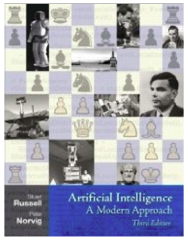
Uncertainty and Bayesian Networks

- Uncertainty
- Probabilities
- Syntax
- Semantics
- Parametrized Distributions



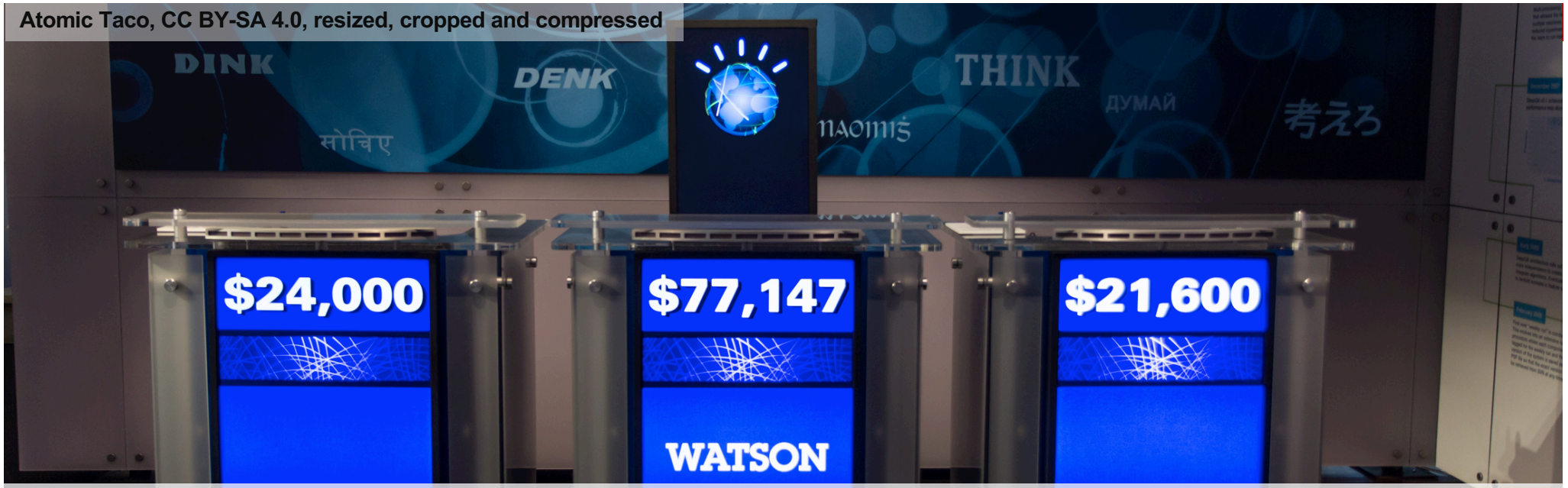
Many slides based on Russell & Norvig's slides Artificial Intelligence: A Modern Approach

Uncertain Actions

- So far, our agents believe that
  - (logical) statements are true or false (maybe unknown)
  - actions will always do what they think they do
- Unfortunately, the real world is not like that
  - agents almost never have access to the whole truth about the world

→ agents must deal with uncertainty

- Example:
  - We have different actions for getting to the airport:
    - action A_t = leave for the airport t minutes before departure
  - Typical problems:
    - Will a given action A_t get me to the airport in time?
    - Which action is the best choice for getting me to the airport?



Atomic Taco, CC BY-SA 4.0, resized, cropped and compressed

Many situations are uncertain.
Agents have to deal with these uncertainties.

Steve Jurvetson, CC BY 2.0, resized, cropped and compressed

Problems with Uncertainty

We leave 90 minutes before departure

- Risks involved in the plan A_{90} will get me to the airport
  - partial observability (road state, other drivers' plans, etc.)
  - noisy sensors (traffic reports may be wrong)
  - uncertainty in action outcomes (flat tire, accident, etc.)
  - immense complexity of modeling and predicting traffic
- A logically correct plan:

A_{90} will get me to the airport as long as my car doesn't break down, I don't run out of gas, no accident, the bridge doesn't fall down, **etc.**

  - impossible to model all things that can go wrong
    - → recall the **qualification problem**
- A more cautious plan:

A_{1440} will get me to the airport

  - will (virtually) certainly succeed, but clearly suboptimal
    - e.g., we have to pay for a night in a hotel

Probabilities

- Probabilities are **one way** of handling uncertainty
  - e.g. A_{90} will get me to the airport with probability 0.5
- The probability **summarizes effects** that are due to
  - Laziness
    - I don't want to list all things that must not go wrong
  - Theoretical Ignorance
    - Some things just can't be known
      - e.g.: We cannot completely model the weather
  - Practical Ignorance
    - Some things might not be known about the particular situation
      - e.g. Is there a traffic jam at A5?

Meaning of Probabilities

- Probabilities that are related to one's **(subjective) beliefs**
  - a probability p attached to a statement means that I believe that the statement will be true in $p \cdot 100\%$ of the cases: there is traffic jam on the A5 in 10% of the cases (meaning: there might be jam, but usually there is none)
  - it does not mean that it is true with $p\%$: the traffic on the A5 is jammed with a degree of 10% (meaning: there's a jam, but it could be worse...)
- A probability may also capture the outcomes of **experiments.** But then consider the probability that the sun will still exist tomorrow; difficult to observe by an experiment
- What is the chance that a patient has a particular disease? Doctor wants to consider other patients who are similar. But if you gather too much information to compare patients, there are no similar patients left

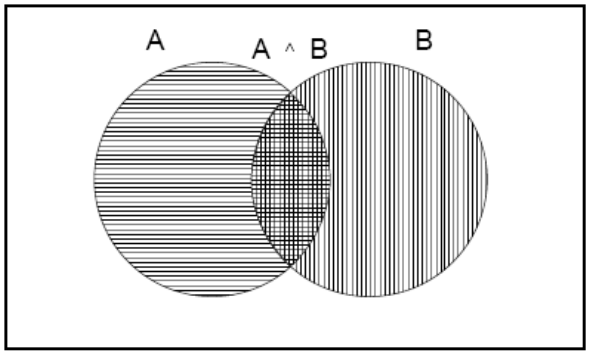
Kolmogorov's Axioms of Probability

1. All probabilities are between 0 and 1

$$0 \leq P(a) \leq 1$$

2. Necessarily true propositions have probability 1, necessarily false propositions have probability 0

$$P(false) = 0 \qquad P(true) = 1$$



3. The probability of a disjunction is

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

4. These axioms restrict the set of probabilistic beliefs that an agent can (reasonably) hold.
similar to logical constraints like A and $\neg A$ can't both be true

Violation of Axioms of Probability

„put its money where its probabilities are“

- Dutch Book Theorem, Bruno de Finetti (1931)
  - an agent (in the example it is Agent 1) who bets according to probabilities that violate the axioms of probability can be forced to bet so as to lose money *regardless of outcome!*

Example:

- suppose Agent 1 believes the following

$$P(a) = 0.4 \qquad P(b) = 0.3 \qquad P(a \vee b) = 0.8$$

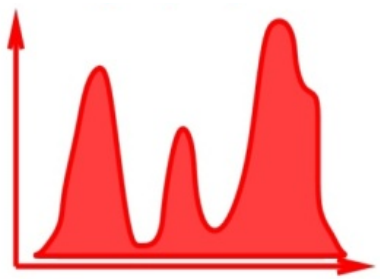
axioms of probability are violated because $P(a \vee b) > P(a) + P(b)$

- Agent 2 can now select a set of events and bet on them according to these probabilities so that she cannot loose

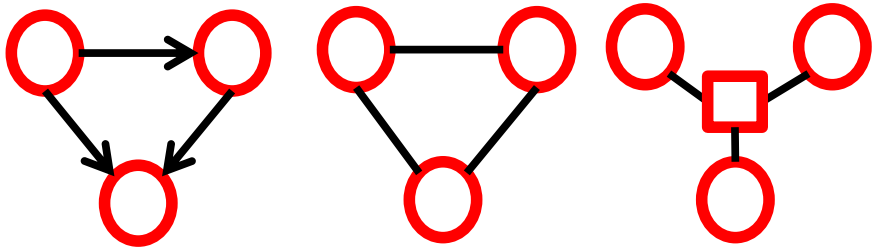
Agent 1		Agent 2		Outcome for Agent 1			
proposition	belief	bet	stakes	$a \wedge b$	$a \wedge \neg b$	$\neg a \wedge b$	$\neg a \wedge \neg b$
a	0.4	a	4:6	-6	-6	4	4
b	0.3	b	3:7	-7	3	-7	3
$a \vee b$	0.8	$\neg(a \vee b)$	2:8	2	2	2	-8
				-11	**-1**	**-1**	**-1**

How do we deal with uncertaints in AI

Probabilsitic graphical models are a graphical notation for (conditional) independency assumptions and therefore a (hopefully) compact specification of probability distributions



Nodes=
Random Variables (RVS)
Edges=
Dependencies among RVs

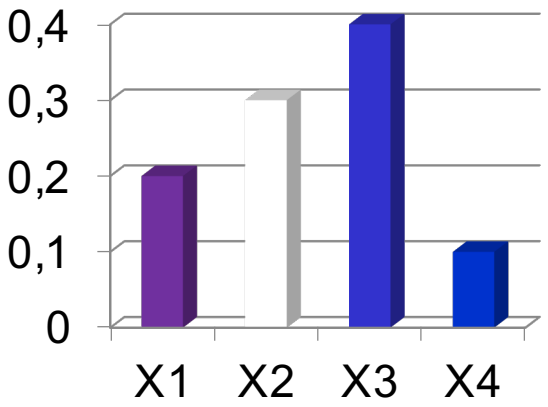


Discrete Random Variables

- Finite set X of possible states

$$X \in \{x_1, x_2, x_3, \ldots, x_n\}$$

$$P(x_i) \geq 0 \quad \sum_{i=1}^{n} P(x_i) = 1$$



OK, but answering the question requires the joint distribution

What is the probability that smoking causes cancer?

Smoking

no	few	many
0.800	0.150	0.050

Cancer

no	benigne	maligne
0.935	0.046	0.019

Joint Distribution

- Probability that X=x and Y=y are “true”

$$P(x, y) \equiv P(X = x \wedge Y = y)$$

Cancer

Smoking	no	benigne	maligne
no	0.768	0.024	0.008
few	0.132	0.012	0.006
many	0.035	0.010	0.005

- The joint distribution allows us to answer any question! But how?

