both call

$\mathbf{P}(B \mid j, m)$

$= \mathbf{P}(B, j, m)/P(j, m)$ (def. of cond. probability)

$= \alpha \mathbf{P}(B, j, m)$ (normalization constant)

$= \alpha \sum_e \sum_a \mathbf{P}(B, e, a, j, m)$ (sum over hidden variables)

where $\sum_a \mathbf{P}(\ldots, a, \ldots)$ means $\mathbf{P}(\ldots, \neg a, \ldots) + \mathbf{P}(\ldots, a, \ldots)$

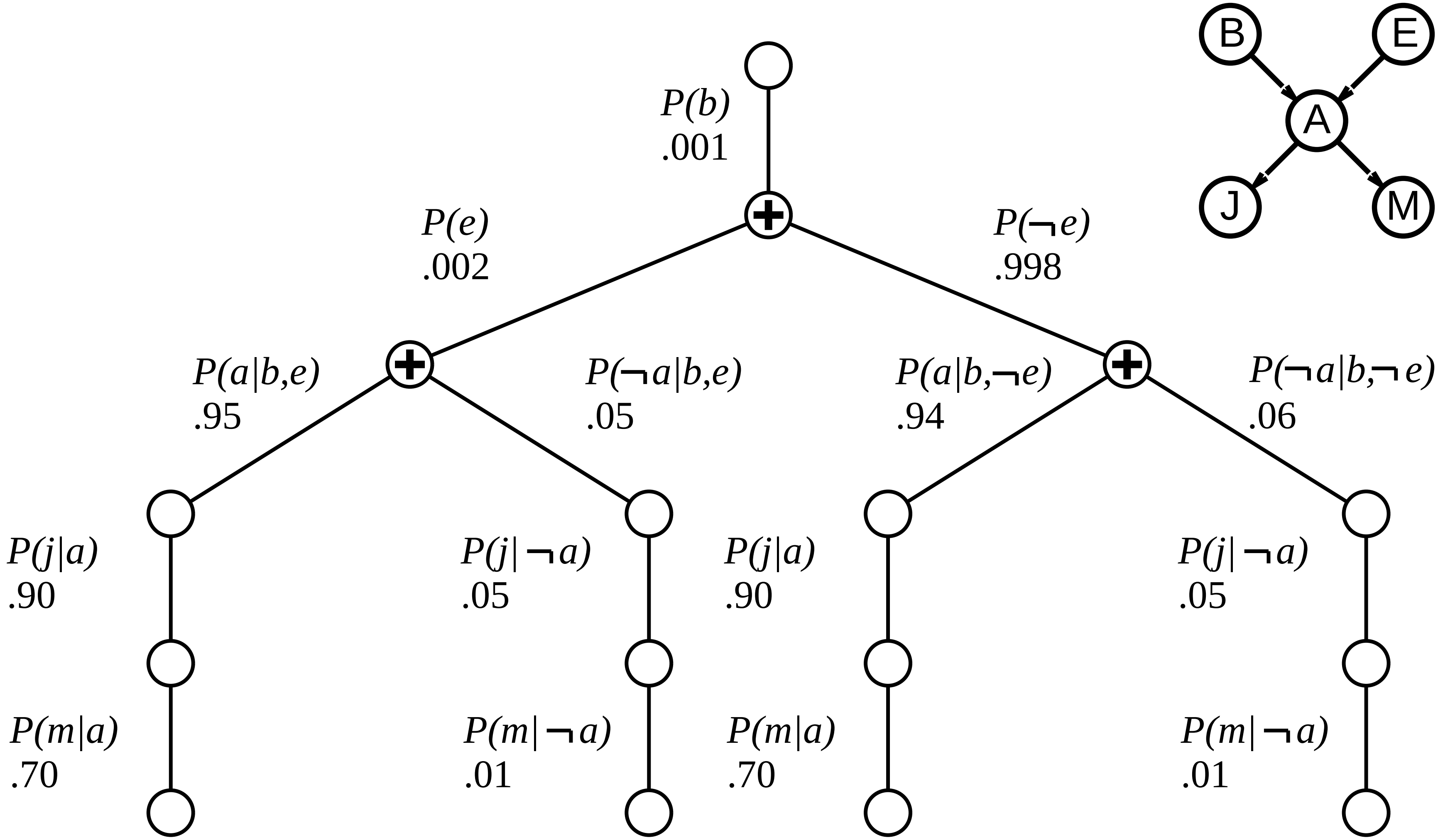
$= \alpha \sum_e \sum_a \; \mathbf{P}(B) \; P(e) \; \mathbf{P}(a|B, e) \; P(j|a) \; P(m|a)$ (cond. indep.)

$= \alpha \mathbf{P}(B) \sum_e P(e) \sum_a \; \mathbf{P}(a|B, e) \; P(j|a) \; P(m|a)$ (move out of Σ)

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

- Algorithm is in the book
- It's like evaluating the tree representation of an arithmetic expression

Inference by enumeration



- ♢ $\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a \mid B, e)\, P(j \mid a)\, P(m \mid a)$
- ♢ Inefficient: computes $P(j \mid a)\, P(m \mid a)$ repeatedly, for each value of e

Inference by variable elimination

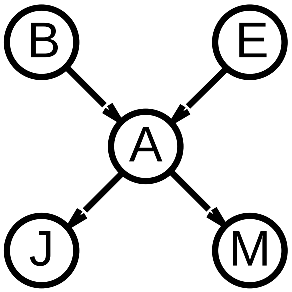
- Variable elimination: carry out summations right-to-left, storing intermediate results (*factors*) to avoid recomputation
  - Below, $\times$ represents pointwise multiplication of tables
    - i.e., multiply the corresponding elements

$$\mathbf{P}(B \mid j, m) = \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B, e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

$$= \alpha \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \underbrace{\sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)}_{\mathbf{f}_6(B,E)}$$

$$= \alpha \mathbf{f}_1(B) \times \underbrace{\sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B, E)}_{\mathbf{f}_7(B)}$$

