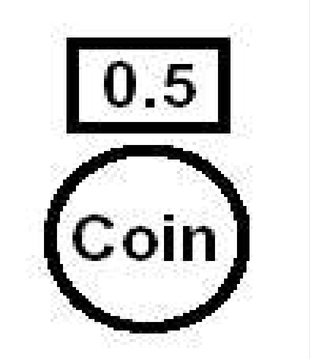
CSC421 Intro to Artificial Intelligence

UNIT 24:
Approximate Inference in Bayesian Networks

Inference by stochastic simulation

- Basic idea
  - Draw N samples from a sampling distribution S
  - Compute an approximate posterior P'
  - Show this converges to true prob. P



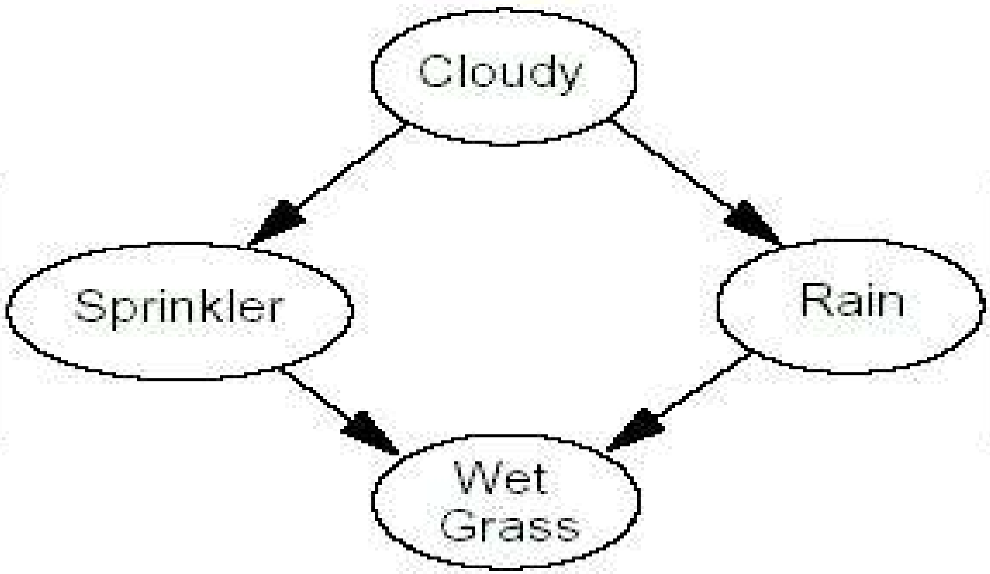
- Outline
  - Sampling from an empty network
  - Rejection sampling: reject samples disagreeing with evidence
  - Likelihood weighting: use evidence to weight samples
  - Markov Chain Monte Carlo (MCMC): sample from a stochastic process whose stationary distribution is the true posterior

Sampling from an empty network

- Generate events from a network with no evidence associated with it
- Idea sample each variable in turn, in topological order, conditioning variables appropriately
- Count actual samples generate
- Frequency converges the more samples we generate
- Consistent estimate = converges to true probability in the large-sample limit