Make use of basic probability theory

- **Marginalization**

$$P(Y) = \sum_{i=1}^{n} P(Y, x_i)$$

sum →

Cancer

Smoking	No	Benigne	Maligne	TOTAL
No	0.768	0.024	0.008	0.800
few	0.132	0.012	0.006	0.150
many	0.035	0.010	0.005	0.050
TOTAL	0.935	0.046	0.019	

P(smoking)

P(cancer)

- Product rule & **conditional probability**

$$P(X, Y) = \boxed{P(X \mid Y)} P(Y) = P(Y \mid X) P(X)$$

Probability that X=x if we have observed Y=y (assuming P(y)>0)

Probably the most important rule: **Bayes' rule**

Product rule $$P(R,K) = P(R \mid K)P(K) = P(K \mid R)P(R)$$

Bayes' rule $$P(R \mid K) = \frac{P(K \mid R)P(R)}{P(K)} = \frac{P(K,R)}{P(K)}$$

cancer

smoking	no	benigne	maligne
no	0.768	0.024	0.008
few	0.132	0.012	0.006
many	0.035	0.010	0.005
TOTAL	0.935	0.046	0.019

P(caner)

Since we know already P(R,K) und auch P(K), just divide them.

Probably the most important rule: **Bayes' rule**

Product rule $$P(R,K) = P(R \mid K)P(K) = P(K \mid R)P(R)$$

Bayes' rule $$P(R \mid K) = \frac{P(K \mid R)P(R)}{P(K)} = \frac{P(K,R)}{P(K)}$$

cancer

smoking	no	benigne	maligne
no	0.768/0.935	0.024/ 0.46	0.008/ 0.019
few	0.132/0.935	0.012/ 0.46	0.006/ 0.019
many	0.035/0.935	0.010/ 0.46	0.005/ 0.019
TOTAL	0.935	0.046	0.019

P(cancer)

Since we know already P(R,K) und auch P(K), just divide them.

Probably the most important rule: **Bayes' rule**

Product rule $$P(R,K) = P(R \mid K)P(K) = P(K \mid R)P(R)$$

Bayes' rule $$P(R \mid K) = \frac{P(K \mid R)P(R)}{P(K)} = \frac{P(K,R)}{P(K)}$$

cancer= …)

P(smoking=…|

	no	benigne	maligne
no	0.821	0.522	0.421
few	0.141	0.261	0.316
many	0.037	0.217	0.263

Since we know already P(R,K) und auch P(K), just divide them.

Example: AIDS-Test

- $Aids$ = a person has Aids or not
- $Positive$ = a person has a positive test result

- Assume the test has the following characteristics:

$$P(positive|aids) = 0.99$$
$$P(negative|aids) = 0.01$$
$$P(positive|\neg aids) = 0.005$$
$$P(negative|\neg aids) = 0.995$$

The test makes 1% mistakes for people that have aids

The test makes 0,5% mistakes for people that don't have aids

- Looks like a pretty reliable test?

Example: AIDS-Test

- $Aids$ = a person has Aids or not
- $Positive$ = a person has a positive test result

- Assume the test has the following characteristics:

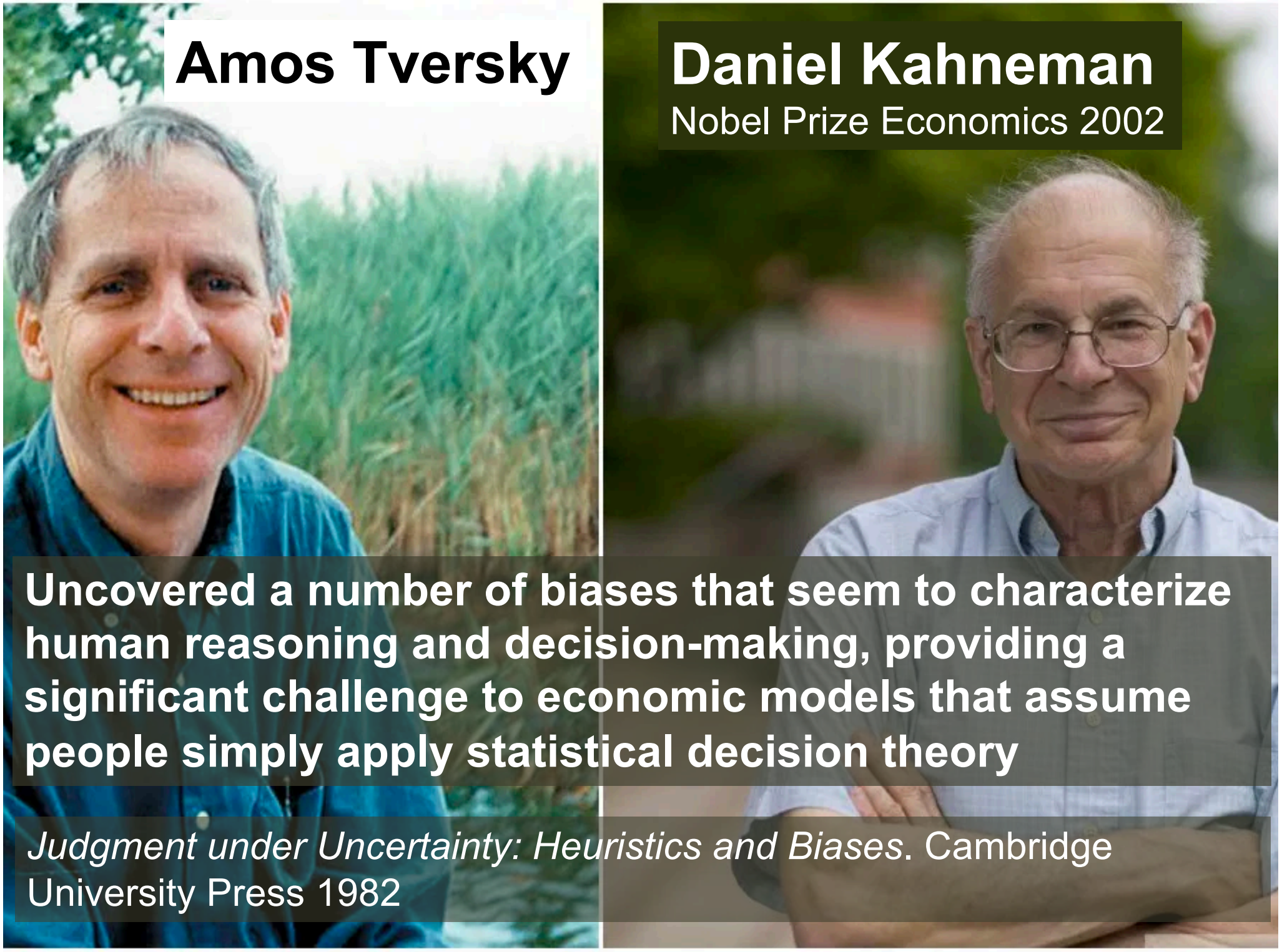
$$P(positive|aids) = 0.99$$
$$P(negative|aids) = 0.01$$
$$P(positive|\neg aids) = 0.005$$
$$P(negative|\neg aids) = 0.995$$

The test makes 1% mistakes for people that have aids

The test makes 0,5% mistakes for people that don't have aids

- Now suppose you are in a low-risk group (low a priori probability of having Aids, say $P(aids) = 0.0001$) and have a positive test result. How much should you be concerned?

$$P(a|p) = \frac{P(p|a)\cdot P(a)}{P(p)} = \frac{P(p|a)\cdot P(a)}{P(p|a)\cdot P(a) + P(p|\neg a)\cdot P(\neg a)} = \frac{0.99\cdot 0.0001}{0.99\cdot 0.0001 + 0.005\cdot 0.9999} = 0.0194$$


Amos Tversky
Daniel Kahneman
Nobel Prize Economics 2002
Uncovered a number of biases that seem to characterize human reasoning and decision-making, providing a significant challenge to economic models that assume people simply apply statistical decision theory
Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press 1982

Anyhow, while the joint distribution allows one to compute everything, a tabular representation is rather inefficient

- Joint distribution is enumerating everything
  - Worst-case run time: $O(2^n)$
    - n = # of RVs
  - Space is $O(2^n)$ too
    - Size of the table of the joint distribution

Main idea: make use of independencies to compress the representation

Indpendency

- (Current) age and the gender of a person are independent

$$P(G, A) = P(G) \cdot P(A)$$

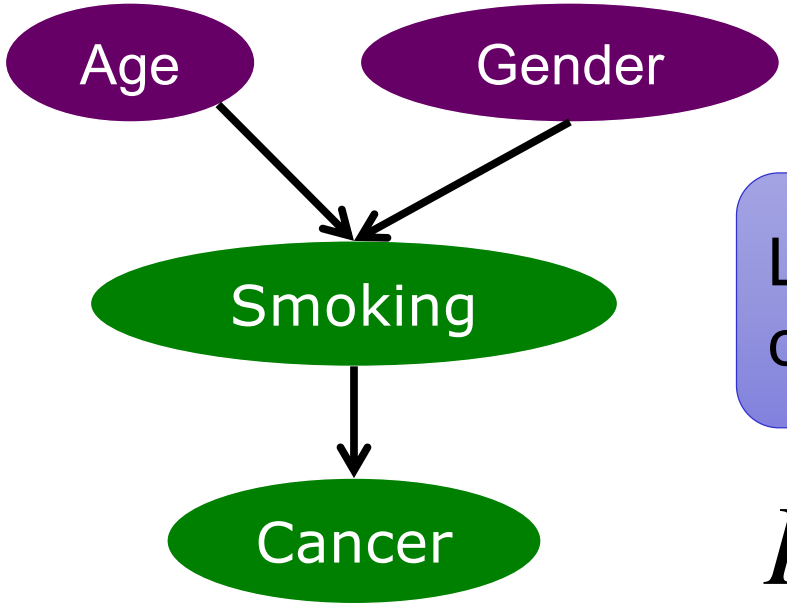
$$P(A \mid G) = P(A)$$

$$P(G \mid A) = P(G)$$

You would not give me money for information on the gender to know the age of a person!