$$= \alpha \mathbf{f}_1(B) \times \mathbf{f}_7(B)$$

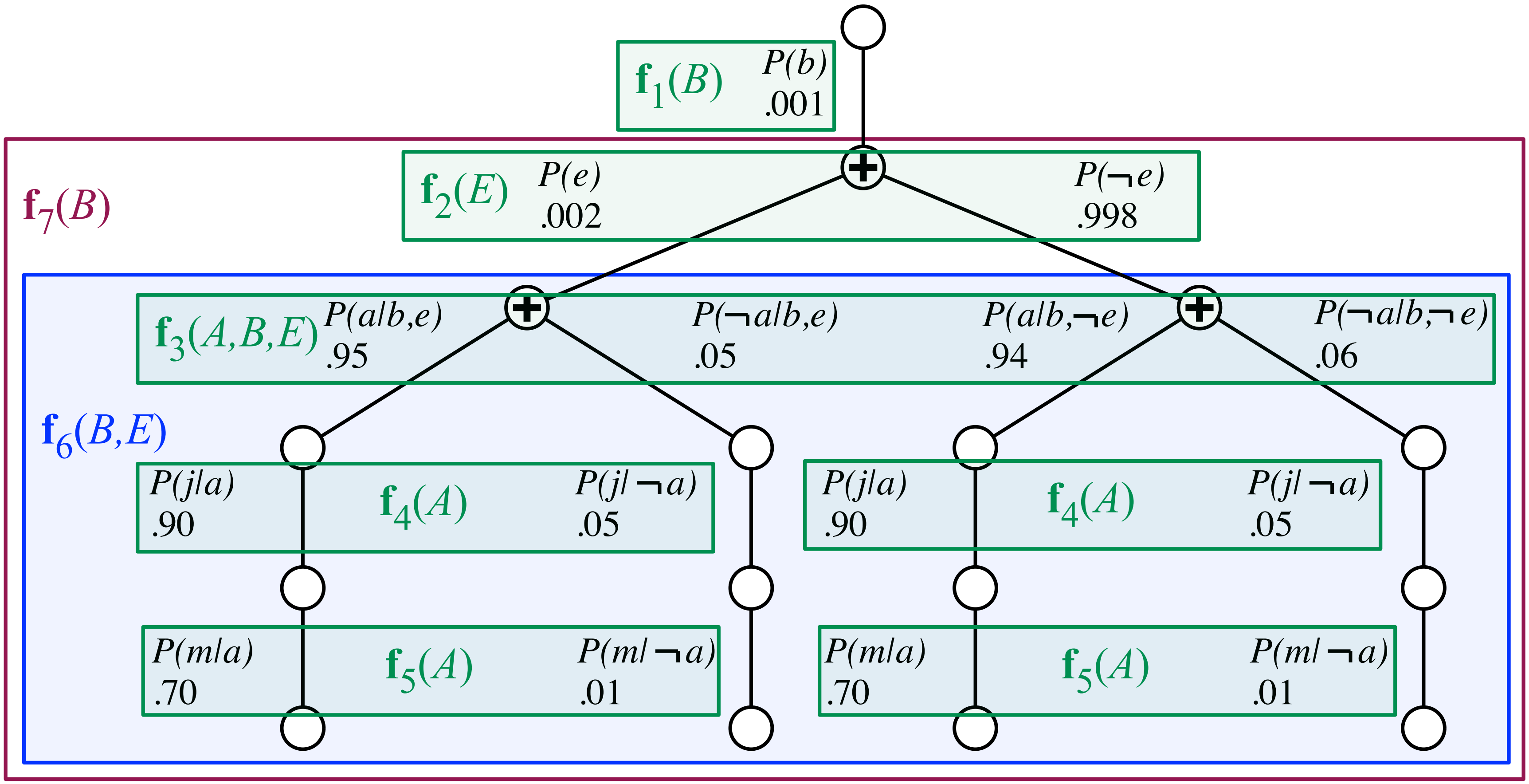


- Less complicated than it looks
  - Just cache the probability tables, going up from the bottom of the tree

Caching the probability tables

$$\mathbf{P}(B \mid j, m) = \alpha \mathbf{f}_1(B) \times \textstyle\sum_e \mathbf{f}_2(E) \times \underbrace{\textstyle\sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)}_{\mathbf{f}_6(B,E)}$$

$$= \alpha \mathbf{f}_1(B) \times \underbrace{\textstyle\sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B, E)}_{\mathbf{f}_7(B)}$$



Irrelevant variables

◇ What's the probability that John calls, given that there's a burglary?

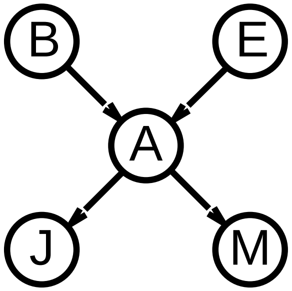
$$P(J \mid b) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e)\ P(J \mid a) \sum_m P(m \mid a)$$

- Sum over m is 1; M is **irrelevant** to the query

◇ **Theorem:** For query X, hidden variable Y is irrelevant unless $Y \in Ancestors(\{X\} \cup \mathbf{E})$

◇ Here, $X = J$,
$\mathbf{E} = \{B\}$,
$Ancestors(\{J, B\}) = \{J, B, A, E\}$

- so M is irrelevant



Complexity of exact inference

♢ *Singly connected* networks (or *polytrees*):

- any two nodes are connected by at most one (undirected) path
- complexity of inference is linear in the size of the network
  - ⋄ size = total number of entries in the probability tables

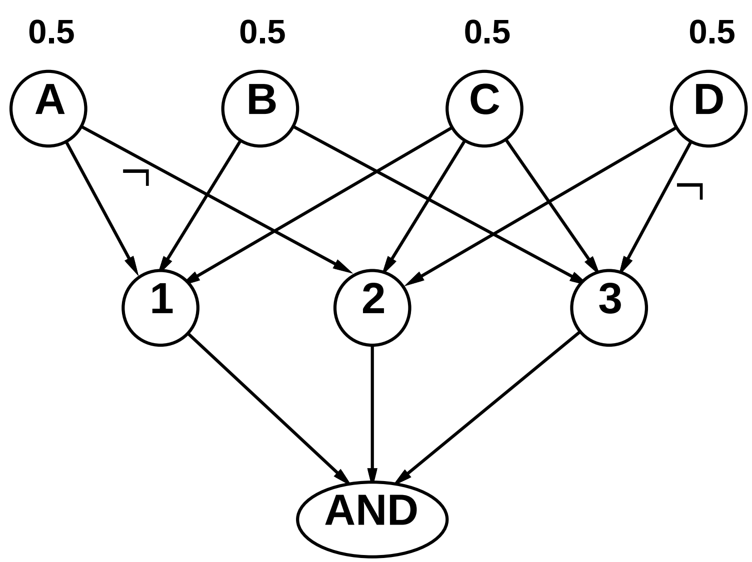
♢ *Multiply connected* networks:

- exponential time and space in the worst case
- includes propositional inference as a special case
- as hard as counting the number of ways to satisfy a propositional formula

1. A v B v C

2. C v D v ¬A

3. B v C v ¬D



Inference by stochastic simulation

Outline:

◇ Sampling from an empty network:
  1) Generate N random samples of events in the network
  2) Average the results
  3) For each event x, this gives us a posterior probability $\hat{P}(x)$
  4) As $N \to \infty$ this converges to x's true probability $P(x)$

◇ Rejection sampling, given evidence e:
  1) Generate N random samples of events in the network
  2) Reject samples that disagree with the evidence e, average the others
  3) For each event x, this gives us a posterior probability $\hat{P}(x \mid e)$
  4) As $N \to \infty$ this converges to x's true conditional probability $P(x \mid e)$

◇ Likelihood weighting: use evidence to weight samples

Sampling from an empty network

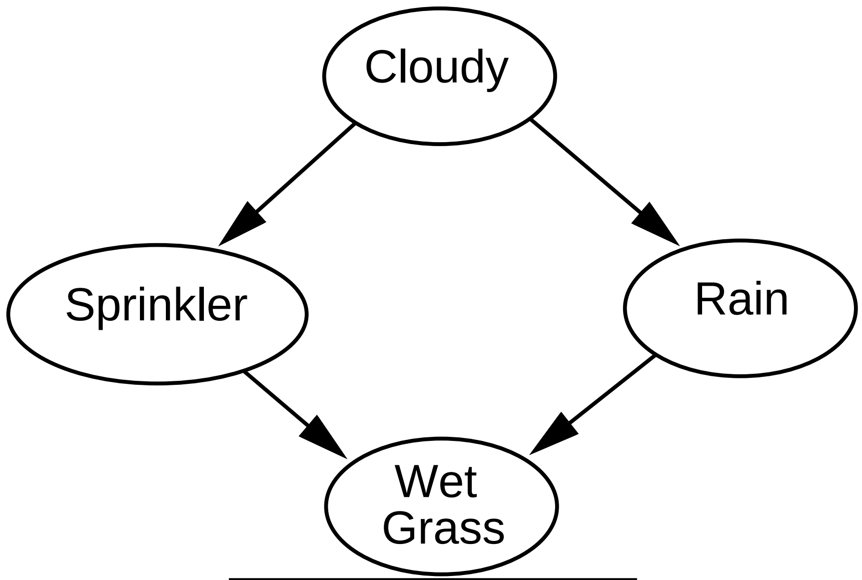
◇ Sampling from an empty network:
1) Generate N random samples of events in the network
2) Average the results
3) For each event x, this gives us a posterior probability $\hat{P}(x)$
4) As $N \to \infty$ this converges to x's true probability $P(x)$

function PRIOR-SAMPLE(bn) **returns** an event sampled from bn
  inputs: bn, a belief network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $\mathbf{x} \leftarrow$ an event with n elements
  for $i = 1$ **to** n **do**
    $x_i \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
      given the values of $Parents(X_i)$ in $\mathbf{x}$
  return $\mathbf{x}$