Example

P(C)
.50

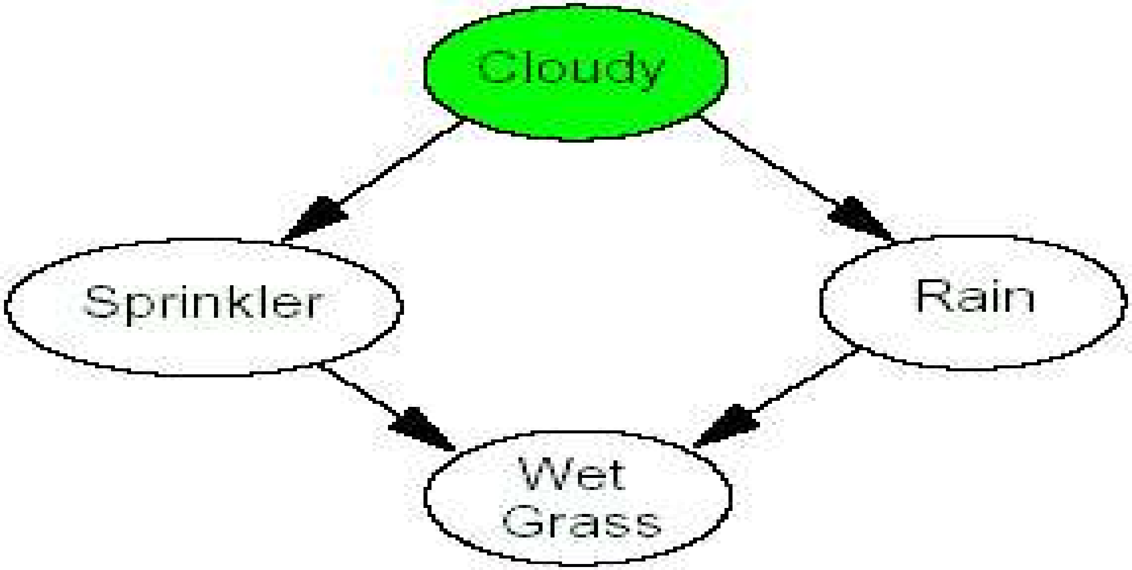


C	P(S|C)
T	.10
F	.50

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

Example



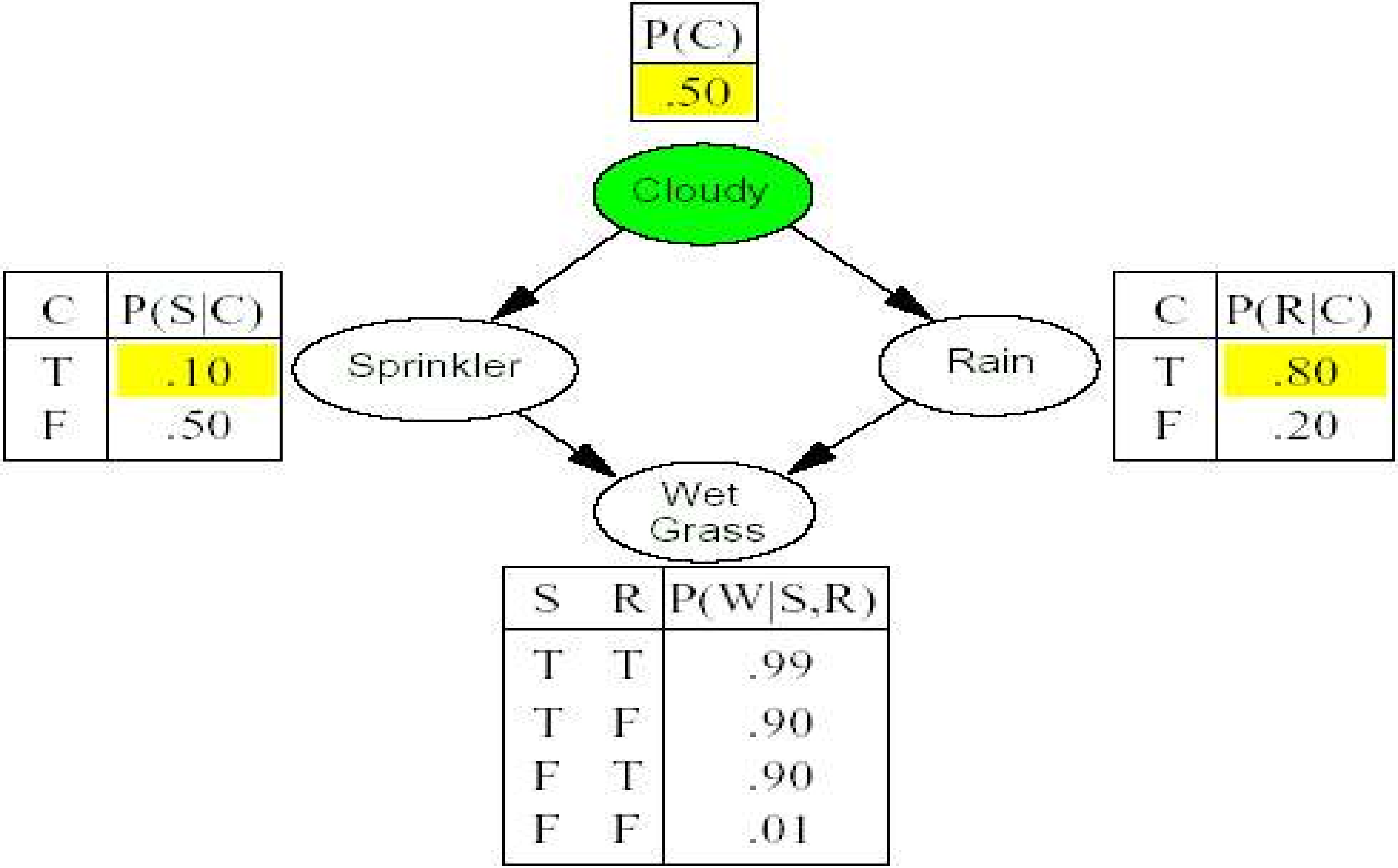
P(C)
.50

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

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

