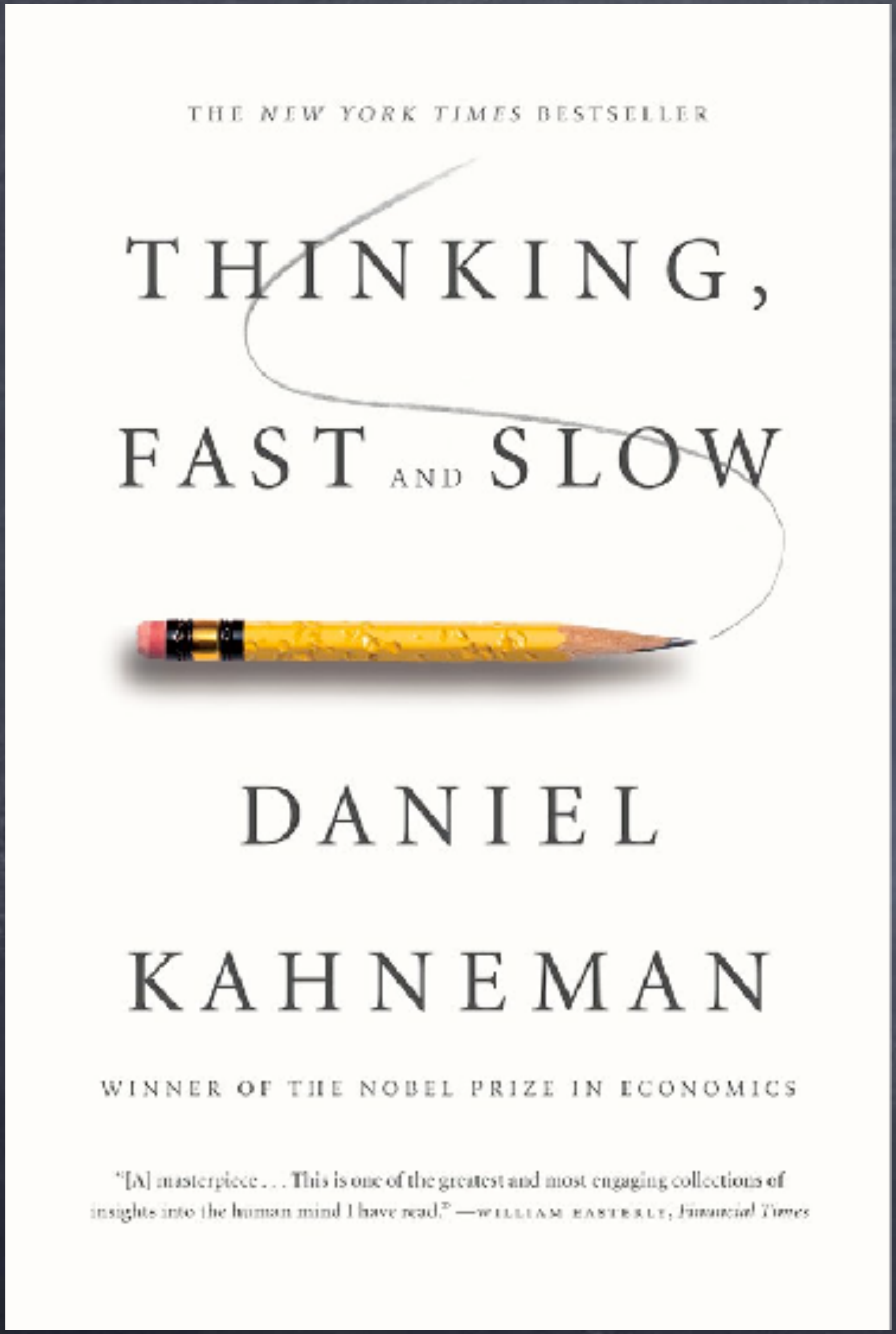
# CSC242: Introduction to Artificial Intelligence

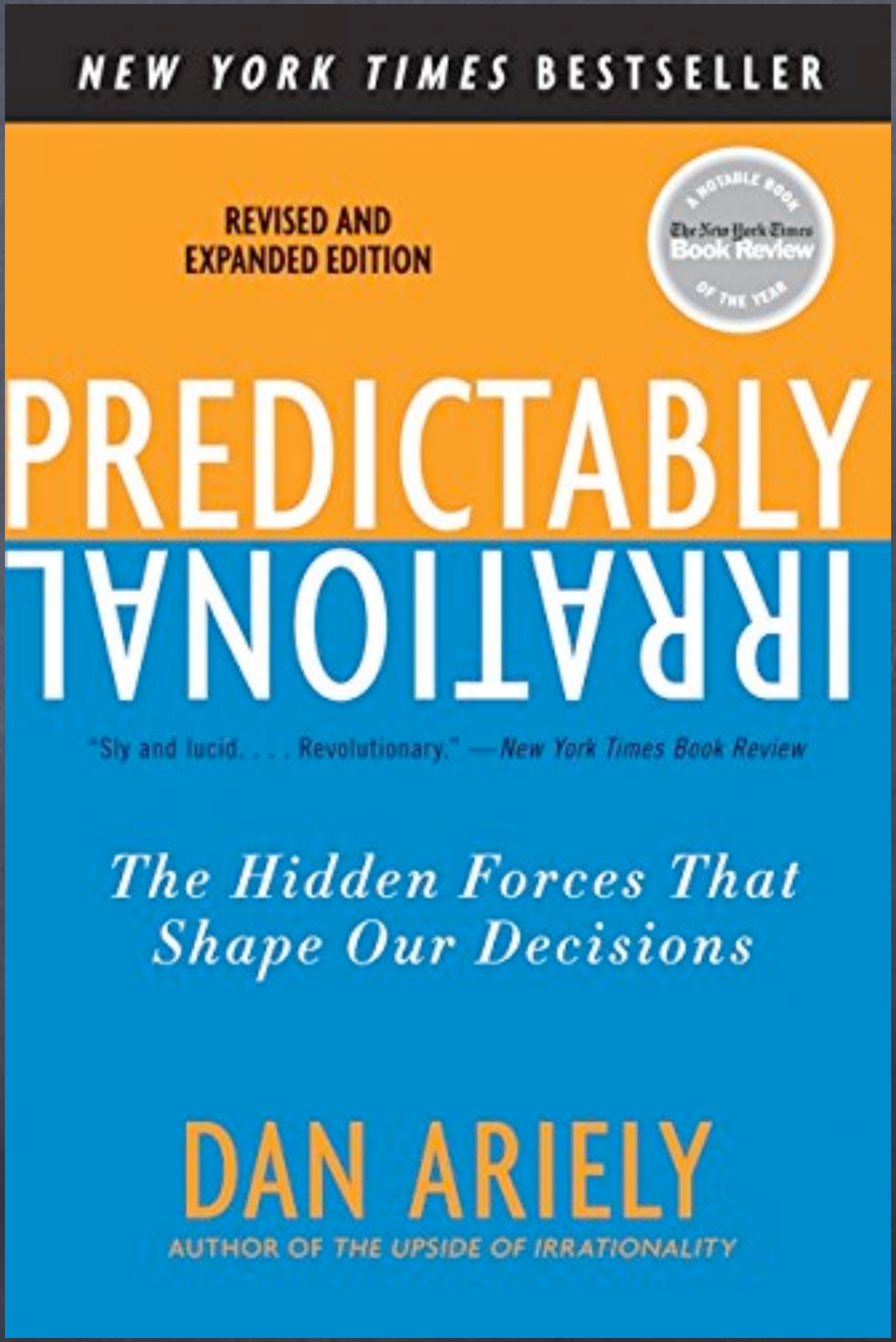
Lecture 3.4

Please put away all electronic devices

# Announcements

- Exam 3 One week from today
- Project 3 due same day 1159PM
  - Don't wait to be finished


THE NEW YORK TIMES BESTSELLER
THINKING,
FAST AND SLOW
DANIEL
KAHNEMAN
WINNER OF THE NOBEL PRIZE IN ECONOMICS
"[A] masterpiece . . . This is one of the greatest and most engaging collections of insights into the human mind I have read." —WILLIAM EASTERLY, Financial Times


NEW YORK TIMES BESTSELLER
REVISED AND
EXPANDED EDITION
A NOTABLE BOOK
The New York Times
Book Review
OF THE YEAR
PREDICTABLY
IRRATIONAL
"Sly and lucid. . . . Revolutionary." —New York Times Book Review
The Hidden Forces That
Shape Our Decisions
DAN ARIELY
AUTHOR OF THE UPSIDE OF IRRATIONALITY

# Bayesian Networks

$\mathbf{P}(Cavity)$

| $cavity$ | $\neg cavity$ |
|---|---|
| | |

$$
\begin{tikzcd}
 & Cavity \arrow[dl] \arrow[dr] & \\
Toothache & & Catch
\end{tikzcd}
$$

$\mathbf{P}(Toothache|Cavity)$

| $Cavity$ | $toothache$ | $\neg toothache$ |
|---|---|---|
| $cavity$ | | |
| $\neg cavity$ | | |

$\mathbf{P}(Catch|Cavity)$

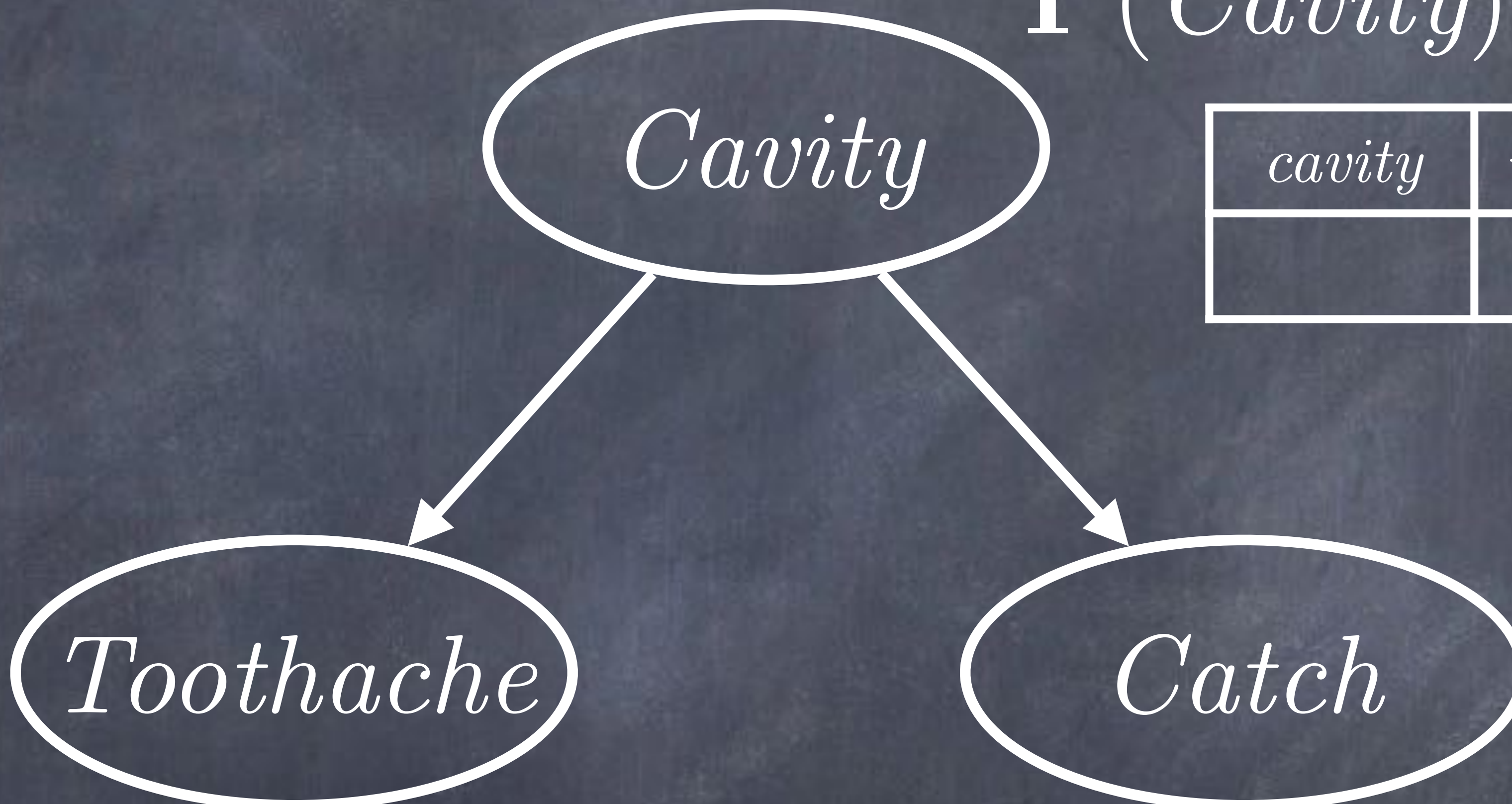
| $Cavity$ | $catch$ | $\neg catch$ |
|---|---|---|
| $cavity$ | | |
| $\neg cavity$ | | |

$$P(toothache, cavity, catch) =$$

$$P(toothache|cavity)P(catch|cavity)P(cavity)$$

$\mathbf{P}(Cavity)$

| $cavity$ | $\neg cavity$ |
|---|---|
| | |




$\mathbf{P}(Toothache|Cavity)$

| $Cavity$ | $toothache$ | $\neg toothache$ |
|---|---|---|
| $cavity$ | | |
| $\neg cavity$ | | |

$\mathbf{P}(Catch|Cavity)$

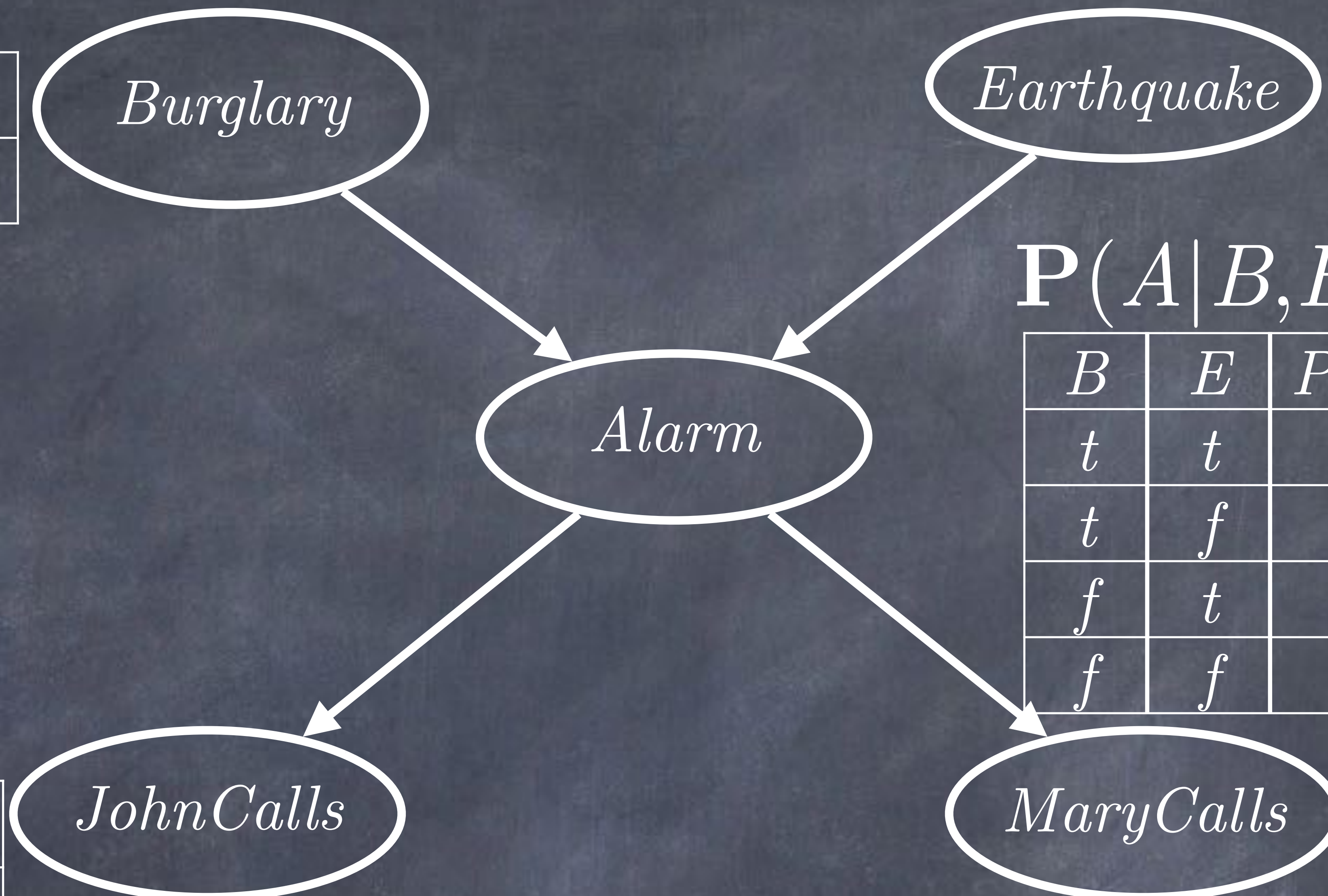
| $Cavity$ | $catch$ | $\neg catch$ |
|---|---|---|
| $cavity$ | | |
| $\neg cavity$ | | |

$\mathbf{P}(B)$

| $P(b)$ |
|---|
| 0.001 |

$\mathbf{P}(E)$

| $P(e)$ |
|---|
| 0.002 |




$\mathbf{P}(A|B,E)$

| $B$ | $E$ | $P(a|B,E)$ |
|---|---|---|
| $t$ | $t$ | 0.95 |
| $t$ | $f$ | 0.94 |
| $f$ | $t$ | 0.29 |
| $f$ | $f$ | 0.001 |

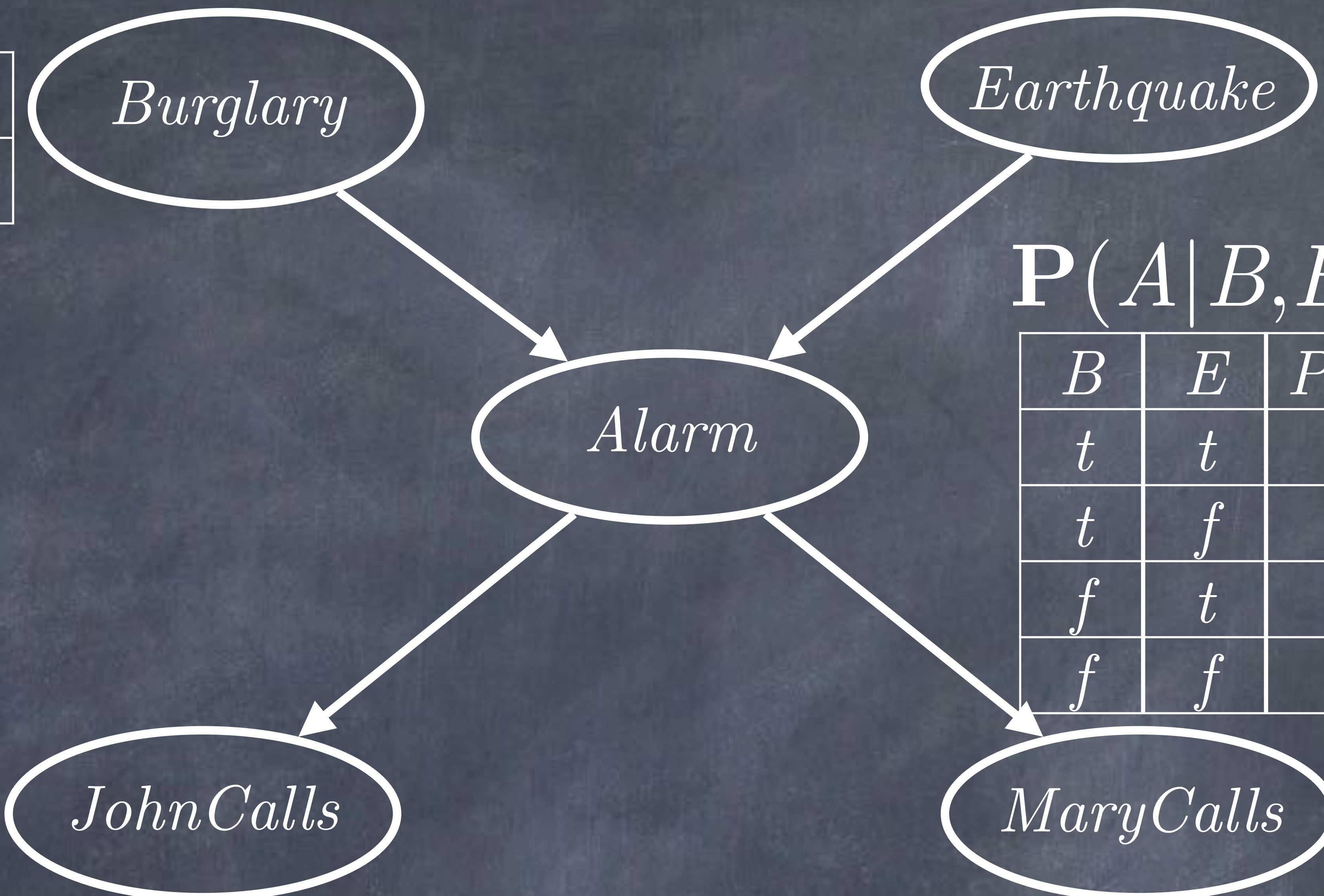
$\mathbf{P}(J|A)$

| $A$ | $P(j|A)$ |
|---|---|
| $t$ | 0.9 |
| $f$ | 0.05 |

$\mathbf{P}(M|A)$

| $A$ | $P(m|A)$ |
|---|---|
| $t$ | 0.7 |
| $f$ | 0.01 |

$$\mathbf{P}(B, E, A, J, M) = \alpha\, \mathbf{P}(B)\, \mathbf{P}(E)\, \mathbf{P}(A \mid B, E)\, \mathbf{P}(J \mid A)\, \mathbf{P}(M \mid A)$$




$\mathbf{P}(B)$

| $P(b)$ |
|---|
| 0.001 |

$\mathbf{P}(E)$

| $P(e)$ |
|---|
| 0.002 |

$\mathbf{P}(A|B,E)$

| $B$ | $E$ | $P(a|B,E)$ |
|---|---|---|
| $t$ | $t$ | 0.95 |
| $t$ | $f$ | 0.94 |
| $f$ | $t$ | 0.29 |
| $f$ | $f$ | 0.001 |

$\mathbf{P}(J|A)$

| $A$ | $P(j|A)$ |
|---|---|
| $t$ | 0.9 |
| $f$ | 0.05 |

$\mathbf{P}(M|A)$

| $A$ | $P(m|A)$ |
|---|---|
| $t$ | 0.7 |
| $f$ | 0.01 |

Full joint distribution: $2^5$=32 entries

Bayesian network: 10 entries

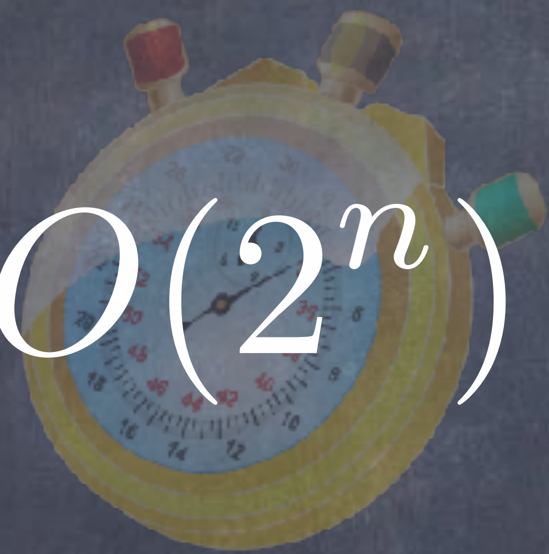
Assuming conditional independences encoded in the network

# Inference in Bayesian Networks

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$
$$= \alpha \sum_{\mathbf{y}} \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

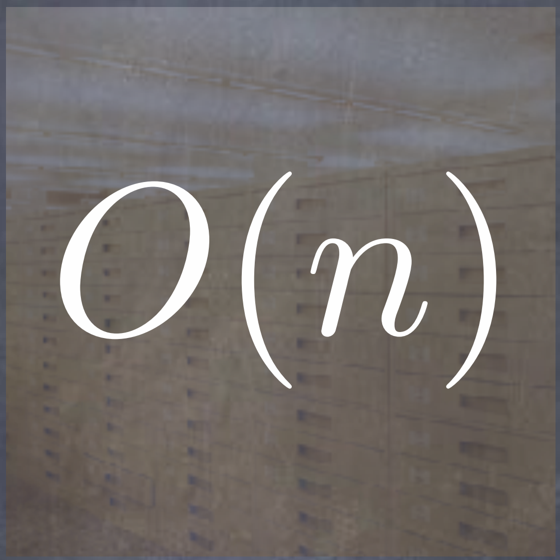
- "A query can be answered from a Bayesian Network by computing sums of products of conditional probabilities from the network."