Example: find prob. of Wet Grass

P(C)
.50



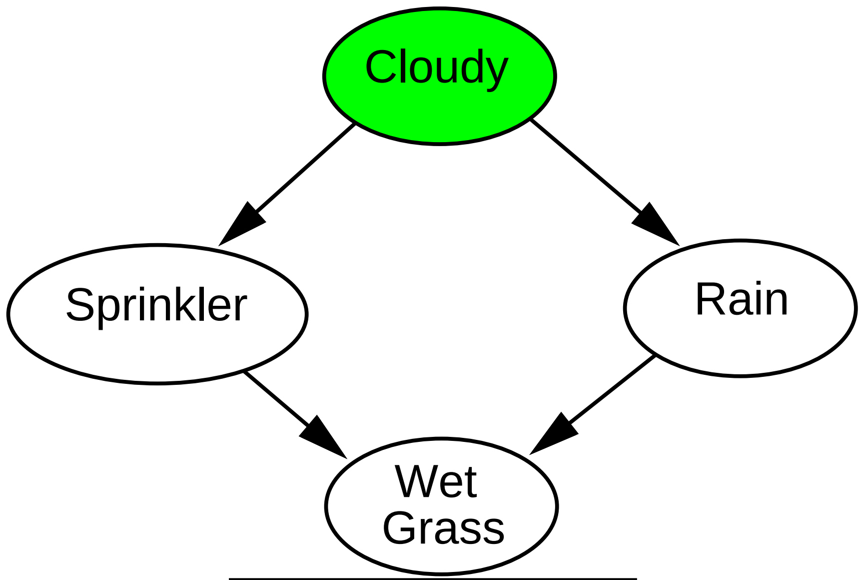
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: find prob. of Wet Grass

P(C)
.50



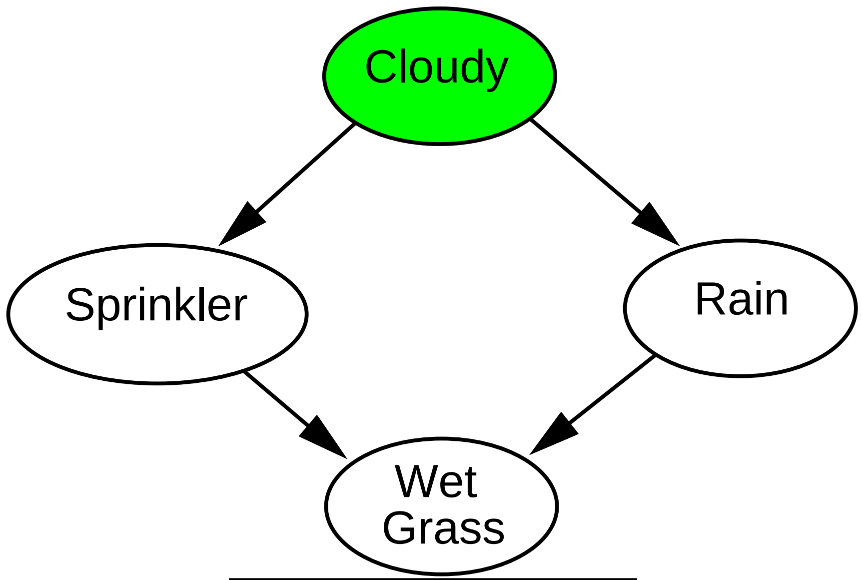
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: find prob. of Wet Grass

P(C)
.50



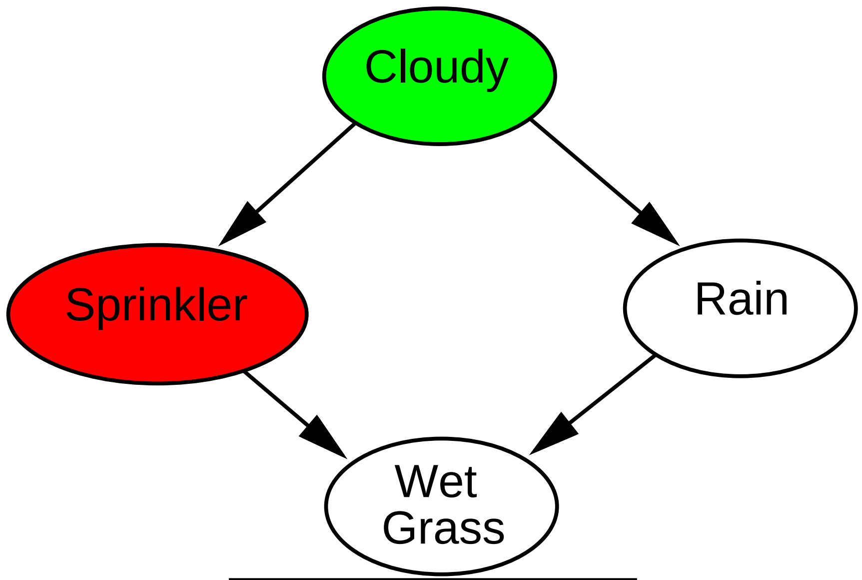
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: find prob. of Wet Grass

P(C)
.50



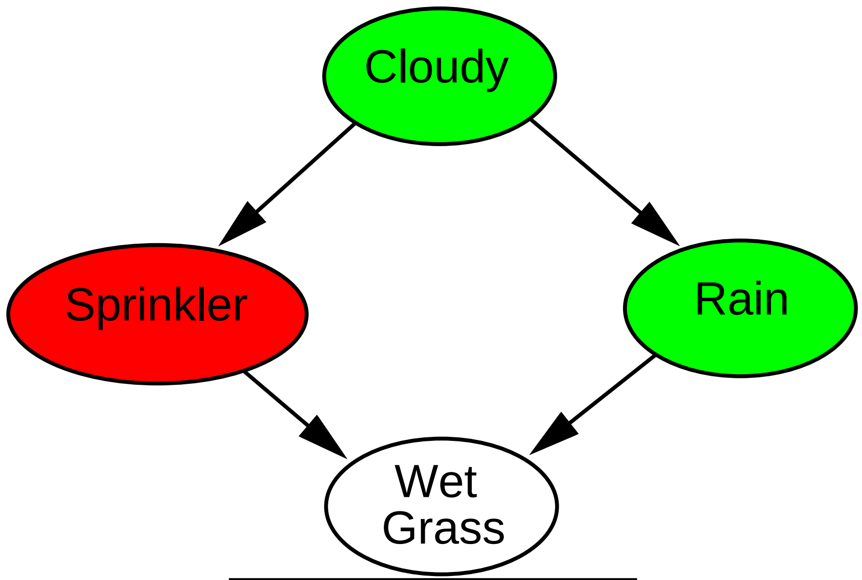
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: find prob. of Wet Grass