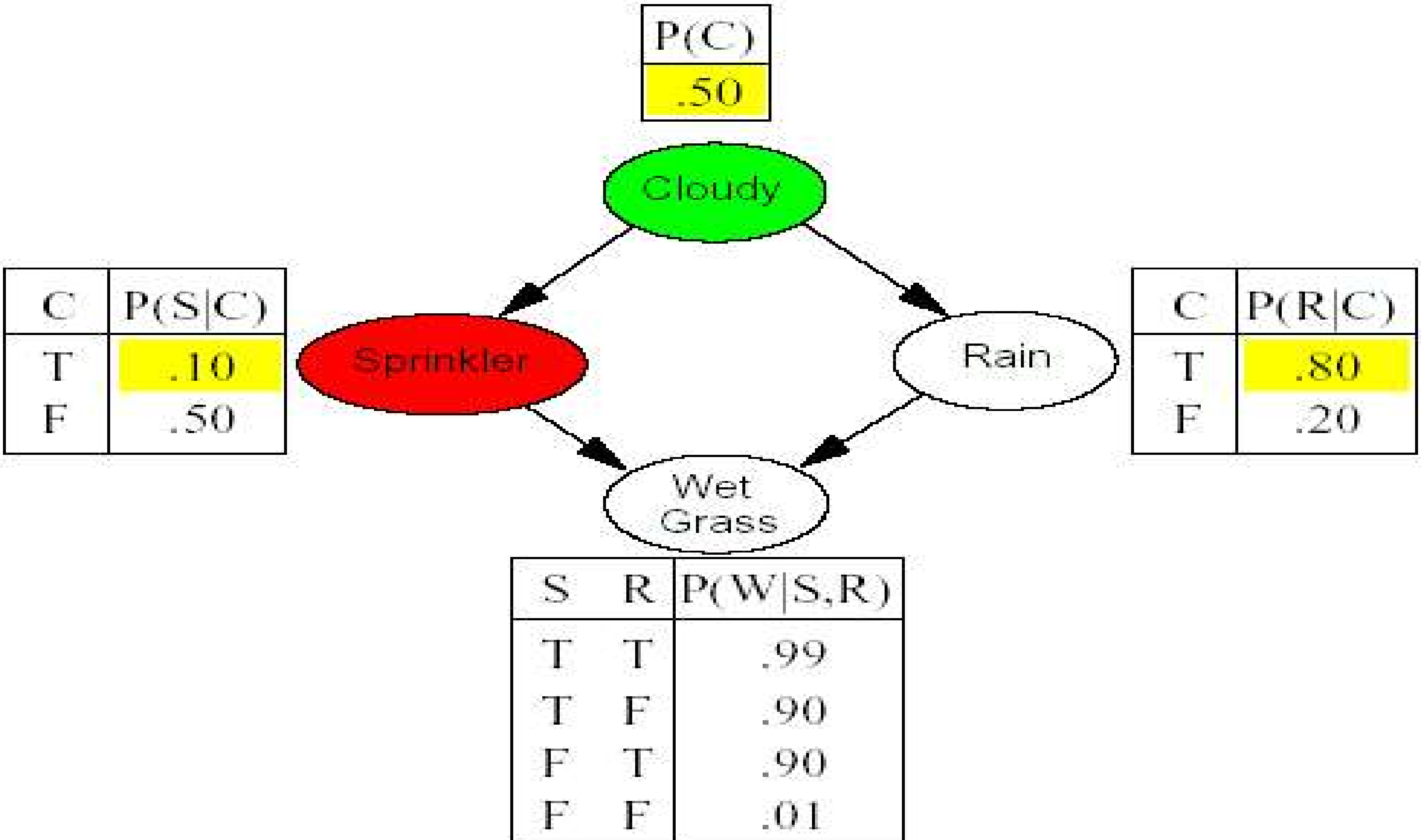
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Example



P(C)
.50
Cloudy
C | P(S|C)
T | .10
F | .50
Sprinkler
Rain
C | P(R|C)
T | .80
F | .20
Wet Grass
S | R | P(W|S,R)
T | T | .99
T | F | .90
F | T | .90
F | F | .01

