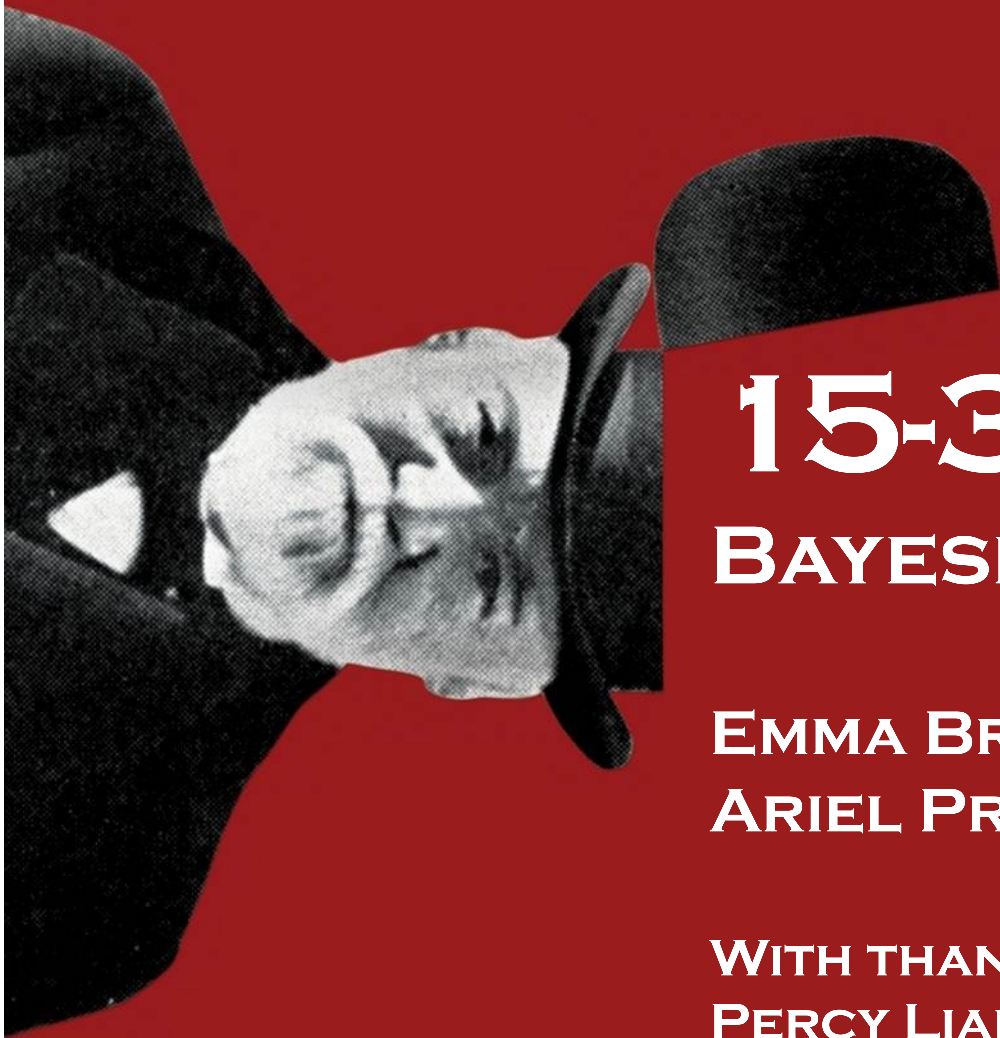


15-381/781

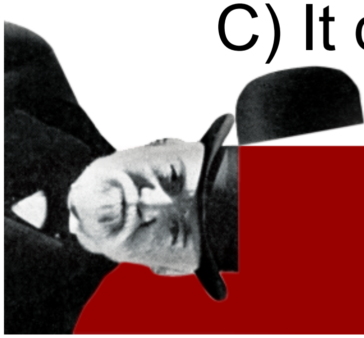
Bayesian Networks

Emma Brunskill (this time)
Ariel Procaccia

With thanks to Dan Klein (Berkeley), Percy Liang (Stanford) and Past 15-381 Instructors for some slide content, and Russell & Norvig

Question

- Eating a poppy seed bagel or taking opium are independent events that can cause a positive drug test.
- John Doe gets a positive drug test.
- How does learning that John Doe ate a bagel earlier today change your beliefs?
  - A) It increases the probability that John took opium
  - B) It decreases the probability than John took opium
  - C) It does not change the probability John took opium



Reasoning & Inference

- Key part of intelligence
- Drawing conclusions based on information
  - All kings are mortal
  - James is a king
  - Is James mortal?
- Logic is one framework
- But real world involves uncertainty
  - Sensors imperfect, actuators imperfect,…

Reasoning Under Uncertainty

- Inference given noisy, uncertain info
- Probability of different conclusions

Probabilistic Inference

- Compute probability of a **query** variable (or variables) taking on a value (or set of values) given some **evidence**

Probabilistic Inference

- Compute probability of a query variable (or variables) taking on a value (or set of values) given some evidence
- Often interested in:
  - **Posterior probability** of taking on any value given some evidence: $Pr[Q \mid E_1=e_1,\ldots,E_k=e_k]$
  - **Most likely explanation** given some evidence: $\mathrm{argmax}_q\ Pr[Q=q \mid E_1=e1,\ldots,E_k=e_k]$

Probabilistic Inference

- Compute probability of a **query** variable (or variables) taking on a value (or set of values) given some **evidence**

- How do we do probabilistic inference in complex domains?
- How can we do this efficiently?

Using the Joint To Answer Queries

- Joint distribution is sufficient to answer any probabilistic inference question involving variables described in joint

Example

- Probability car is red given that it's a sedan?
- What rules can we use?