P(C)
.50



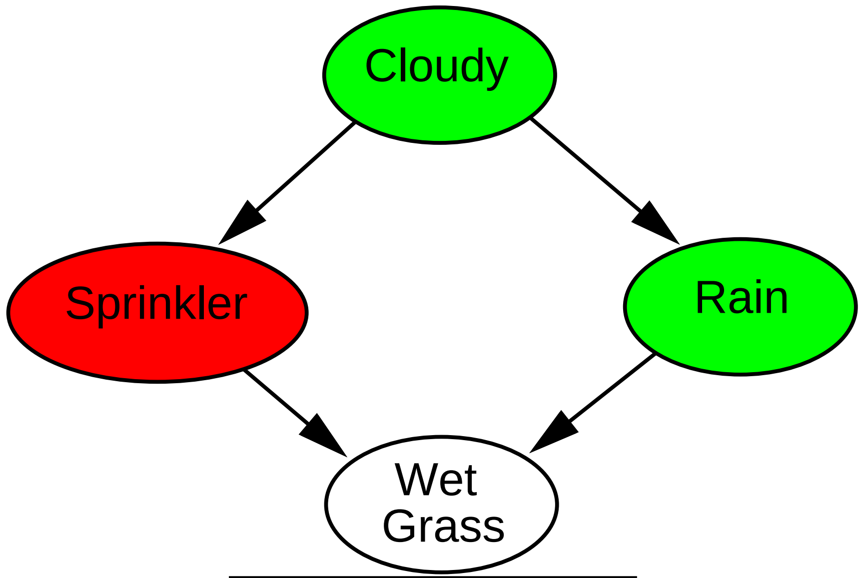
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: find prob. of Wet Grass

P(C)
.50



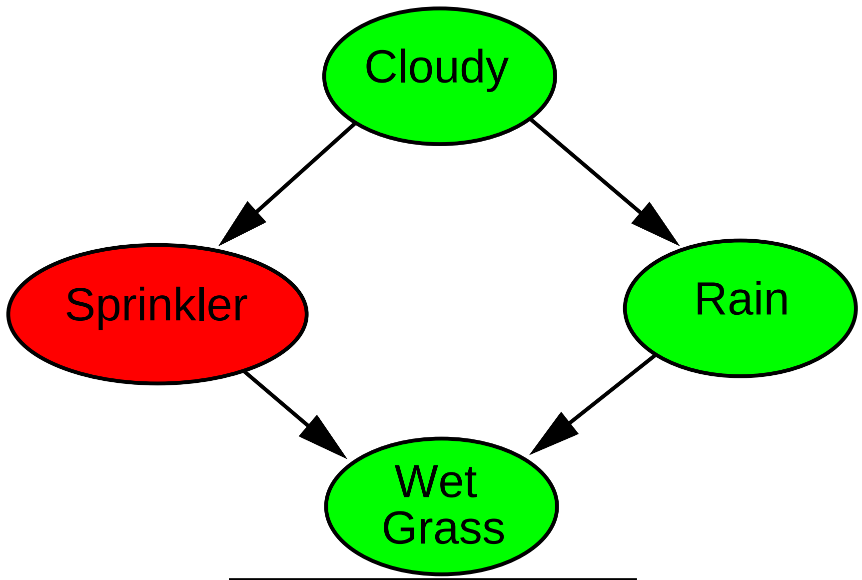
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: find prob. of Wet Grass

P(C)
.50



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Sampling from an empty network, continued

◇ Probability that PriorSample generates a particular set of events

- $S_{PS}(x_1 \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i)) = P(x_1 \ldots x_n)$

◇ i.e., the true prior probability of $x_1, \ldots, x_n$

- E.g., $S_{PS}(t, f, t, t) = 0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324 = P(t, f, t, t)$

◇ Suppose we collect N samples. Let $N_{PS}(x_1 \ldots x_n)$ be the number of samples in which $x_1, \ldots, x_n$ occurred

◇ Then we have

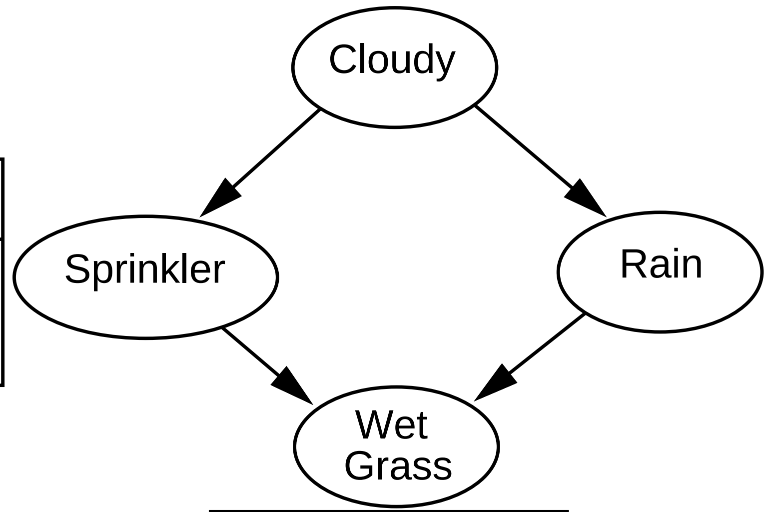
$$\begin{aligned}\lim_{N\to\infty} \hat{P}(x_1, \ldots, x_n) &= \lim_{N\to\infty} N_{PS}(x_1, \ldots, x_n)/N \\ &= S_{PS}(x_1, \ldots, x_n) \\ &= P(x_1 \ldots x_n)\end{aligned}$$

- That is, estimates derived from PriorSample are *consistent*

◇ Shorthand: $\hat{P}(x_1, \ldots, x_n) \approx P(x_1 \ldots x_n)$

Example: prob. of Wet Grass, given Sprinkler

P(C)
.50



C	P(S|C)
T	.10
F	.50

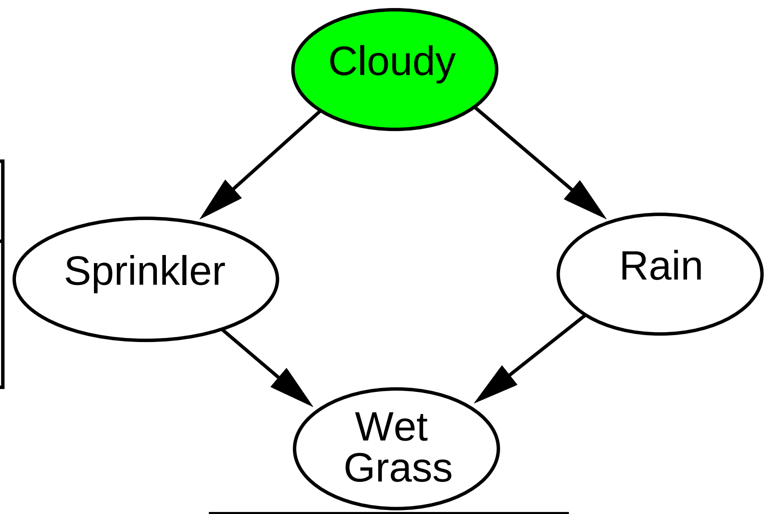
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: prob. of Wet Grass, given Sprinkler

P(C)
.50

C	P(S|C)
T	.10
F	.50



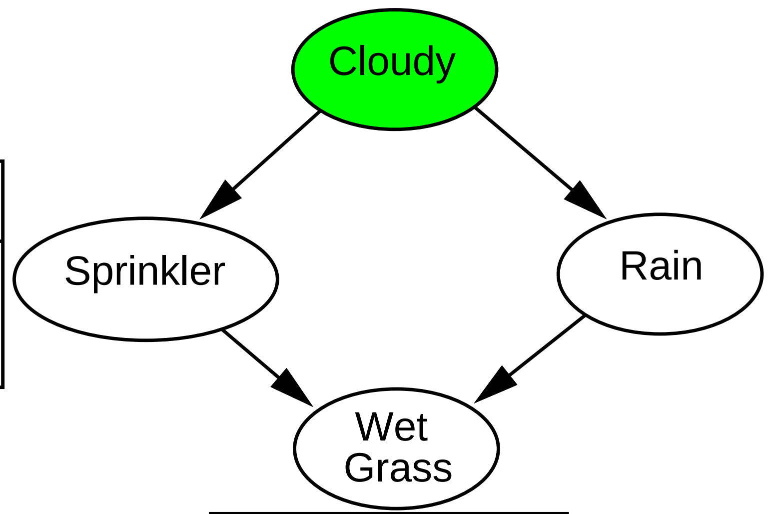
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: prob. of Wet Grass, given Sprinkler