Example



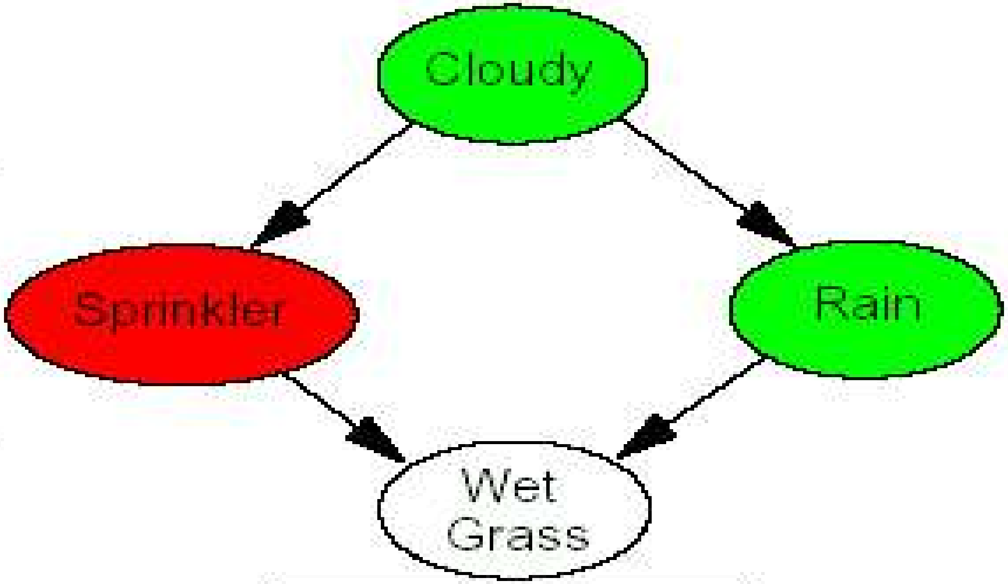
P(C)
.50

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

Example



P(C)
.50

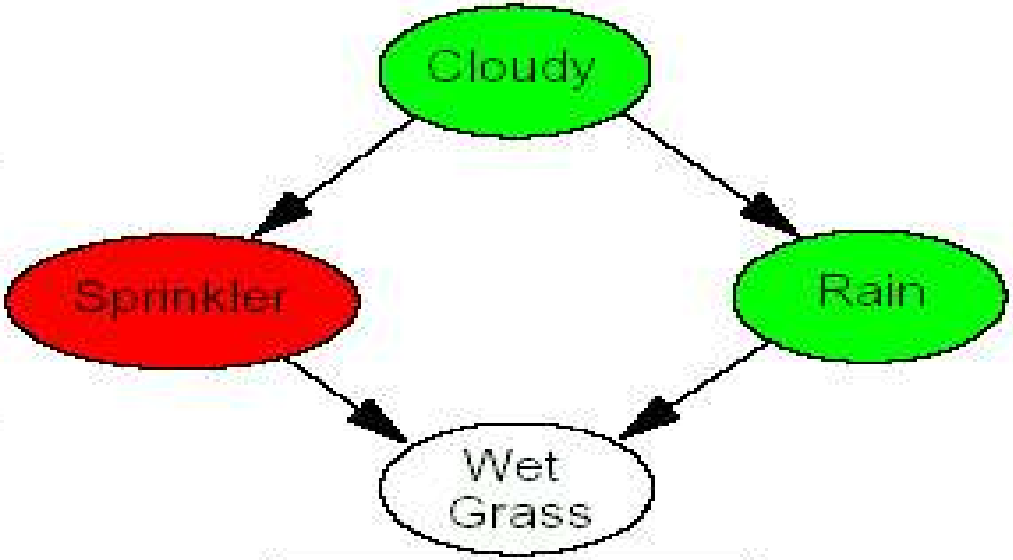
C	P(S|C)
T	.10
F	.50

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

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

Sampling empty network

- Probability that prior-samples generates a particular event
  - $SPS(x_1, \ldots, x_n) = ? = P(x_1, \ldots, x_n)$
  - e.g. SPS(t, f, t, t) = 0.5 * 0.9 * 0.8 * 0.9 = 0.234 = P (t,f,t,t)
- Let $N(x_1, \ldots, x_n)$ be the number of samples generated for event $(x_1, \ldots, x_n)$ then
  - $\lim_{n\to\infty} N(x_1, \ldots, x_n) / N = SPS(x_1, \ldots, x_n) = P(x_1, \ldots, x_n)$
- Estimates derived with PriorSample are consistent