	Sedan	SUV	Coupe	Truck
Red	.05	.2	0	.1
White	.1	0	.1	0
Blue	0	.1	0.05	.1
Beige	.1	0	.1	0

Example

$$P(Color = Red \mid Type = Sedan)$$
$$= \frac{P(Color = Red \,\&\, Type = Sedan)}{P(Sedan)}$$

Use Bayes rule and Sum rule

	Sedan	SUV	Coupe	Truck
Red	.05	.2	0	.1
White	.1	0	.1	0
Blue	0	.1	0.05	.1
Beige	.1	0	.1	0

Example

$$P(Color = \text{Red} \mid Type = Sedan)$$

$$= \frac{P(Color = \text{Red} \,\&\, Type = Sedan)}{P(Sedan)}$$

Use Bayes rule and Sum rule

= 0.2

	Sedan	SUV	Coupe	Truck
Red	.05	.2	0	.1
White	.1	0	.1	0
Blue	0	.1	0.05	.1
Beige	.1	0	.1	0

Patient Visit Inference

- Joint distribution over:
  - LastName, FirstName, Gender, Height, Birthdate, Weight, Fever, Subcounty, HIV status, HIV assay, Headache, UTI diagnosis, Vomiting, Diarrhea, Malaria, Cipro, Productive cough, Civil Status, TransportMode

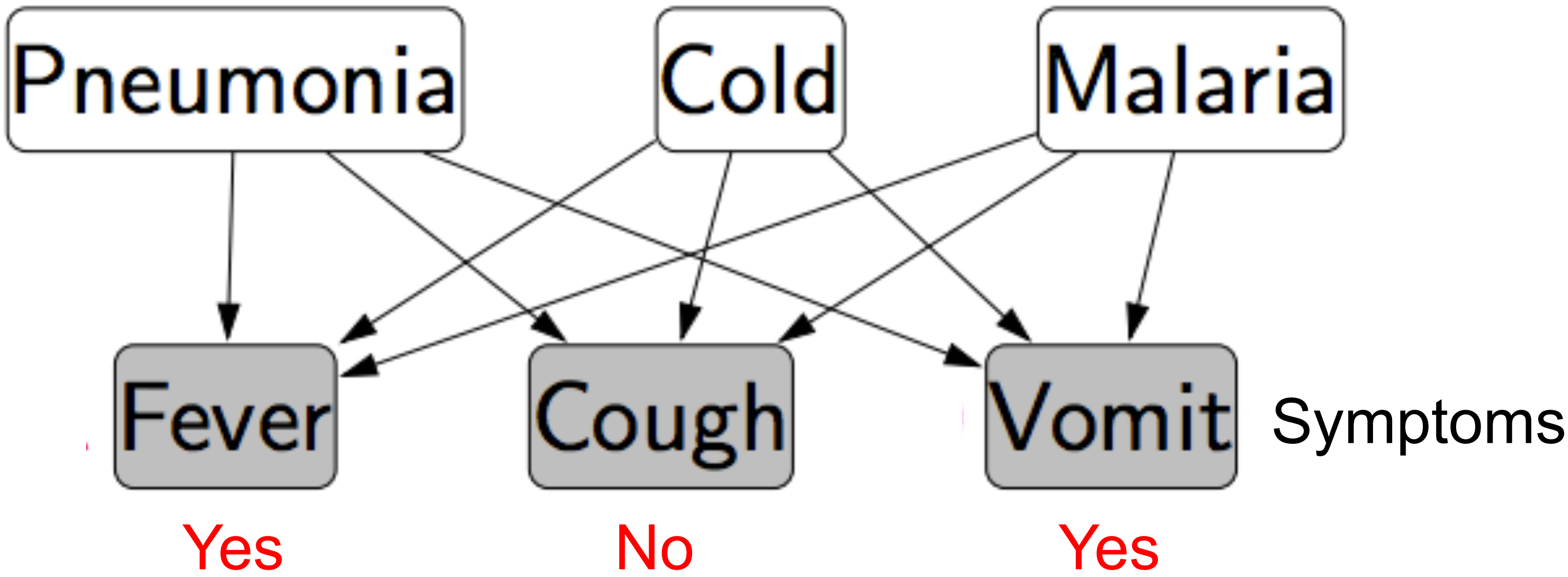
- How many parameters need to represent joint?
- Potential computational cost?

Bayes Networks

- Compact representation of the joint distribution
- Make conditional independence relationships explicit

Medical Diagnosis

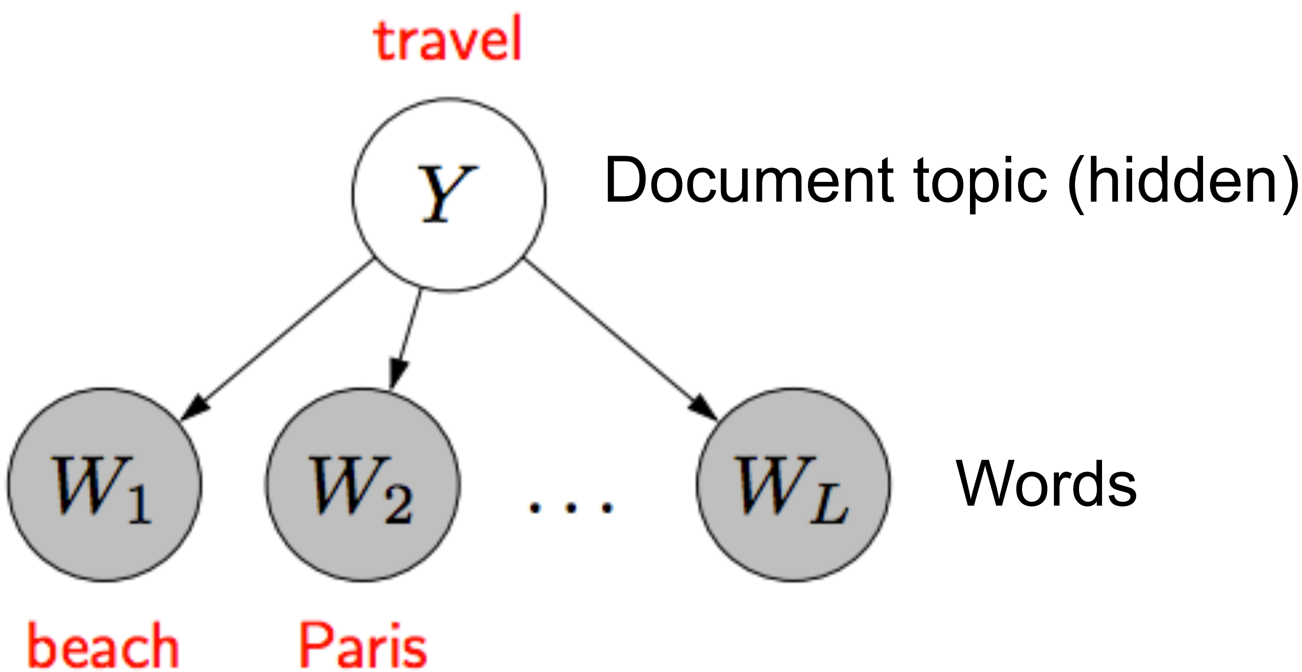
- Given a patient's symptoms, what might conditions or diseases might he have?