P(C)
.50

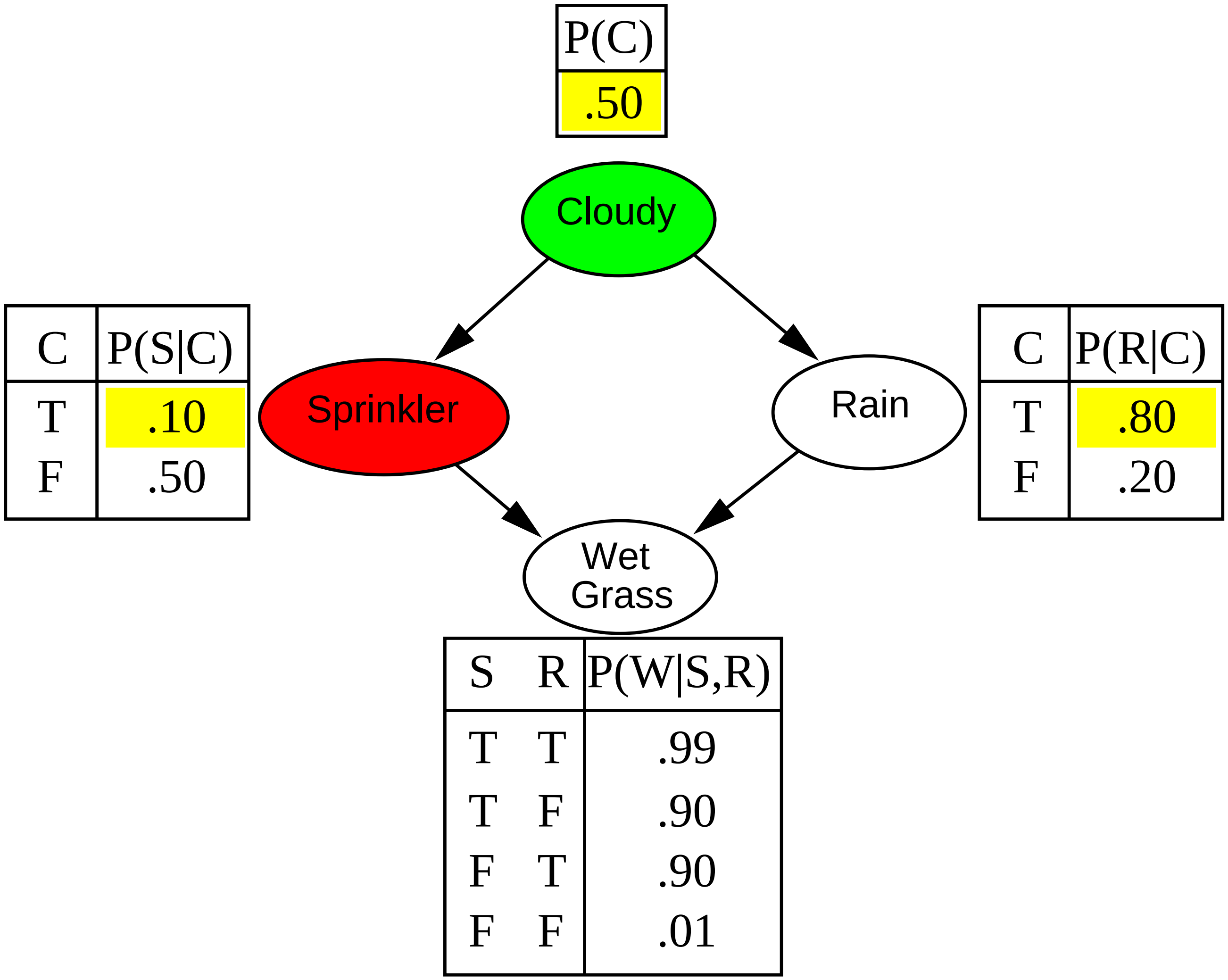
C	P(S|C)
T	.10
F	.50



C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example: prob. of Wet Grass, given Sprinkler



◇ Reject this sample, start running the next one

Rejection sampling

◇ $\hat{\mathbf{P}}(X \mid \mathbf{e})$ estimated from samples agreeing with $\mathbf{e}$

```
function REJECTION-SAMPLING(X, e, bn, N) returns an estimate of P(X|e)
   local variables: N, a vector of counts over X, initially zero

   for j = 1 to N do
       x ← PRIOR-SAMPLE(bn)
       if x is consistent with e then
           N[x] ← N[x]+1 where x is the value of X in x
   return NORMALIZE(N[X])
```

◇ E.g., estimate $\mathbf{P}(Rain \mid Sprinkler = true)$ using 100 samples

- 27 samples have $Sprinkler = true$
- Of these, 8 have $Rain = true$ and 19 have $Rain = false$.
- $\hat{\mathbf{P}}(Rain \mid Sprinkler = true) = \text{NORMALIZE}(\langle 8, 19 \rangle) = \langle 0.296, 0.704 \rangle$

◇ Similar to a basic real-world empirical estimation procedure

Analysis of rejection sampling

$$\begin{aligned}\hat{\mathbf{P}}(X \mid \mathbf{e}) &= \alpha \mathbf{N}_{PS}(X, \mathbf{e}) && \text{(algorithm def.)}\\ &= \mathbf{N}_{PS}(X, \mathbf{e})/N_{PS}(\mathbf{e}) && \text{(normalized by } N_{PS}(\mathbf{e})\text{)}\\ &\approx \mathbf{P}(X, \mathbf{e})/P(\mathbf{e}) && \text{(property of PriorSample)}\\ &= \mathbf{P}(X \mid \mathbf{e}) && \text{(def. of conditional probability)}\end{aligned}$$

◇ Hence rejection sampling returns consistent posterior estimates

◇ Problem:

- if $P(\mathbf{e})$ is small, this is hopelessly expensive to compute:
  - ◇ must reject most of the samples because they disagree with $\mathbf{e}$
- $P(\mathbf{e})$ drops off exponentially with number of evidence variables!

Likelihood weighting

◇ Idea: fix evidence variables, sample only nonevidence variables, and weight each sample by the likelihood it accords the evidence