Conditional Independence

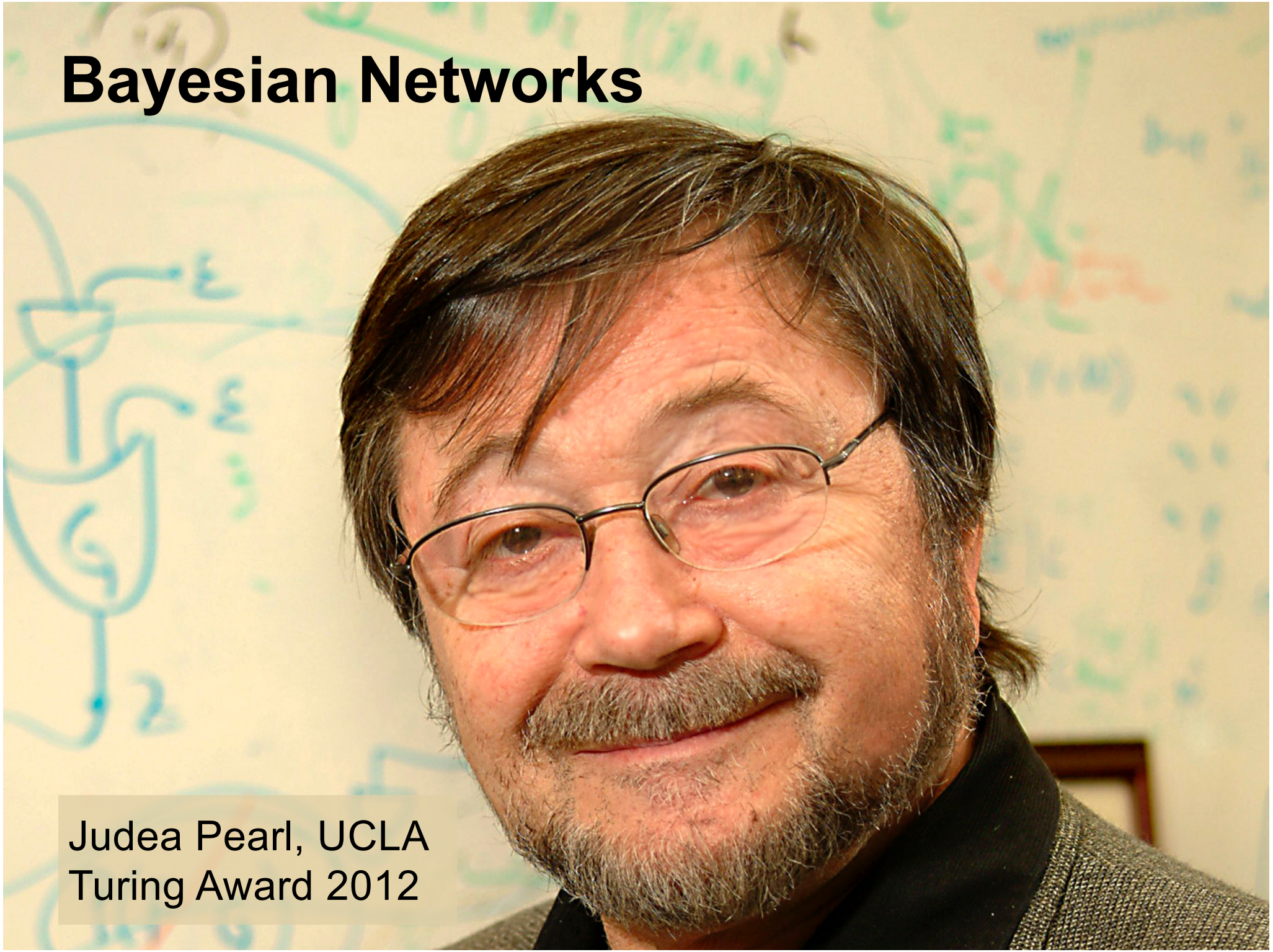
- Cance is independent of age and gender, if the person smokes.
- If you have not observed anything,age and gender are independent.



Less entries, therefore lower complexity

$$P(C \mid S, G, A) = P(C \mid S)$$

Bayesian Networks

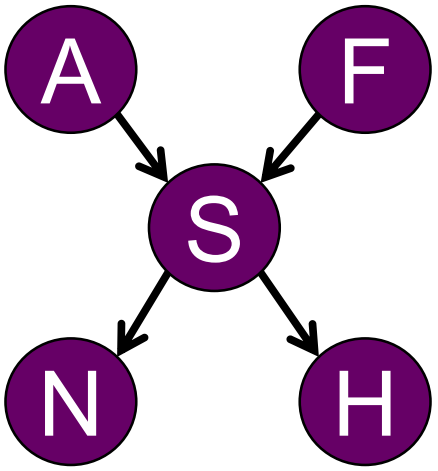


Judea Pearl, UCLA
Turing Award 2012

Bayesian Networks [Pearl 1989]

- Set of random variables $\{X_1,\ldots,X_n\}$
- **Directed, acyclic graph (DAG)**
- To each RV Xi we associate the
- **conditional probabilitiy distribution:**

$$P\left(X_i \mid \mathrm{Pa}\left(X_i\right)\right)$$



- The **joint distribution** is

$$P\left(X_1,\ldots,X_n\right)=\prod_{i=1}^{n} P\left(X_i \mid \mathrm{Pa}\left(X_i\right)\right)$$

BN semantics

- **Local Markov Assumption**

Each RV X is independent of ist „non-descendant“ given ist parents (Xi ⊥ nonDescendants| Pa_{Xi})

A very simple example

But how do we do inference?

$R \in \{no, few, many\}$

P(S=n)	0.80
P(S=f)	0.15
P(S=m)	0.05

$K \in \{no, benigne, maligne\}$

Smoking=	n	f	m
P(C=n)	0.96	0.88	0.60
P(C=b)	0.03	0.08	0.25
P(C=m)	0.01	0.04	0.15

What is Probabilistic Inference?

- **Query:** $P(X \mid e)$
- **Definiton of conditional probability** $P(X \mid e) = \frac{P(X, e)}{P(e)}$
- **Up to normalization** $P(X \mid e) \propto P(X, e)$

Hence, this rewrites to

$$P(\mathbf{Y}) = \sum_{X_i \notin \mathbf{Y}} \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i))$$

BN semantics

Marginalization

Main observation: Σ and ∏ commute

$\Sigma_a(P_1 \times P_2) = (\Sigma_a P_1) \times P_2$ if A is not in P_2

Variable Elimination

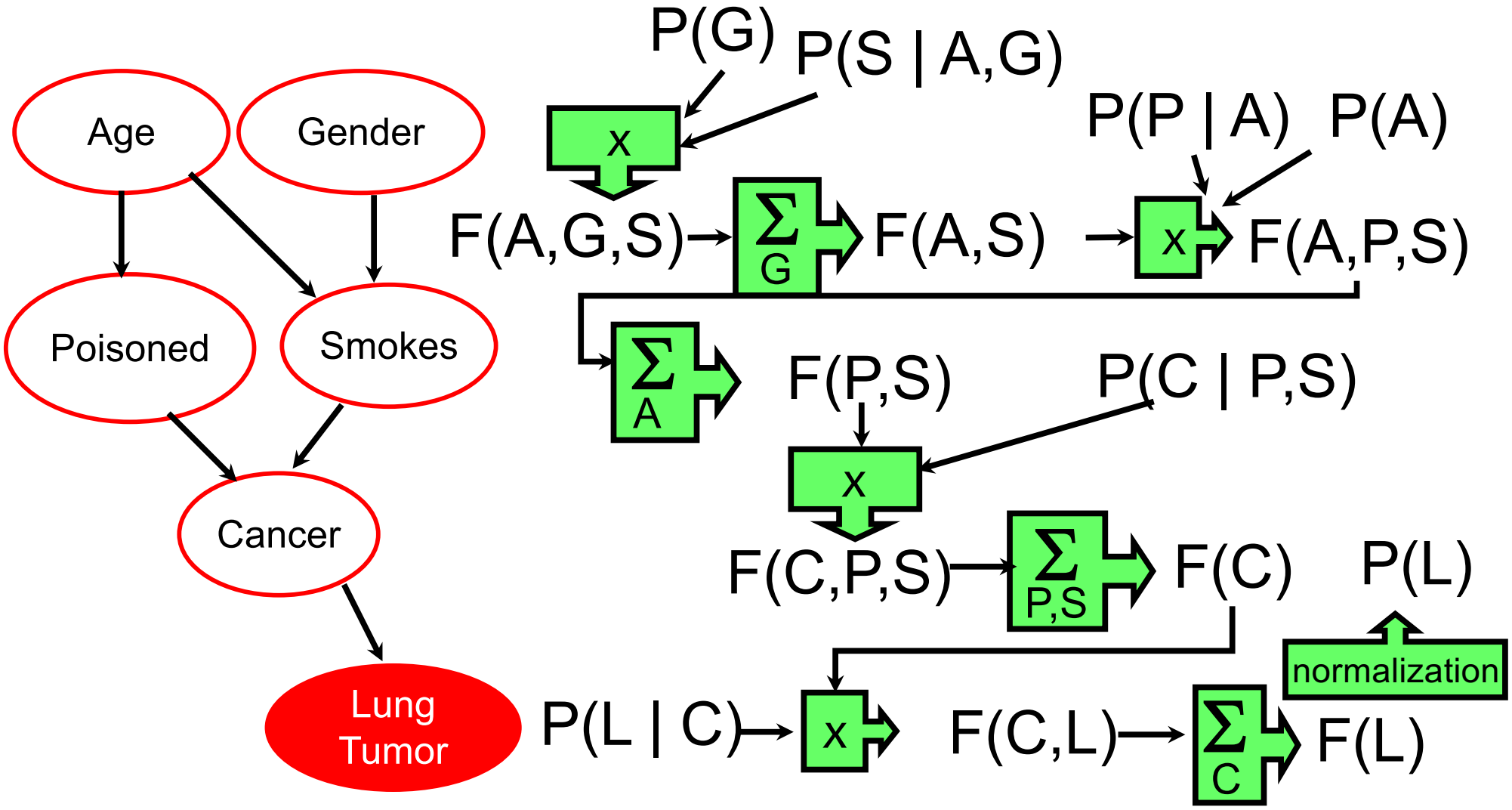
Move the sums into the products

- Key idea:
  - Do not multiply left-to-right but right-to-left.
  - Thus, terms that appear inside sums are evaluated first
    - intermediate results are stored as so-called **factors**
    - factors can be re-used several times in the same computation
  - is a form of **dynamic programming**
- Example:

$$\begin{aligned}
&\mathbf{P}(B|j,m) \\
&= \alpha \underbrace{\mathbf{P}(B)}_{B} \Sigma_e \underbrace{P(e)}_{E} \Sigma_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M} \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a \mathbf{P}(a|B,e) P(j|a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a \mathbf{P}(a|B,e) f_J(a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a f_A(a,b,e) f_J(a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) f_{\bar{A}JM}(b,e) \text{ (sum out } A\text{)} \\
&= \alpha \mathbf{P}(B) f_{\bar{E}\bar{A}JM}(b) \text{ (sum out } E\text{)} \\
&= \alpha f_B(b) \times f_{\bar{E}\bar{A}JM}(b)
\end{aligned}$$