# Exact Inference in BNs



$O(2^n)$

Time Complexity



$O(n)$

Space Complexity

# Approximate Inference in Bayesian Networks

# Exact Inference

- Query variable $X$
- Evidence variables $E_1, \ldots, E_m$
  - Observed values: $\boldsymbol{e} = \langle e_1 , \ldots, e_m \rangle$
- Non-evidence, non-query ("hidden") variables: $\boldsymbol{Y}$
- Compute: $\mathbf{P}(X \mid \mathbf{e})$

# Approximate Inference

- Query variable $X$
- Evidence variables $E_1, ..., E_m$
  - Observed values: $\boldsymbol{e} = \langle e_1 , ..., e_m \rangle$
- Non-evidence, non-query ("hidden") variables: $Y$
- Approximate (estimate): $\mathbf{P}(X \mid \mathbf{e})$

# Unconditional Approximation

- Query variable $X$
- Evidence variables $E_1, ..., E_m$
  - Observed values: $\boldsymbol{e} = \langle e_1 , ..., e_m \rangle$
- Non-evidence, non-query ("hidden") variables: $Y$
- Approximate (estimate): $\mathbf{P}(X)$

*Heads*

$Heads$

Goal: $\mathbf{P}(Heads)$

$\mathrm{P}(Heads{=}true)$

$\mathrm{P}(heads)$

$Heads$

# of flips: $N$

# of heads: $N_{heads}$

$$Heads$$

$$P(heads) \approx \frac{N_{heads}}{N}$$

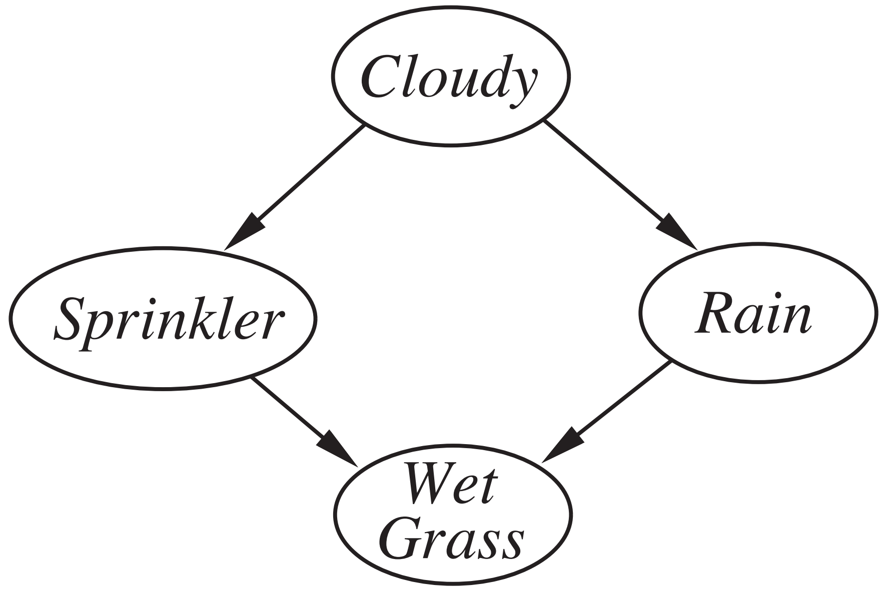
$$Heads$$

$$P(heads) = \lim_{N \to \infty} \frac{N_{heads}}{N}$$

# Sampling

- Generating events (possible worlds) from a distribution
- Estimating probabilities as ratio of observed events to total events
  - Consistent estimate: becomes exact in the large-sample limit

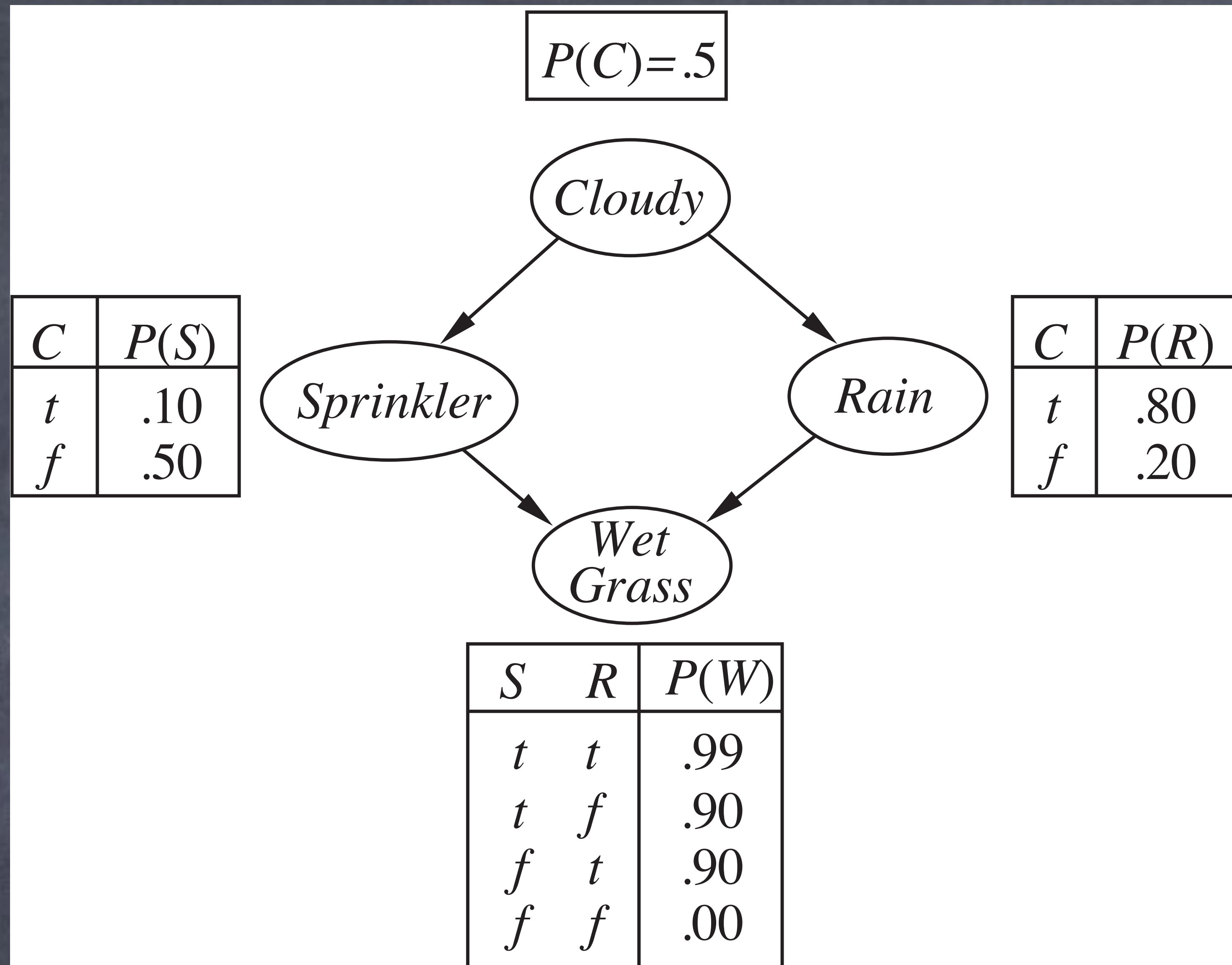
$P(C) = .5$




| $C$ | $P(S)$ |
|---|---|
| $t$ | .10 |
| $f$ | .50 |

| $C$ | $P(R)$ |
|---|---|
| $t$ | .80 |
| $f$ | .20 |

| $S$ | $R$ | $P(W)$ |
|---|---|---|
| $t$ | $t$ | .99 |
| $t$ | $f$ | .90 |
| $f$ | $t$ | .90 |
| $f$ | $f$ | .00 |




$P(C)=.5$

| $C$ | $P(S)$ |
|---|---|
| $t$ | .10 |
| $f$ | .50 |

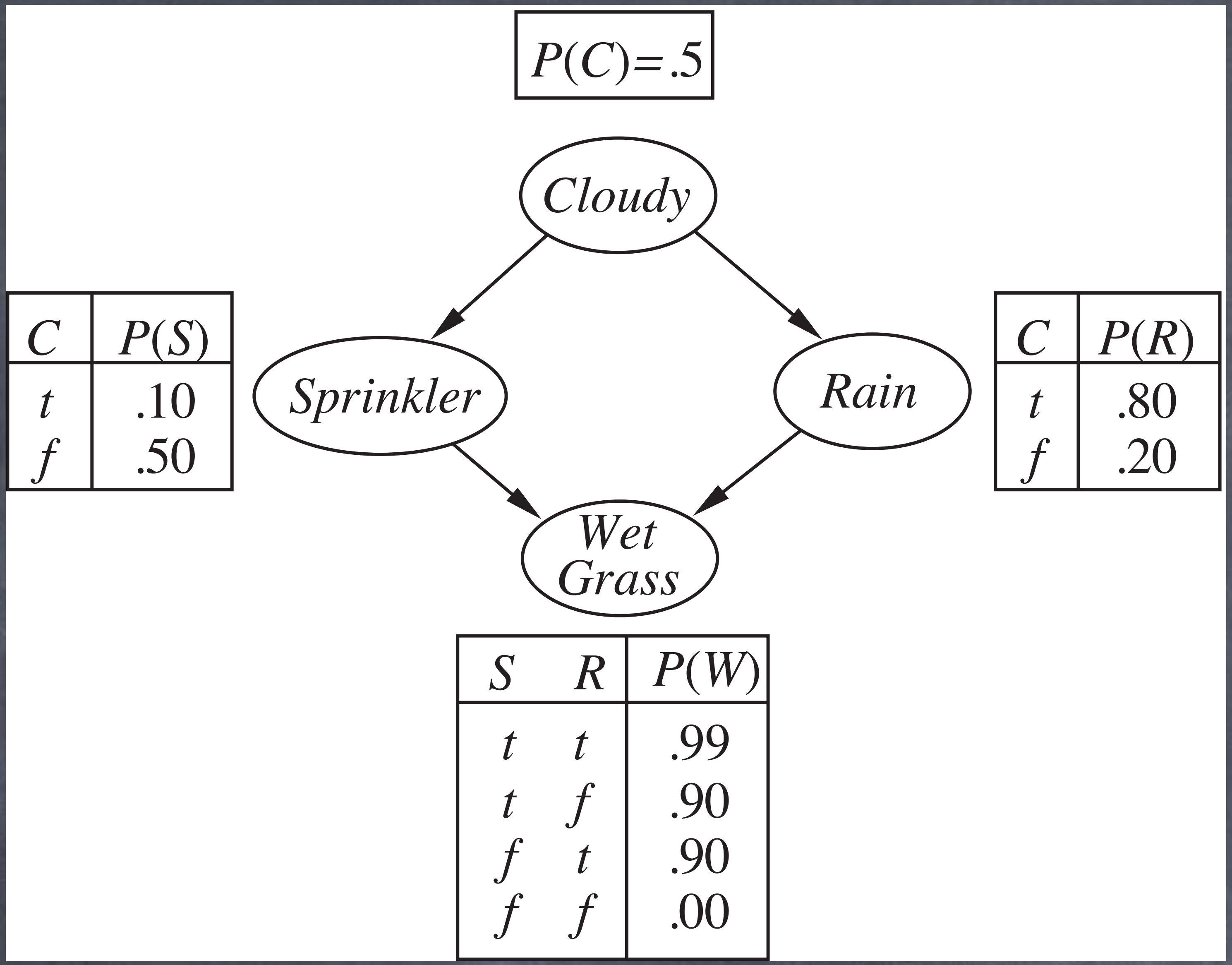
| $C$ | $P(R)$ |
|---|---|
| $t$ | .80 |
| $f$ | .20 |

| $S$ | $R$ | $P(W)$ |
|---|---|---|
| $t$ | $t$ | .99 |
| $t$ | $f$ | .90 |
| $f$ | $t$ | .90 |
| $f$ | $f$ | .00 |

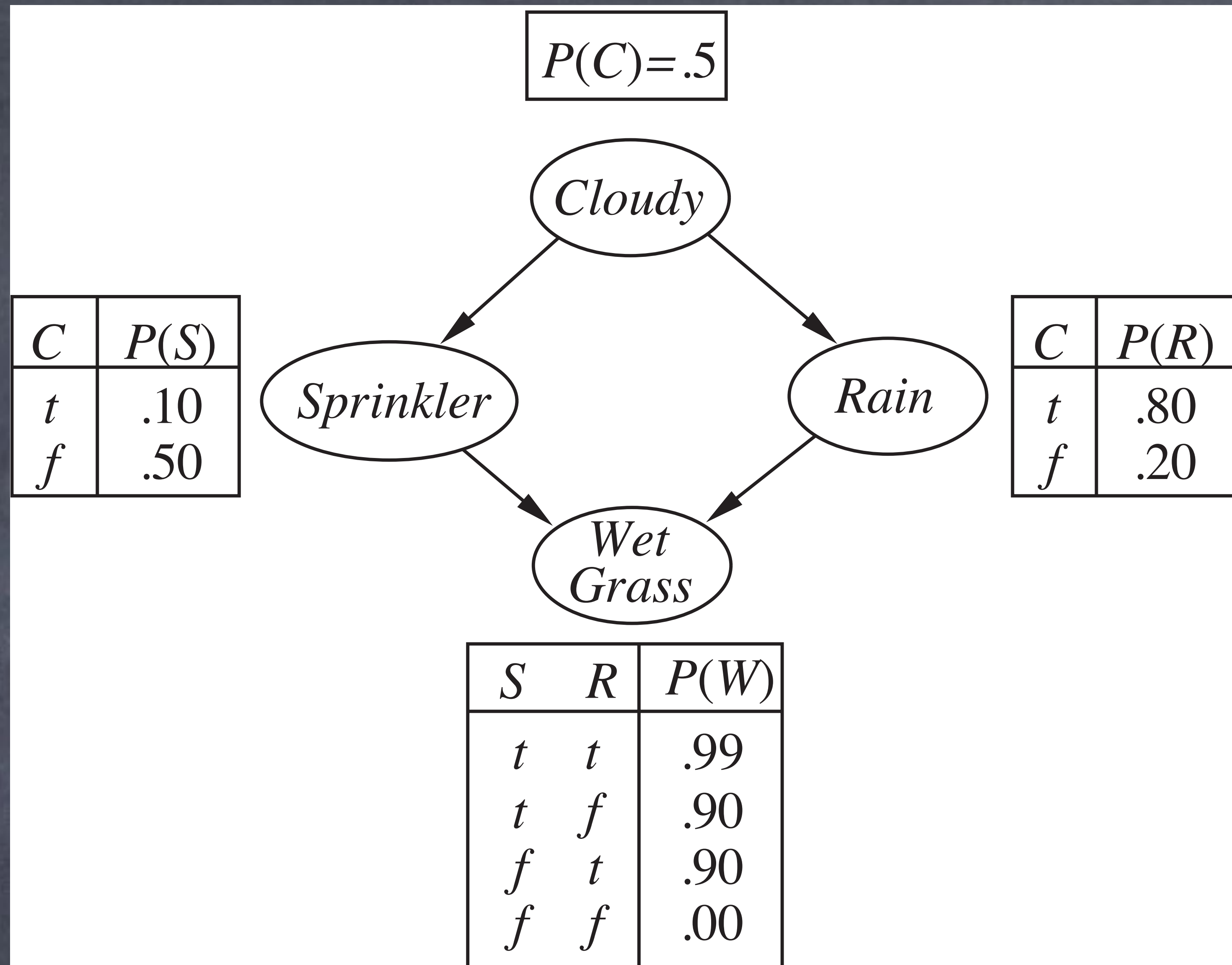
$$P(Rain = true)$$

# Sampling

- Generate assignments of values to the random variables
- That are consistent with the full joint distribution encoded in the network
  - In the sense that in the limit, the frequency of occurrence of any event (possible world) is equal to its probability




$$\langle Cloudy{=}true,\ Sprinkler{=}false,\ Rain{=}true,\ WetGrass{=}true\rangle$$




| $C$ | $P(S)$ |
|---|---|
| $t$ | .10 |
| $f$ | .50 |

| $C$ | $P(R)$ |
|---|---|
| $t$ | .80 |
| $f$ | .20 |

| $S$ | $R$ | $P(W)$ |
|---|---|---|
| $t$ | $t$ | .99 |
| $t$ | $f$ | .90 |
| $f$ | $t$ | .90 |
| $f$ | $f$ | .00 |

$$\langle Cloudy{=}true,\ Sprinkler{=}false,\ Rain{=}true,\ WetGrass{=}true\rangle$$

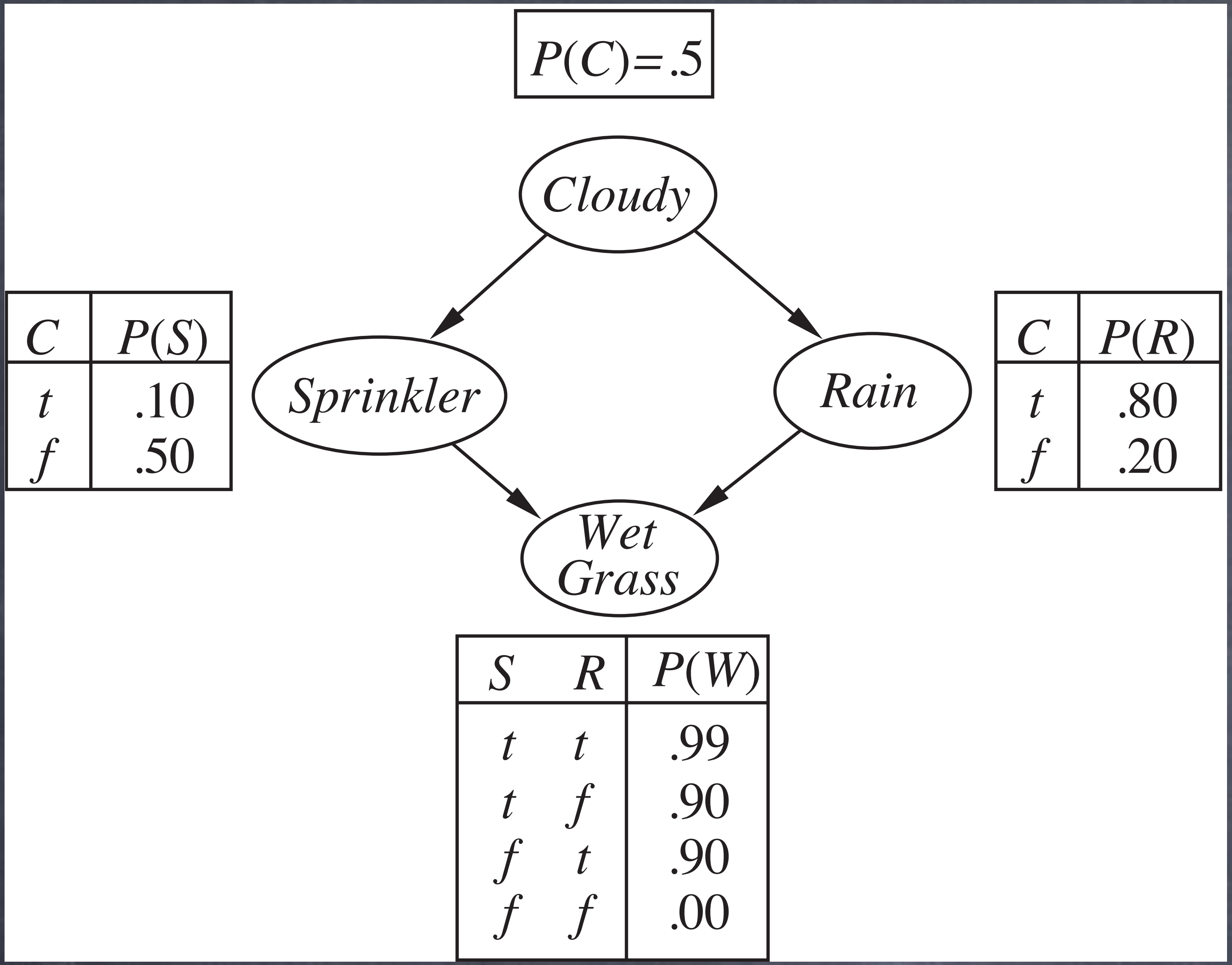
$$\langle Cloudy{=}false,\ Sprinkler{=}false,\ Rain{=}false,\ WetGrass{=}false\rangle$$

$$\langle Cloudy{=}true,\ Sprinkler{=}true,\ Rain{=}true,\ WetGrass{=}true\rangle$$

...