Rejection sampling

- P'(X/e) estimated from samples agreeing with e
- e.g estimate P(Rain | Sprinkler = true) using 100 samples
- Using direct sampling gerate 100 samples
  - 27 samples have Sprinkler= true
  - Of these 8 have Rain = true and 19 have Rain = false
- P'(Rain| Sprinkler = true) = Norm(<8, 19>) = < 0.296, 0.704>
- Similar to basic real world estimation procedure - any problems ?

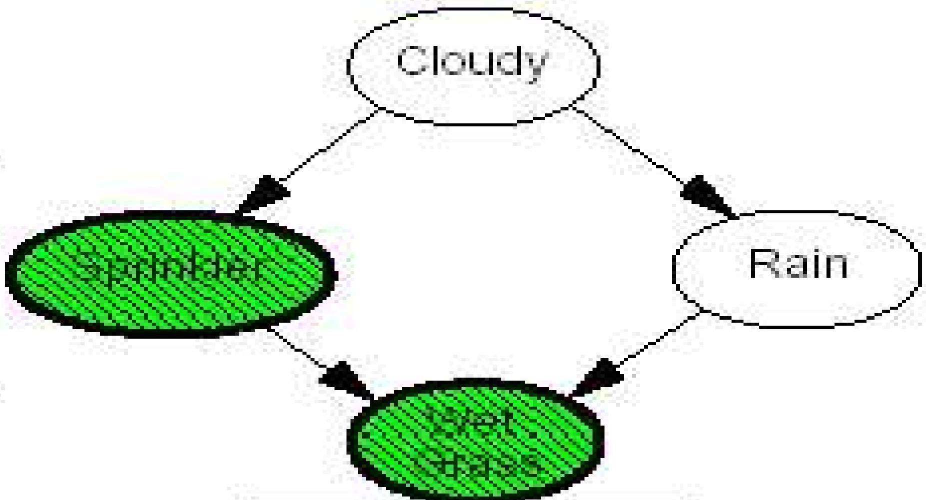
Analysis of rejection sampling

- Rejection sampling returns consistent posterior estimates
- Problem:
  - Hopelessly expensive if P(e) is small
  - Why ?
- P(e) drops exponentially with number of evidence variables

Likelihood Weighting

- Idea: fix evidence variables, sample only nonevidence variables and weight each sample with the likelihood it affords the evidence
- Produces consistent estimates (details in book)