Example from Percy Liang

Document Classification

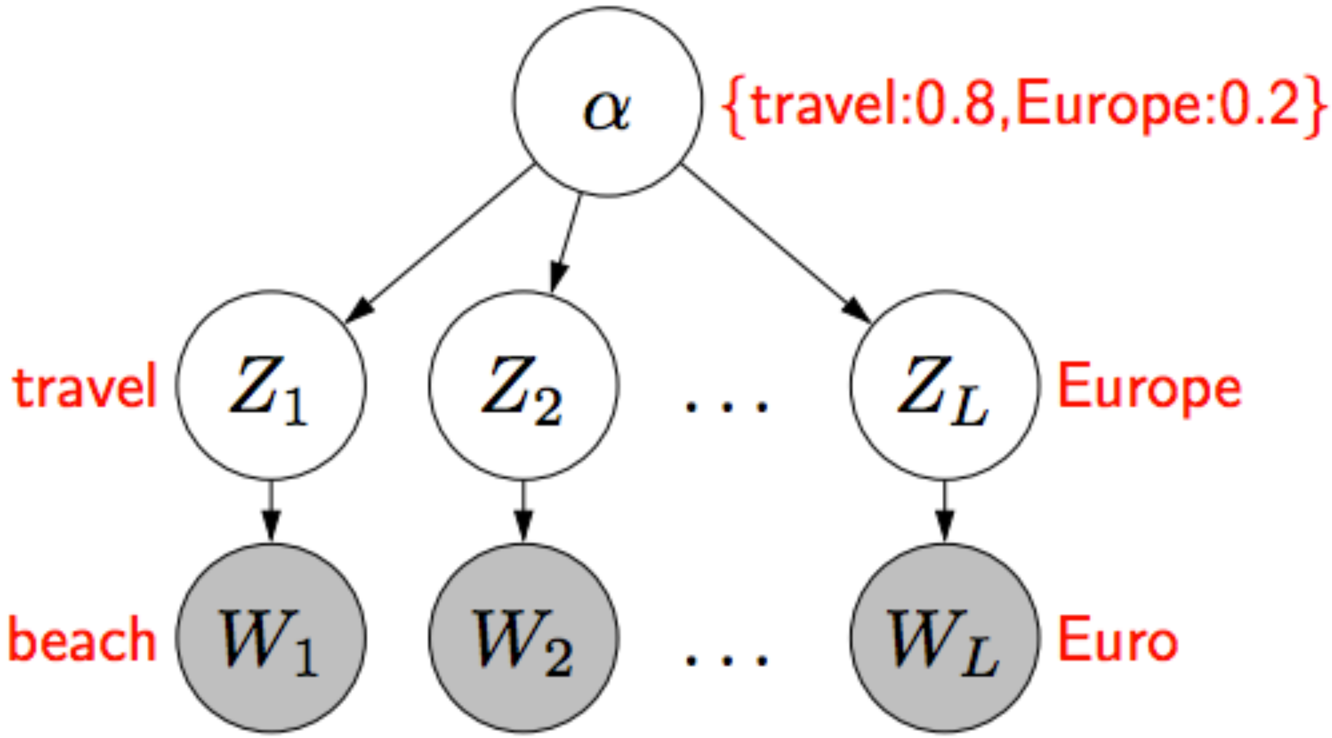
- Given the words in a document, what is it about?



Example from Percy Liang

Topic Modeling

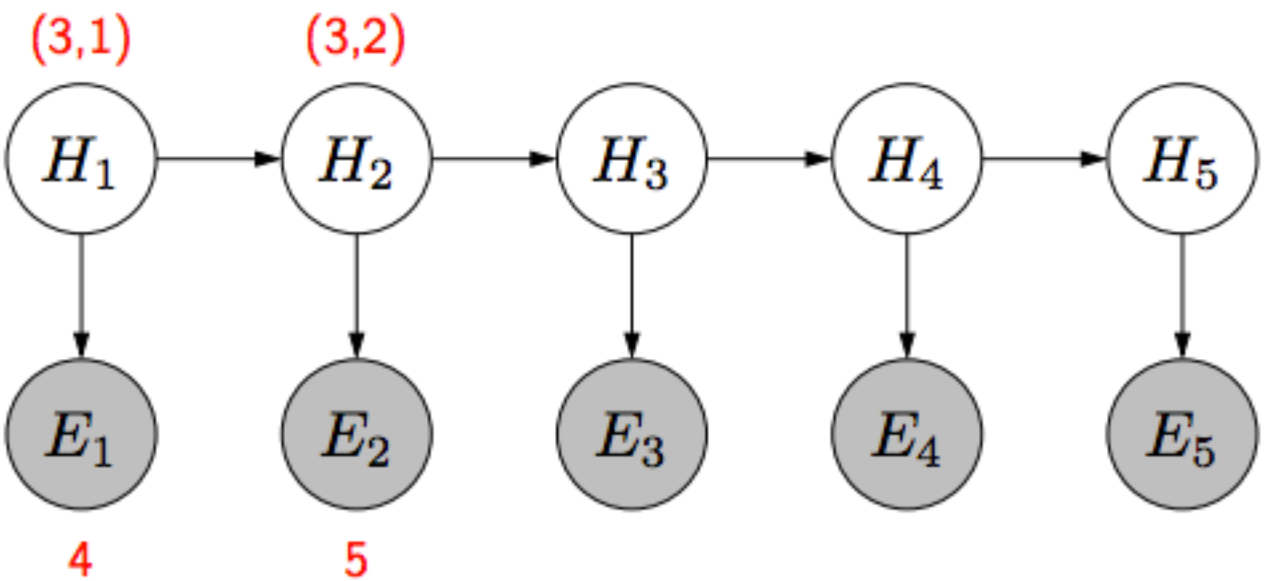
- Given the words in a document, what topics is it about?