# Generating Samples

- Sample each variable in topological order
  - Child appears after its parents
- Choose the value for that variable conditioned on the values already chosen for its parents

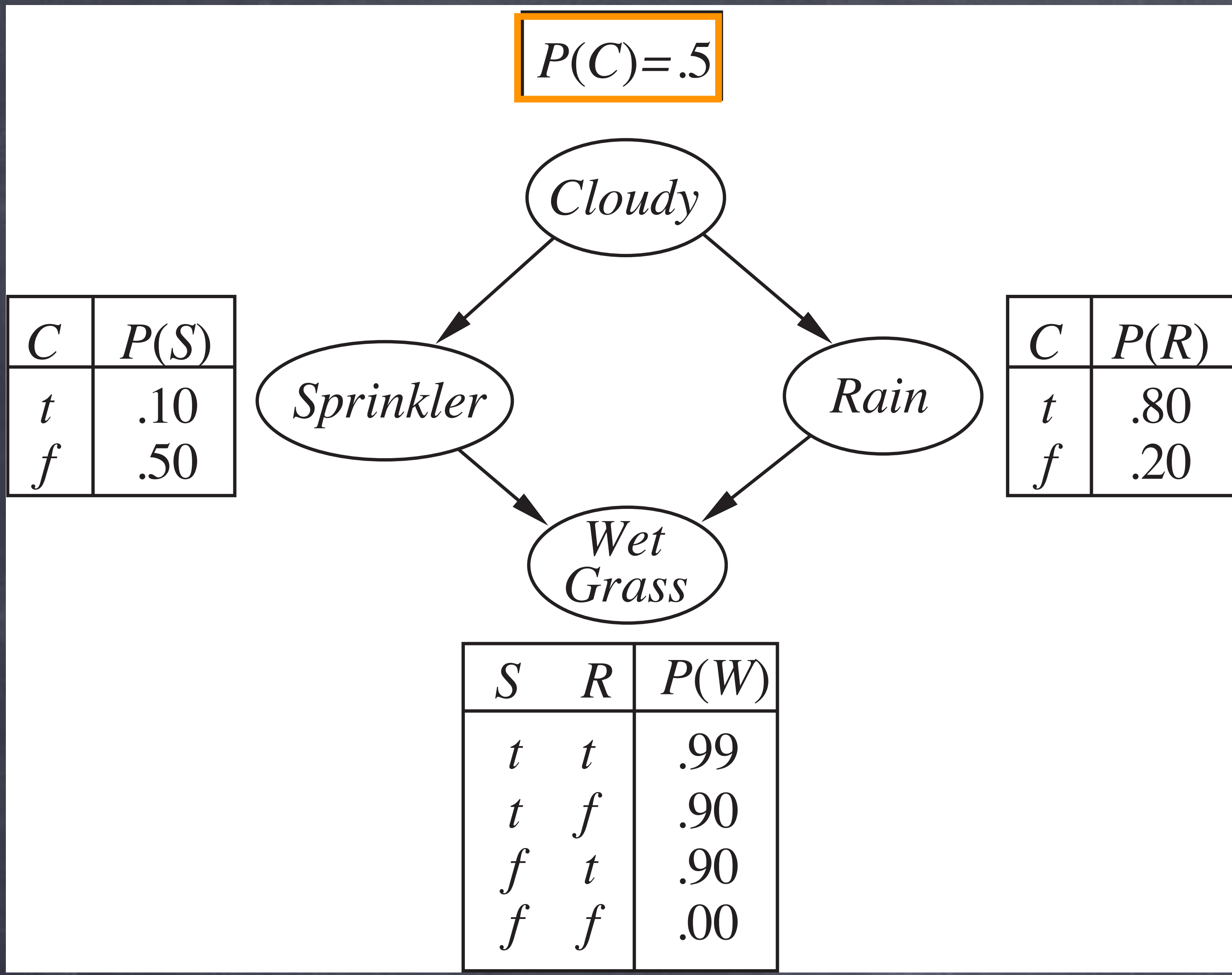



*Cloudy*

*Sprinkler*

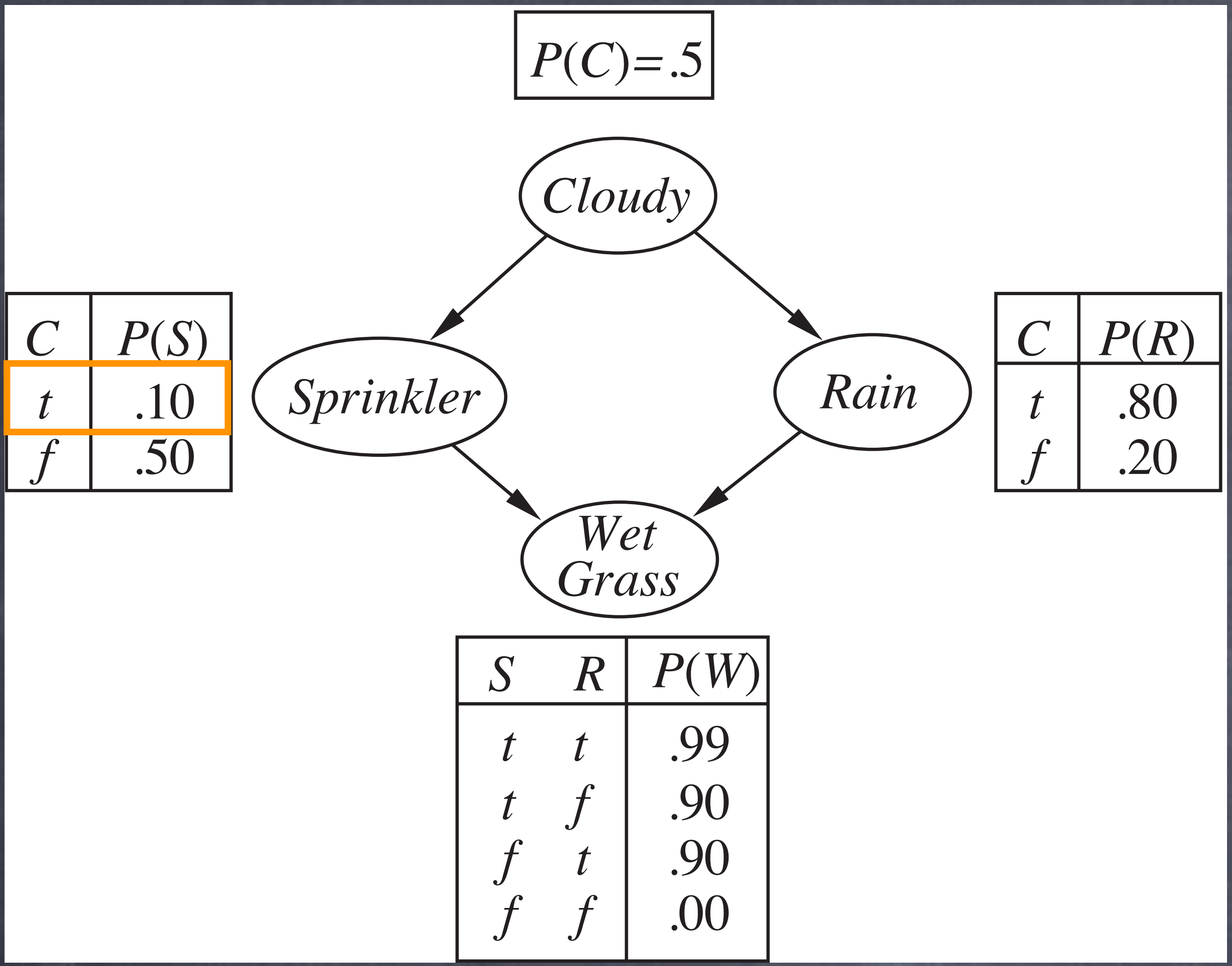
*Rain*

*WetGrass*



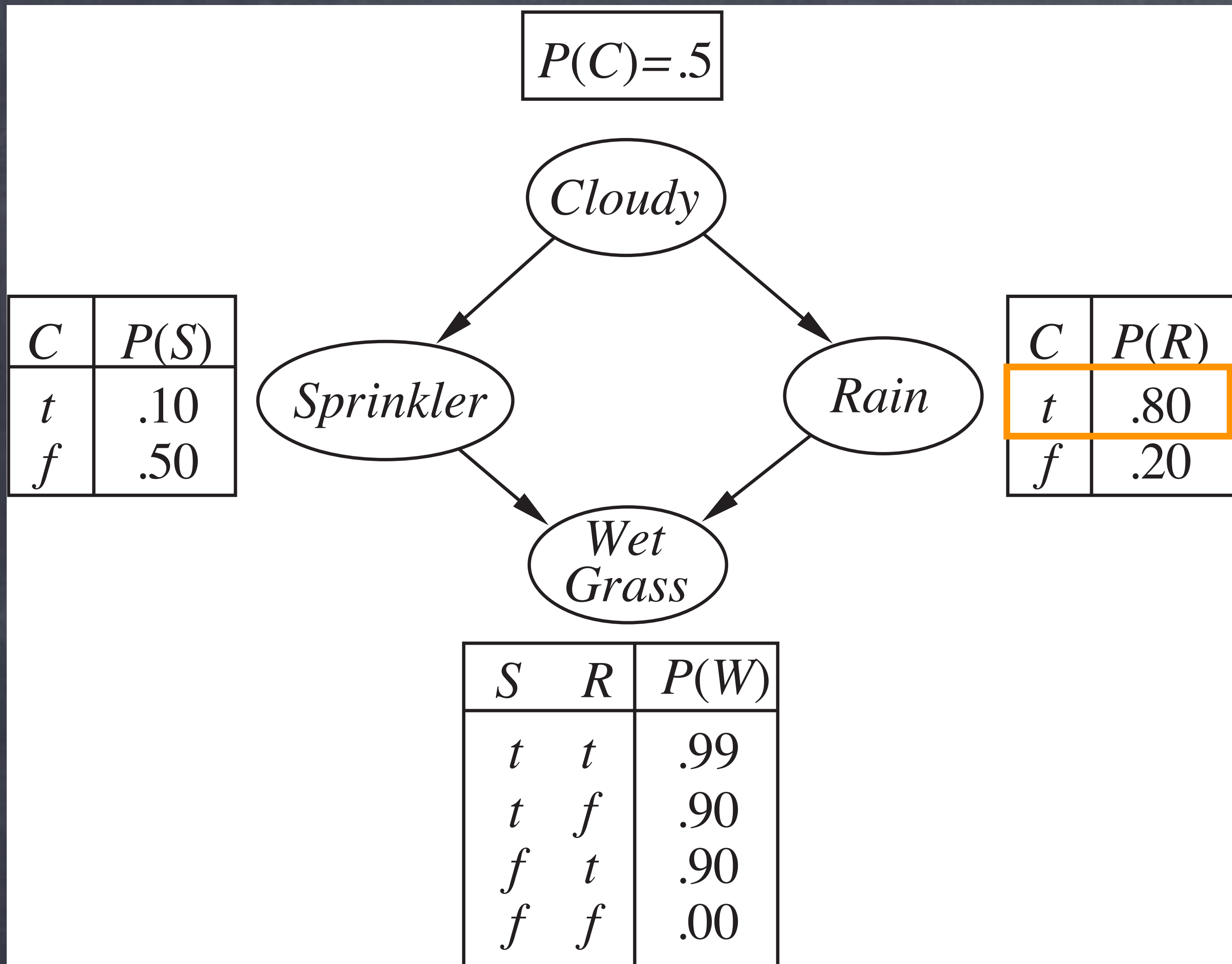

| | |
|---|---|
| *Cloudy* | *true* |
| *Sprinkler* | |
| *Rain* | |
| *WetGrass* | |

$$\mathbf{P}(Cloudy) = \langle 0.5, 0.5 \rangle$$



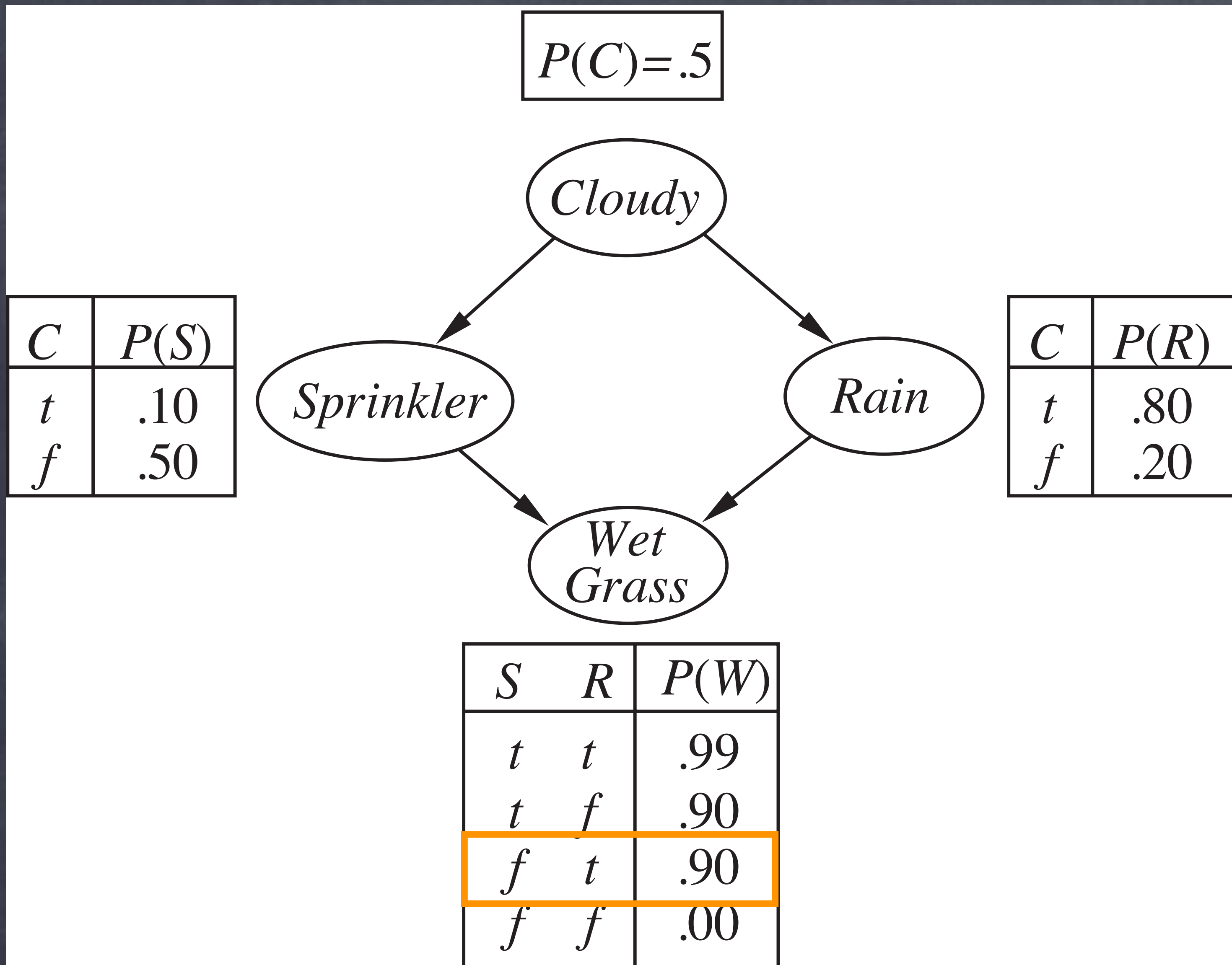

| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $false$ |
| $Rain$ | |
| $WetGrass$ | |

$$\mathbf{P}(Sprinkler \mid Cloudy = true) = \langle 0.1, 0.9 \rangle$$



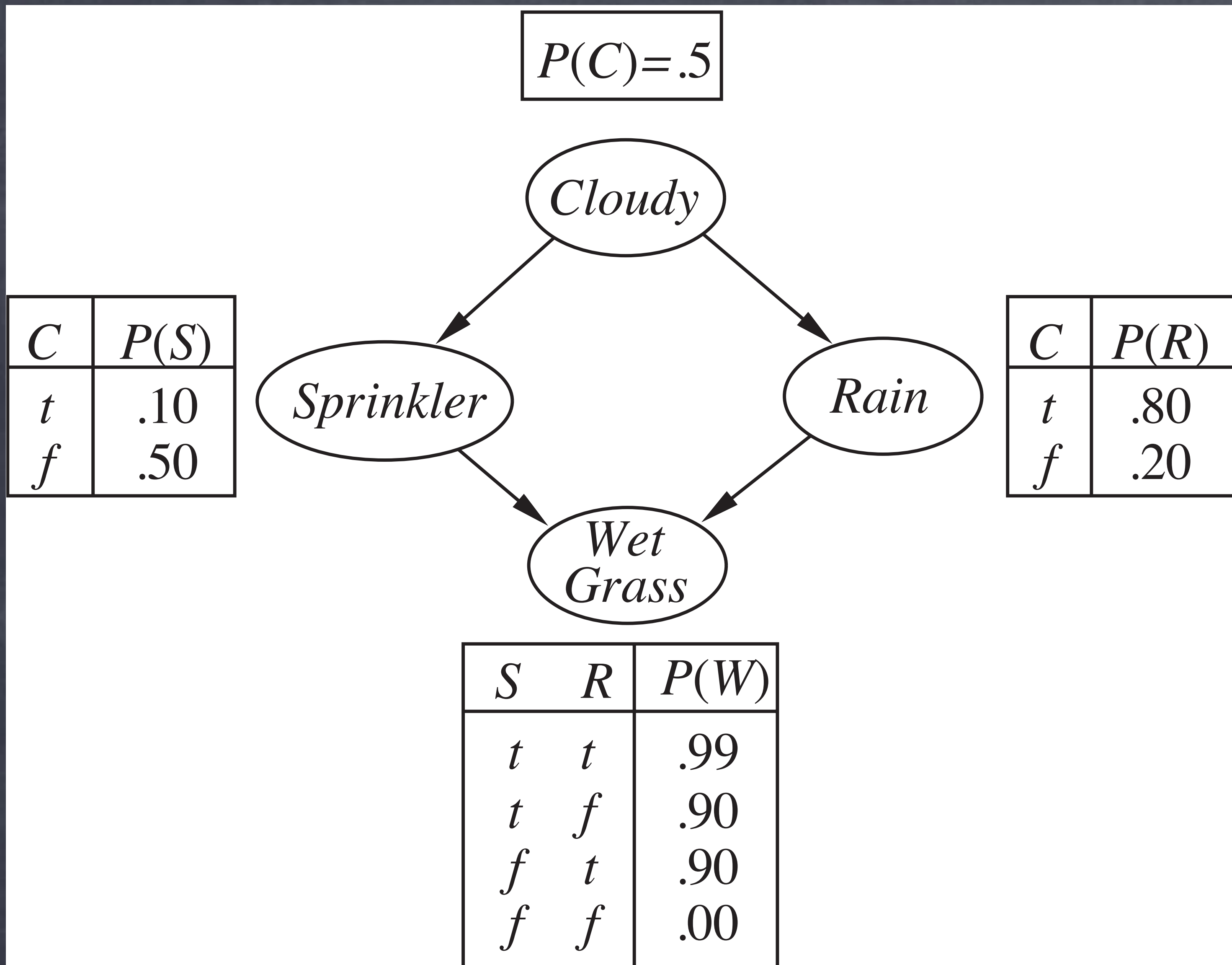

| | |
|---|---|
| *Cloudy* | *true* |
| *Sprinkler* | *false* |
| *Rain* | *true* |
| *WetGrass* | |

$$\mathbf{P}(Rain \mid Cloudy = true) = \langle 0.8, 0.2 \rangle$$




| | |
|---|---|
| *Cloudy* | *true* |
| *Sprinkler* | *false* |
| *Rain* | *true* |
| *WetGrass* | *true* |

$$\mathbf{P}(\,WetGrass \mid Sprinkler = false, Rain = true) = \langle 0.9, 0.1 \rangle$$




| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $false$ |
| $Rain$ | $true$ |
| $WetGrass$ | $true$ |

$$\langle Cloudy{=}true,\ Sprinkler{=}false,\ Rain{=}true,\ WetGrass{=}true\rangle$$