Complete example: Let's comput P(L)

elimination order is : G,A,P,S,C



Exponentiall in the size of the largest (induced) factor (F) (table) also called treewidth: 2^3 vs 2^7

As an algorithm, this is called: Variable elimination

- Given a BN and a query P(X|e) / P(X,e)
- Instantiate evidence **e**
- Choose an elimination order over the variables, e.g., $X_1, \ldots, X_n$
- Initial *factors* $\{f_1,\ldots,f_n\}$: $f_i = P(X_i|\mathbf{Pa}_{Xi})$ (CPT for X_i)

- For i = 1 to n, if $X_i \notin \{X,\mathbf{E}\}$
  - Collect factors $f_1,\ldots,f_k$ that include X_i
  - Generate a new factor by eliminating X_i from these factors

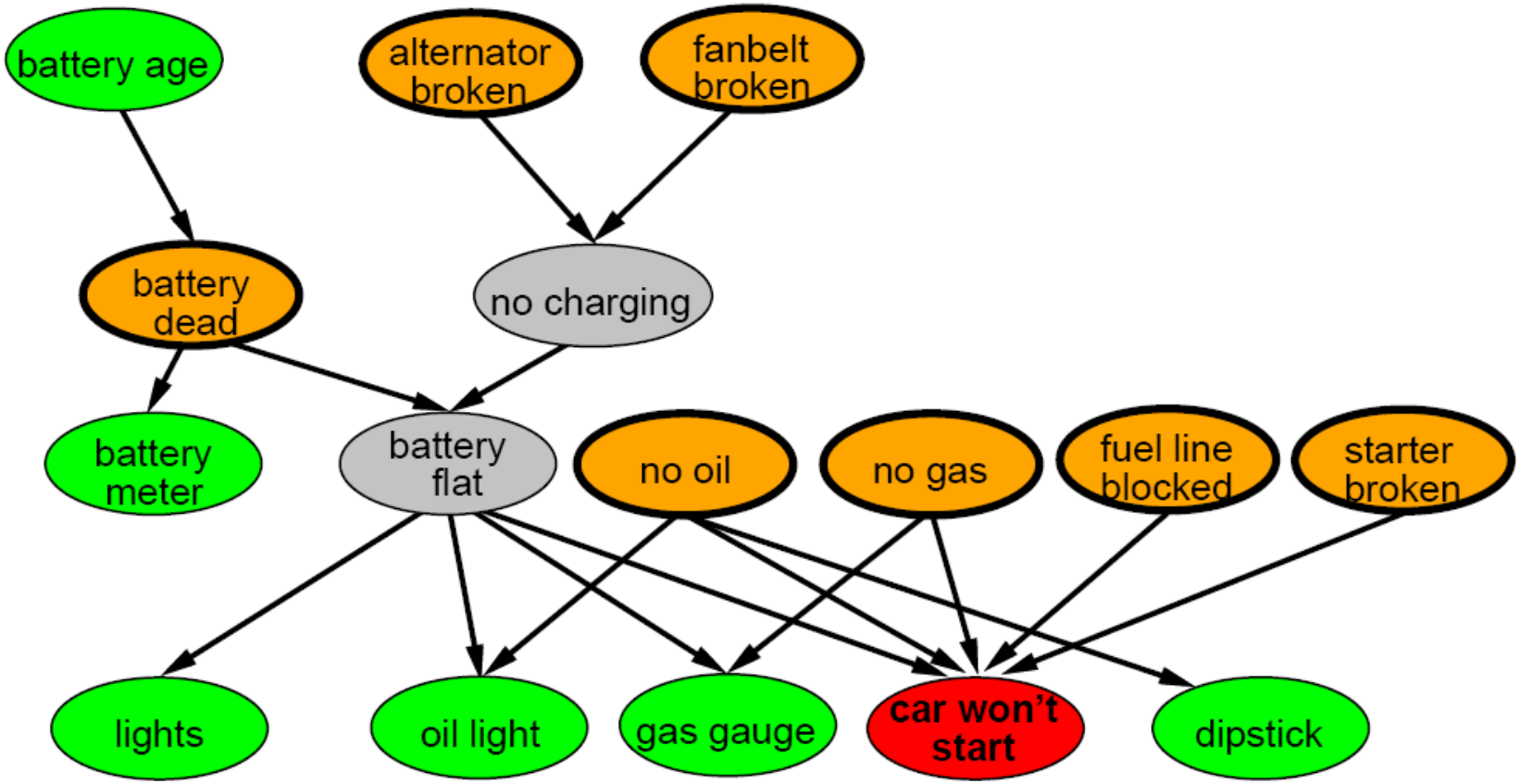
$$g = \sum_{X_i} \prod_{j=1}^{k} f_j$$

  - Variable X_i has been eliminated! Add g to the set of factors

- Normalize P(X,**e**) to obtain P(X|**e**)

More Complex Example: Car Diagnosis

- Initial evidence: Car does not start
- Test variables
- Variables for possible failures
- Hidden variables: ensure spare structure, reduce parameters



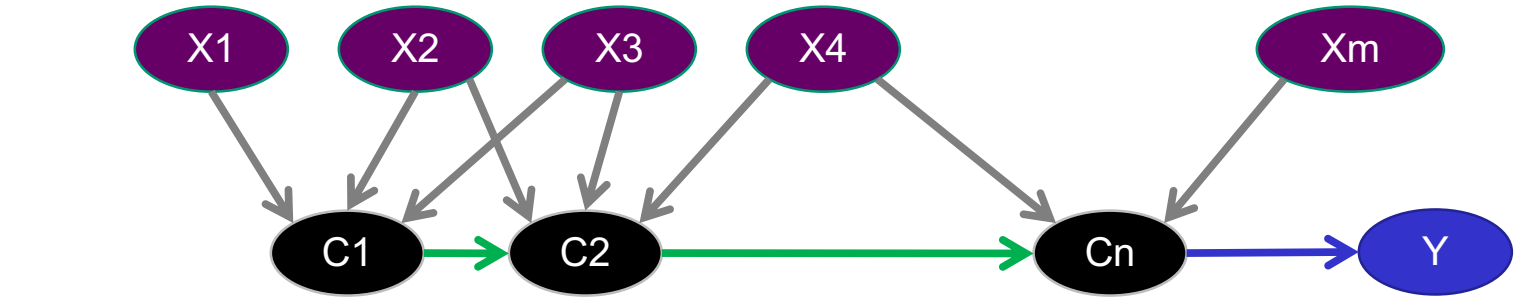
What have we learned so far?

- Uncertainty is omnipresent
- Uncertainty can be captured using probability distributions
- Graphical models are compact encodings of probability distributions
- They lead to „efficient“ algorithms for inference such as Variablen-Elimination

Mission Completed? No …

Theorem: Inference (even approximate) in Bayesian networks is NP-hard (#P; via reduction to 3-SAT)

$$\underbrace{(\overline{X}_1 \vee X_2 \vee X_3)}_{C_1} \wedge \underbrace{(\overline{X}_2 \vee X_3 \vee X_4)}_{C_2} \wedge \ldots \underbrace{\quad}_{C_n}$$

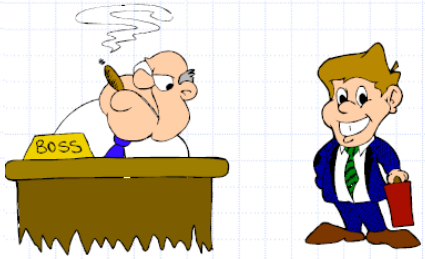


P(Y=t)>0 iff 3-SAT formula is satisfiable

- What to do when we find a problem that looks hard...

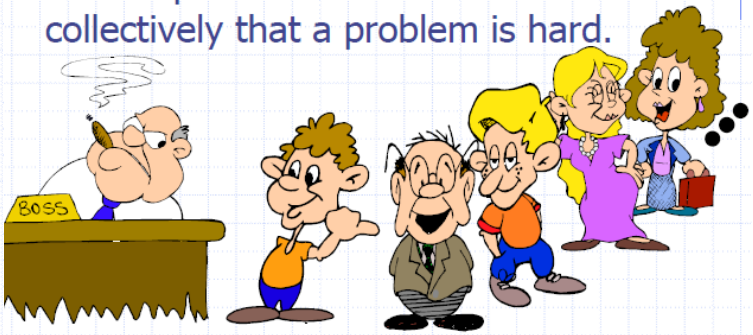
I couldn't find a polynomial-time algorithm;
I guess I'm too dumb.

- Sometimes we can prove a strong lower bound... (but not usually)

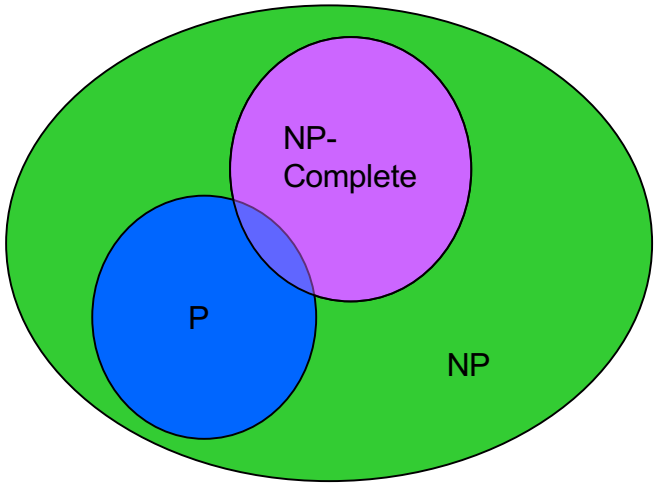


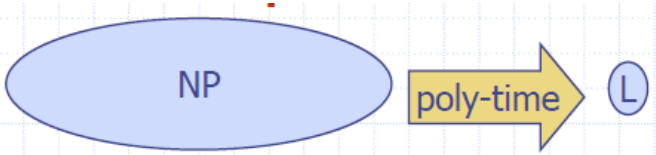
I couldn't find a polynomial-time algorithm,
because no such algorithm exists!