Example from Percy Liang

Object Tracking

- Given some observations, what was the path the agent went through?



Example from Percy Liang

Probabilistic Inference

- Compute probability of a **query** variable (or variables) taking on a value (or set of values) given some **evidence**

Exact Inference is NP-Hard

- Consider the 3-SAT clause:

$$(x_1 \vee x_2 \vee \neg x_3) \wedge (\neg x_1 \vee x_3 \vee \neg x_4) \wedge (x_2 \vee \neg x_2 \vee x_4) \wedge (\neg x_3 \vee \neg x_4 \vee \neg x_5) \wedge (x_2 \vee x_5 \vee x_7) \wedge (x_4 \vee x_5 \vee x_6) \wedge (\neg x_5 \vee x_6 \vee \neg x_7) \wedge (\neg x_5 \vee \neg x_6 \vee x_7)$$

which can be encoded by the following Bayes' net:

$P(X_i = 0) = P(X_i = 1) = 0.5$

$Y_1 = X_1 \vee X_2 \vee \neg X_3$

…

$Y_8 = \neg X_5 \vee X_6 \vee X_7$

$Y_{1,2} = Y_1 \wedge Y_2$

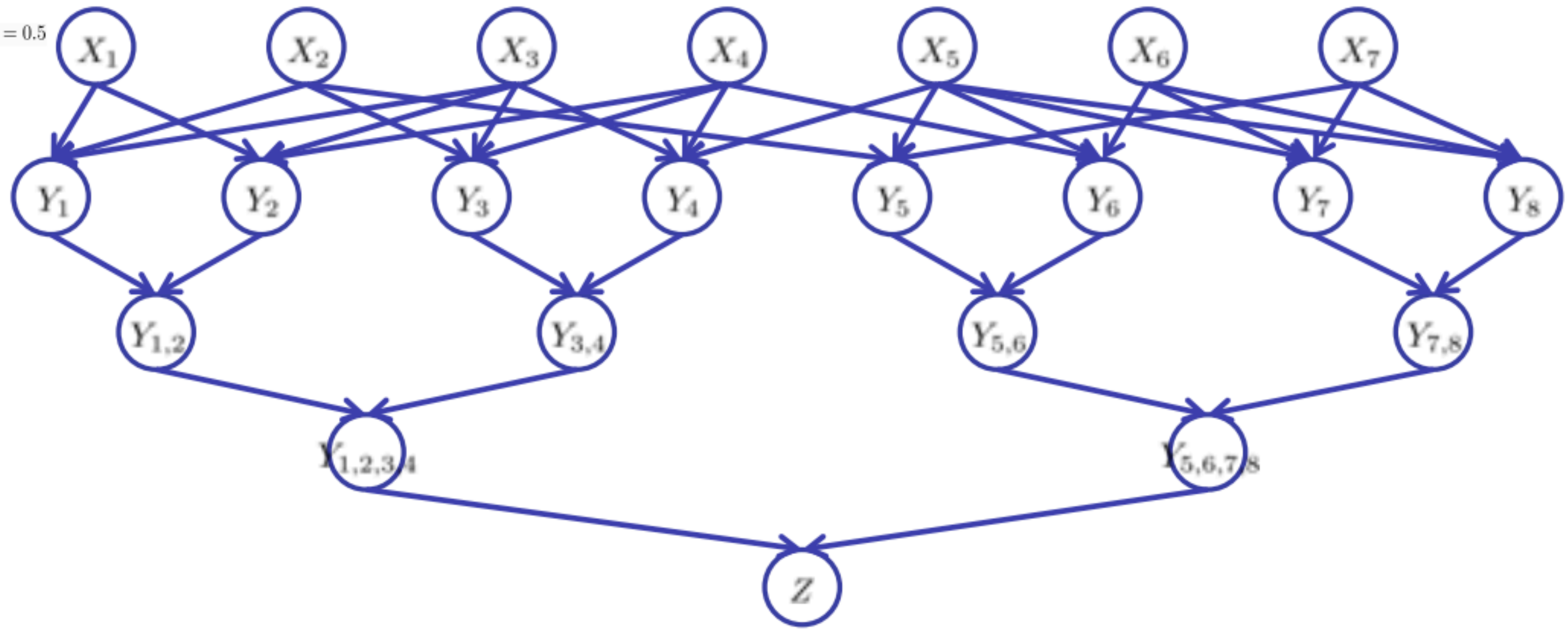
…

$Y_{7,8} = Y_7 \wedge Y_8$

$Y_{1,2,3,4} = Y_{1,2} \wedge Y_{3,4}$

$Y_{5,6,7,8} = Y_{5,6} \wedge Y_{7,8}$

$Z = Y_{1,2,3,4} \wedge Y_{5,6,7,8}$



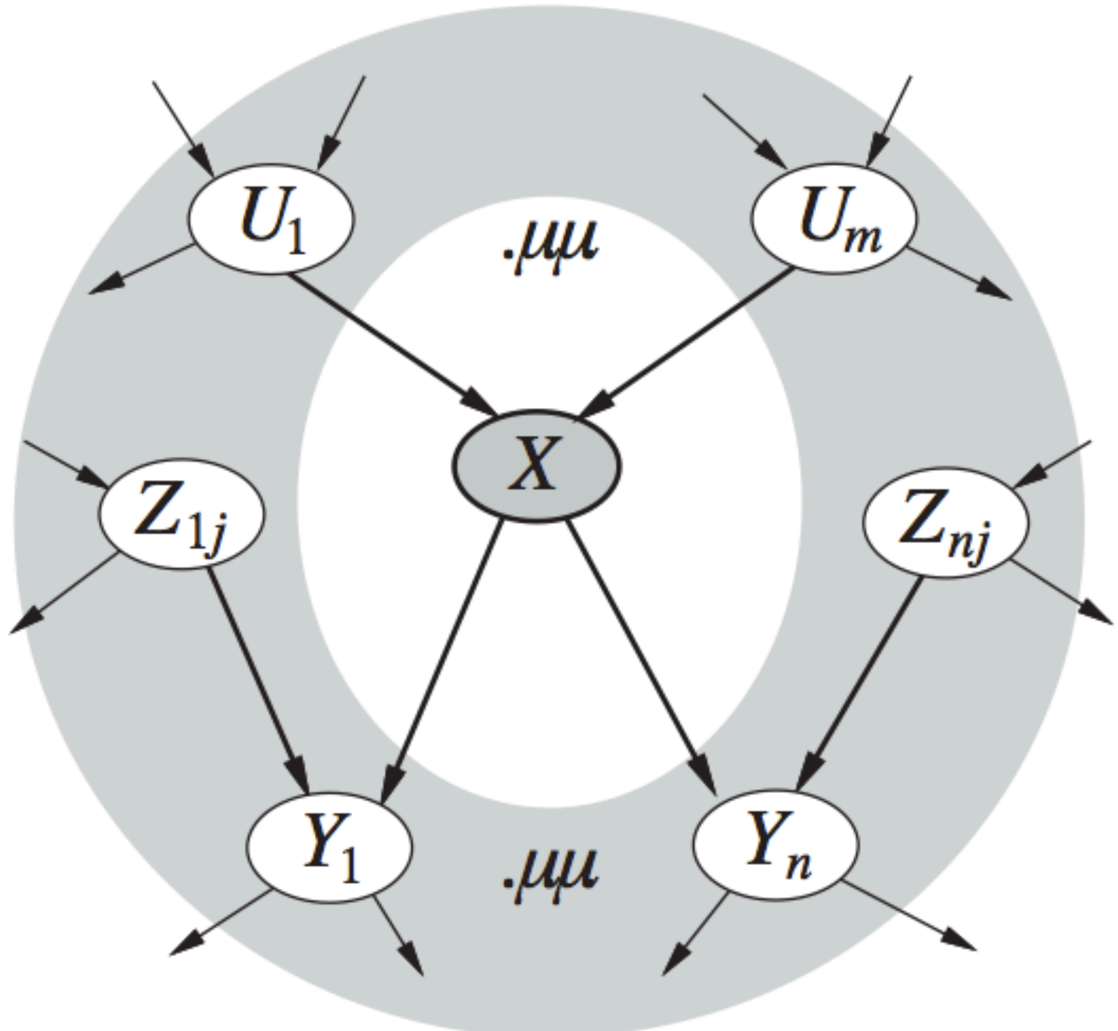
If we can answer P(z) equal to zero or not, we answered whether the 3-SAT problem has a solution.

Probabilistic Inference

- Compute probability of a query variable(s) given some evidence
- But in large networks, exact inference is often computationally intractable

Markov Blanket

- Markov blanket
  - Parents
  - Children
  - Children's parents
- Variable conditionally independent of all other nodes given its Markov Blanket

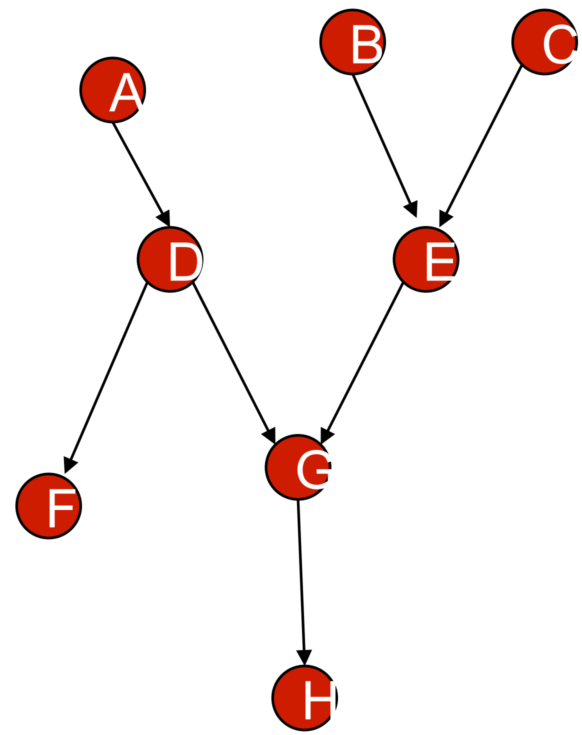


Markov Blanket Poll

- Markov blanket
  - Parents
  - Children
  - Children's parents

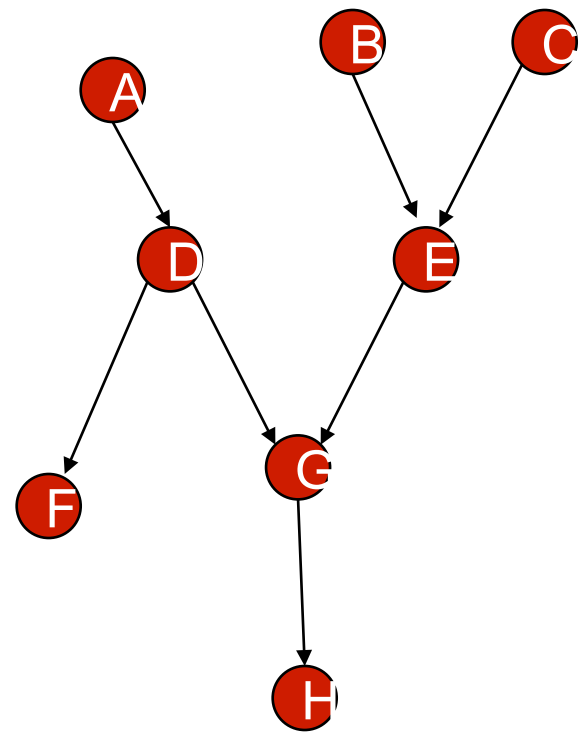
What is the Markov blanket of D?