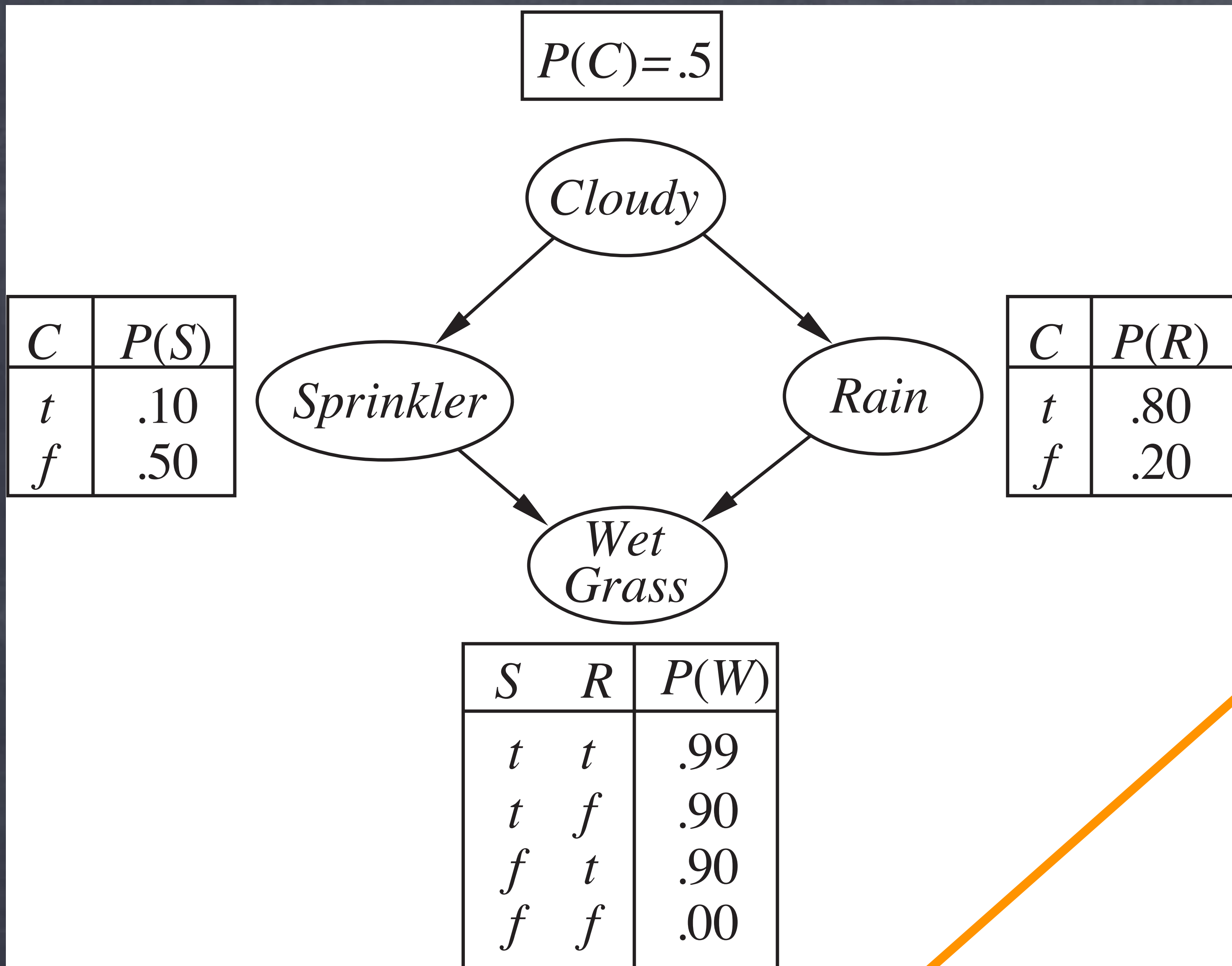
Guaranteed to be a consistent estimate (becomes exact in the large-sample limit)

# Sampling

- Generate assignments of values to the random variables
- That are consistent with the full joint distribution encoded in the network
  - In the sense that in the limit, the frequency of occurrence of any event (possible world) is equal to its probability

# Approximate Inference

- Query variable $X$
- Evidence variables $E_1, ..., E_m$
  - Observed values: $\boldsymbol{e} = \langle e_1 , ..., e_m \rangle$
- Non-evidence, non-query (“hidden”) variables: $Y$
- Approximate (estimate): $\mathbf{P}(X \mid \mathbf{e})$

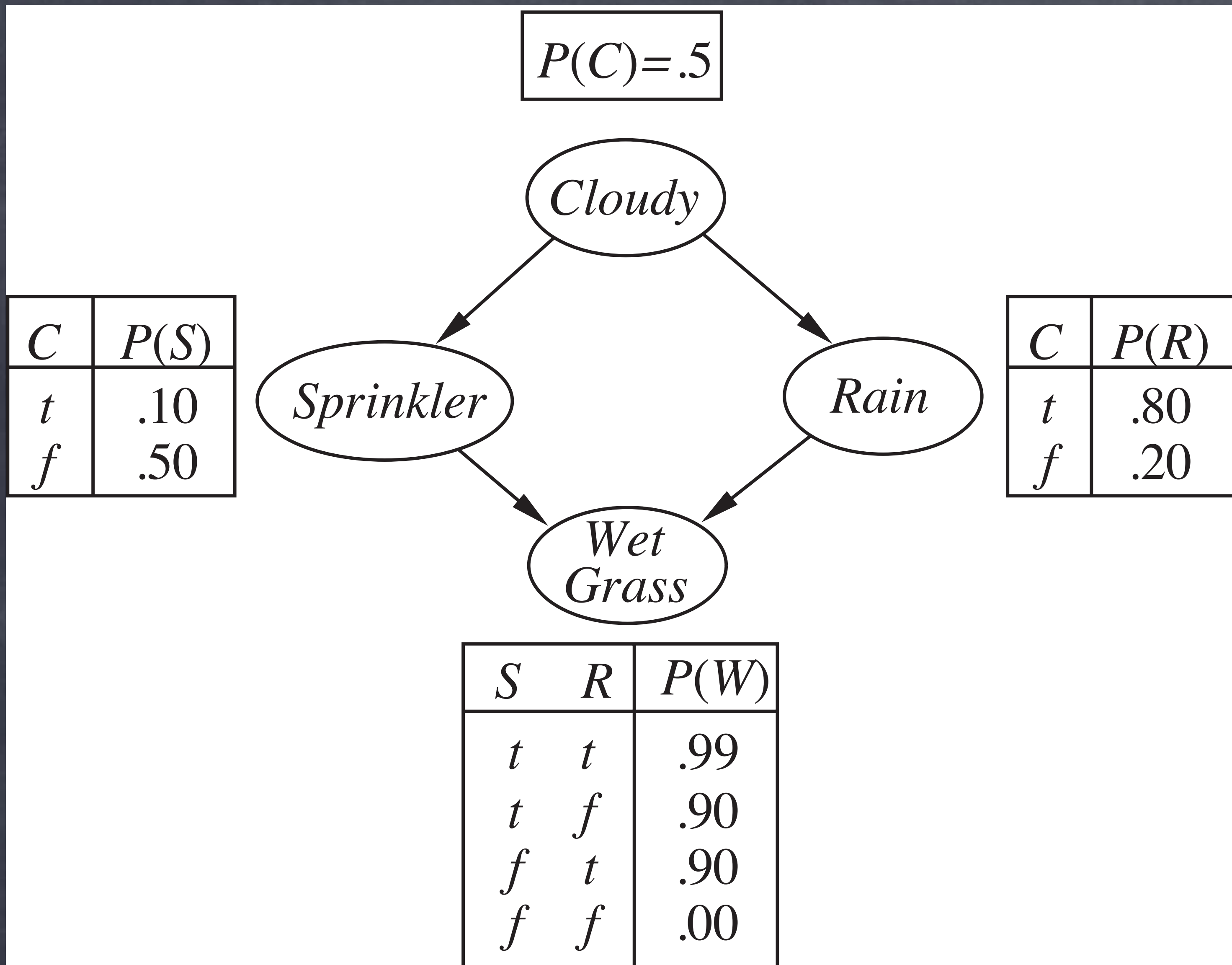



$$\mathbf{P}(Rain \mid Sprinkler = true)$$

$$\langle Cloudy{=}true,\ Sprinkler{=}false,\ Rain{=}true,\ WetGrass{=}true \rangle$$

# Rejection Sampling

- Generate sample from the prior distribution specified by the network
- Reject sample if inconsistent with the evidence
- Use remaining samples to estimate probability of event

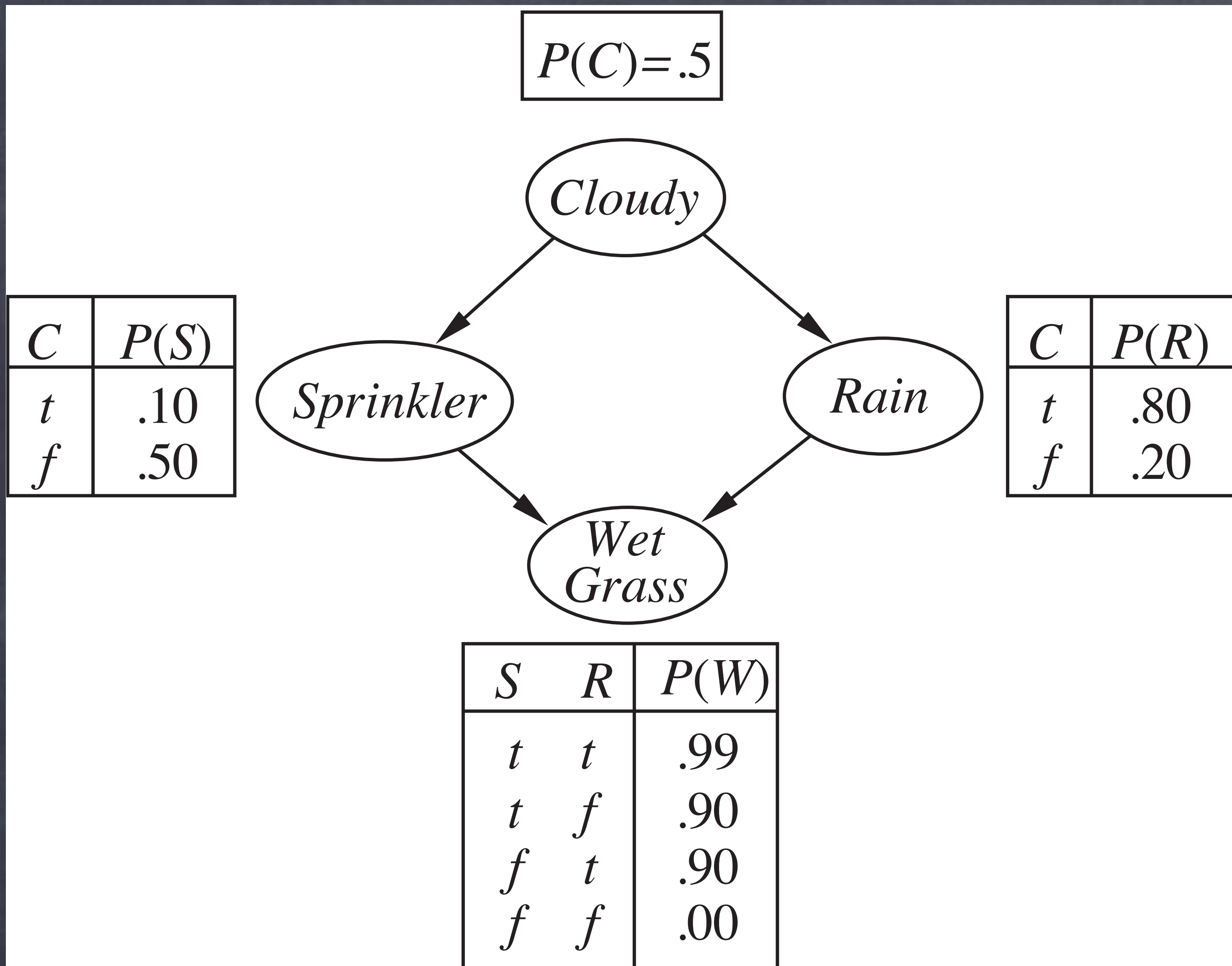



$\mathbf{P}(Rain \mid Sprinkler = true)$

100 samples

$Sprinkler{=}false$: 73

$Sprinkler{=}true$: 27




$\mathbf{P}(Rain \mid Sprinkler = true)$

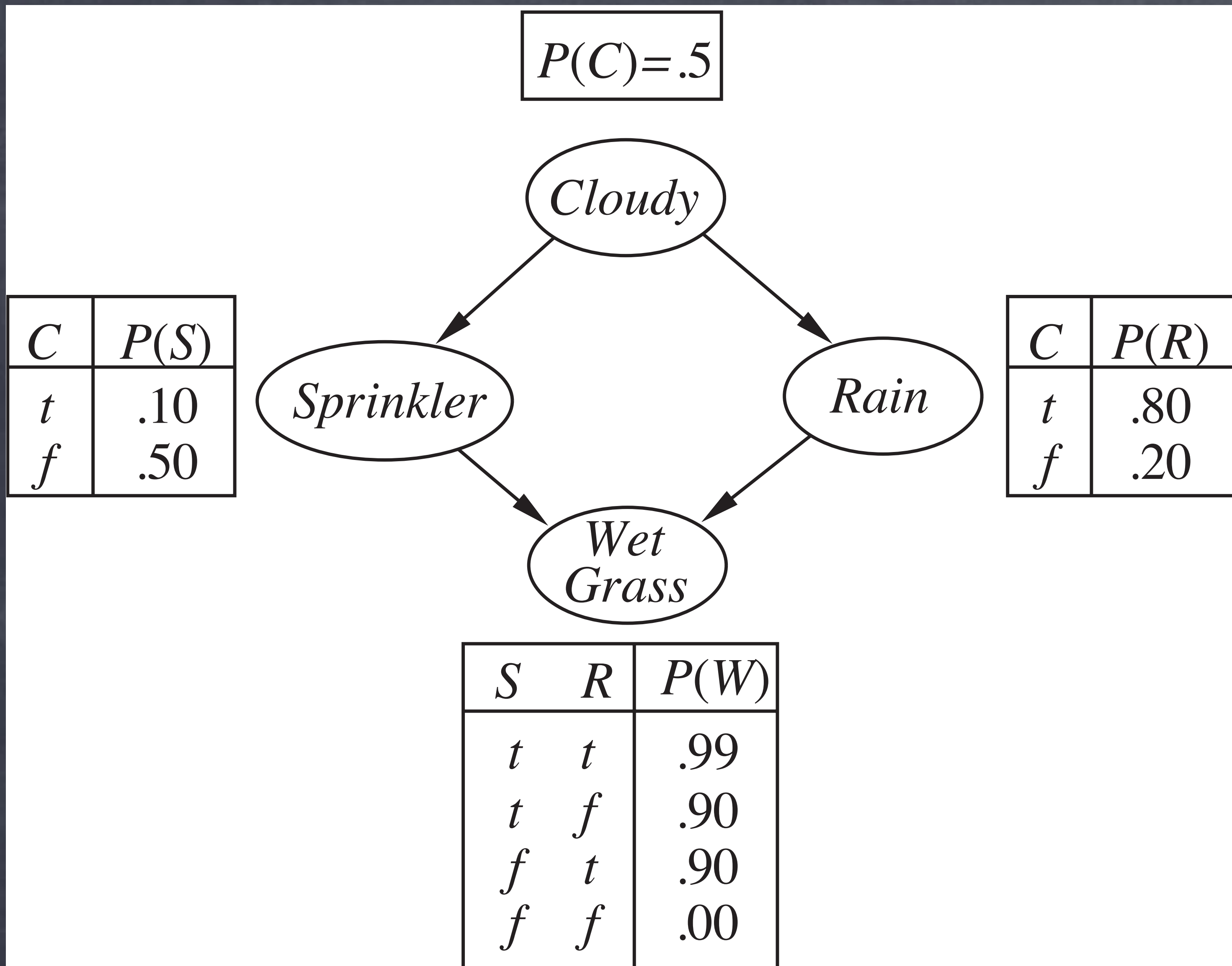
100 samples

$Sprinkler{=}false$: 73

$Sprinkler{=}true$: 27

$Rain{=}true$: 8

$Rain{=}false$: 19




$\mathbf{P}(Rain \mid Sprinkler = true)$

100 samples

$Sprinkler{=}false$: 73

$Sprinkler{=}true$: 27

$Rain{=}true$: 8

$Rain{=}false$: 19

$$\mathbf{P}(Rain \mid Sprinkler = true) \approx \alpha \, \langle \frac{8}{27}, \frac{19}{27} \rangle = \langle 0.296, 0.704 \rangle$$

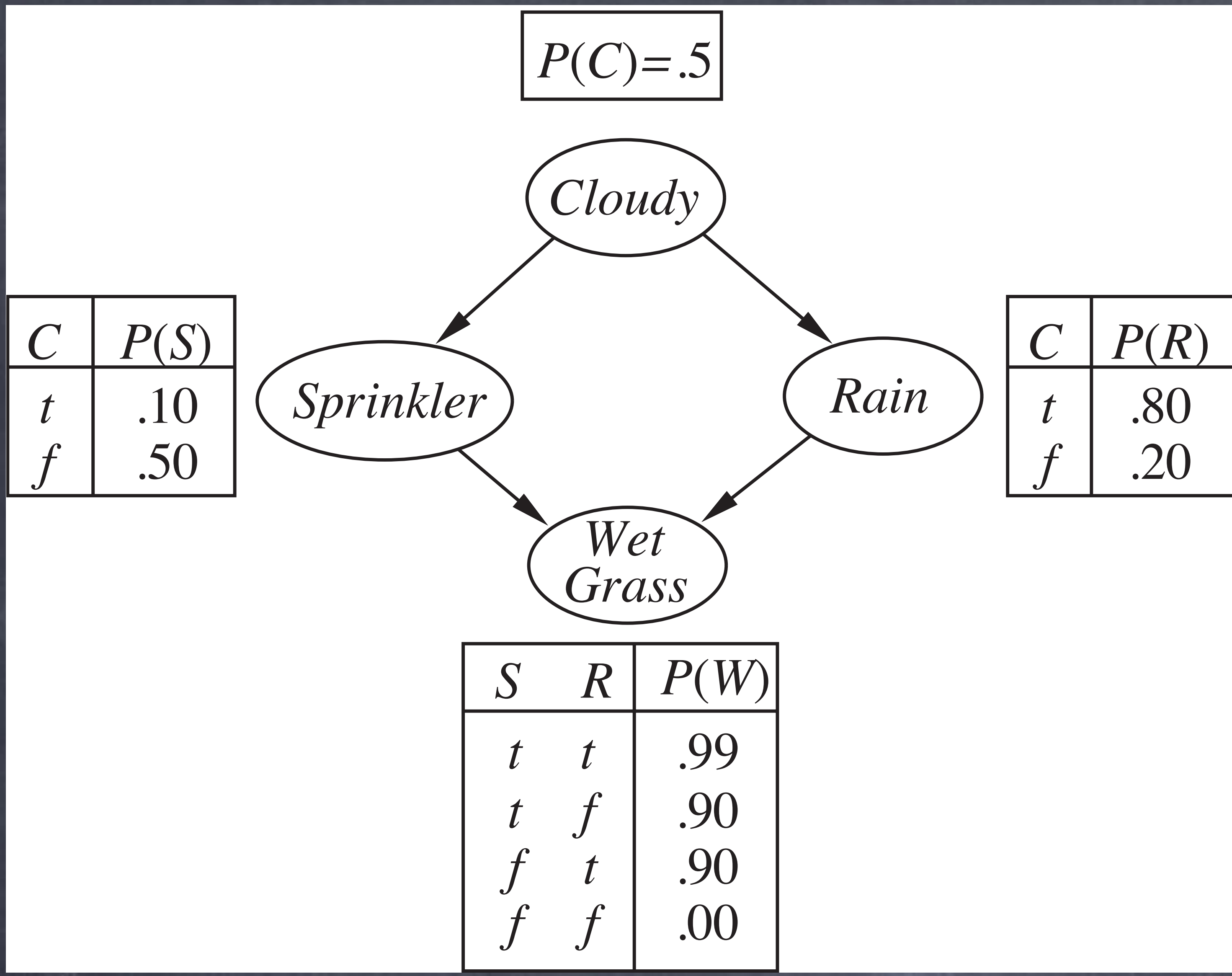
# Rejection Sampling

- Generate sample from the prior distribution specified by the network
- Reject sample if inconsistent with the evidence
- Use remaining samples to estimate probability of event

# Rejection Sampling

- Generate sample from the prior distribution specified by the network
- Reject sample if inconsistent with the evidence
- Use remaining samples to estimate probability of event

Fraction of samples consistent with the evidence drops exponentially with number of evidence variables

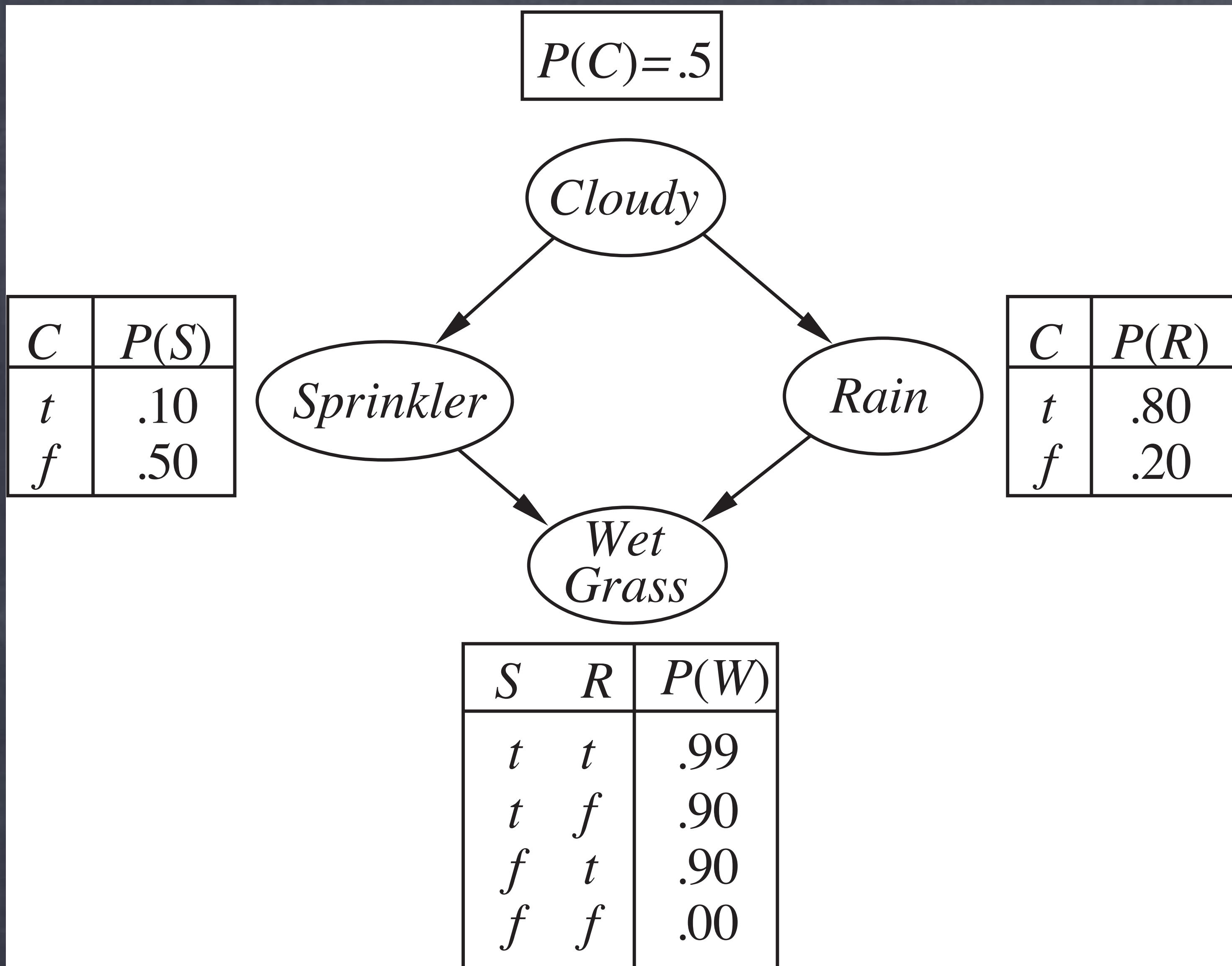



$Cloudy$

$Sprinkler$

$Rain$

$WetGrass$



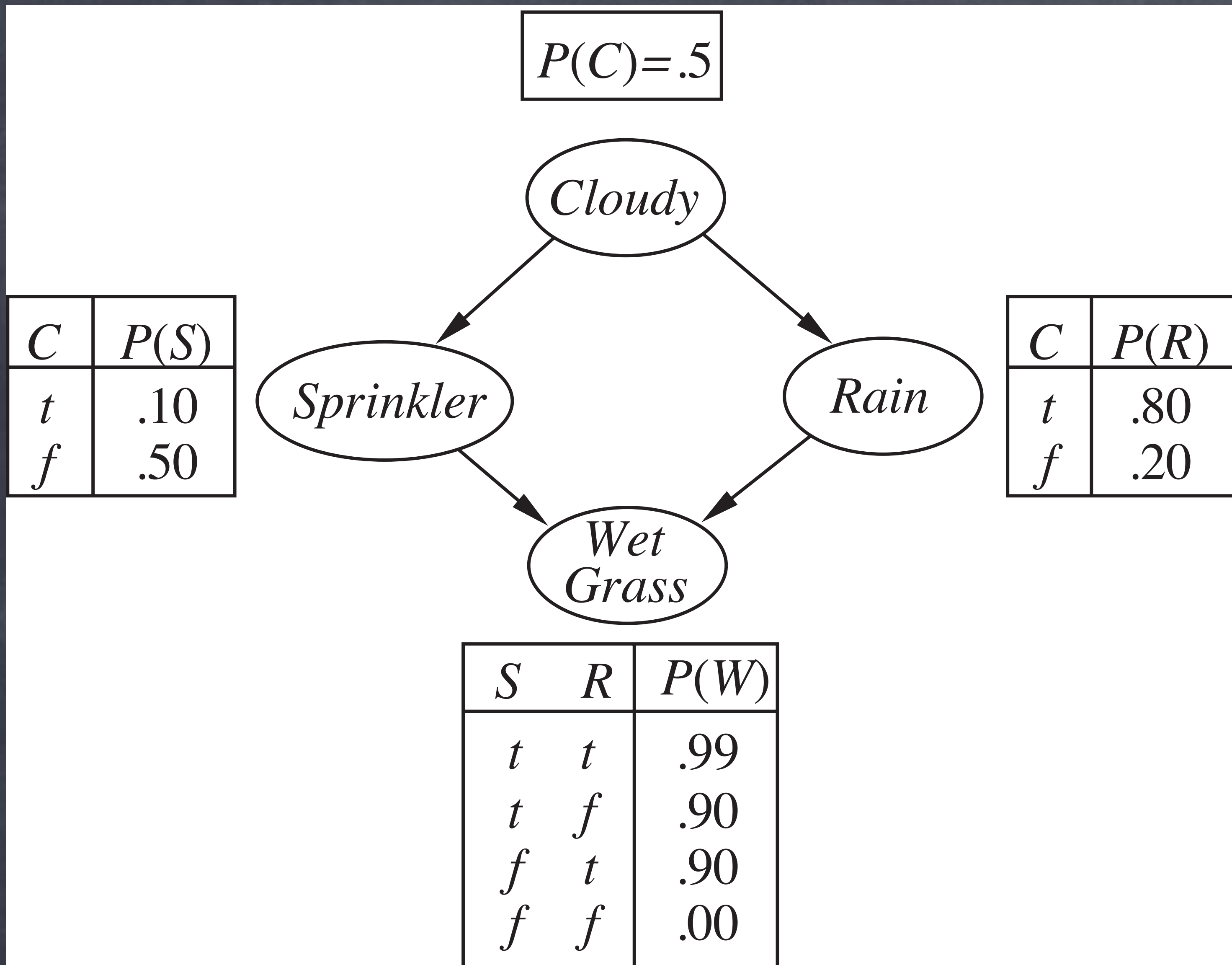

*Cloudy*

*Sprinkler*

*Rain*

*WetGrass*

$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$




$Cloudy$

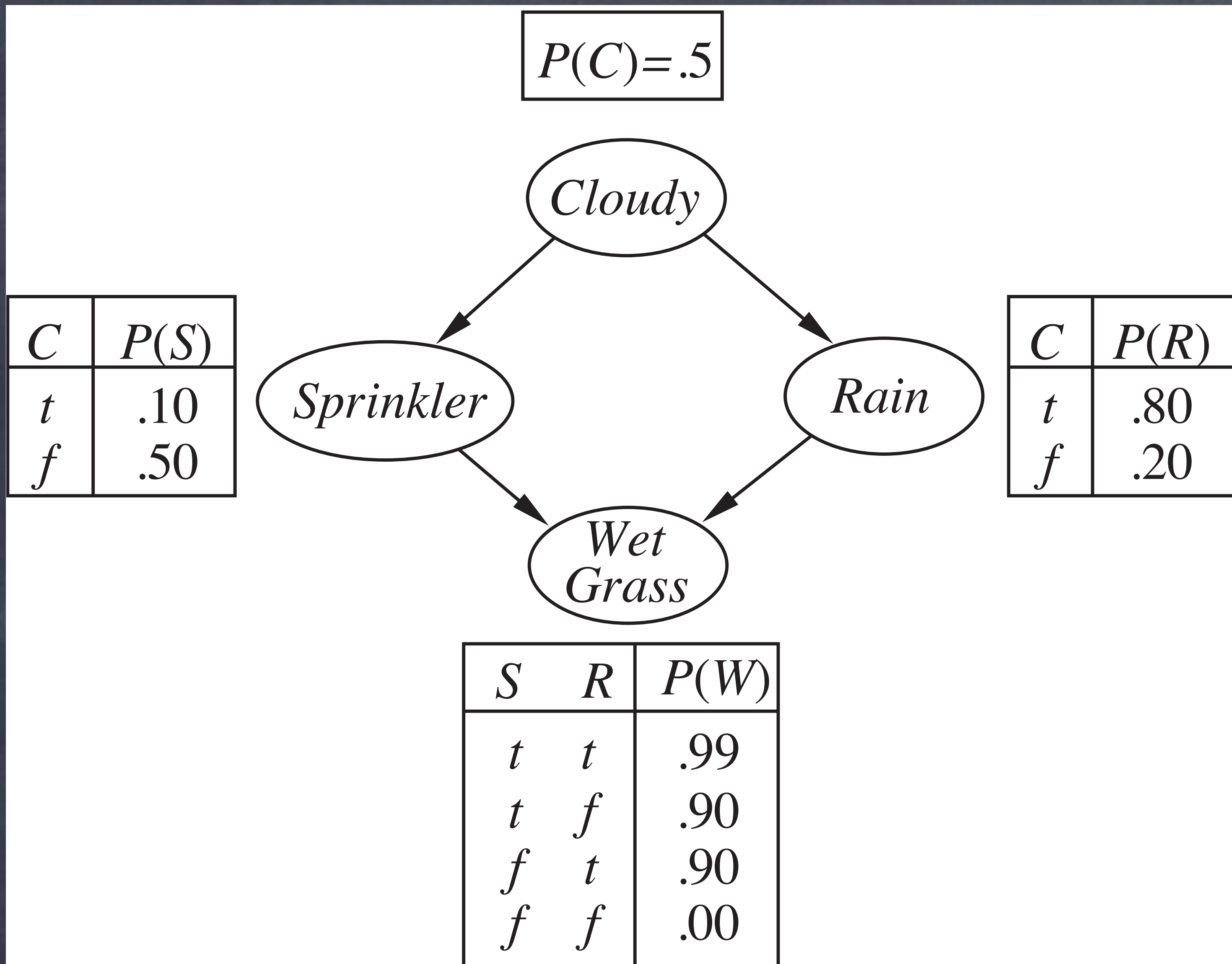
$Sprinkler$

$Rain$

$WetGrass$

$w = 1.0$

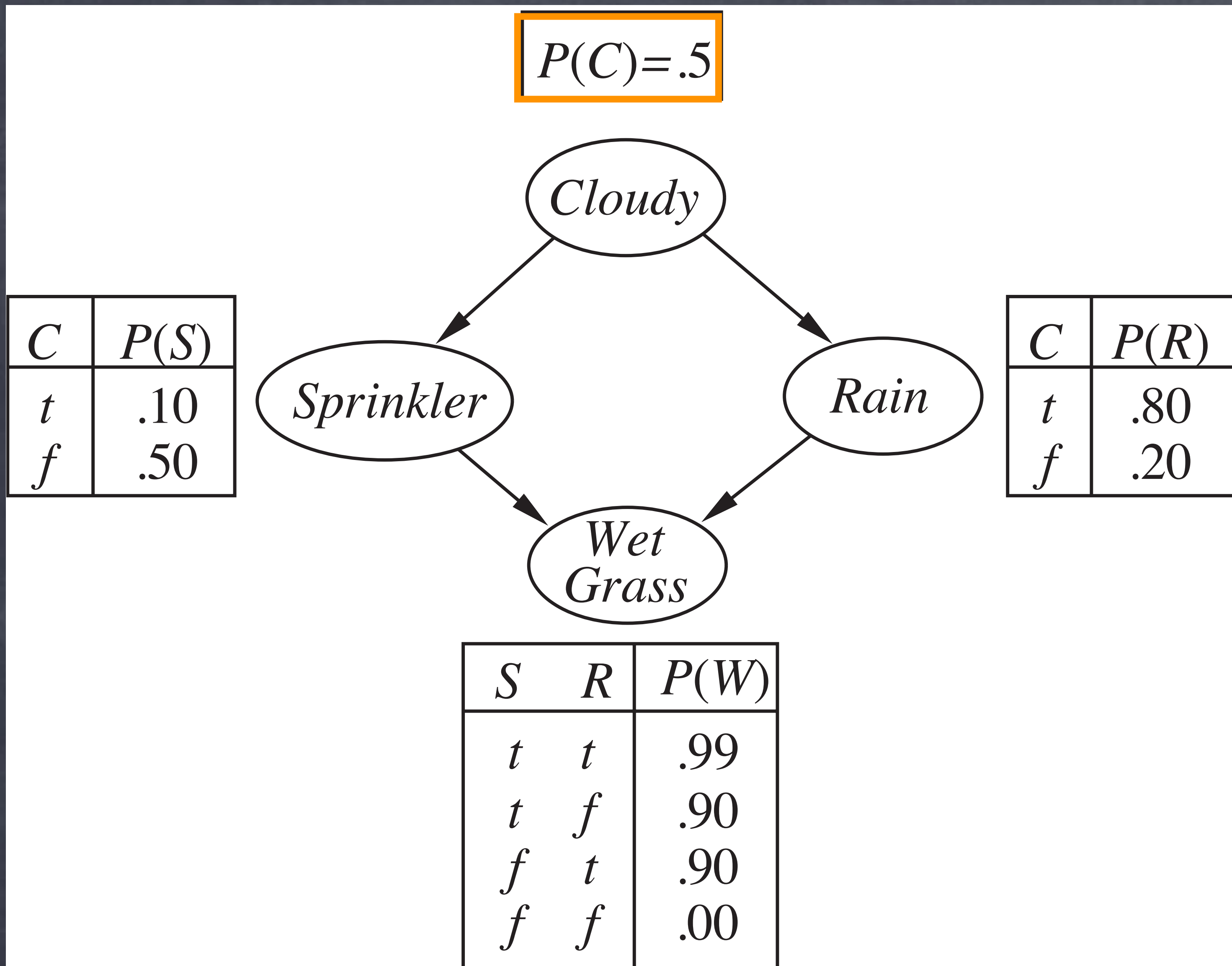
$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$




| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | |
| $Rain$ | |
| $WetGrass$ | |

$$w = 1.0$$

$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$




$Cloudy$ $true$

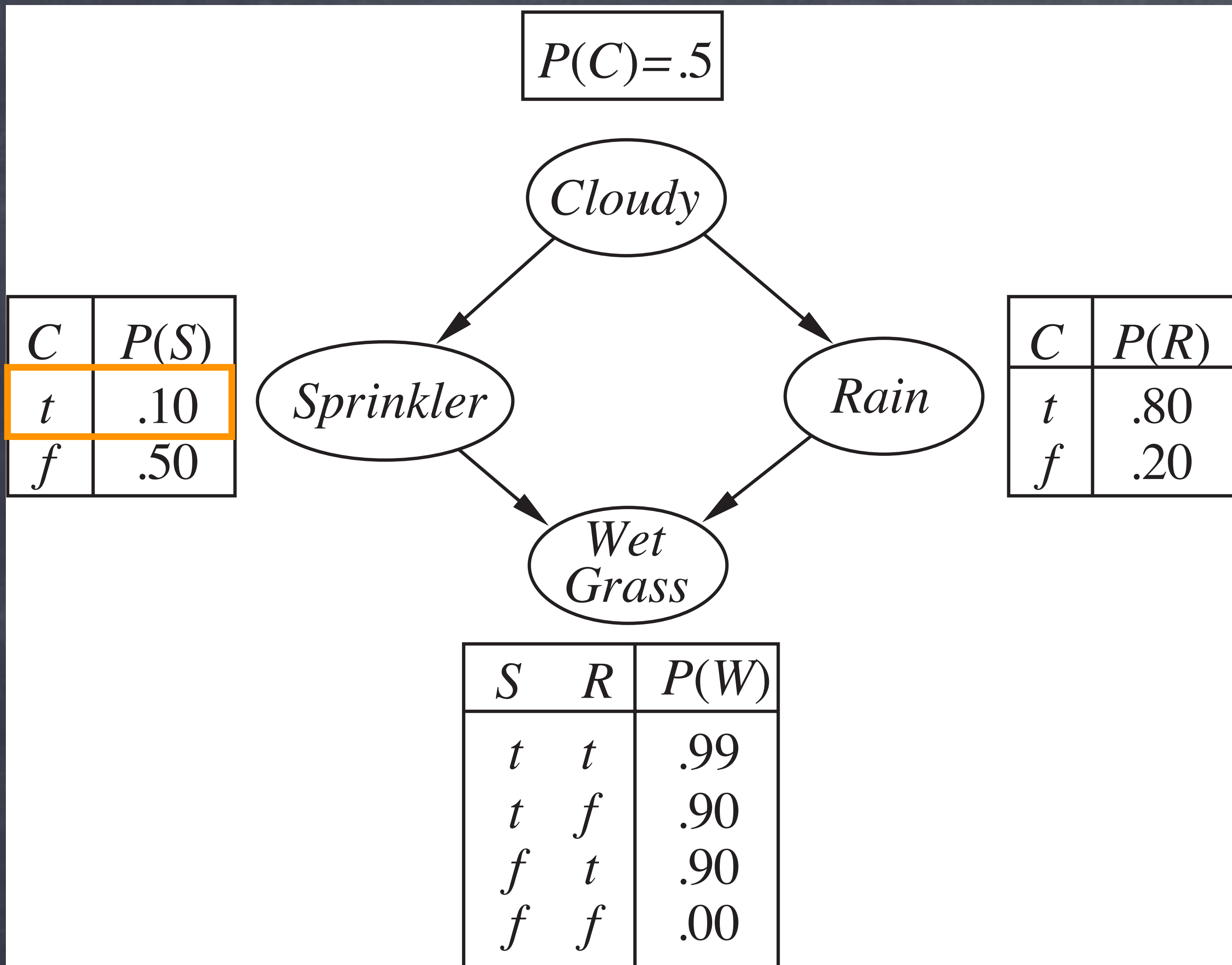
$Sprinkler$

$Rain$

$WetGrass$

$w = 0.5$

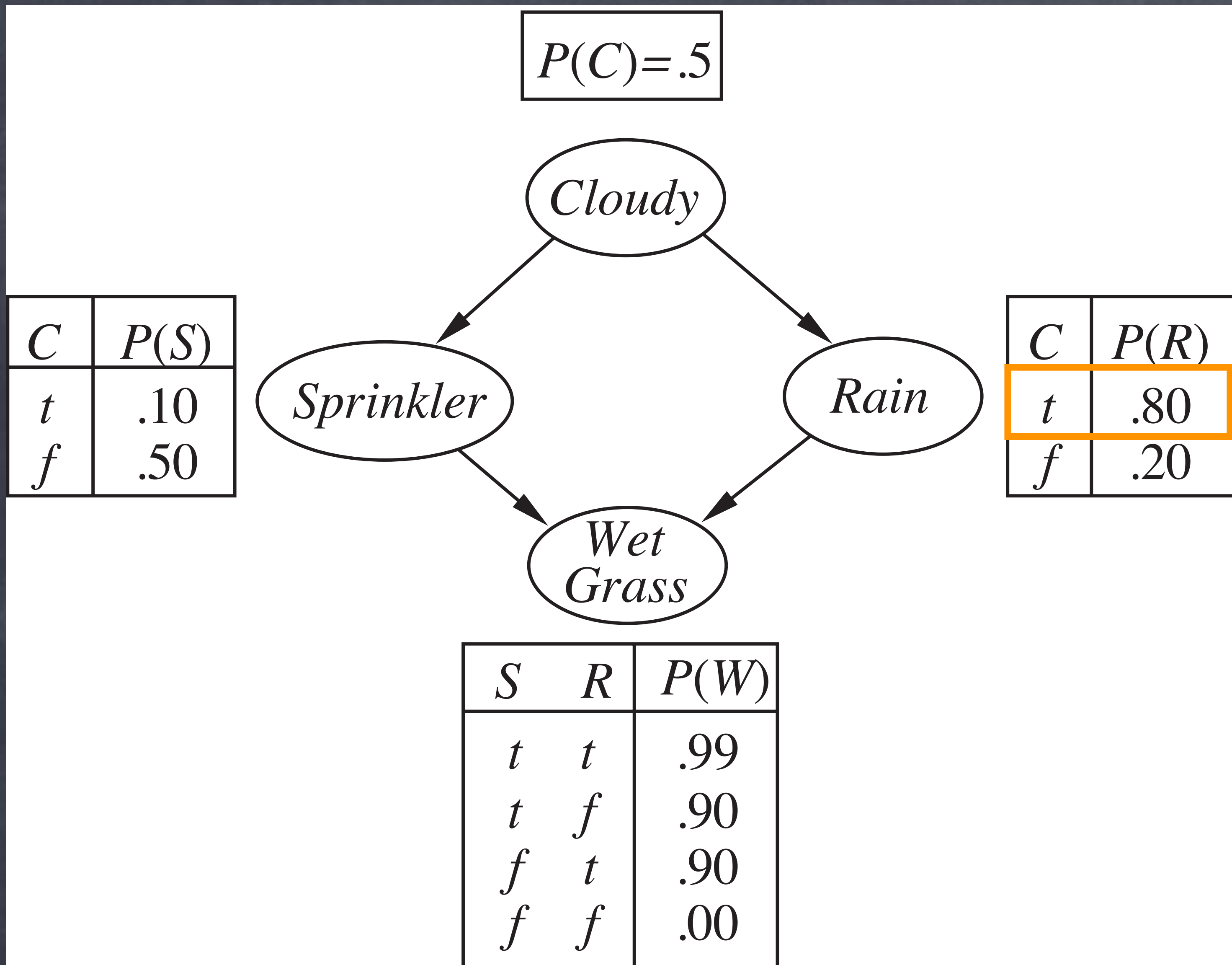
$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$




| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $false$ |
| $Rain$ | |
| $WetGrass$ | |

$$w = 0.5$$

$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$




| C | P(S) |
|---|---|
| t | .10 |
| f | .50 |

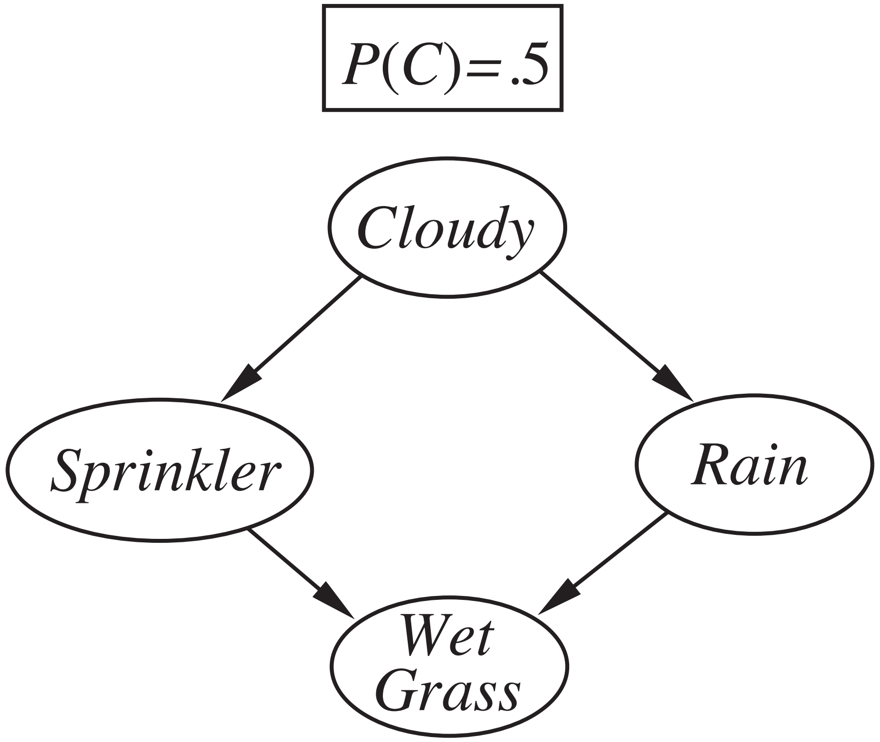
| C | P(R) |
|---|---|
| t | .80 |
| f | .20 |

| S | R | P(W) |
|---|---|---|
| t | t | .99 |
| t | f | .90 |
| f | t | .90 |
| f | f | .00 |

| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $false$ |
| $Rain$ | $true$ |
| $WetGrass$ | |

