INFERENCE IN BAYESIAN NETWORKS

Today

- Reading
  - AIMA 14.4 – 14.5

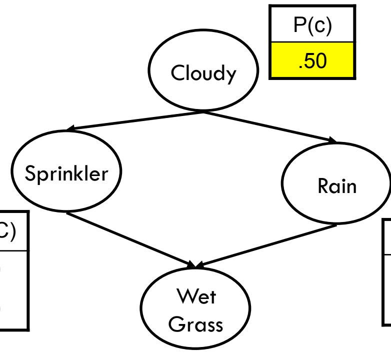
- Goals
  - Approximate inference
  - (Case Study: Latent Dirichlet Allocation)

Approximate Inference

- Analogous to uninformed/informed search algorithms that use an incremental formulation
  - Direct sampling
  - Rejection sampling
  - Likelihood weighting

- Analogous to local search algorithms that use a complete-state formulation and make local modifications
  - Gibbs sampling (special case of MCMC methods)

Direct Sampling: no evidence



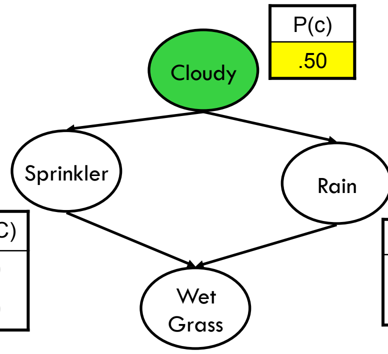
P(c)
.50

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

Direct Sampling: no evidence



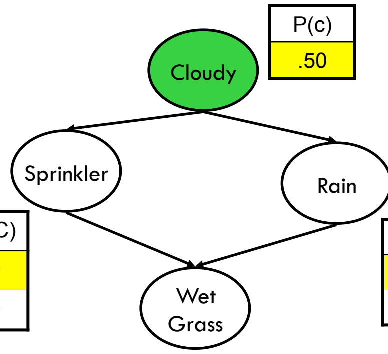
P(c)
.50

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

Direct Sampling: no evidence



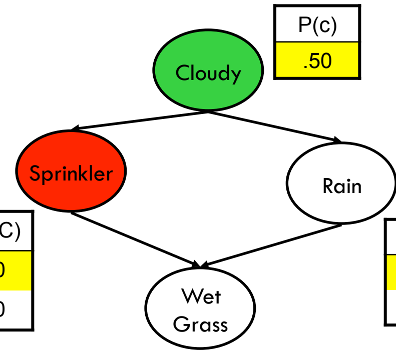
P(c)
.50

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

Direct Sampling: no evidence



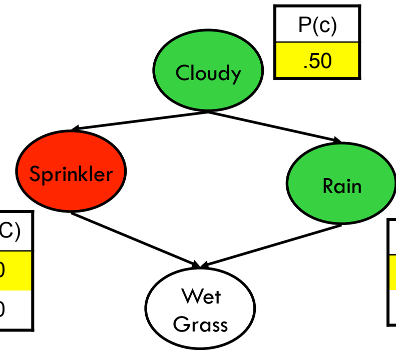
P(c)
.50

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

Direct Sampling: no evidence



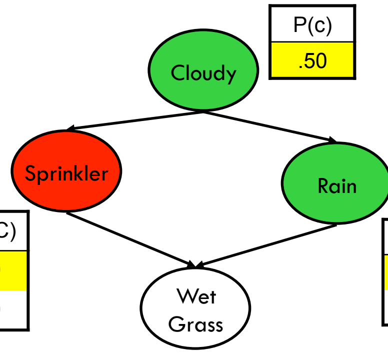
P(c)
.50

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

Direct Sampling: no evidence



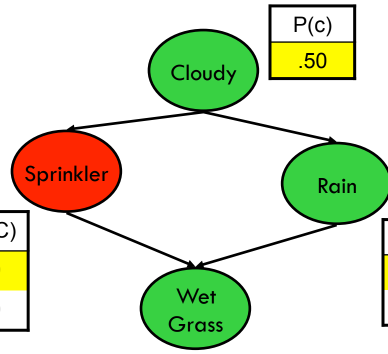
P(c)
.50

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

Direct Sampling: no evidence



P(c)
.50

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

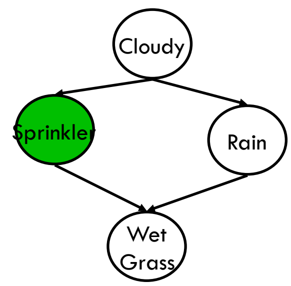
C	P(r|C)
T	.80
F	.20

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

Sample: [T, F, T, T]