1. A,F,G
2. **A,F,G,E**
3. A,F,G,E,B,C
4. Not sure



Markov Blanket & Independence

- Markov blanket: Parents, Children, Children's parents
- Variable conditionally independent of all other nodes given its Markov Blanket
- Ex: Evidence is G=True. Is E conditionally independent of A given G=True?
- Not necessarily
- Variable conditionally independent of all other nodes given know *values of all variables in its Markov Blanket*



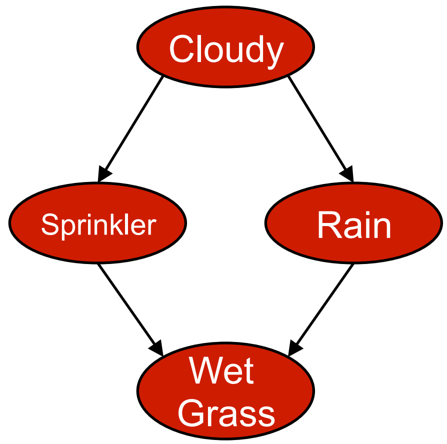
Overview

- Approximate inference through sampling
  - Direct
  - Rejection
  - Likelihood weighting
  - Gibbs sampling
- Know why each approach is consistent
- Be able to analyze cost of generating a sample in each method
- Tradeoffs in efficiency (# of samples need to get a good estimate)

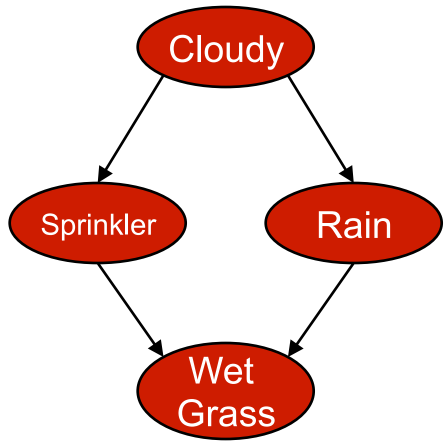
Approximate Inference

- Often interested in:
  - **Posterior probability** of taking on any value given some evidence: $Pr[Q \mid E_1=e_1,...,E_k=e_k]$
  - **Most likely explanation** given some evidence: $argmax_q\ Pr[Q=q \mid E_1=e1,...,E_k=e_k]$
- Imagine we could get samples from the posterior distribution of the query variable given some evidence
- Could use these samples to approximate posterior distribution and/or most likely explanation

Wet Grass Example



Pr(Cloudy | Sprinkler=T,Rain=T)?



- Samples *of Cloudy given Sprinkler=T & Rain=T*): 1 0 1 1 0 1 1 1 1 0
- Posterior probability of taking on any value given some evidence: Pr[Q | $E_1=e_1,\ldots,E_k=e_k$]
  - Pr(Cloudy = T | Sprinkler=T, Rain=T) ≈ .7
  - Pr(Cloudy = F | Sprinkler=T, Rain=T) ≈ .3

Sampling As Approximate Inference

- http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Sampling From a Distribution

- We'll spend time today talking about different ways to obtain samples from posterior distribution from a Bayes Net
- But first, how to sample the value of a single variable

Sampling Single Variable

- Consider when have a CPT (conditional probability table) that specifies the probability of C being true or false
- Want to sample values from this distribution
- Simple approach
  - r = random # generator between (0,1)
  - If(r < 0.5) sample = c+ (c=true)
  - Else sample = c- (c=false)

P(C)

+c	0.5
-c	0.5

Sampling Single Variable 2

- Want to sample s when C=-c (c is false)
- Simple approach
  - r = random # generator between (0,1)
  - If(r < 0.5) sample = s+
  - Else sample = s-

P(S|C)

+c	+s	0.90
+c	-s	0.10
-c	+s	0.5
-c	-s	0.5

Note: can be a bit more complicated for certain parametric distributions

Sampling

- Have some method for generating samples given a known probability distribution
- Sample will be an assignment of values to each variable in the network
  - Generally will only be interested in query variables after finish sampling
- **Use samples to approximately compute posterior probabilities**

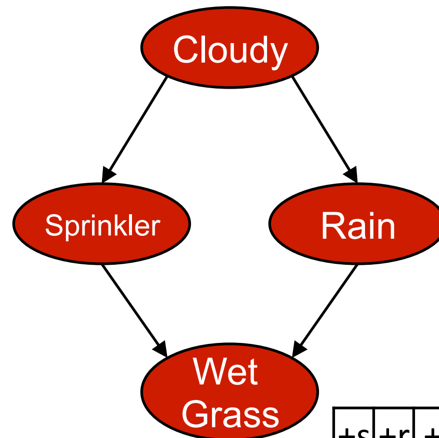
Direct Sampling

- Generate samples from a network with no evidence
- Create a topological order of the variables in the Bayes Net
- Sample each variable conditioned on the values of its parents

Direct Sampling

- Sample Pr[C]=(.5,.5)
  ⇒ true

+c	0.5
-c	0.5



+c	+s	.1
+c	-s	.9
-c	+s	.5
-c	-s	.5

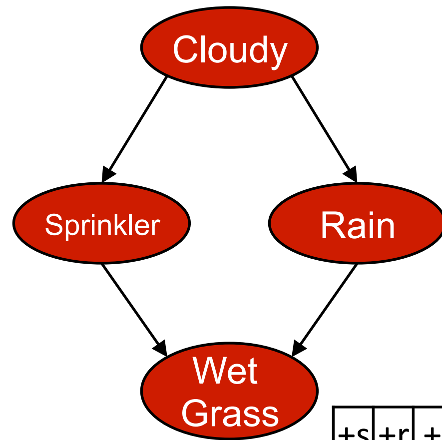
+c	+r	.8
+c	-r	.2
-c	+r	.2
-c	-r	.8

+s	+r	+w	.99
+s	+r	-w	.01
+s	-r	+w	.90
+s	-r	-w	.10
-s	+r	+w	.90
-s	+r	-w	.10
-s	-r	+w	0
-s	-r	-w	1.0

Direct Sampling

- Sample Pr[C]=(.5,.5)
  ⇒ true
- Sample Pr[S|C=t]=(.1,.9)
  ⇒ false

+c	0.5
-c	0.5



+c	+s	.1
+c	-s	.9
-c	+s	.5
-c	-s	.5

+c	+r	.8
+c	-r	.2
-c	+r	.2
-c	-r	.8

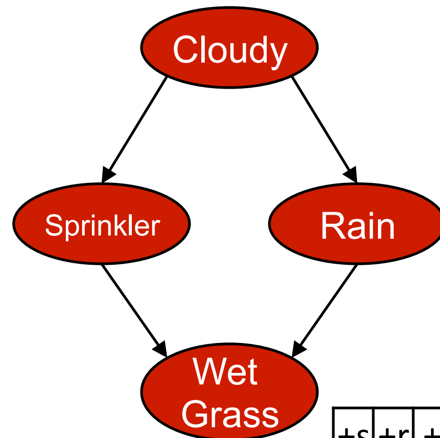
+s	+r	+w	.99
+s	+r	-w	.01
+s	-r	+w	.90
+s	-r	-w	.10
-s	+r	+w	.90
-s	+r	-w	.10
-s	-r	+w	0
-s	-r	-w	1.0