$$w = 0.5$$

$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$




| C | P(S) |
|---|---|
| t | .10 |
| f | .50 |

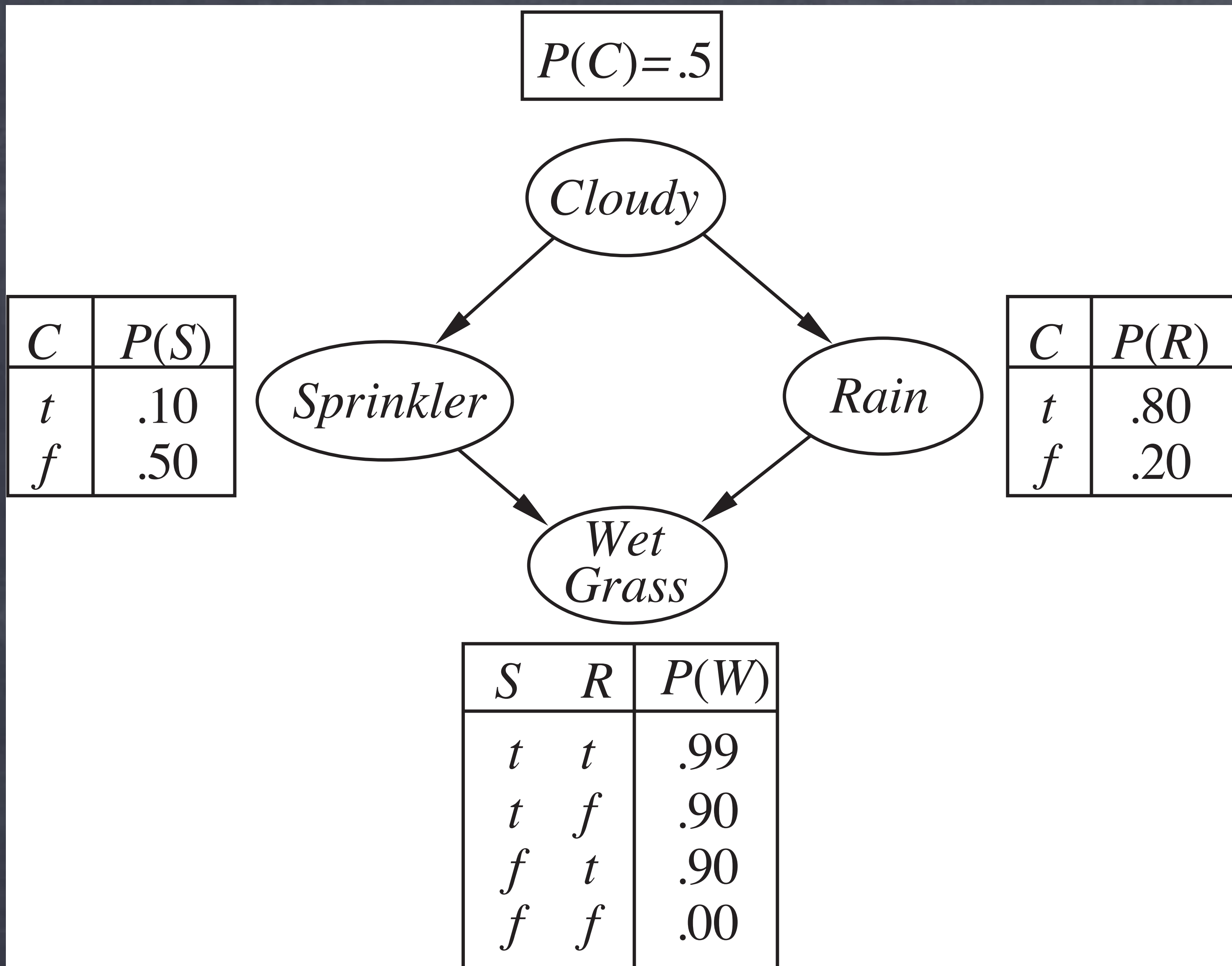
| C | P(R) |
|---|---|
| t | .80 |
| f | .20 |

| S | R | P(W) |
|---|---|---|
| t | t | .99 |
| t | f | .90 |
| f | t | .90 |
| f | f | .00 |

| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $false$ |
| $Rain$ | $true$ |
| $WetGrass$ | $true$ |

$$w = 0.5$$

$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$




| C | P(S) |
|---|---|
| t | .10 |
| f | .50 |

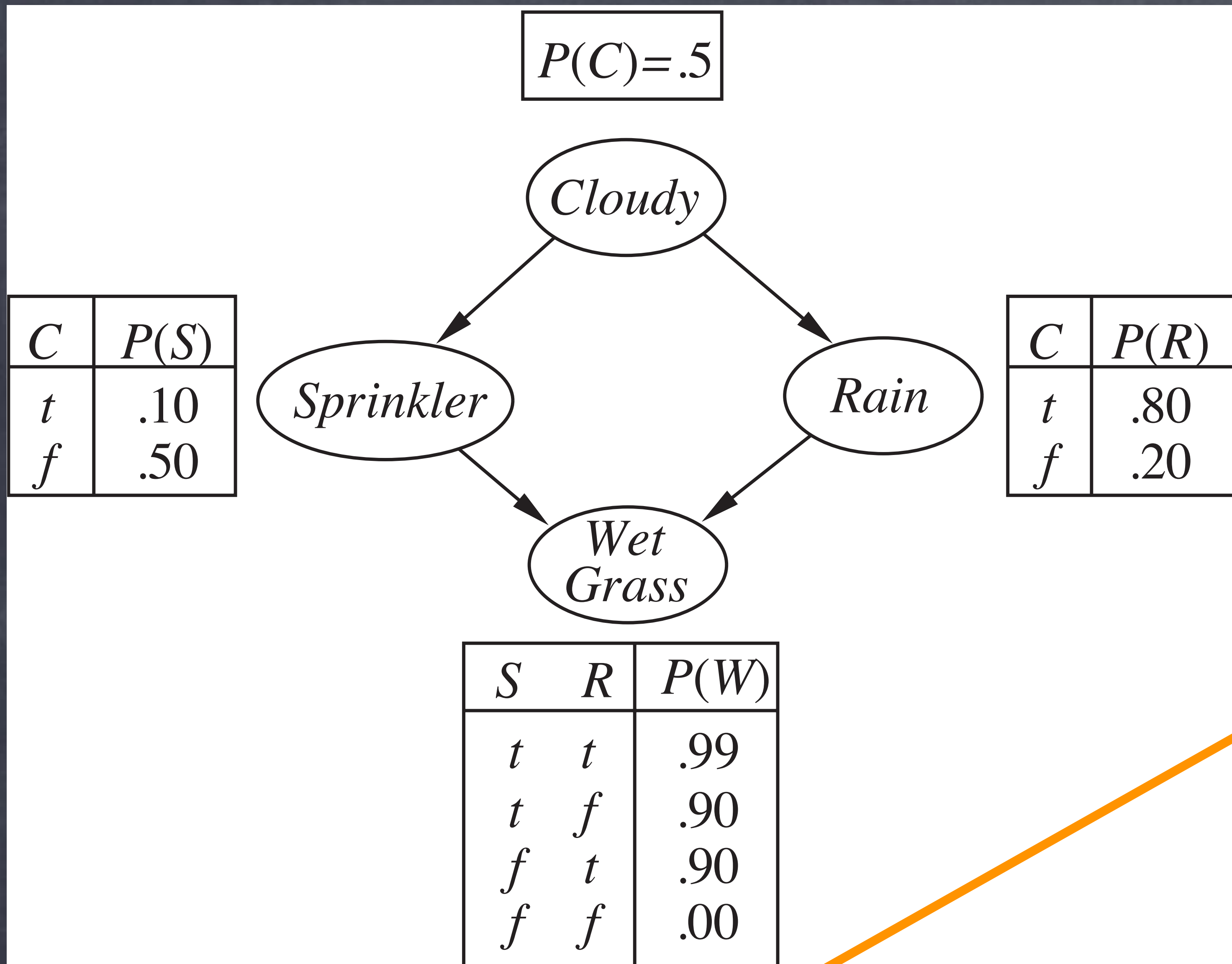
| C | P(R) |
|---|---|
| t | .80 |
| f | .20 |

| S | R | P(W) |
|---|---|---|
| t | t | .99 |
| t | f | .90 |
| f | t | .90 |
| f | f | .00 |

| | |
|---|---|
| *Cloudy* | *true* |
| *Sprinkler* | *false* |
| *Rain* | *true* |
| *WetGrass* | *true* |

$$w = 0.45$$

$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$

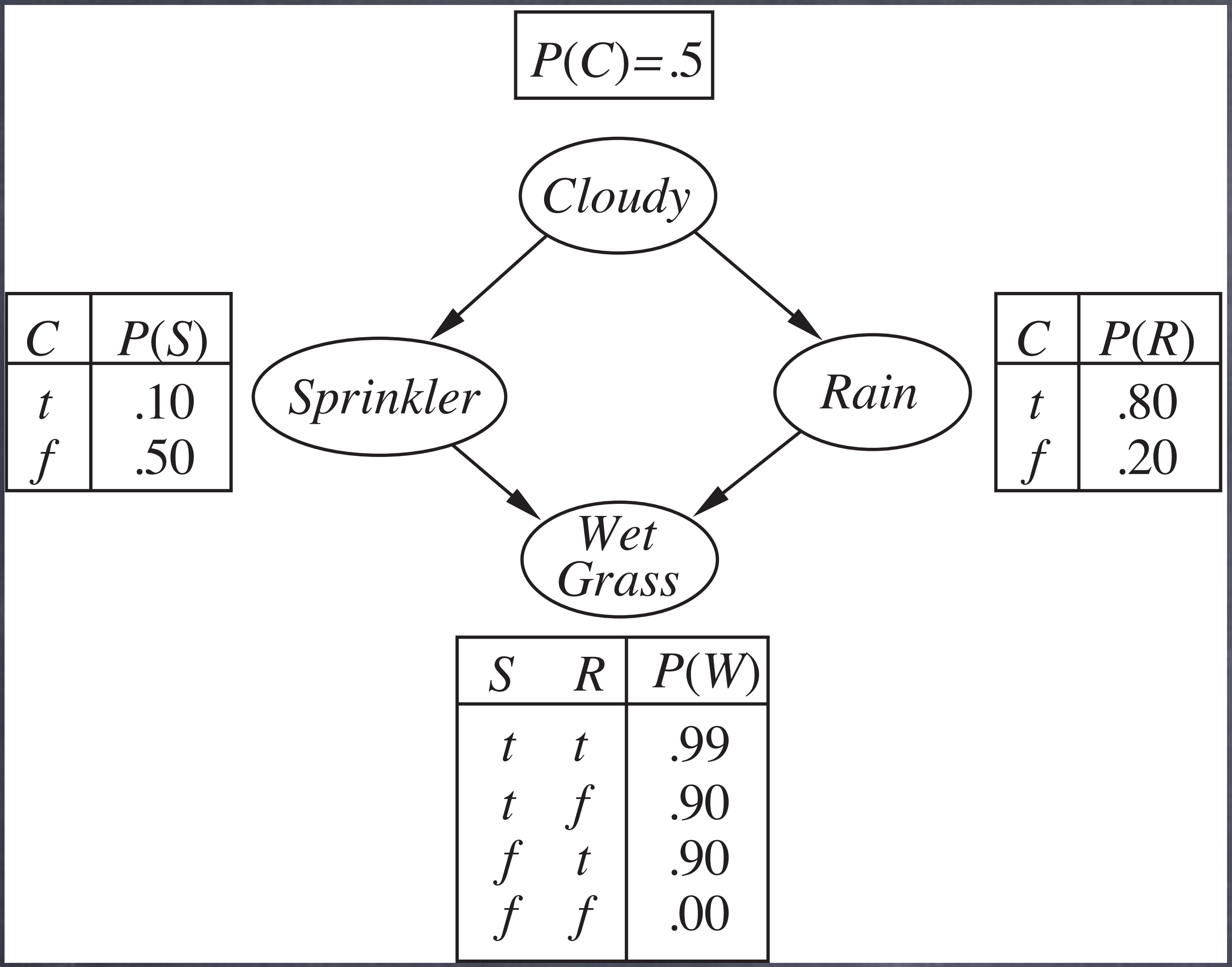



| | |
|---|---|
| *Cloudy* | *true* |
| *Sprinkler* | *false* |
| *Rain* | *true* |
| *WetGrass* | *true* |

$w = 0.45$

$$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$$

$N_{Rain=true}$ **+=**




| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $false$ |
| $Rain$ | $true$ |
| $WetGrass$ | $true$ |

$w = 0.45$

$\mathbf{P}(Rain \mid Cloudy = true, WetGrass = true)$

$N_{Rain=true}$ += $0.45$

# Likelihood Weighting

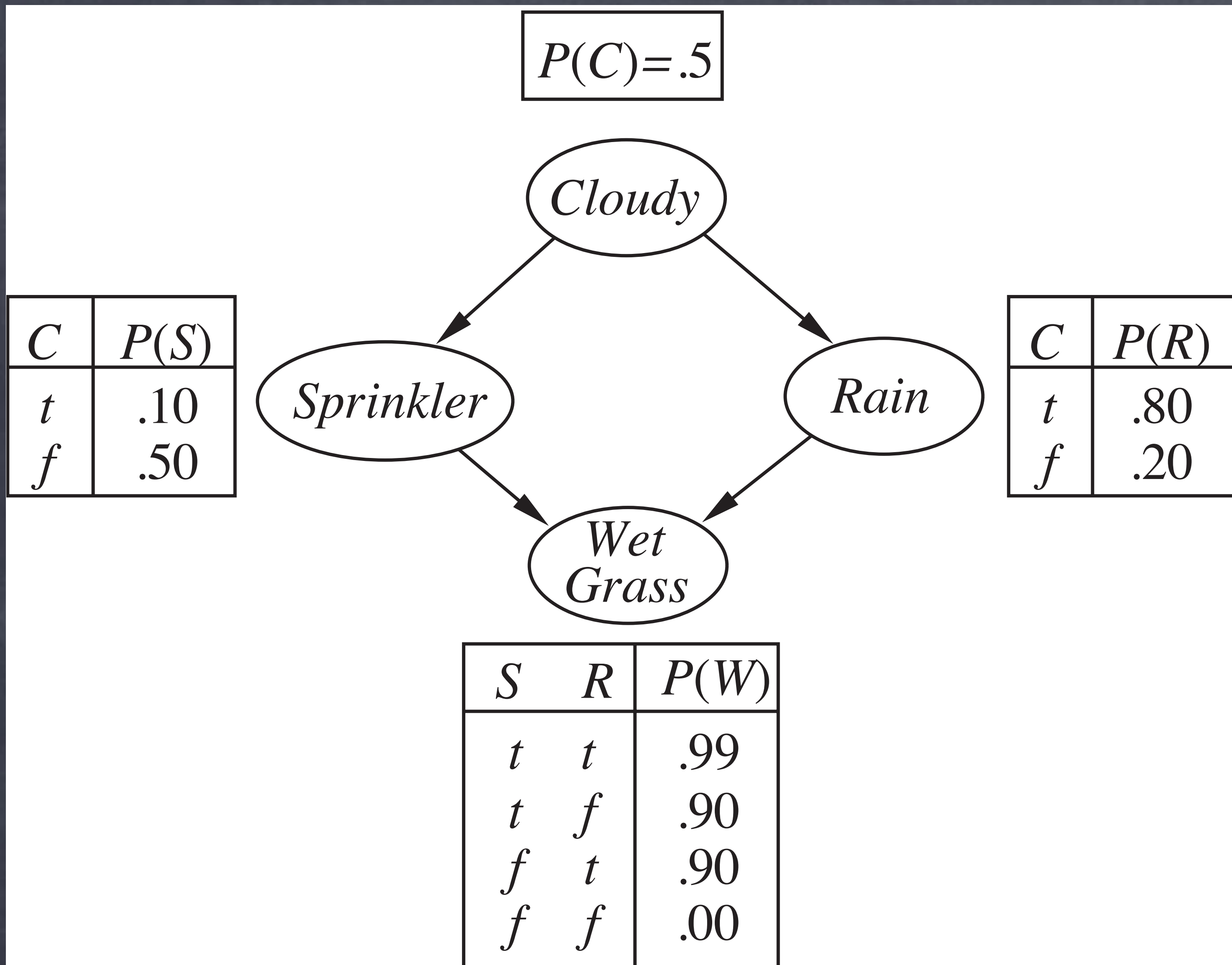
- Generate only samples consistent with the evidence
  - Generate sample using topological order
  - Evidence variable: Value fixed, update weight
  - Non-evidence variable: Sample value from network
- Compute weighted sum of samples for estimate

# Likelihood Weighting

- Pros:
  - Doesn’t reject any samples
- Cons:
  - More evidence $\Rightarrow$ lower weight
  - Affected by order of evidence vars in topsort (later = worse)

# Approximate Inference in Bayesian Networks

- Rejection Sampling
- Likelihood Weighting




| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $false$ |
| $Rain$ | $true$ |
| $WetGrass$ | $true$ |

$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$

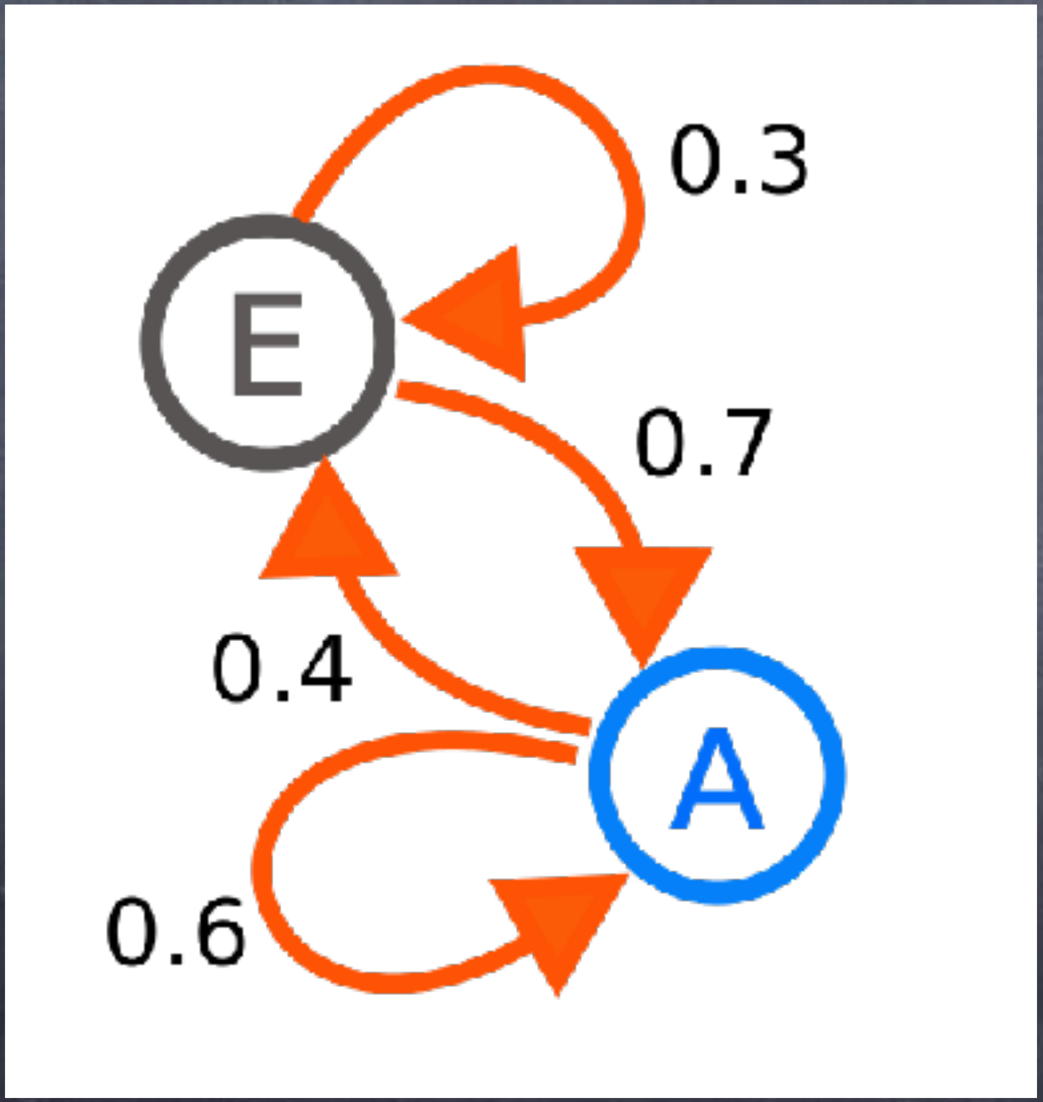
- To approximate: $\mathbf{P}(X \mid \mathbf{e})$
- Generate a sequence of states
  - Values of evidence variables are fixed
  - Values of other variables appear in the right proportion given the distribution encoded by the network

# Markov Chain Monte Carlo Simulation

- To approximate: $\mathbf{P}(X \mid \mathbf{e})$
- Generate a sequence of states
  - Values of evidence variables are fixed
  - Values of other variables appear in the right proportion given the distribution encoded by the network

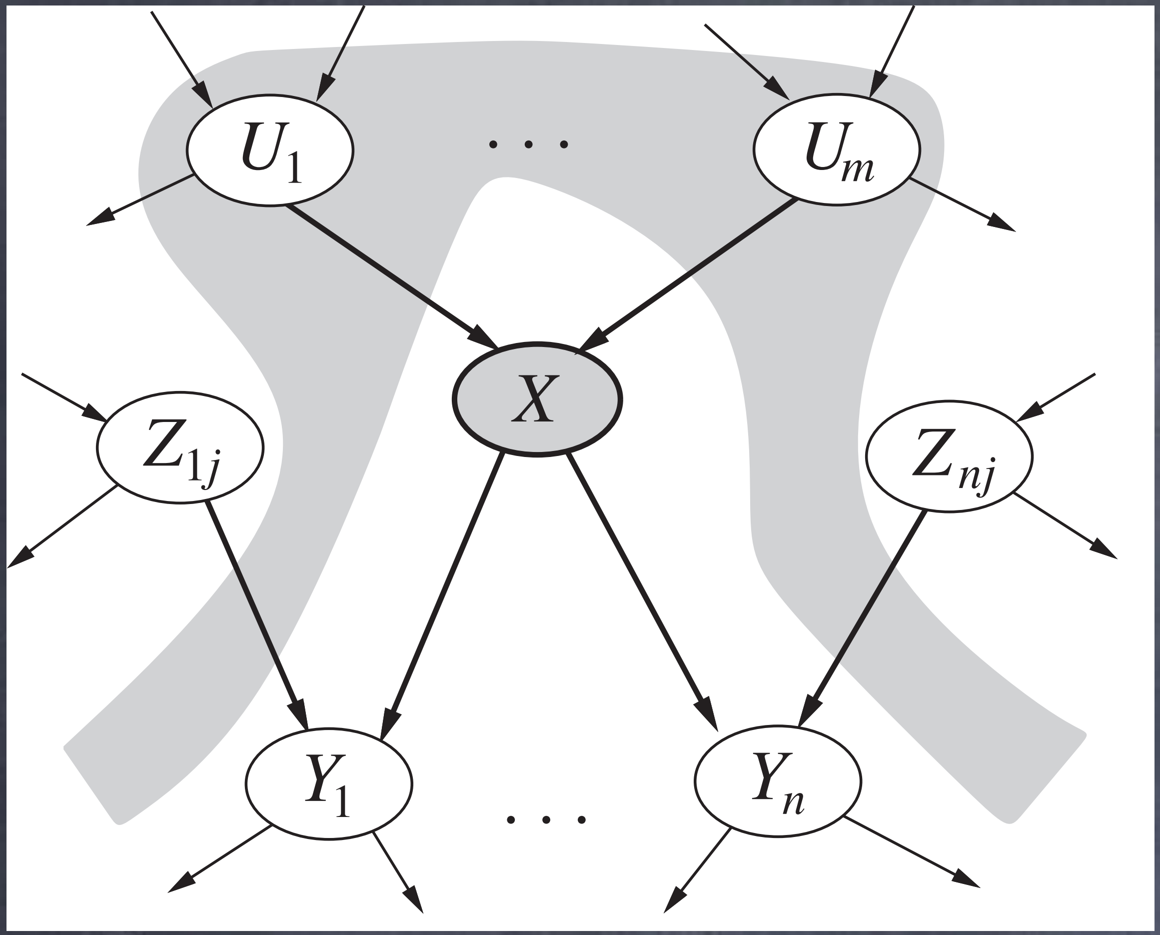
# Monte Carlo

# Markov Chain



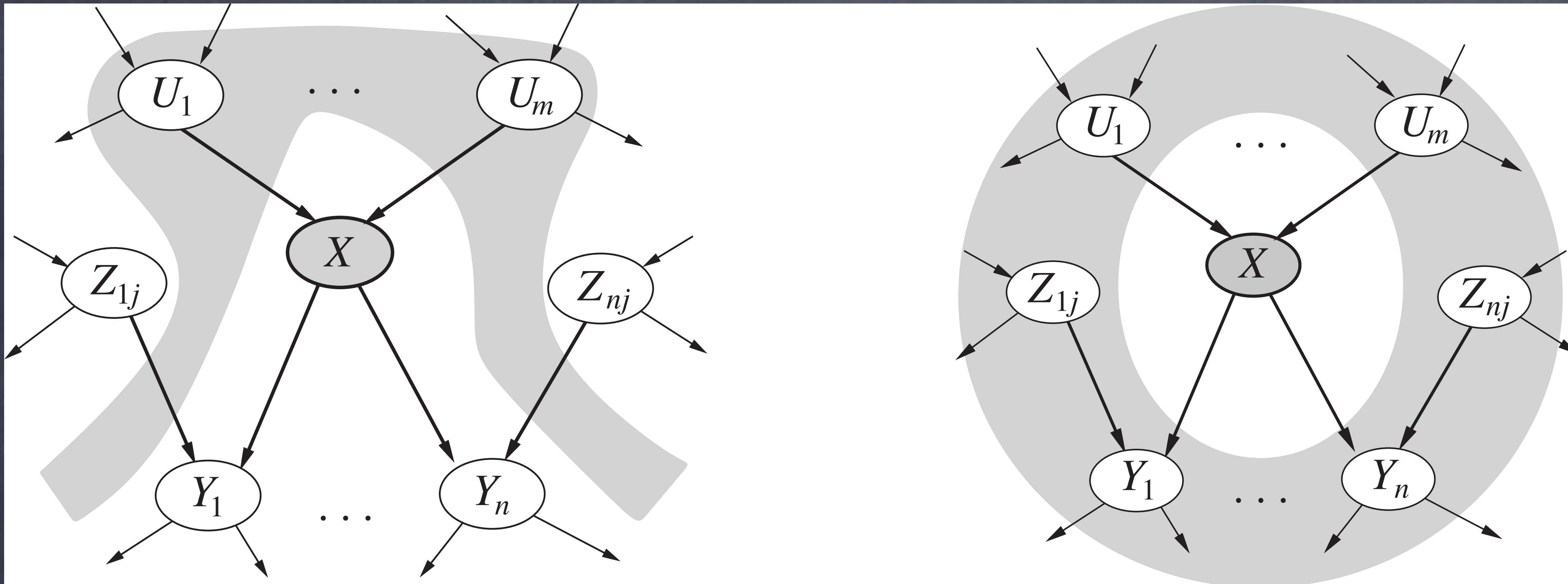
# Markov Chain Monte Carlo Simulation

- To approximate: $\mathbf{P}(X \mid \mathrm{e})$
- Generate a sequence of states
  - Values of evidence variables are fixed
  - Values of other variables appear in the right proportion given the distribution encoded by the network




Conditional Independence

$X$ conditionally independent of $Z$s given $U$s




Conditional Independence

Markov Blanket

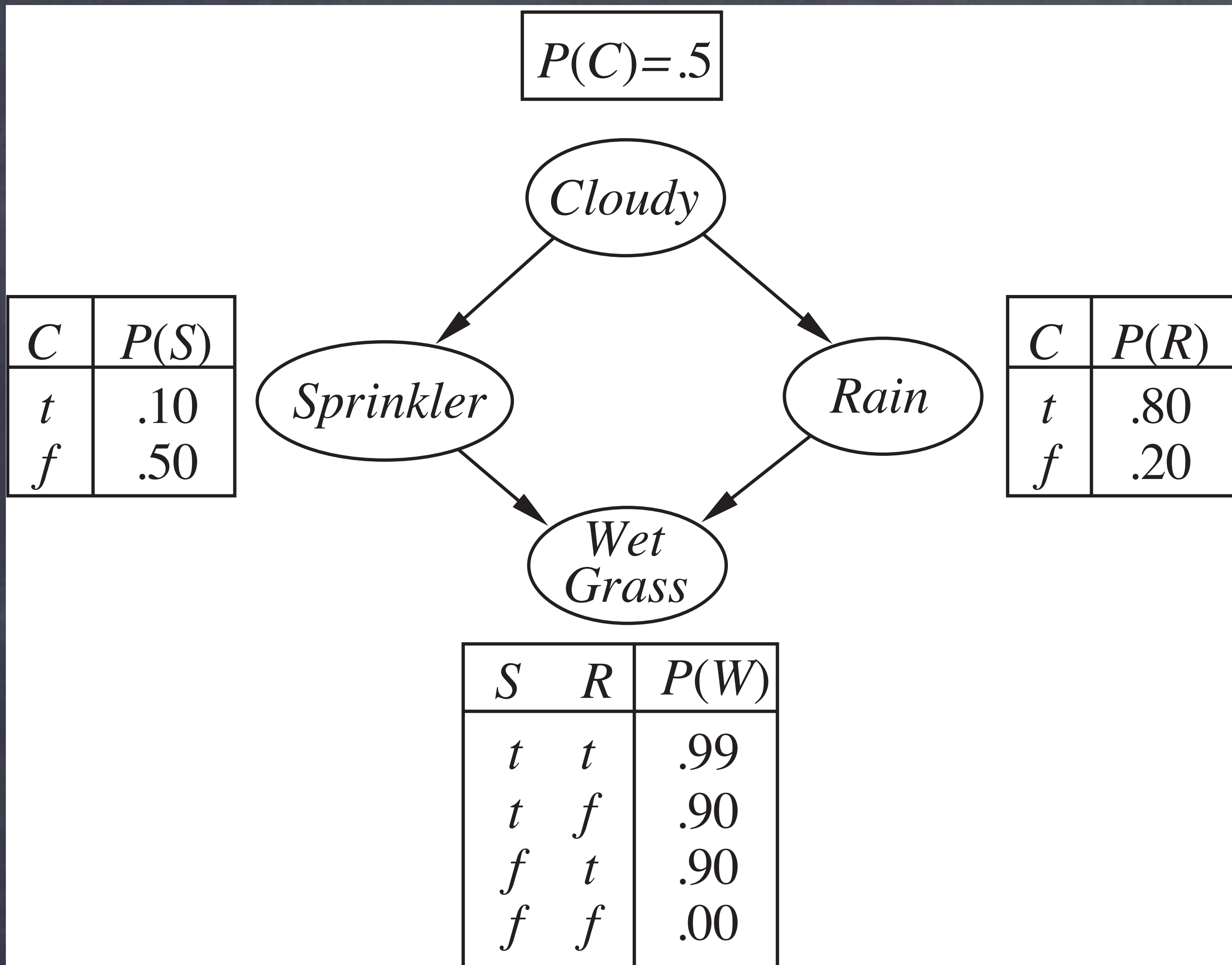
$X$ conditionally independent of all $V$s given $U$s, $Y$s, and $Z$s

# Markov Blanket

- The Markov Blanket of a node is its parents, its children, and its children's parents.
- A node is conditionally independent of all other nodes in the network given its Markov Blanket

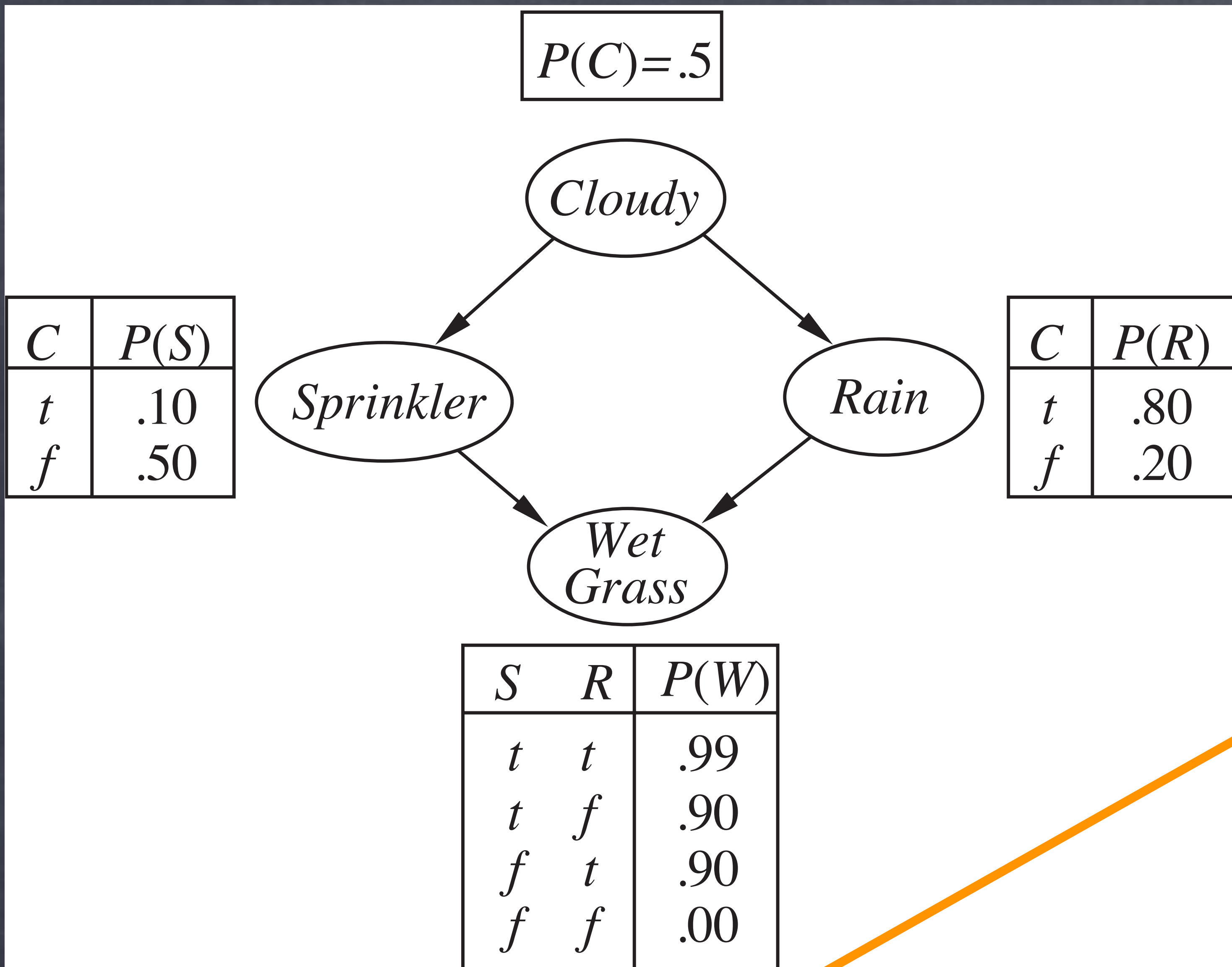
# MCMC Simulation
# Gibbs Sampling

- To approximate: $\mathbf{P}(X \mid \mathbf{e})$
- Start in a state with evidence variables set to evidence values (others arbitrary)
- On each step, sample the non-evidence variables conditioned on the values of the variables in their Markov Blankets
  - Order irrelevant




| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $true$ |
| $Rain$ | $false$ |
| $WetGrass$ | $true$ |

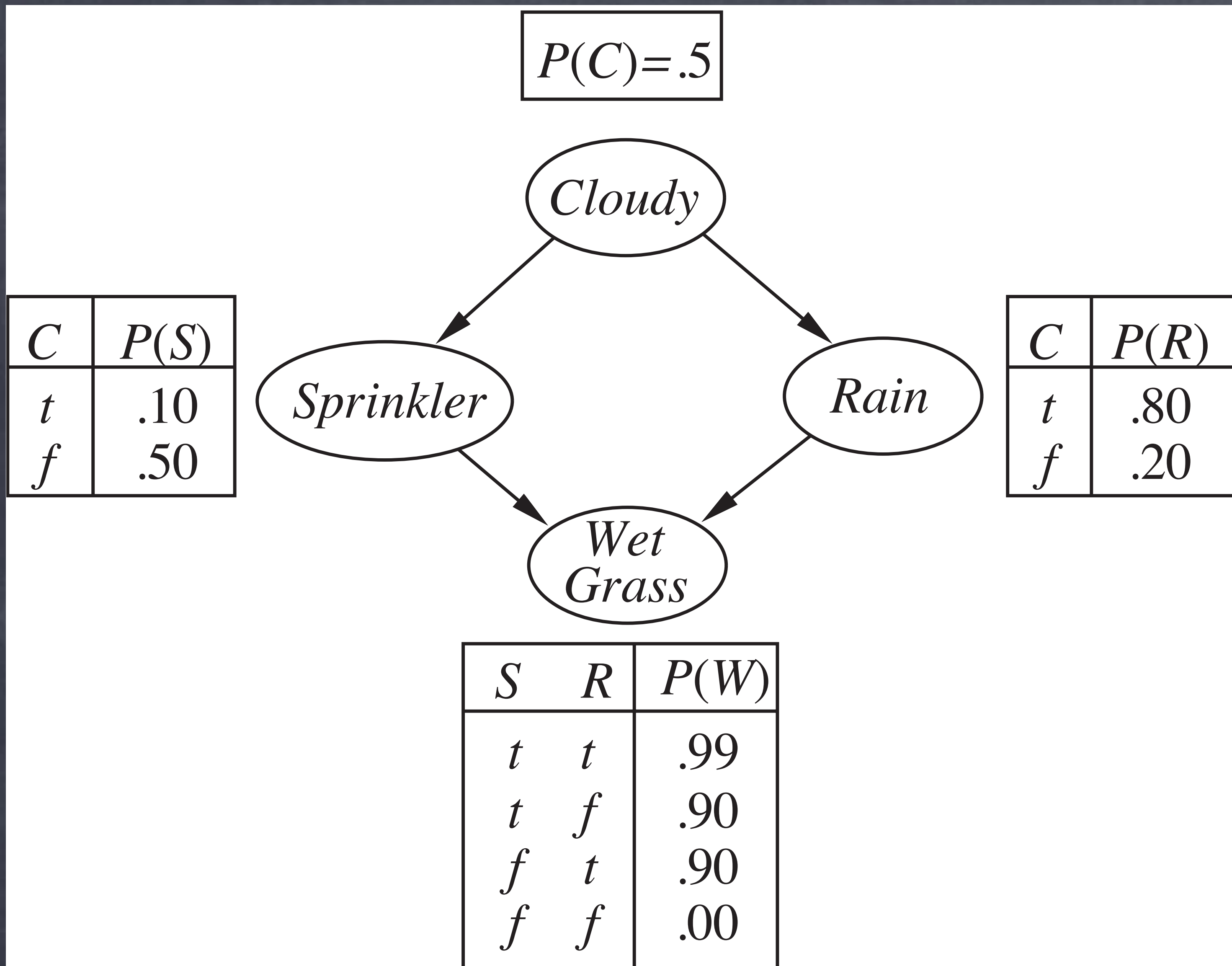
$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$




| | |
|---|---|
| $Cloudy$ | $true$ |
| $Sprinkler$ | $true$ |
| $Rain$ | $false$ |
| $WetGrass$ | $true$ |

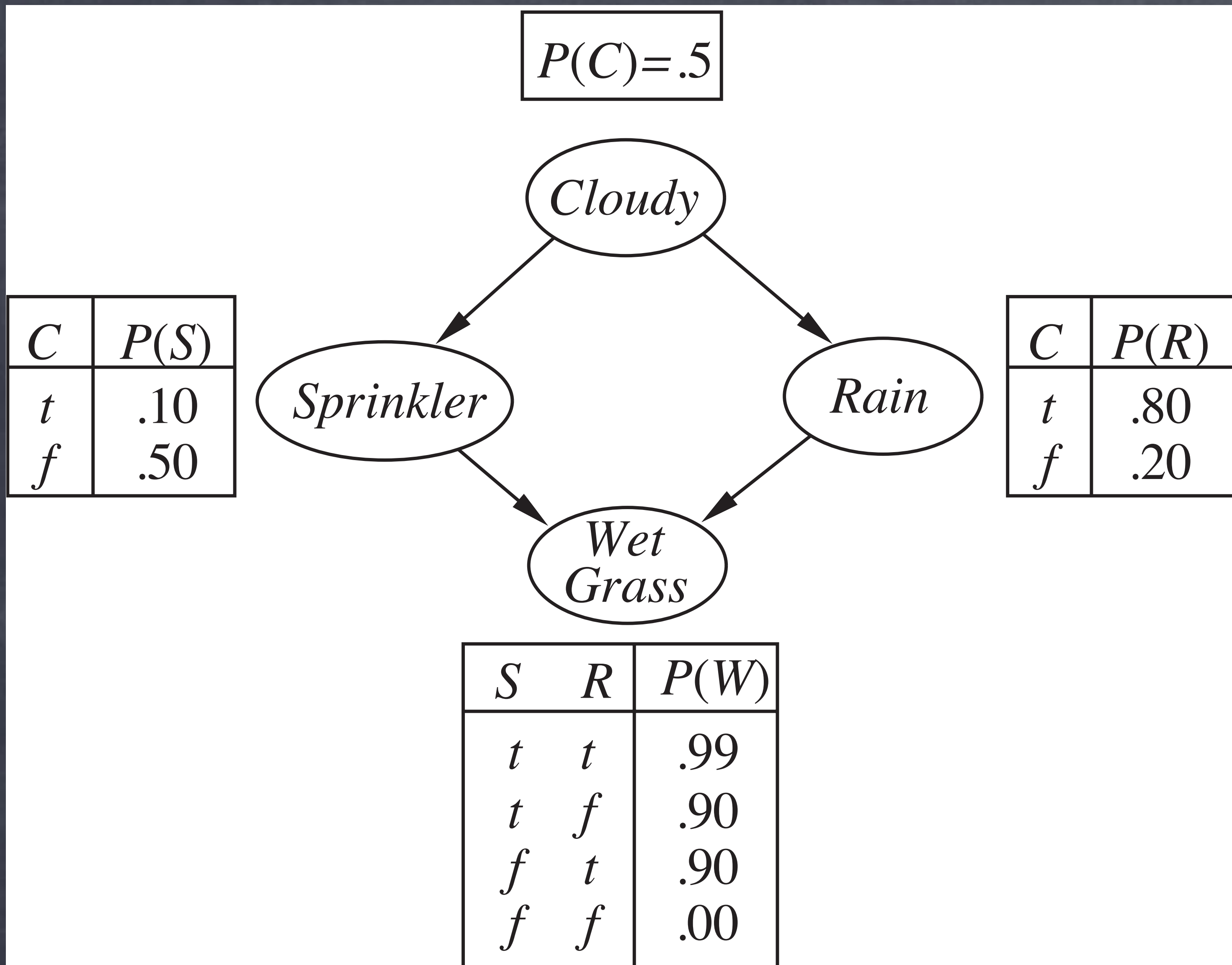
$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$