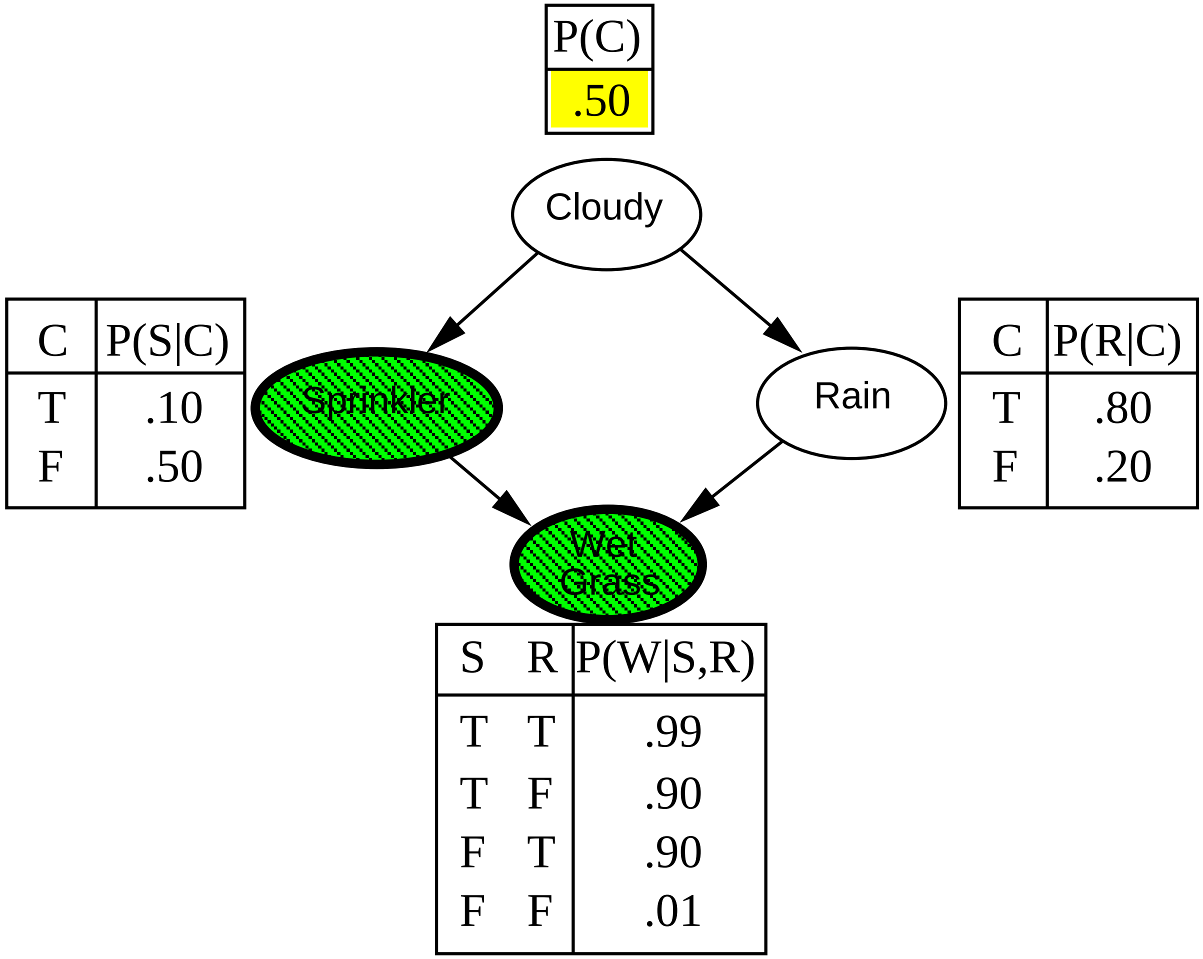
```
function LIKELIHOOD-WEIGHTING(X, e, bn, N) returns an estimate of P(X|e)
   local variables: W, a vector of weighted counts over X, initially zero

   for j = 1 to N do
      x, w ← WEIGHTED-SAMPLE(bn)
      W[x] ← W[x] + w where x is the value of X in x
   return NORMALIZE(W[X])
```

```
function WEIGHTED-SAMPLE(bn, e) returns an event and a weight

   x ← an event with n elements; w ← 1
   for i = 1 to n do
      if X_i has a value x_i in e
         then w ← w × P(X_i = x_i | parents(X_i))
         else x_i ← a random sample from P(X_i | parents(X_i))
   return x, w
```

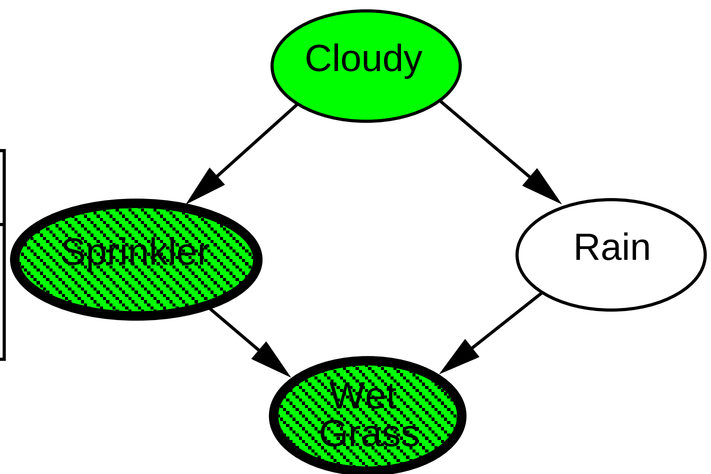
Likelihood weighting example



$w = 1.0$

Likelihood weighting example

P(C)
.50



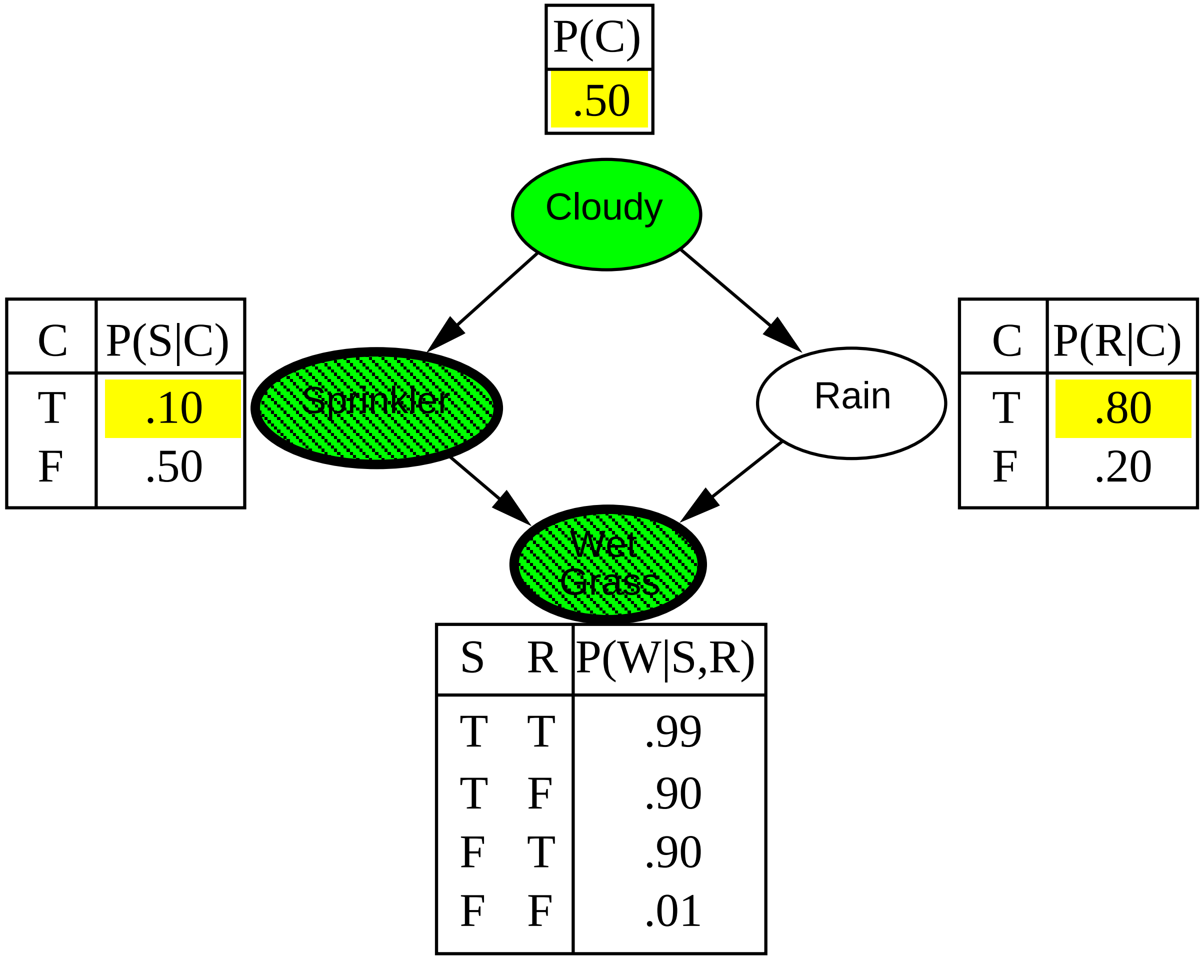
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