Direct Sampling

- Sample Pr[C]=(.5,.5)
  ⇒ true
- Sample Pr[S|C=t]=(.1,.9)
  ⇒ false
- Sample Pr[R|C=t]=(.8,.2)
  ⇒ true

+c	0.5
-c	0.5

+c	+s	.1
+c	-s	.9
-c	+s	.5
-c	-s	.5



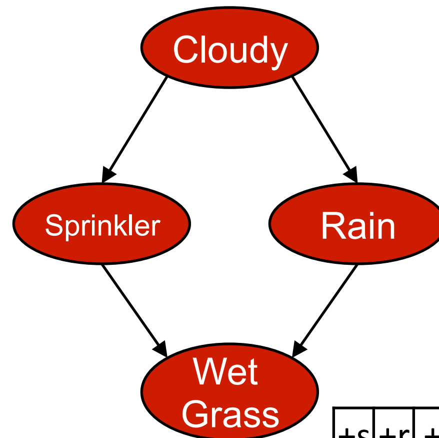
+c	+r	.8
+c	-r	.2
-c	+r	.2
-c	-r	.8

+s	+r	+w	.99
+s	+r	-w	.01
+s	-r	+w	.90
+s	-r	-w	.10
-s	+r	+w	.90
-s	+r	-w	.10
-s	-r	+w	0
-s	-r	-w	1.0

Direct Sampling

- Sample Pr[C]=(.5,.5)
  ⇒ true
- Sample Pr[S|C=t]=(.1,.9)
  ⇒ false
- Sample Pr[R|C=t]=(.8,.2)
  ⇒ true
- Sample Pr[W|S=f,R=t]=(.9,.1)
  ⇒ true
- Sampled [t,f,t,t]

+c	0.5
-c	0.5



+c	+s	.1
+c	-s	.9
-c	+s	.5
-c	-s	.5

+c	+r	.8
+c	-r	.2
-c	+r	.2
-c	-r	.8

+s	+r	+w	.99
+s	+r	-w	.01
+s	-r	+w	.90
+s	-r	-w	.10
-s	+r	+w	.90
-s	+r	-w	.10
-s	-r	+w	0
-s	-r	-w	1.0

Direct Sampling

- Sampling process generates samples from prior joint distribution specified by BN
- Use samples to estimate probability of a specific event
  - Reminder: event is assignment of values to variables
- $\Pr[X_1=x_1,\ldots,X_5=x_5] \approx \#(x_1,\ldots,x_5)/\#\text{samples}$
  - $\approx$ means becomes exact in large-sample limit
  - Implies estimate is consistent

Rejection sampling

- What about when we have evidence?
- Want to estimate Pr[Rain=t|Sprinkler=t] using 100 direct samples
- 73 have S=f, of which 12 have R=t
- 27 have S=t, of which 8 have R=t

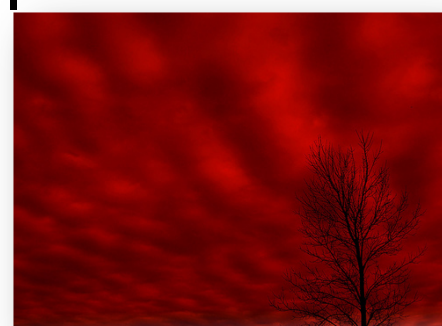
What's the estimate?

A) 20/100 B) 12 / 73

C**) 8 / 27** D) Not sure

Rejection sampling

- What about when we have evidence?
- Use direct sampling
- Reject all samples inconsistent with evidence, and estimate probability of events in remaining samples
- Problem: try to estimate Pr[Rain| RedSkyAtNight=t]!



Solution: Likelihood weighting

- Current approach: generate samples until have many that agree with evidence
- Proposed approach:
  - Generate only samples that agree with evidence
  - Weight them according to likelihood of evidence

Generating a Sample Using Likelihood Weighting

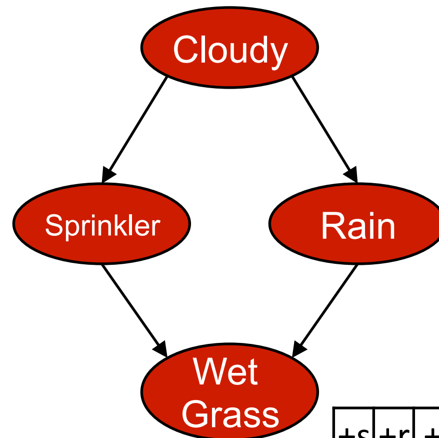
- Select a topological ordering of variables
- Set w = 1
- **x** ← event with evidence variables set
- For each variable X_i in order $(X_1, X_2, \ldots)$:
  - If X_i is an evidence variable

    Update w ← w * $P(X_i = e_i \mid Parents(X_i) = \mathbf{x}(Parents(X_i)))$
  - Else **x**[i] ← sample from $P(X_i \mid Parents(X_i) = \mathbf{x}(Parents(X_i)))$

Likelihood weighting

- Evidence: C=t,W=t
- C is evidence var
  $\Rightarrow$ w = 1·Pr[C=t] = 0.5

+c	0.5
-c	0.5



+c	+s	.1
+c	-s	.9
-c	+s	.5
-c	-s	.5

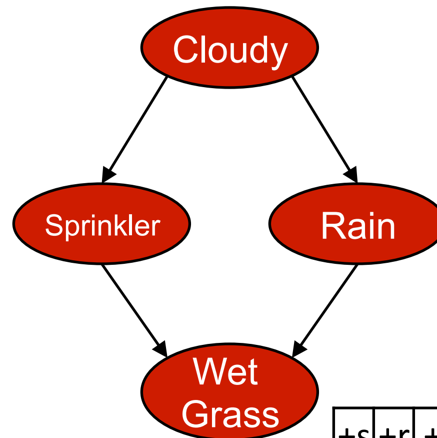
+c	+r	.8
+c	-r	.2
-c	+r	.2
-c	-r	.8

+s	+r	+w	.99
+s	+r	-w	.01
+s	-r	+w	.90
+s	-r	-w	.10
-s	+r	+w	.90
-s	+r	-w	.10
-s	-r	+w	0
-s	-r	-w	1.0