Example



P(C)
.50

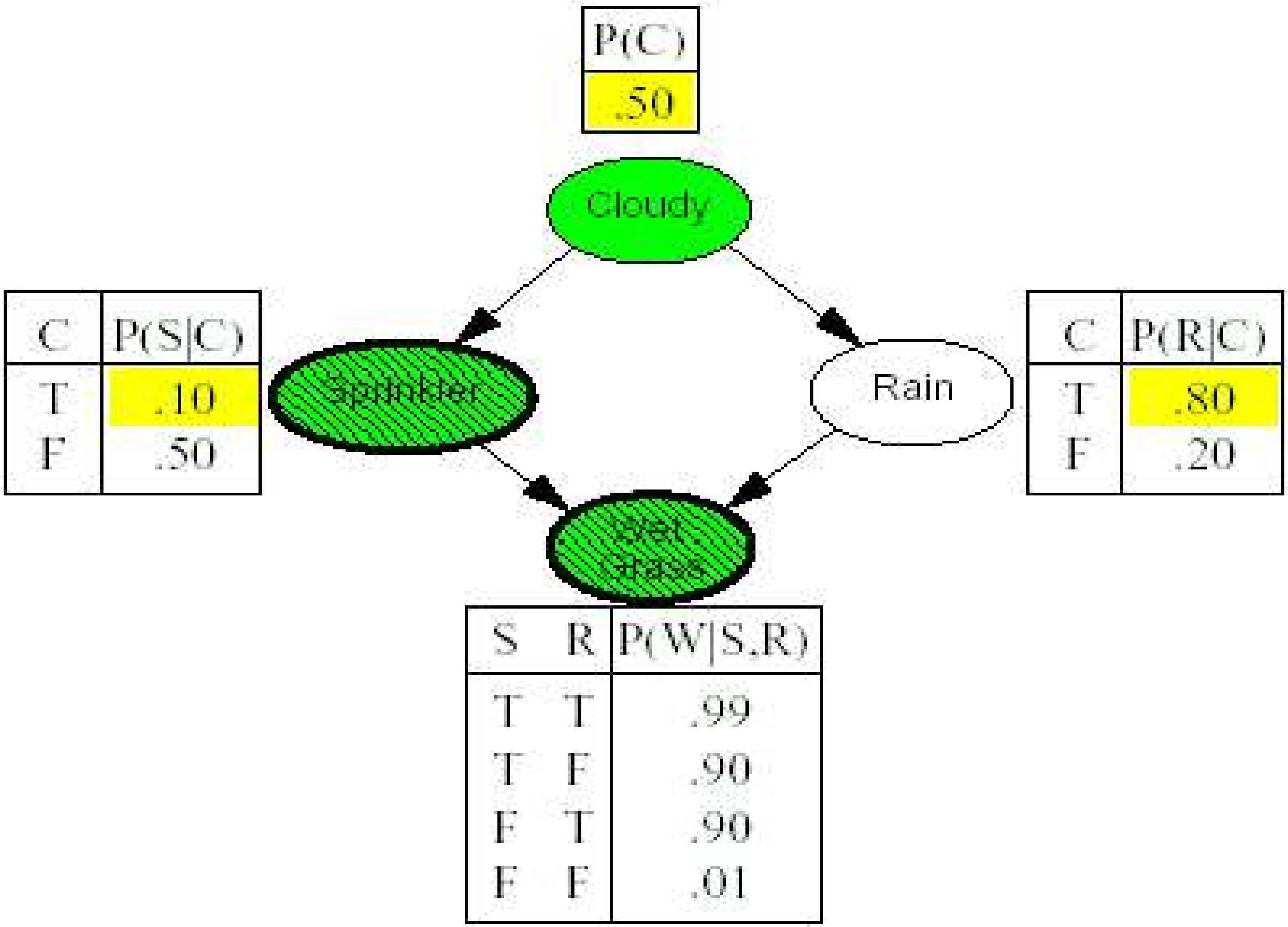
C	P(S|C)
T	.10
F	.50

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

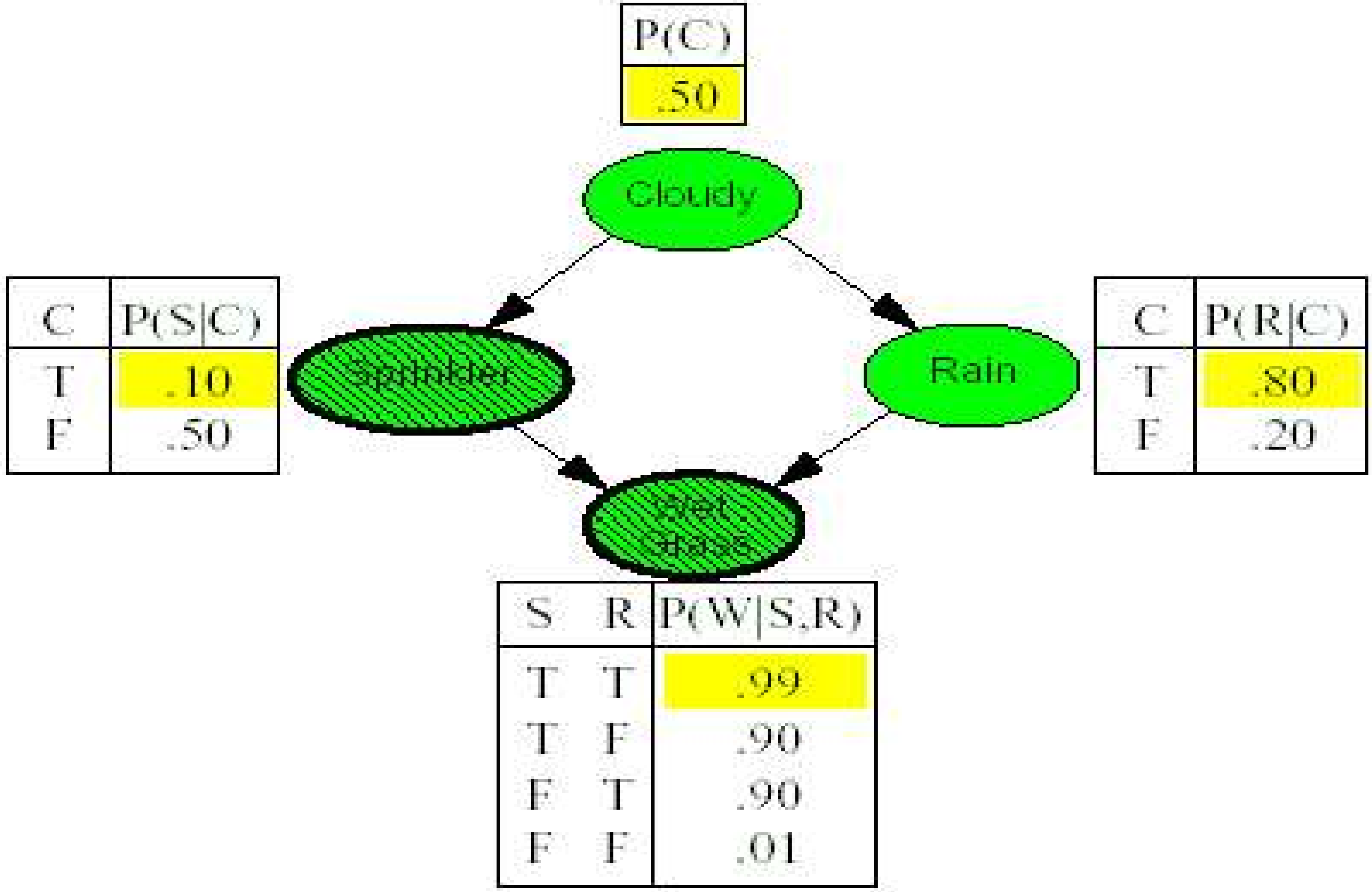
$w = 1.0$

Example



$w = 1.0 \times 0.1$