Rejection Sampling: evidence

- Perform direct sampling
- "Reject", i.e. remove, any samples that are inconsistent with the evidence



[C, S, R, W]	
[T, T, F, T]	[T, T, F, T]
[F, F, F, F]	~~[F, F, F, F]~~
[F, T, F, T]	[F, T, F, T]
[F, F, T, T]	~~[F, F, T, T]~~
[T, F, F, F]	~~[T, F, F, F]~~
[T, T, F, T]	[T, T, F, T]
[F, T, F, T]	[F, T, F, T]
[T, F, F, F]	~~[T, F, F, F]~~
[F, T, T, F]	[F, T, T, F]
[T, T, F, F]	[T, T, F, F]

p(R | S = true)
p(R = true | S = true) ≈ 1/6
p(R = false | S = true) ≈ 5/6

Likelihood weighting

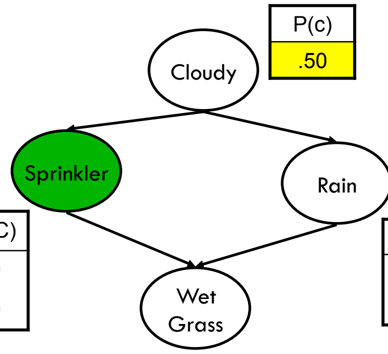
- Fixes the values for the evidence so there are no wasted samples
- Sample only the non-evidence variables
- Not every sample is created equal
  - Need to weight each sample by how likely the evidence is given the sampled values
  - Compute the product of the conditional distribution of the evidence given the sampled values of its parents

$$\text{weight} = p(e_1 \mid \text{Parents}(e_1)) * p(e_2 \mid \text{Parents}(e_2)) \ldots$$

Likelihood weighting

weight = 1

P(c)
.50



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

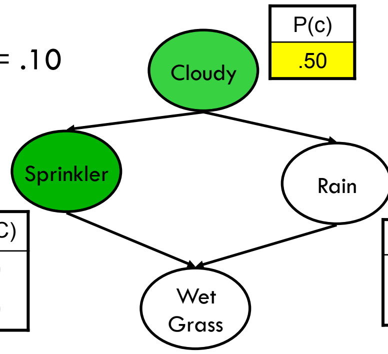
C	P(r|C)
T	.80
F	.20

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

Likelihood weighting

weight = $p(s|c)$ = .10

P(c)
.50



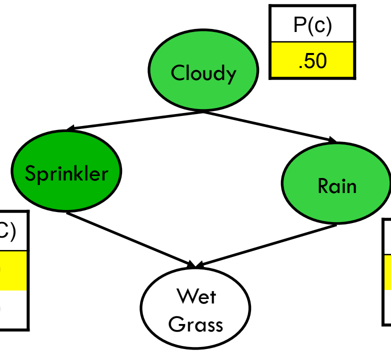
C	P(s|C)
T	.10
F	.50

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

Likelihood weighting

weight = .10



P(c)
.50

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

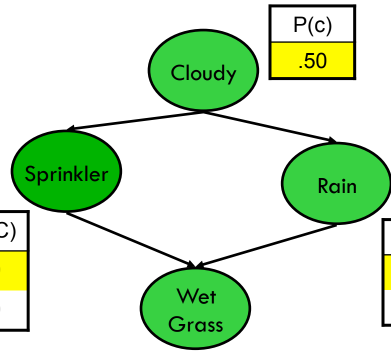
C	P(r|C)
T	.80
F	.20

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

Likelihood weighting

weight = .10

Sample: [T, T, T, T]



P(c)
.50

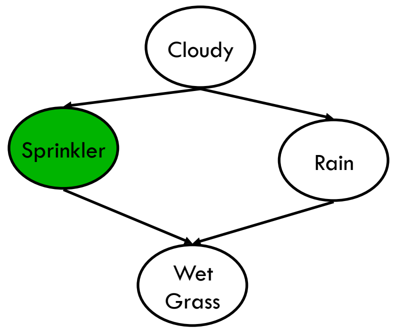
C	P(s|C)
T	.10
F	.50

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

Likelihood weighting

Sample [C,S,R,W]	Weight
[T, T, F, T]	p(s|c) = .10
[F, T, F, T]	p(s|-c) = .50
[T, T, F, T]	p(s|c) = .10
[F, T, T, F]	p(s|-c) = .50
[T, T, T, T]	p(s|c) = .10
[F, T, F, T]	p(s|-c) = .50