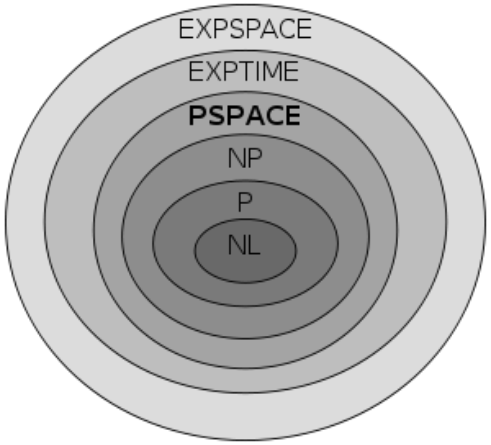
- NP-completeness let's us show collectively that a problem is hard.



I couldn't find a polynomial-time algorithm,
but neither could all these other smart people.







Complexity of Inference

Theorem:

Inference in Bayesian networks (even approximate, without proof) is NP-hard

Inference by Stochastic Simulation
(Sampling from a Bayesian Network)

Basic idea:

1) Draw N samples from a sampling distribution S
2) Compute an approximate posterior probability $\hat{P}$
3) Show this converges to the true probability 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

How to draw a sample ?

- Given random variable X, D(X)={0, 1}
- Given P(X) = {0.3, 0.7}

How to draw a sample ?

- Given random variable X, D(X)={0, 1}
- Given P(X) = {0.3, 0.7}

- **Sample X ← P (X)**
  - **draw random number r ∈ [0, 1]**
  - **If (r < 0.3) then set X=0**
  - **Else set X=1**

- Can generalize for any domain size

Sampling from an "Empty" Network

- Generating samples from a network that has no evidence associated with it (*empty* network)
- Basic idea
  - sample a value for each variable in topological order
  - using the specified conditional probabilities

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

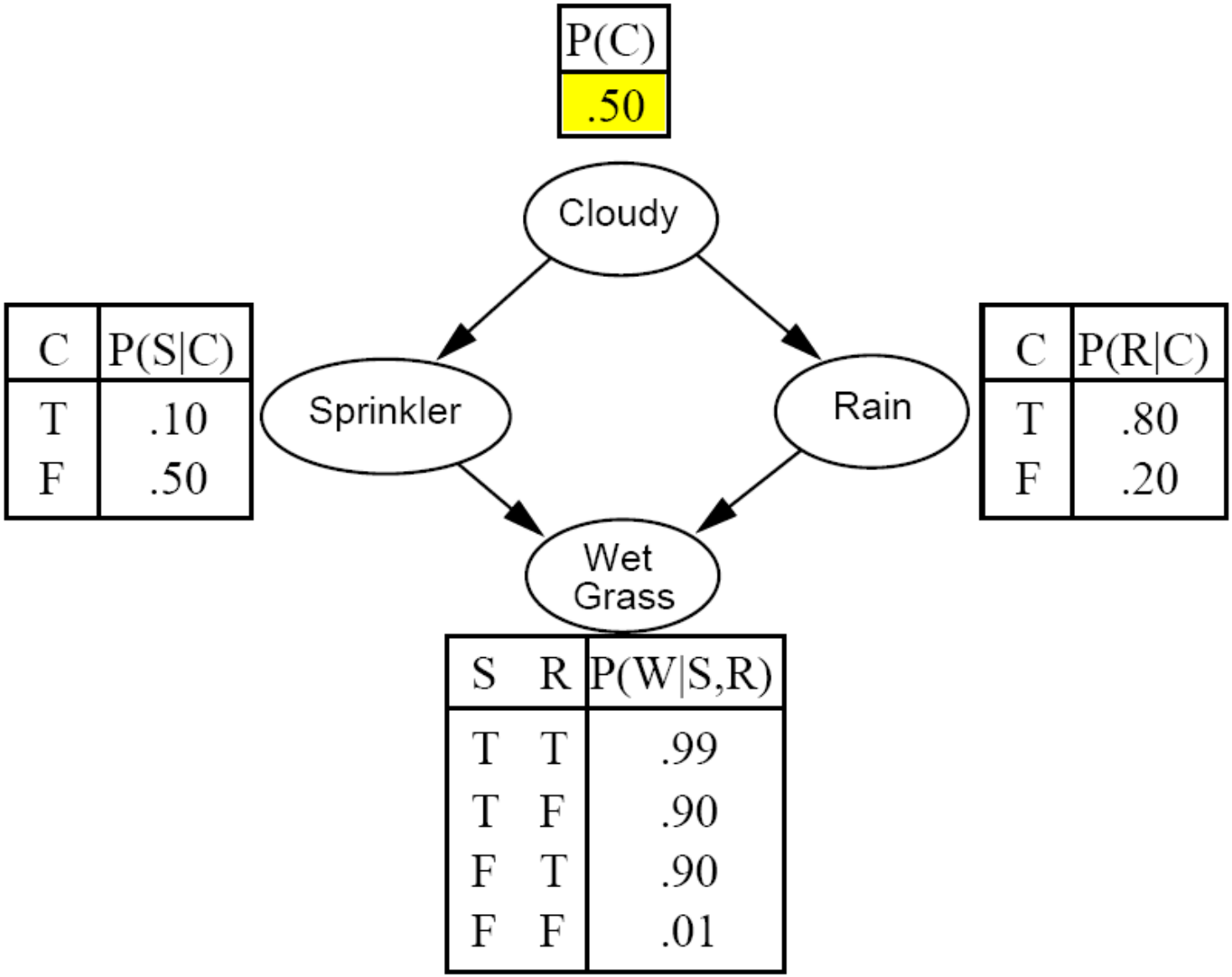
for $i = 1$ **to** n **do**

$x_i \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$

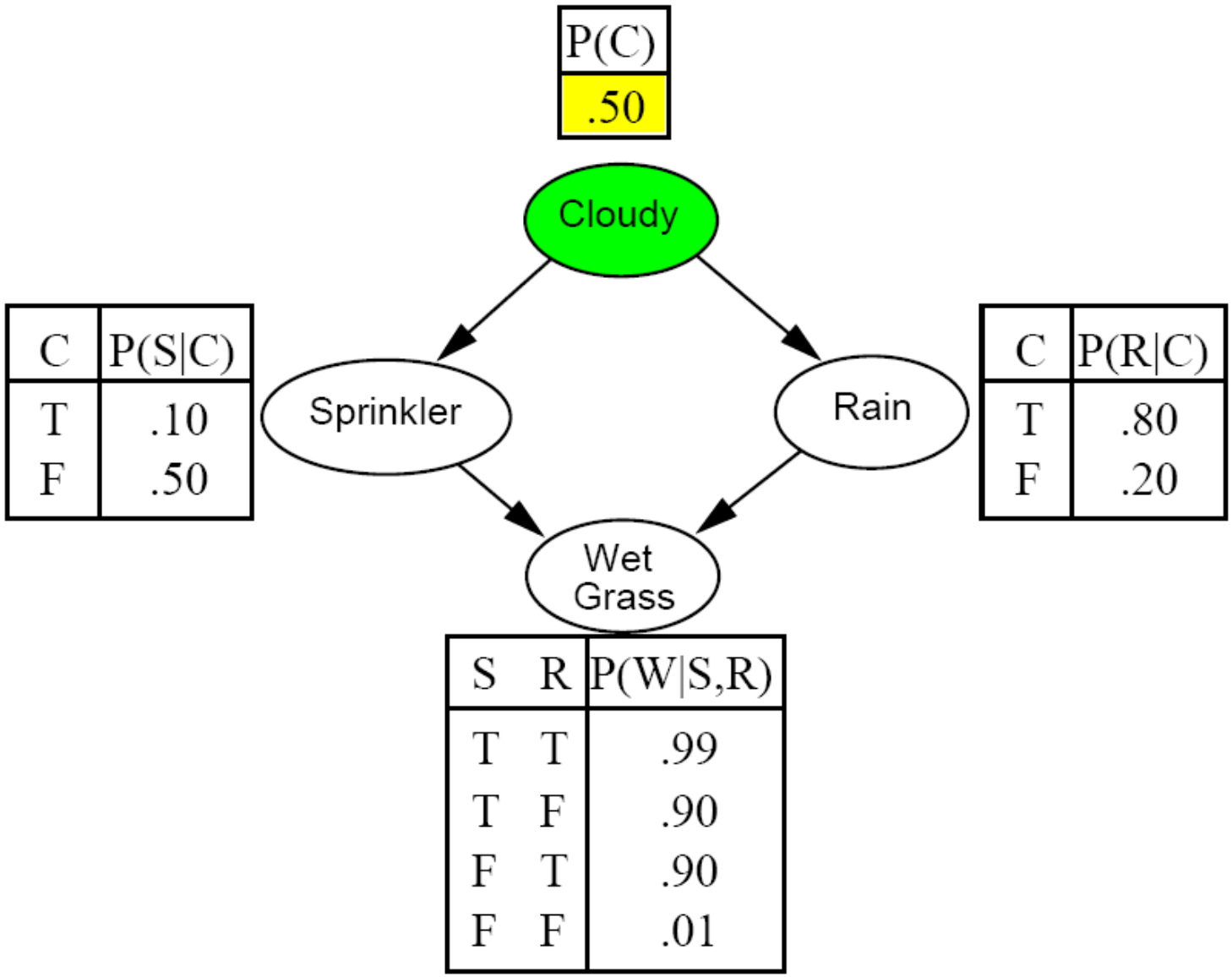
given the values of $Parents(X_i)$ in $\mathbf{x}$