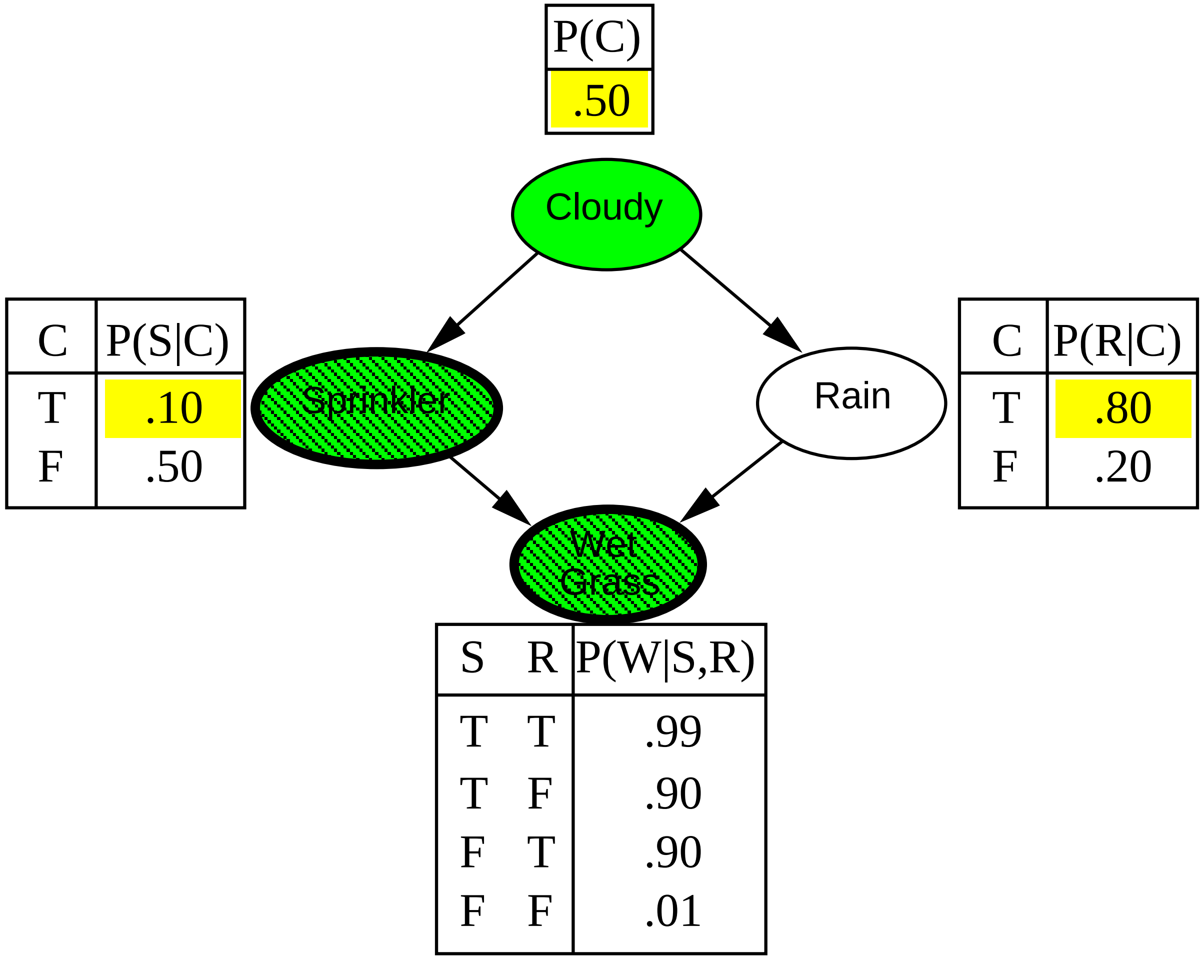
$w = 1.0$

Likelihood weighting example



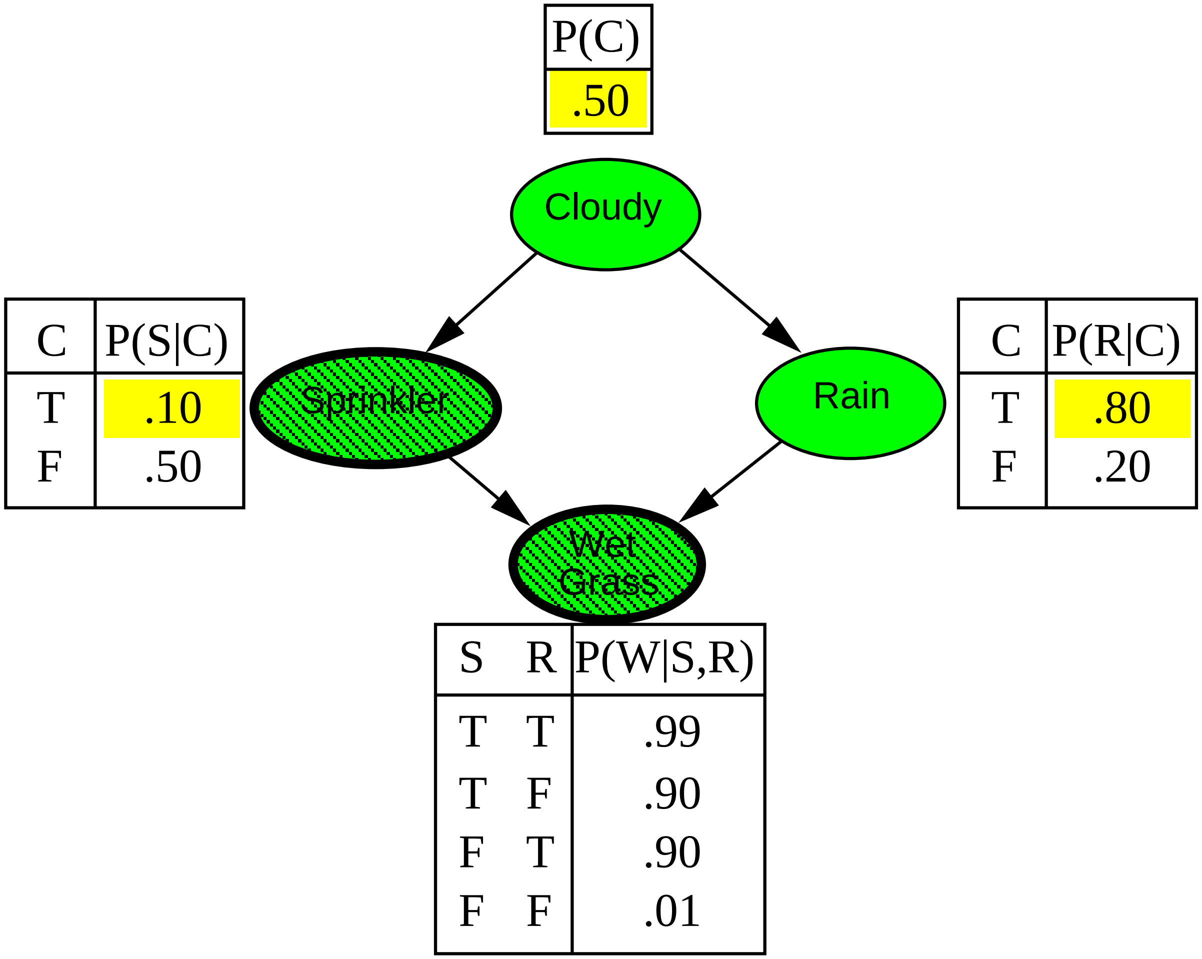
$w = 1.0$

Likelihood weighting example



$w = 1.0 \times 0.1$

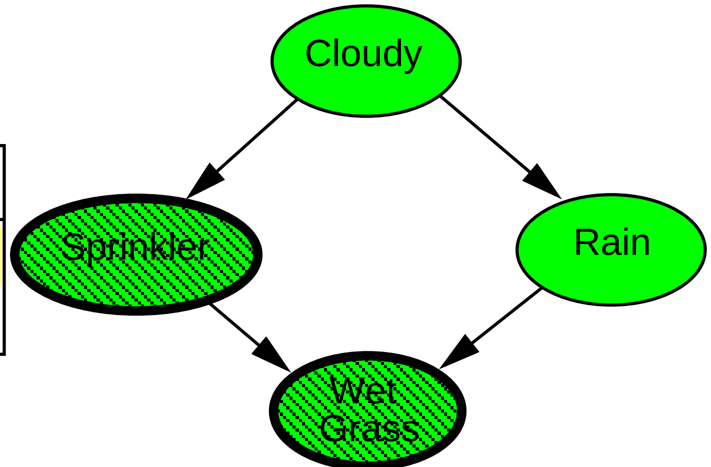
Likelihood weighting example



$w = 1.0 \times 0.1$

Likelihood weighting example

P(C)
.50



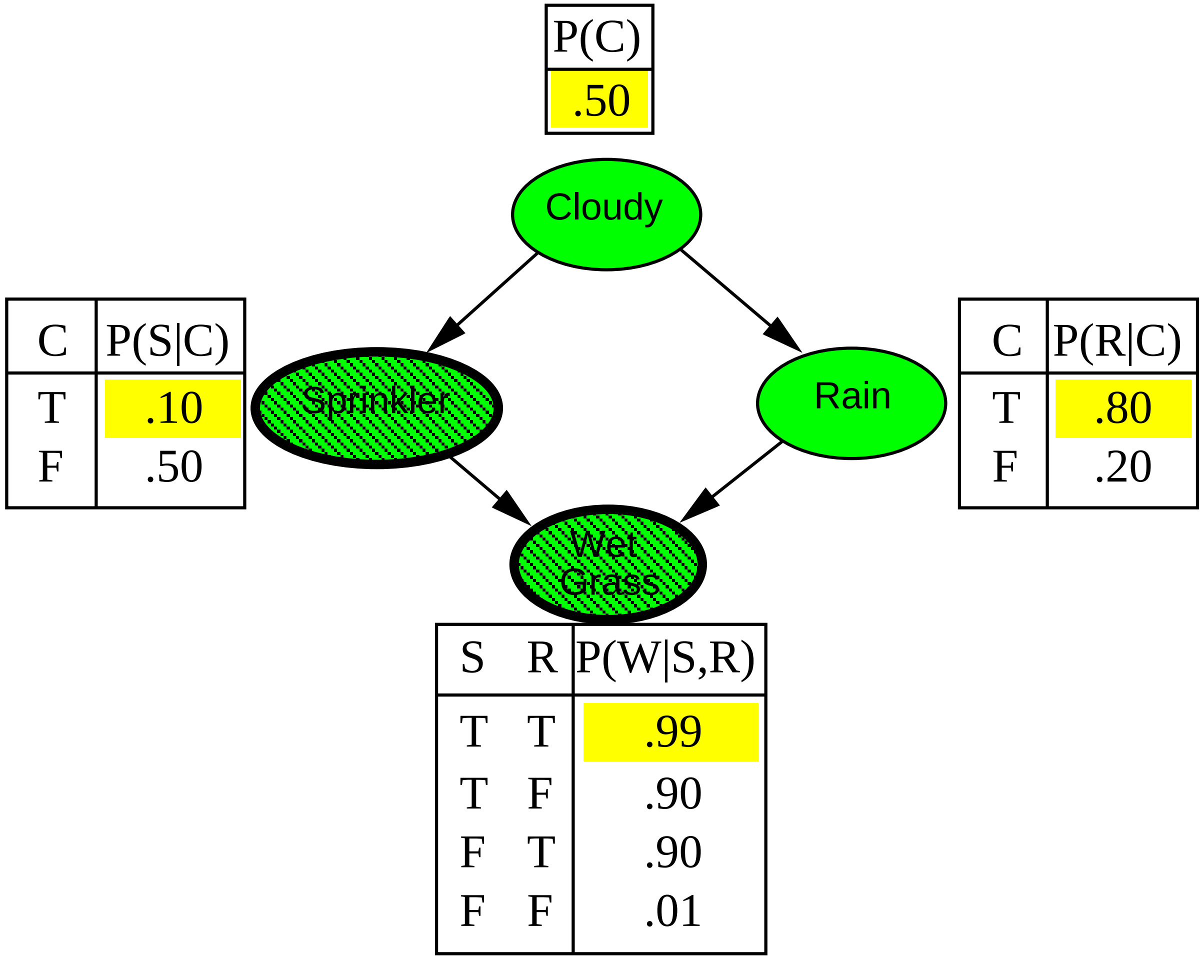
C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$w = 1.0 \times 0.1$

Likelihood weighting example

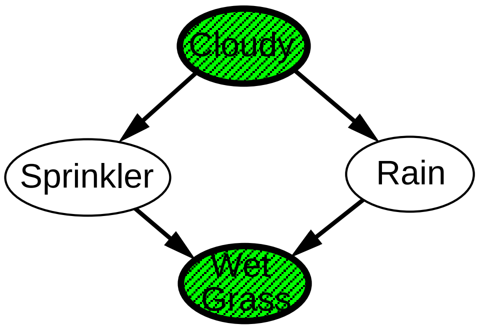


$w = 1.0 \times 0.1 \times 0.99 = 0.099$

Likelihood weighting analysis

◇ Sampling probability for WEIGHTEDSAMPLE is

- $S_{WS}(\mathbf{z}, \mathbf{e}) = \prod_{i=1}^{l} P(z_i \mid parents(Z_i))$
- Note: pays attention to evidence in **ancestors** only $\Rightarrow$ somewhere "in between" prior and posterior distribution



◇ Weight for a given sample $\mathbf{z}, \mathbf{e}$ is

- $w(\mathbf{z}, \mathbf{e}) = \prod_{i=1}^{m} P(e_i \mid parents(E_i))$

◇ Weighted sampling probability is

- $S_{WS}(\mathbf{z}, \mathbf{e}) w(\mathbf{z}, \mathbf{e})$
  $= \prod_{i=1}^{l} P(z_i \mid parents(Z_i)) \;\; \prod_{i=1}^{m} P(e_i \mid parents(E_i))$
  $= P(\mathbf{z}, \mathbf{e})$ (by standard global semantics of network)

◇ Hence likelihood weighting returns consistent estimates

Summary

◇ Bayes nets provide a natural representation for (causally induced) conditional independence

◇ Topology + CPTs = compact representation of joint distribution

◇ Generally easy for (non)experts to construct

◇ Canonical distributions (e.g., noisy-OR) = compact representation of CPTs

◇ Continuous variables $\Rightarrow$ parameterized distributions (e.g., linear Gaussian)

◇ Exact inference by variable elimination:

- polytime on polytrees, NP-hard on general graphs
- space = time, very sensitive to topology

◇ Approximate inference by stochastic simulation

- Convergence can be very slow with probabilities close to 1 or 0

Homework assignment

◇ Here is Homework 6, the last homework assignment of the semester:

- Problems 13.8, 13.17, 13.21, 13.24, and 14.14.
- 10 points each, 50 points total.

◇ Due date: Dec 11

◇ No late date!