$N_{Rain=false}$ **+=** $1$



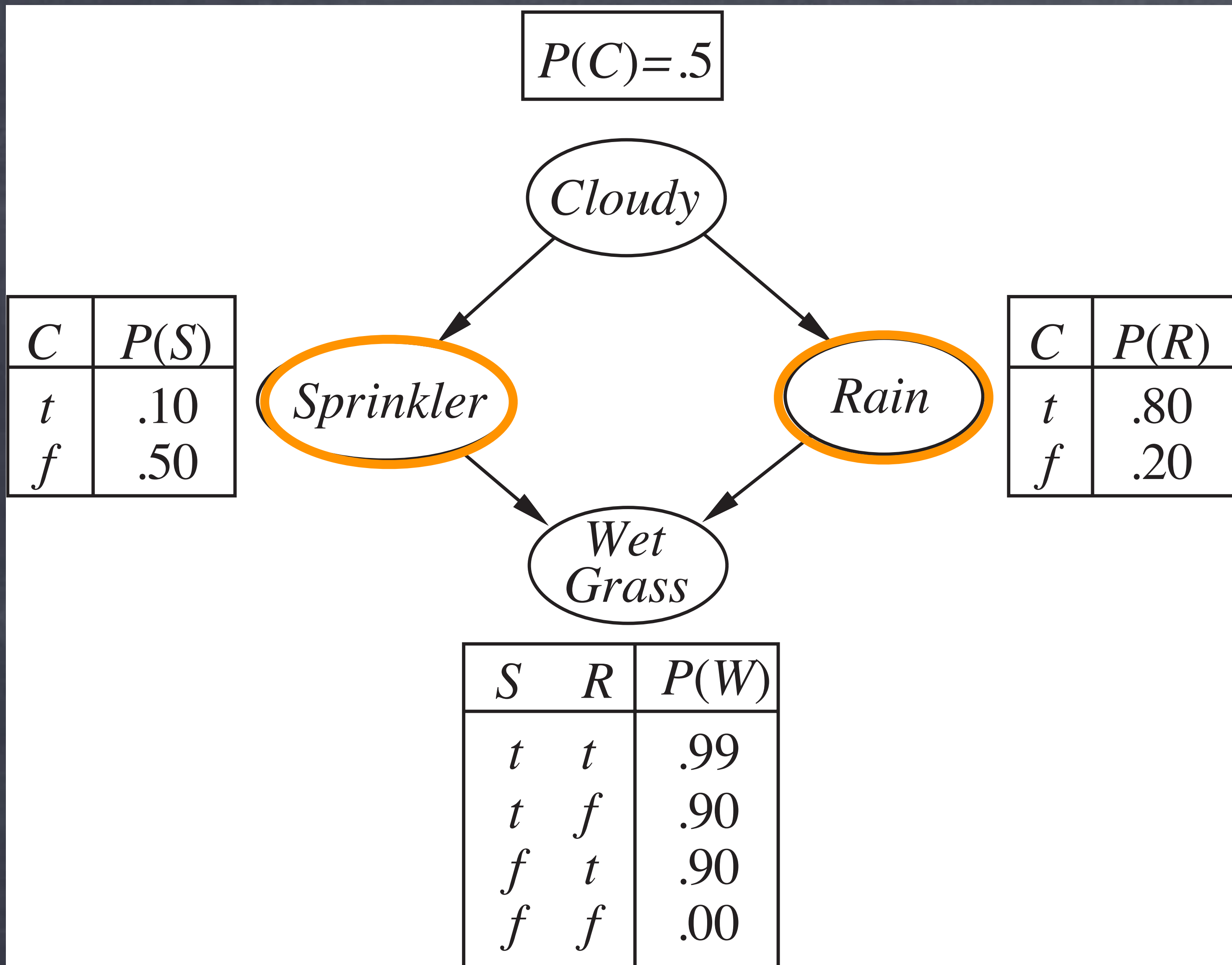

| | |
|---|---|
| *Cloudy* | *true* |
| *Sprinkler* | *true* |
| *Rain* | *false* |
| *WetGrass* | *true* |

$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$




| | |
|---|---|
| *Cloudy* | *true* |
| *Sprinkler* | *true* |
| *Rain* | *false* |
| *WetGrass* | *true* |

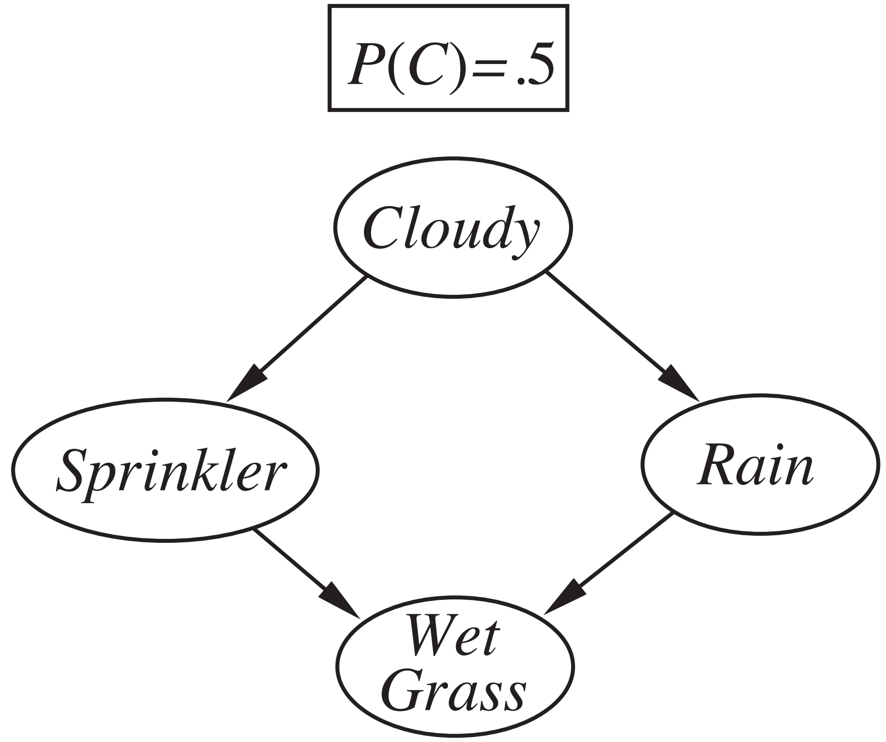
$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$




| | |
|---|---|
| *Cloudy* | *true* |
| *Sprinkler* | *true* |
| *Rain* | *false* |
| *WetGrass* | *true* |

$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$

$$\mathbf{P}(Cloudy \mid Sprinkler{=}true,\ Rain{=}false)$$



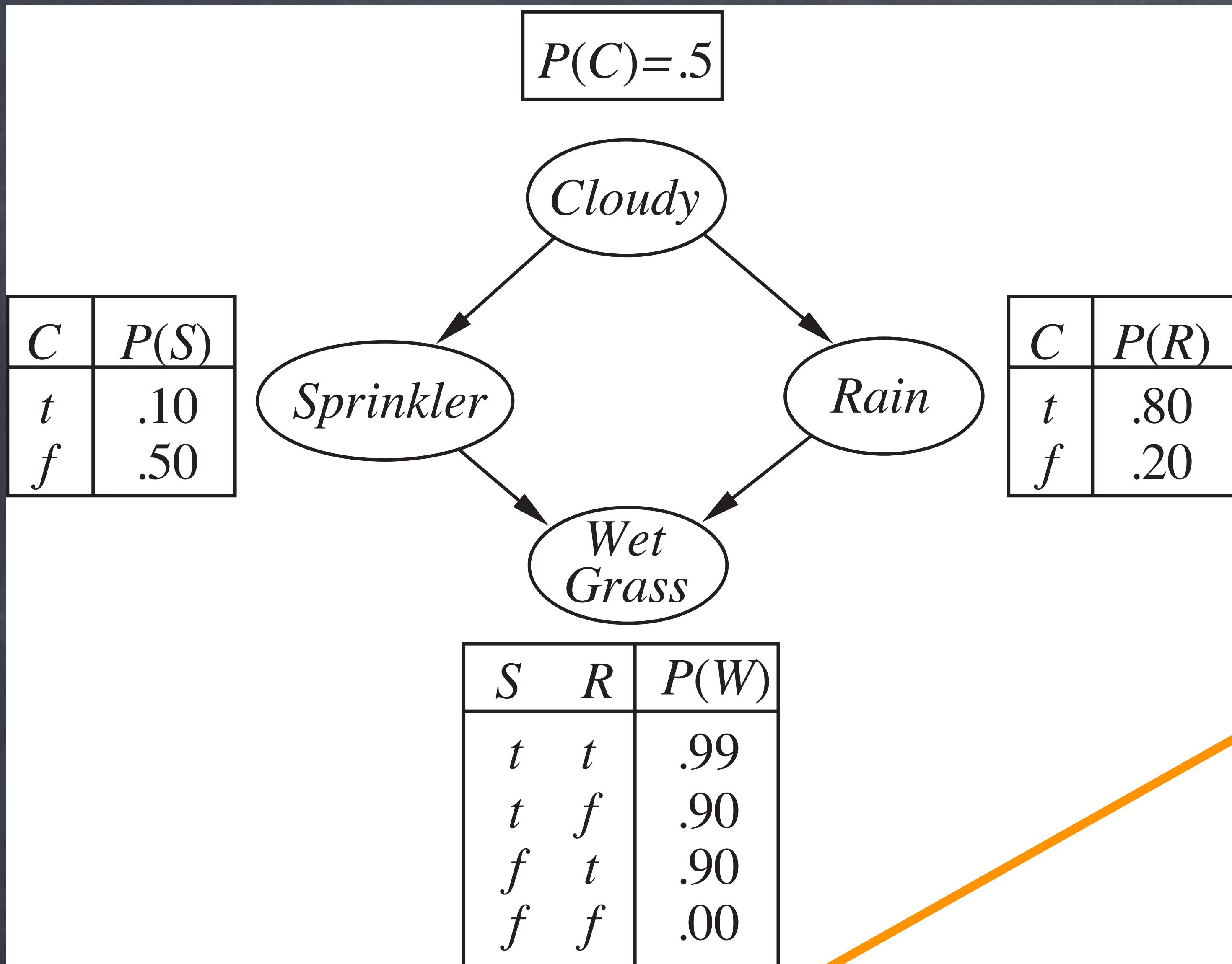

| C | P(S) |
|---|---|
| t | .10 |
| f | .50 |

| C | P(R) |
|---|---|
| t | .80 |
| f | .20 |

| S | R | P(W) |
|---|---|---|
| t | t | .99 |
| t | f | .90 |
| f | t | .90 |
| f | f | .00 |

| | |
|---|---|
| *Cloudy* | *false* |
| *Sprinkler* | *true* |
| *Rain* | *false* |
| *WetGrass* | *true* |

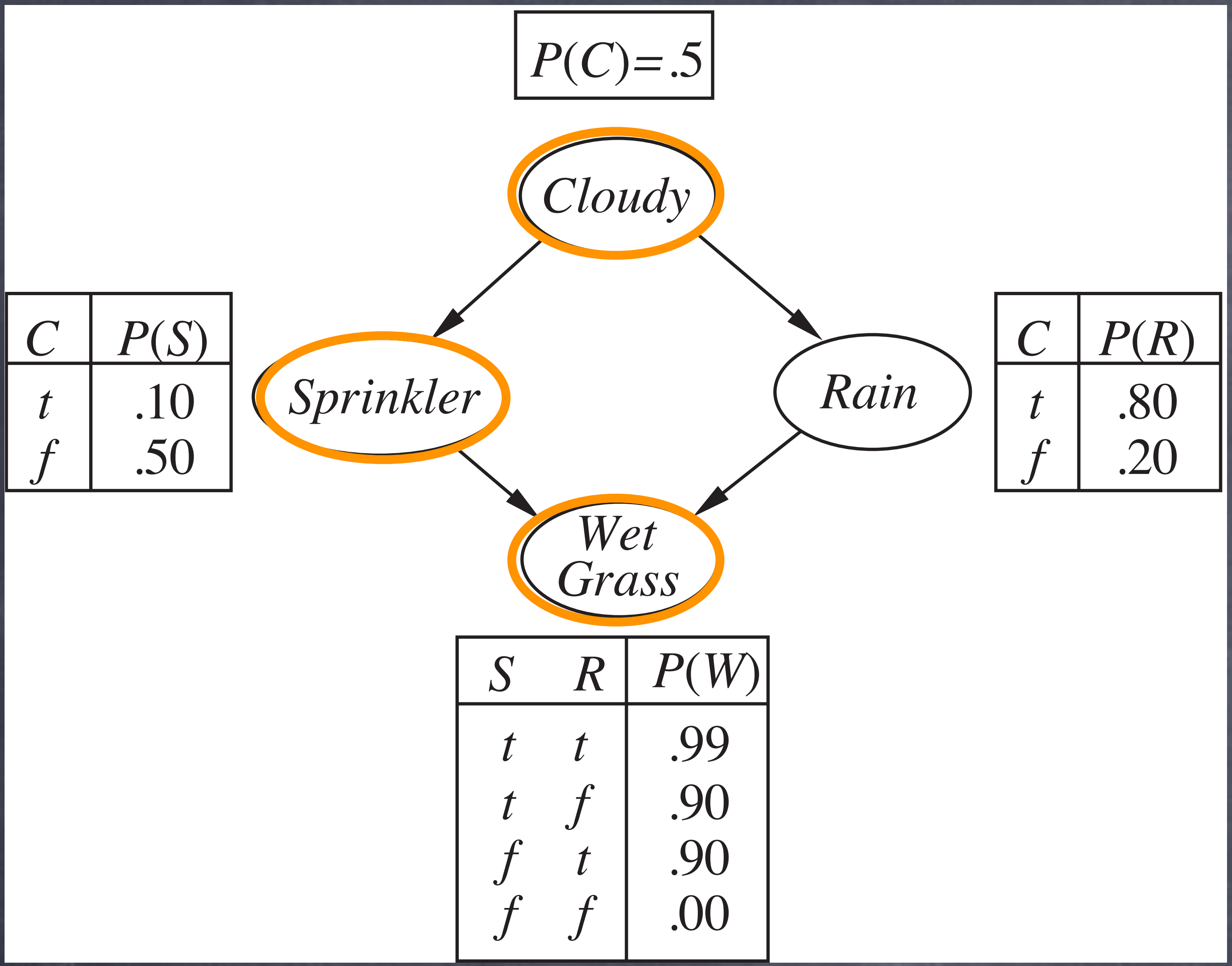
$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$




| | |
|---|---|
| $Cloudy$ | $false$ |
| $Sprinkler$ | $true$ |
| $Rain$ | $false$ |
| $WetGrass$ | $true$ |

$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$

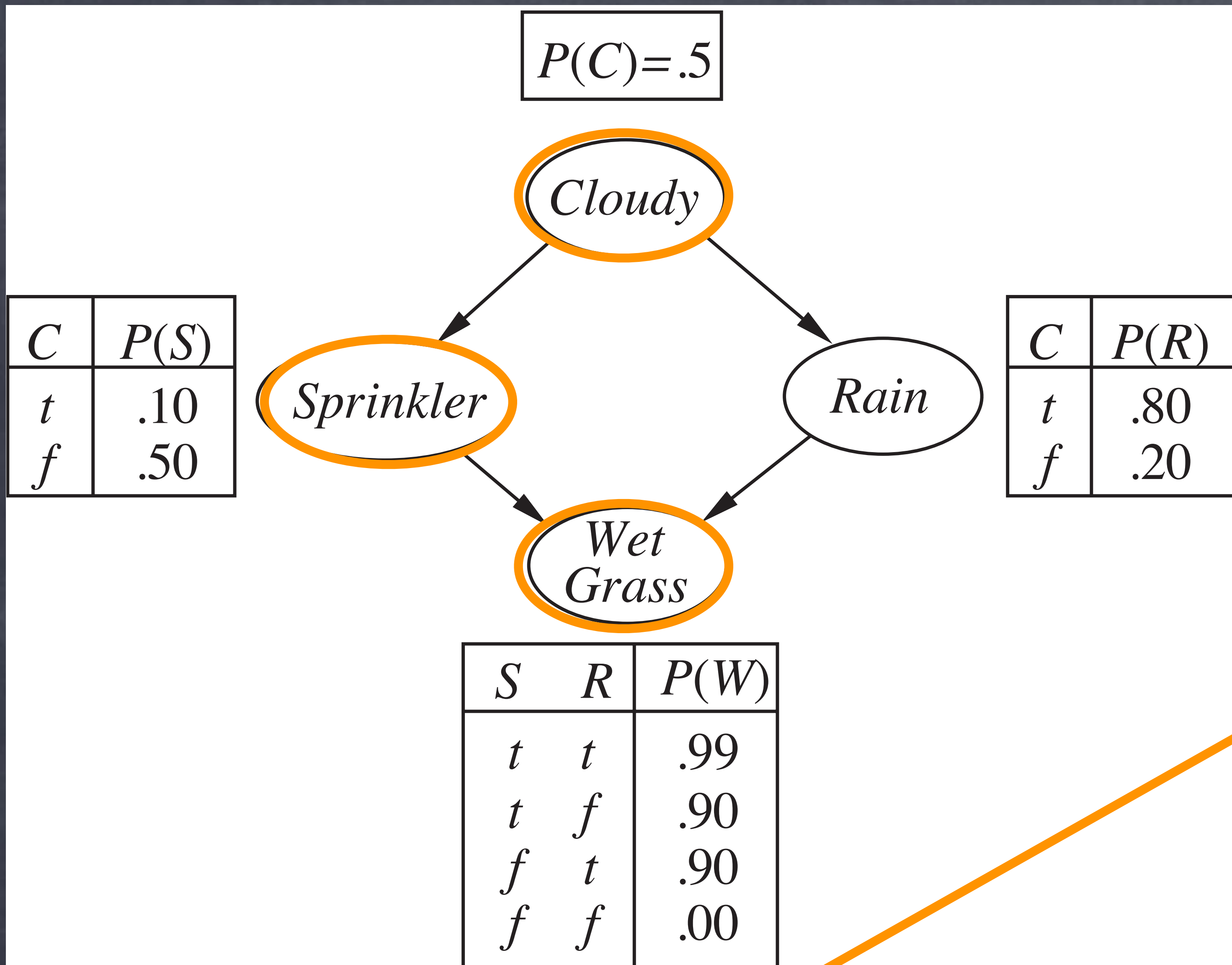
$$N_{Rain=false} \mathrel{+}= 1$$




| | |
|---|---|
| *Cloudy* | *false* |
| *Sprinkler* | *true* |
| *Rain* | *false* |
| *WetGrass* | *true* |

$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$

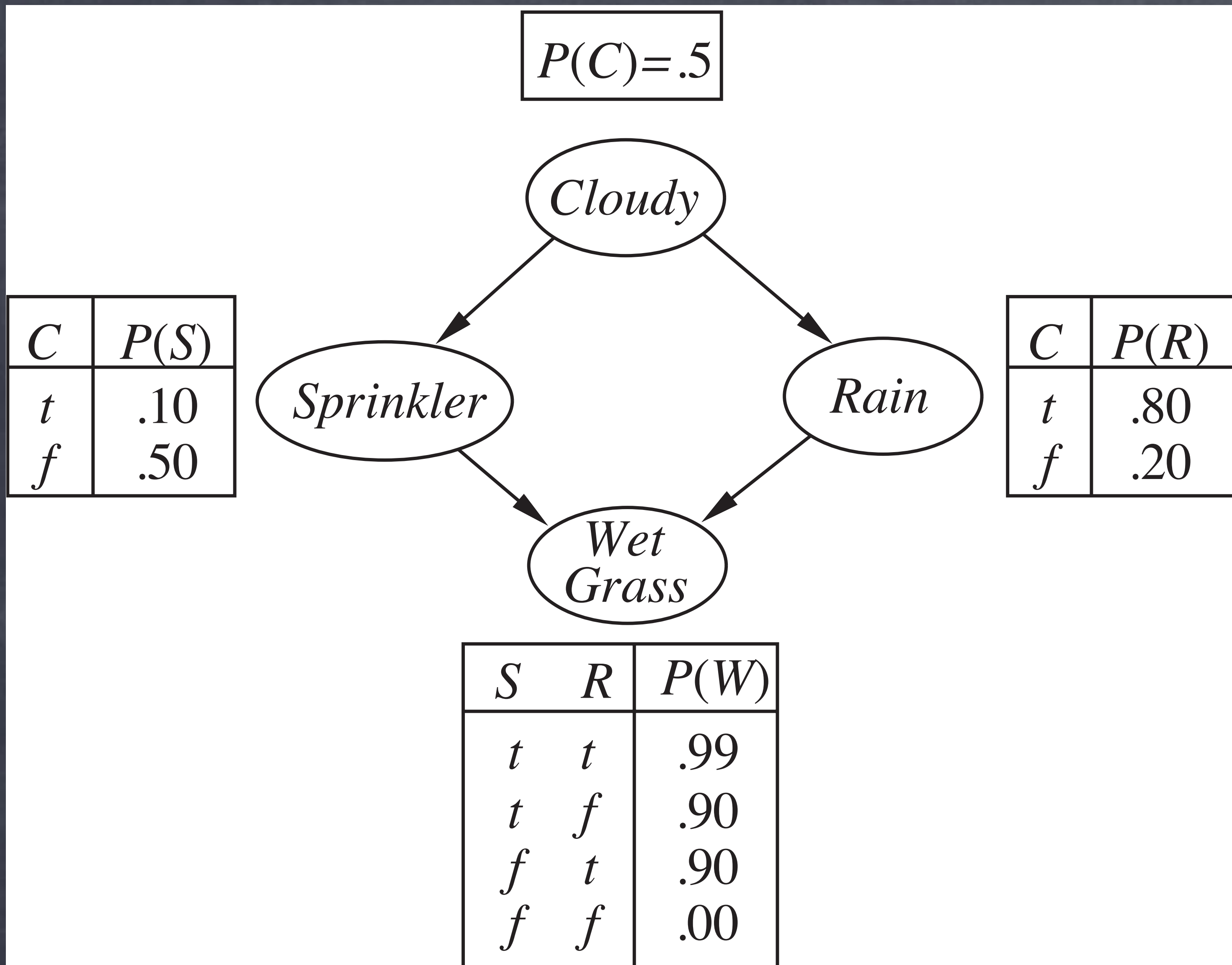
$$\mathbf{P}(Rain \mid Sprinkler{=}true,\ Rain{=}false,\ Cloudy{=}false)$$




| | |
|---|---|
| $Cloudy$ | $false$ |
| $Sprinkler$ | $true$ |
| $Rain$ | $true$ |
| $WetGrass$ | $true$ |

$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$

$$N_{Rain=true} \mathrel{+}= 1$$




| | |
|---|---|
| $Cloudy$ | $false$ |
| $Sprinkler$ | $true$ |
| $Rain$ | $true$ |
| $WetGrass$ | $true$ |

$$\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$$

# Gibbs Sampling

- To approximate: $\mathbf{P}(X \mid \mathrm{e})$
- Start in a state with evidence variables set to evidence values (others arbitrary)
- On each step, sample non-evidence variables conditioned on the values of the variables in their Markov Blanket
  - Order irrelevant
- A form of local search!

# Exact Inference in Bayesian Networks

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\,\mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$
$$= \alpha \sum_{\mathbf{y}} \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- Intractable (#P-Hard)

# Approximate Inference in Bayesian Networks

- Sampling consistent with a distribution
- Rejection Sampling: rejects too much
- Likelihood Weighting: weights get too small
- Gibbs Sampling: MCMC algorithm
  - Similar to local search
- All generate consistent estimates (equal to exact probability in the large-sample limit)

For Next Time:

AIMA 15.0-15.3