return $\mathbf{x}$

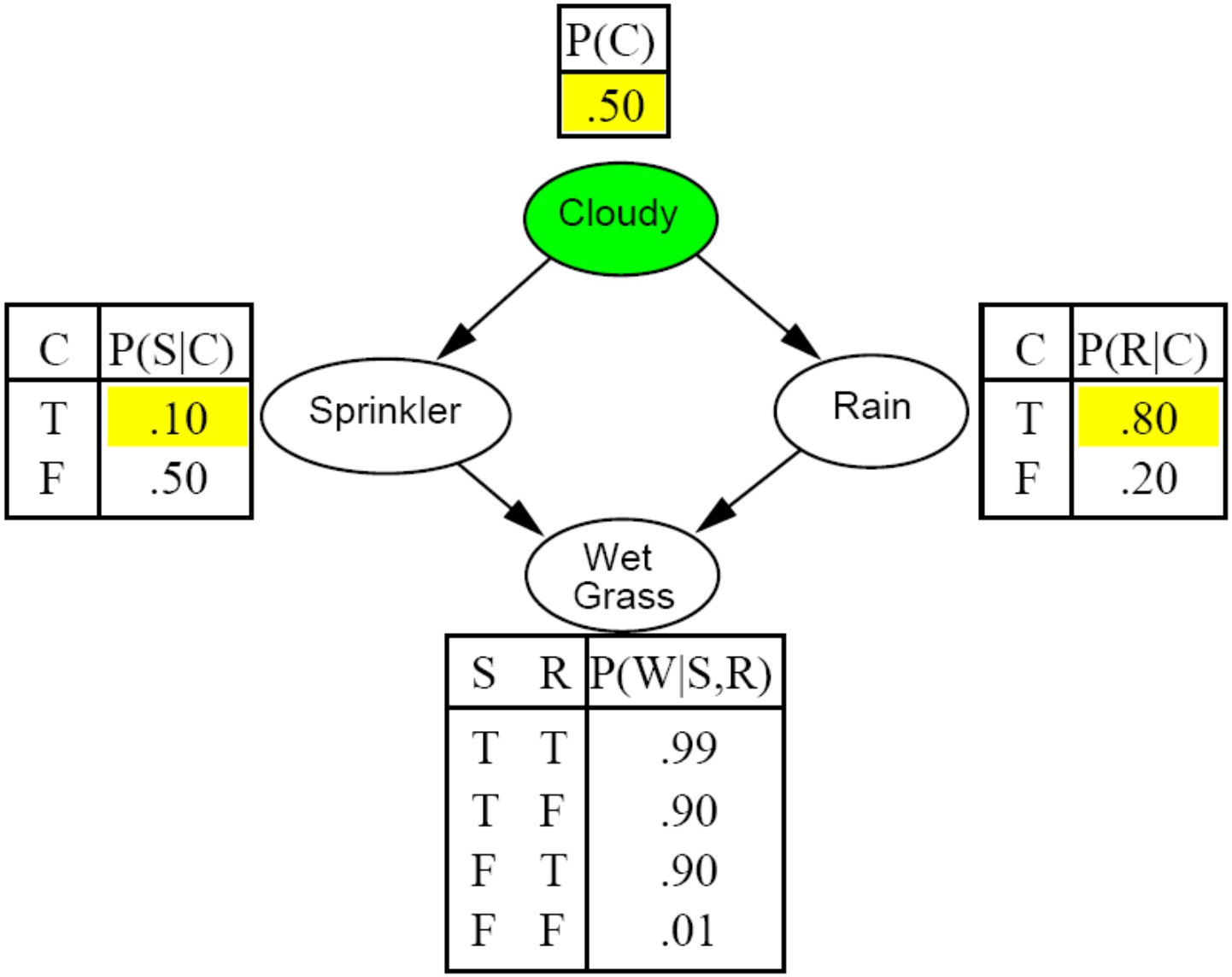
Example


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

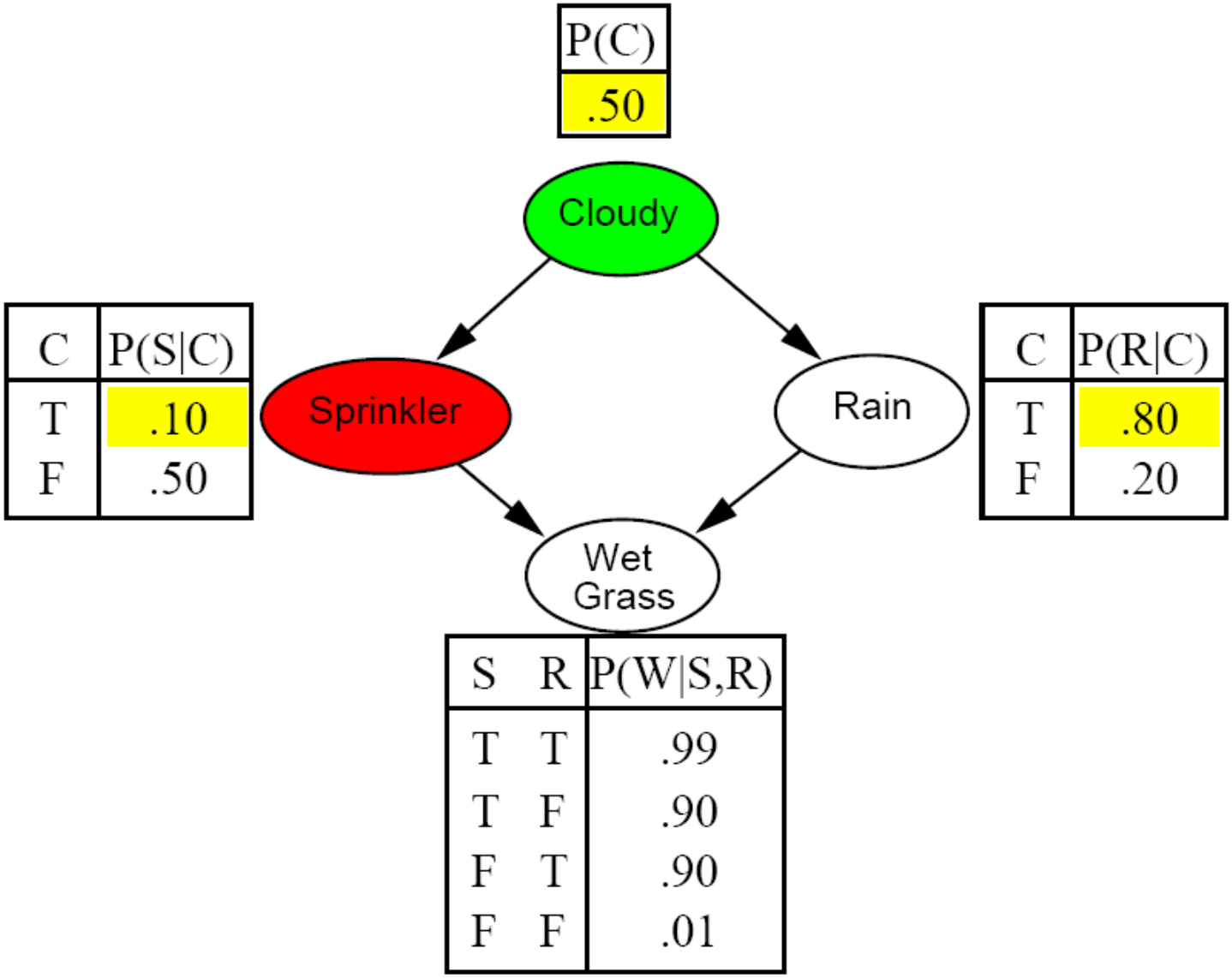
Example


P(C)
.50
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

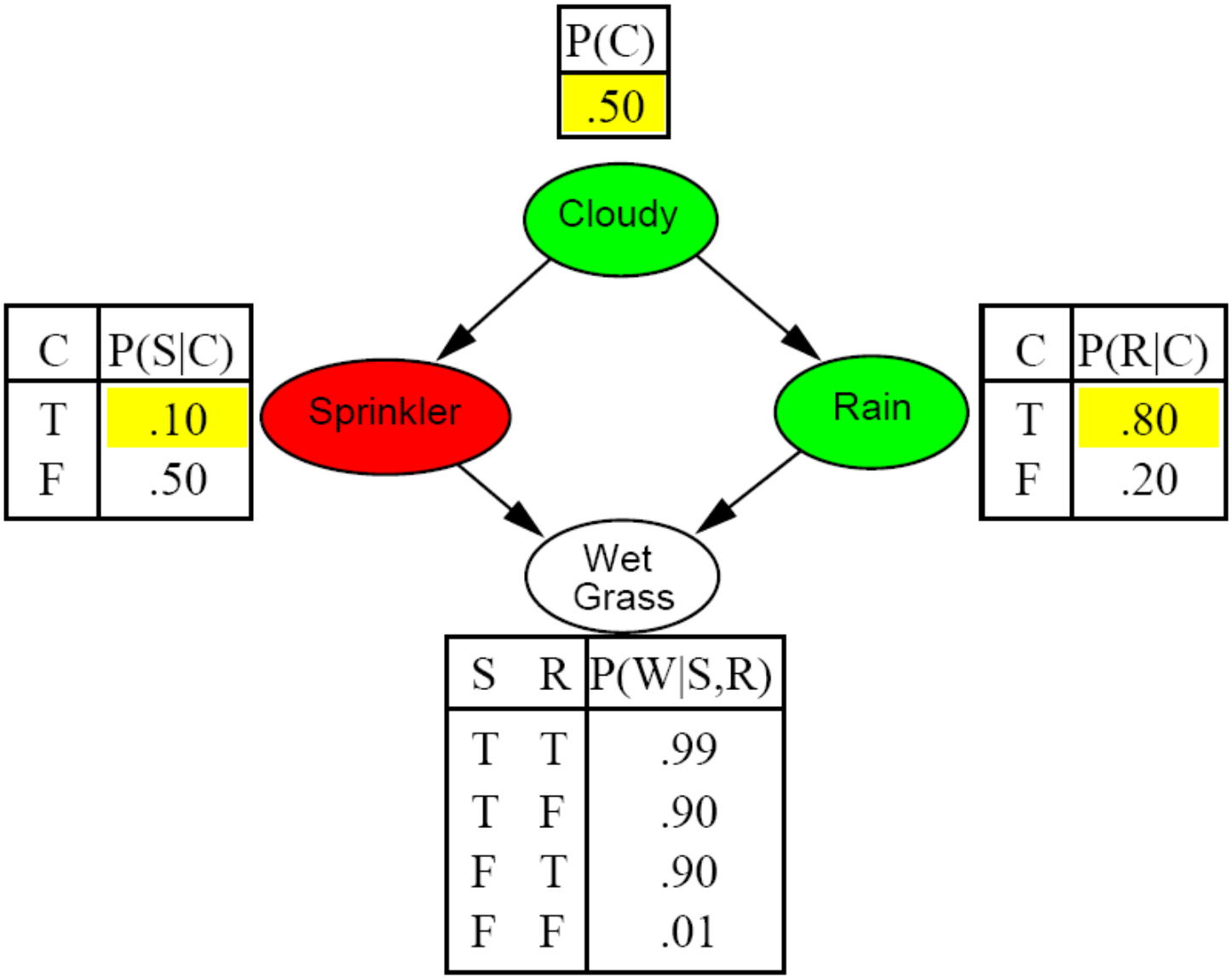
Example


P(C)
.50
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

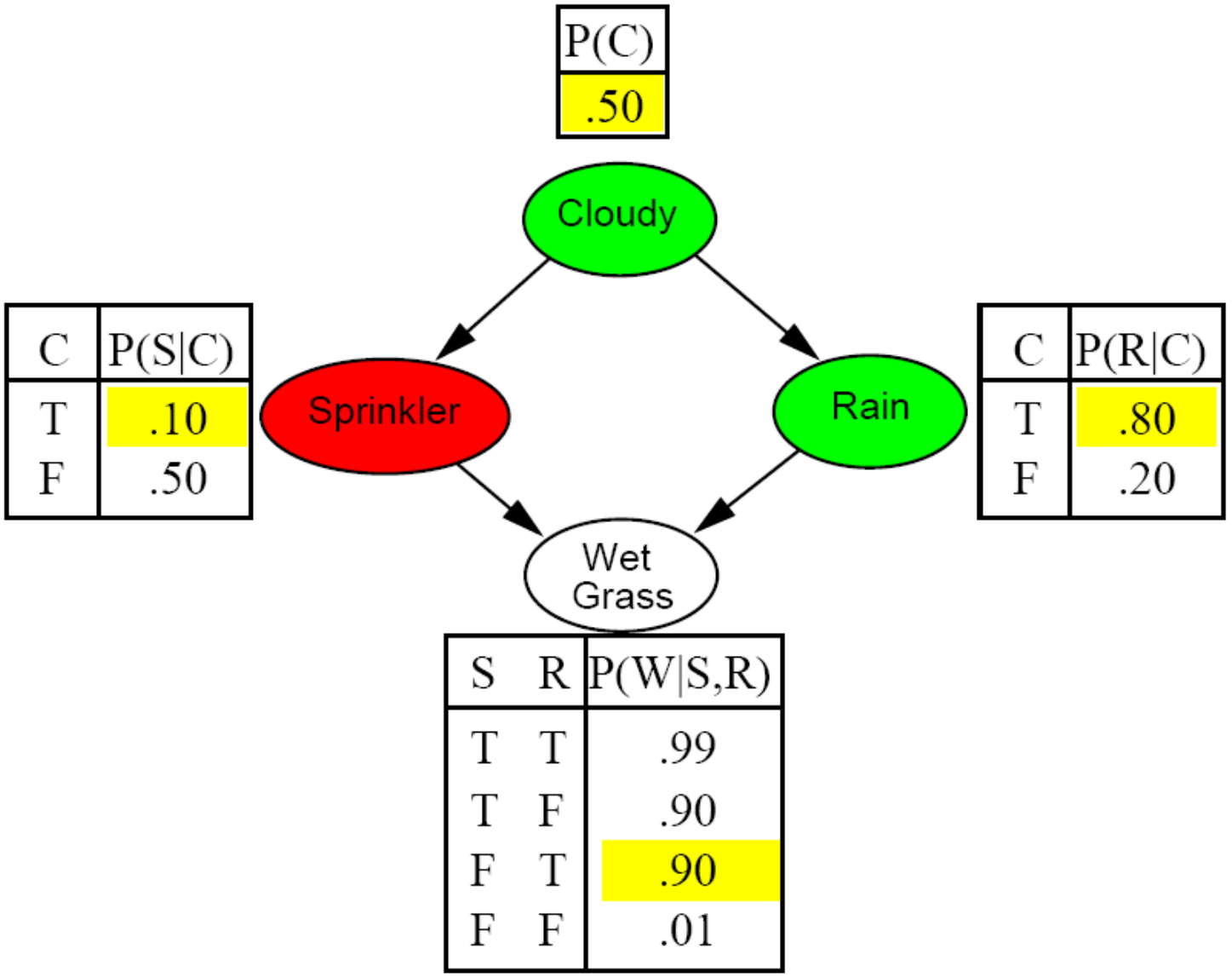
Example


P(C)
.50
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

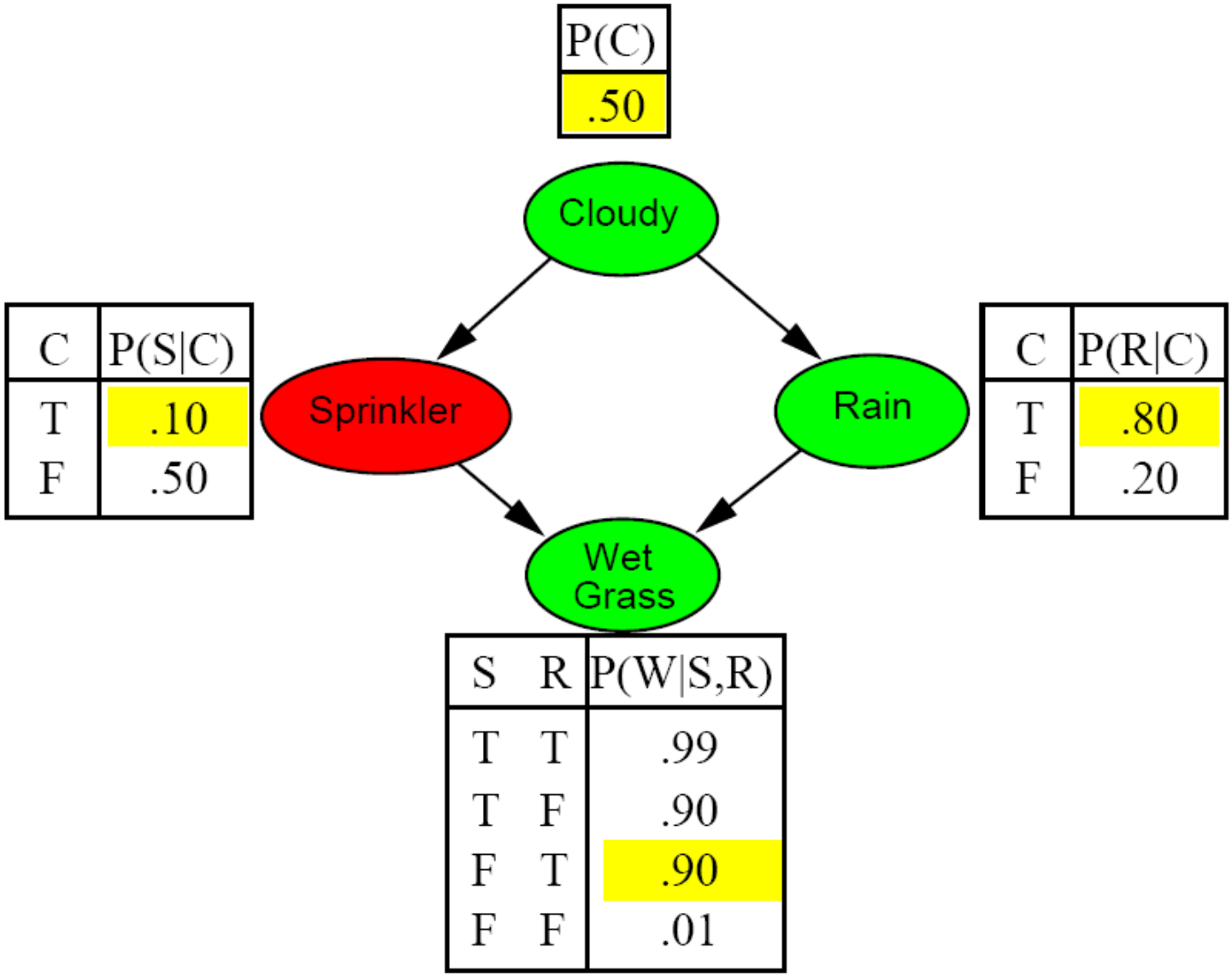
Example


P(C)
.50
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
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
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

Probability Estimation using Sampling

- sample many points using the above algorithm
- count how often each possible combination $x_1, x_2, \ldots, x_n$ appears
  - increment counters $N_{PS}(x_1 \ldots x_n)$
- estimate the probability by the observed percentages

$$\hat{P}_{PS}(x_1 \ldots x_n) = N_{PS}(x_1 \ldots x_n)/N$$

- **This converges towards the joint probability function!**

Markov Chain Monte Carlo (MCMC) Sampling

"State" of network = current assignment to all variables.

Generate next state by sampling one variable given Markov blanket
Sample each variable in turn, keeping evidence fixed

function MCMC-ASK($X, \mathbf{e}, bn, N$) **returns** an estimate of $P(X|\mathbf{e})$
  local variables: $\mathbf{N}[X]$, a vector of counts over X, initially zero
    $\mathbf{Z}$, the nonevidence variables in bn
    $\mathbf{x}$, the current state of the network, initially copied from $\mathbf{e}$