Example



P(C)
.50

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

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

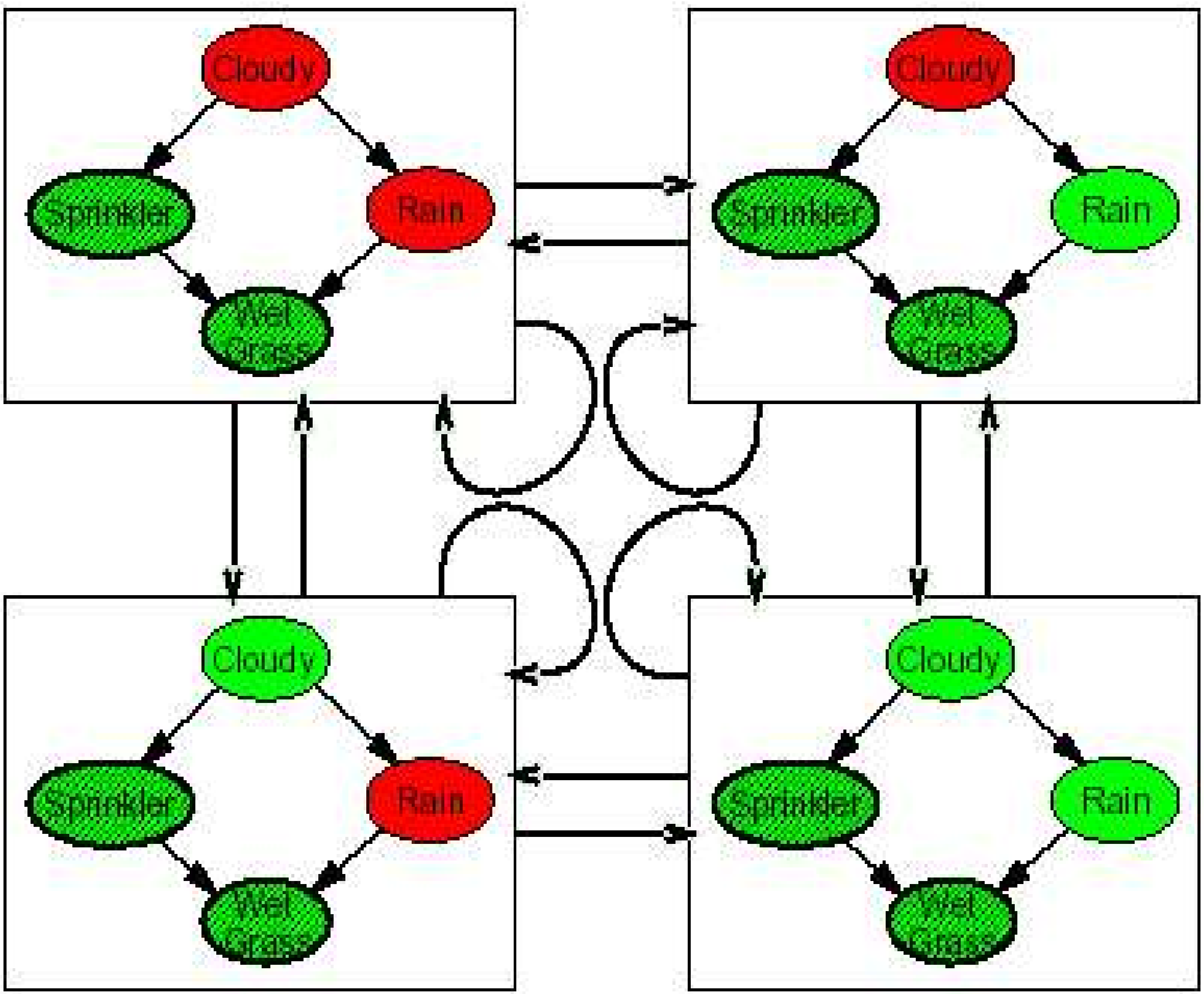
S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

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

Approximate inference using MCMC

- State of network = current assignment of all network variables
- Generate next state by sampling one variable given Markov Blanket
- Sample each variable in turn keeping evidence fixed
- Can also choose a variable to sample at random each time

Markov Chain



Wander
around a bit
average
what you see

Transition
probabilities

Every time
a variable
is sampled
new state

Markov Blanket

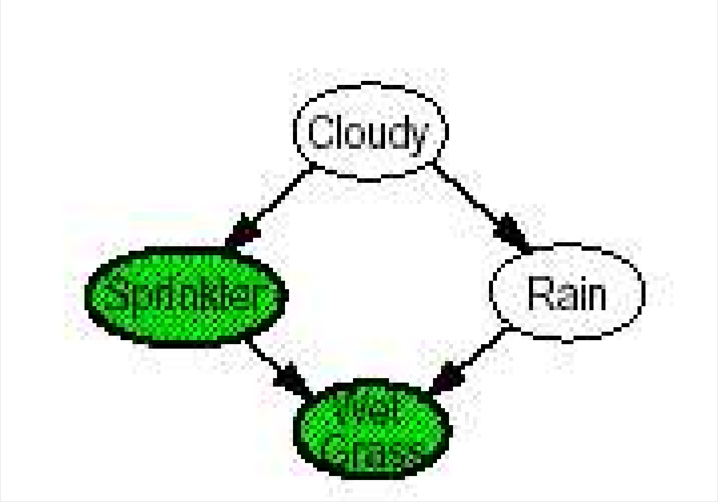
- Markov Blanket = the parents, children and children's parents
- Example:
  - P(Rain | Sprinkler = true, WetGrass = true)
  - Cloudy , Rainy initialized randomly [t,f]
  - [t, t, f, t]
- Cloudy is sampled from P(Cloudy | s, r). Suppose Cloudy = false. Then the new current state is [f, t, f,t].
- Repeat wandering around
- Each state is a sample is an estimate.

Why this works ?

- Theorem:
  - Chain approaches stationary distribution: long run fraction of time spent in each state is exactly proportional to its posterior probability

Markov Blanket Sampling

- Markov Blanket of Cloudy is Sprinkler and Rain
- Markov Blanket of Rain is Cloudy, Sprinkler,and Wet Grass
- Prob. given Markov blanket = $P(x \mid MB(X)) = P(x \mid Parents(X)) * \prod_{z \in children(X)} P(z \mid parents(Z))$



A little sidenote on Financial Engineering

- Financial instruments
  - Ways to manage risk
- Example: livestock producer purchase futures (type of instrument) to cover feed costs, so that they can plan on a fixed cost of feed. Typically the price is higher than the current market price
- If in future price drops bank is profitable if price goes up banks might loose money
- Question: how much to charge extra for the risk protection ?

Summary

- Exact inference by variable elimination
  - Polytime on polytrees, NP-hard on general graphs
  - Space = time, very sensitive to topology
- Approximate inference by LW, MCMC
  - LW, MCMC generally insensitive to topology
  - Convergence can be slow with probabilities close to 0 or 1
  - Can handle arbitrary combinations of discrete and continuous random variables
  - USED HEAVILY IN PRACTICE