Likelihood weighting

- Evidence: C=t,W=t
- C is evidence var
  ⇒ w = 1·Pr[C=t] = 0.5
- Sample Pr[S|C=t]=(.1,.9)
  ⇒ false

+c	0.5
-c	0.5



+c	+s	.1
+c	-s	.9
-c	+s	.5
-c	-s	.5

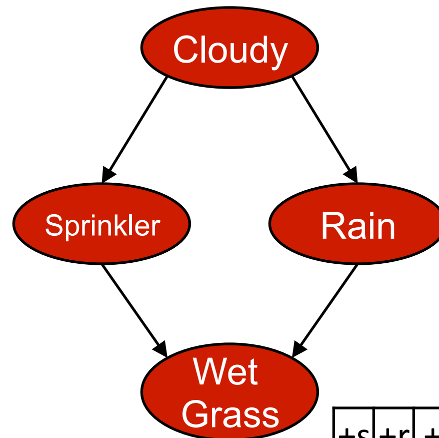
+c	+r	.8
+c	-r	.2
-c	+r	.2
-c	-r	.8

+s	+r	+w	.99
+s	+r	-w	.01
+s	-r	+w	.90
+s	-r	-w	.10
-s	+r	+w	.90
-s	+r	-w	.10
-s	-r	+w	0
-s	-r	-w	1.0

Likelihood weighting

- Evidence: C=t,W=t
- C is evidence var
  $\Rightarrow$ w = 1·Pr[C=t] = 0.5
- Sample Pr[S|C=t]=(.1,.9)
  $\Rightarrow$ false
- Sample Pr[R|C=t]=(.8,.2)
  $\Rightarrow$ true

+c	0.5
-c	0.5



+c	+s	.1
+c	-s	.9
-c	+s	.5
-c	-s	.5

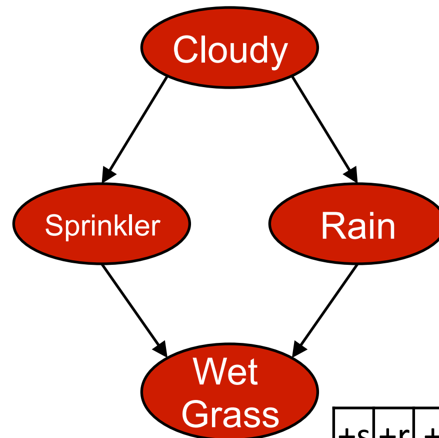
+c	+r	.8
+c	-r	.2
-c	+r	.2
-c	-r	.8

+s	+r	+w	.99
+s	+r	-w	.01
+s	-r	+w	.90
+s	-r	-w	.10
-s	+r	+w	.90
-s	+r	-w	.10
-s	-r	+w	0
-s	-r	-w	1.0

Likelihood weighting

- Evidence: C=t,W=t
- C is evidence var
  $\Rightarrow$ w = 1·Pr[C=t] = 0.5
- Sample Pr[S|C=t]=(.1,.9)
  $\Rightarrow$ false
- Sample Pr[R|C=t]=(.8,.2)
  $\Rightarrow$ true
- W is evidence var
  $\Rightarrow$ w = 0.5·Pr[W=t|S=f,R=t] = .45
- Sampled [t,f,t,t] with weight .45, tallied under R=t

+c	0.5
-c	0.5



+c	+s	.1
+c	-s	.9
-c	+s	.5
-c	-s	.5

+c	+r	.8
+c	-r	.2
-c	+r	.2
-c	-r	.8

+s	+r	+w	.99
+s	+r	-w	.01
+s	-r	+w	.90
+s	-r	-w	.10
-s	+r	+w	.90
-s	+r	-w	.10
-s	-r	+w	0
-s	-r	-w	1.0

Likelihood Weighting: Computing P(X|e)

inputs: X, the query variable
  e, observed values for variables **E**
  bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$
  N, the total number of samples to be generated
local variables: **W**, a vector of weighted counts for each value of X, initially zero

for j = 1 to N **do**
  $\mathbf{x}, w \leftarrow$ WEIGHTED-SAMPLE(bn, **e**)
  $\mathbf{W}[x] \leftarrow \mathbf{W}[x] + w$ where x is the value of X in $\mathbf{x}$
return NORMALIZE(**W**)

Consistency

- Samples each non-evidence variable z in a sample according to

$$S_{WS}(\mathbf{z}, \mathbf{e}) = \prod_{i=1}^{l} P(z_i \mid parents(Z_i))$$

- Is this the true posterior distribution P(z|e)?
  - No, but weights fix this!

Weighted Probability

- Samples each non-evidence variable z according to

$$S_{WS}(\mathbf{z},\mathbf{e}) = \prod_{i=1}^{l} P(z_i \,|\, parents(Z_i))$$

- Weight of a sample is

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

- Weighted probability of a sample is

$$\begin{aligned} S_{WS}(\mathbf{z},\mathbf{e})w(\mathbf{z},\mathbf{e}) &= \prod_{i=1}^{l} P(z_i \,|\, parents(Z_i)) \prod_{i=1}^{m} P(e_i \,|\, parents(E_i)) \\ &= P(\mathbf{z},\mathbf{e}) \end{aligned}$$

Does Likelihood Weighting Produce Consistent Estimates?

- Yes, see book

Example

- When sampling S and R the evidence W=t is ignored
    - Samples with S=f and R=f although evidence rules this out
- Weight makes up for this difference
    - above weight would be 0
- If we have 100 samples with R=t and **total weight 1**, and 400 samples with R=f **and total weight 2**, what is estimate of R=t?
    - = 1/ 3

Limitations of Likelihood Weighting

- Poor performance if evidence vars occur later in ordering