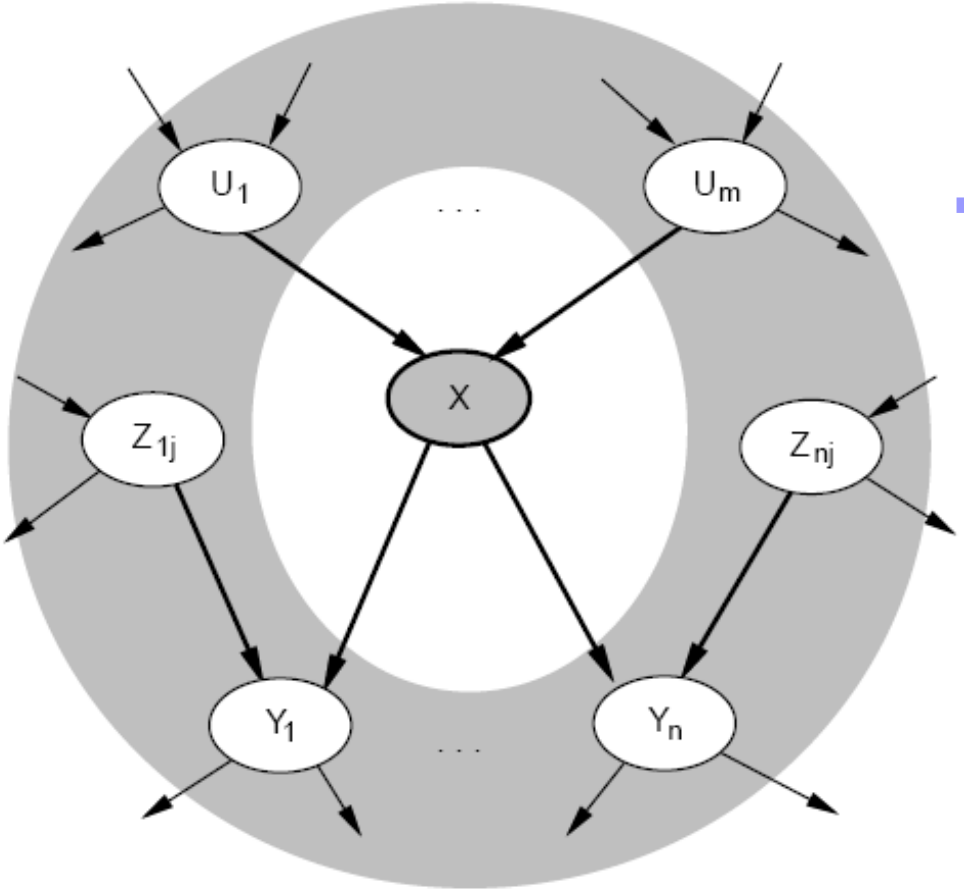
  initialize $\mathbf{x}$ with random values for the variables in $\mathbf{Y}$
  for $j = 1$ to N **do**
    for each Z_i in $\mathbf{Z}$ **do**
      sample the value of Z_i in $\mathbf{x}$ from $\mathbf{P}(Z_i|mb(Z_i))$
        given the values of $MB(Z_i)$ in $\mathbf{x}$
      $\mathbf{N}[x] \leftarrow \mathbf{N}[x] + 1$ where x is the value of X in $\mathbf{x}$
  return NORMALIZE($\mathbf{N}[X]$)

Gibbs Sampling

Can also choose a variable to sample at random each time

(Directed) Markov Blanket

- **Markov Blanket:**
  - parents + children + children's parents



- Each node is conditionally independent of all other nodes given its markov blanket

$$\mathbf{P}\left(X \mid U_{1,\ldots}, U_m, Y_{1,\ldots}, Y_n, Z_{1\mathrm{j}}, \ldots, Z_{nj}\right) = \mathbf{P}\left(X \mid \textit{all variables}\right)$$

Ordered Gibbs Sampler

- Generate sample x^{t+1} from x^t :

Process all variables in some order

$$X_1 = x_1^{t+1} \leftarrow P(x_1 \mid x_2^t, x_3^t, \ldots, x_N^t, e)$$

$$X_2 = x_2^{t+1} \leftarrow P(x_2 \mid x_1^{t+1}, x_3^t, \ldots, x_N^t, e)$$

...

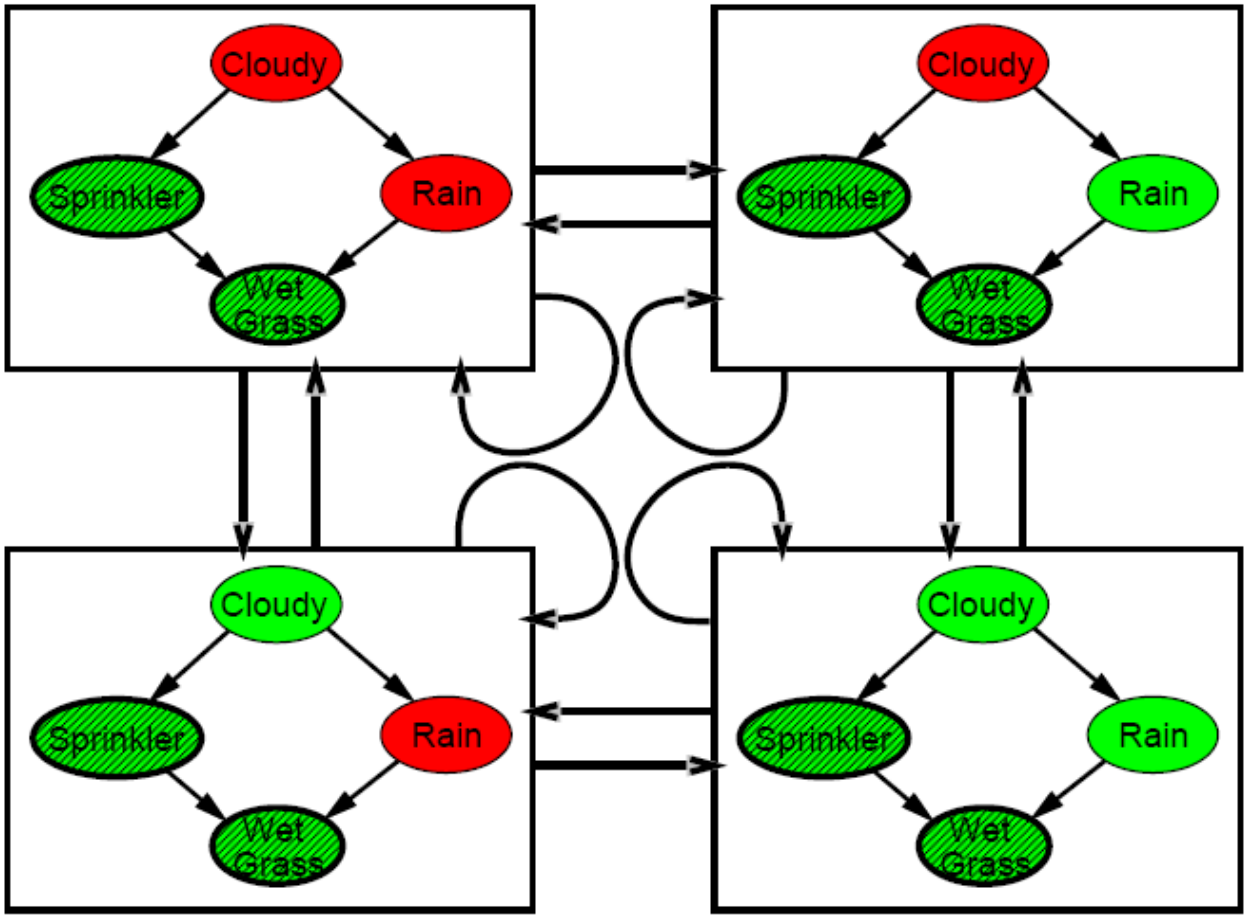
$$X_N = x_N^{t+1} \leftarrow P(x_N \mid x_1^{t+1}, x_2^{t+1}, \ldots, x_{N-1}^{t+1}, e)$$

- In short, for i=1 to N:

$$X_i = x_i^{t+1} \leftarrow \textbf{sampled from } P(x_i \mid x^t \setminus x_i, e)$$

The Markov Chain

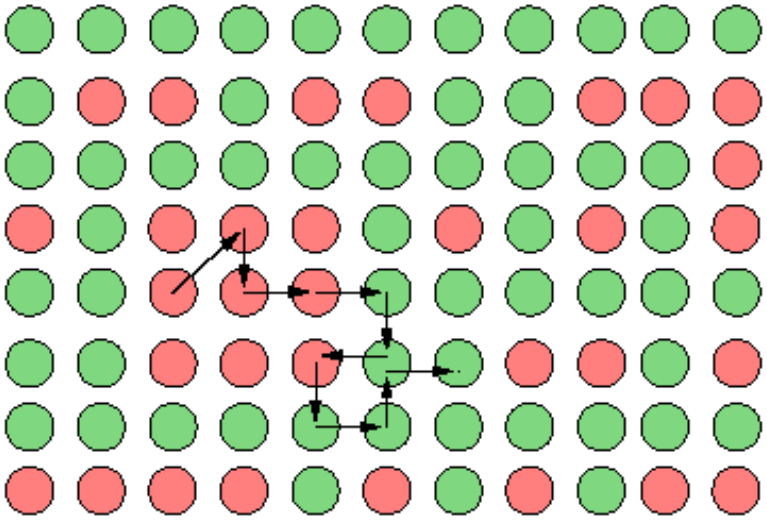
With $Sprinkler = true, WetGrass = true$, there are four states:



Wander about for a while, average what you see

Gibbs Sampling: Illustration

The process of Gibbs sampling can be understood as a *random walk* in the space of all instantiations with $\boldsymbol{Y} = \boldsymbol{u}$:



Reachable in one step: instantiations that differ from current one by value assignment to at most one variable (assume randomized choice of variable X_k).

Guaranteed to converge iff chain is :

irreducible (every state reachable from every other state)
aperiodic (returns to state i can occur at irregular times)
ergodic (returns to every state with probability 1)

Example

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

Sample $Cloudy$ or $Rain$ given its Markov blanket, repeat.
Count number of times $Rain$ is true and false in the samples.

E.g., visit 100 states
31 have $Rain = true$, 69 have $Rain = false$

$\hat{\mathbf{P}}(Rain|Sprinkler = true, WetGrass = true)$
$= \textsc{Normalize}(\langle 31, 69 \rangle) = \langle 0.31, 0.69 \rangle$

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

What have we learned?

- Exact inference via Variable Elimination (VE)
- Inference in Bayesian networks is **NP-hard**, even when approximating. Still, for many distributions, sampling is the only option

- Forward sampling
- MCMC sampling (GIBBS sampling)

- Overall, we now know:
  - Very Basics of probability theory
  - Arguments why to follow probabilitiy theory
  - Bayesian networks (representation and semantics)
  - Inference in Bayesian networks

Real-World Applications of BN

- Industrial
- Processor Fault Diagnosis - by Intel
- Auxiliary Turbine Diagnosis - GEMS by GE
- Diagnosis of space shuttle propulsion systems - VISTA by NASA/Rockwell
- Situation assessment for nuclear power plant – NRC

- Military
- Automatic Target Recognition - MITRE
- Autonomous control of unmanned underwater vehicle - Lockheed Martin
- Assessment of Intent

Real-World Applications of BN

- Medical Diagnosis (also at TUDA)
- Internal Medicine
- Pathology diagnosis - Intellipath by Chapman & Hall
- Breast Cancer Manager with Intellipath

- NLP, CV, Cognitive Science, Plant Phenotpying (also at TUDA)

- Commercial
- Financial Market Analysis
- Information Retrieval
- Software troubleshooting and advice - Windows 95 & Office 97
- Pregnancy and Child Care – Microsoft
- Software debugging - American Airlines' SABRE online reservation system