- Estimate probability of query using a weighted average

Gibbs Sampling

- Analogous to a local search algorithm where we make local modifications to our current state
  - Initial state = random assignment of non-evidence variables
  - States = complete assignment of values to variables
  - Transition = sample a new value for each variable in turn

Draw state space for WetGrass example on board

Gibbs Sampling

- Analogous to a local search algorithm where we make local modifications to our current state
  - Initial state = random assignment of non-evidence variables
  - States = complete assignment of values to variables
  - Transition = sample a new value for each variable in turn
- Each step to a new state is recorded as a sample
- In the limit, the probability of being in a state is proportional to that state's posterior probability

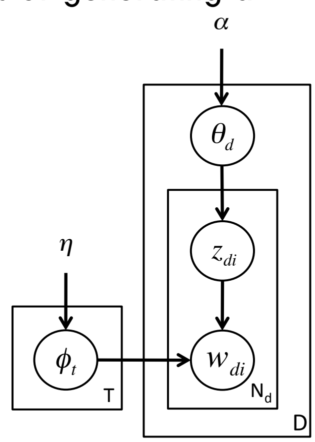
Gibbs Sampling

- Gibbs sampling is an instance of a more general class of algorithms known as Markov Chain Monte Carlo (MCMC) algorithms
  - Note the use of the phrase "Markov chain" which we saw an example of earlier
- Other methods you might hear mentioned
  - Metropolis-Hastings (a generalization of Gibbs sampling)
  - Variational method
  - Belief propagation

Case Study: Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a Bayesian network that describes a hypothetical process of generating a document

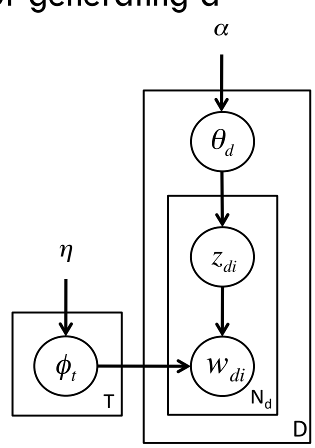
Plate notation is a compact representation of a BN where boxes (i.e. plates) are analogous to for-loops



Case Study: LDA

Latent Dirichlet Allocation is a Bayesian network that describes a hypothetical process of generating a document

- Similarities/differences to past examples?
- What are the independencies encoded in the Bayesian Network?



Case Study: Inference in LDA

- Marginalize out θ and ϕ
- Use Gibbs sampling to draw samples from the posterior distribution:

$$p(z|w) \propto p(z,w)$$

- Each sample is an assignment of words to topics
- We want the most likely assignment, i.e. the assignment of words to topics that has the highest probability

Case Study: Latent Dirichlet Allocation

Topic 247

word	prob.
DRUGS	.069
DRUG	.060
MEDICINE	.027
EFFECTS	.026
BODY	.023
MEDICINES	.019
PAIN	.016
PERSON	.016
MARIJUANA	.014
LABEL	.012
ALCOHOL	.012
DANGEROUS	.011
ABUSE	.009
EFFECT	.009
KNOWN	.008
PILLS	.008

Topic 5

word	prob.
RED	.202
BLUE	.099
GREEN	.096
YELLOW	.073
WHITE	.048
COLOR	.048
BRIGHT	.030
COLORS	.029
ORANGE	.027
BROWN	.027
PINK	.017
LOOK	.017
BLACK	.016
PURPLE	.015
CROSS	.011
COLORED	.009

Topic 43

word	prob.
MIND	.081
THOUGHT	.066
REMEMBER	.064
MEMORY	.037
THINKING	.030
PROFESSOR	.028
FELT	.025
REMEMBERED	.022
THOUGHTS	.020
FORGOTTEN	.020
MOMENT	.020
THINK	.019
THING	.016
WONDER	.014
FORGET	.012
RECALL	.012

Topic 56

word	prob.
DOCTOR	.074
DR.	.063
PATIENT	.061
HOSPITAL	.049
CARE	.046
MEDICAL	.042
NURSE	.031
PATIENTS	.029
DOCTORS	.028
HEALTH	.025
MEDICINE	.017
NURSING	.017
DENTAL	.015
NURSES	.013
PHYSICIAN	.012
HOSPITALS	.011

Figure 1. An illustration of four (out of 300) topics extracted from the TASA corpus.

Case Study: Latent Dirichlet Allocation

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI