Advanced Artificial Intelligence - Bayesian Networks II

Dr Nicola Bellotto

Exact vs Appropriate Inference

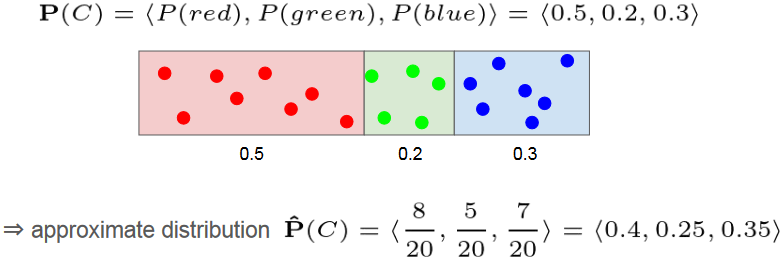
Exact inference become problematic though in case of large probability models, mainly because of computational complexity

Approximate inference methods are therefore a viable alternative to give reasonable answers in case of large models.

**Probabilistic Sampling**

The main idea is to draw N samples from the probability distribution of the network, and from those derive the distribution of the query P(X | evidence)

Example: sampling 20 times from a real (unknown) distribution



It can be shown that, as the number of sample increases, the approximation converges to the true probability distribution.

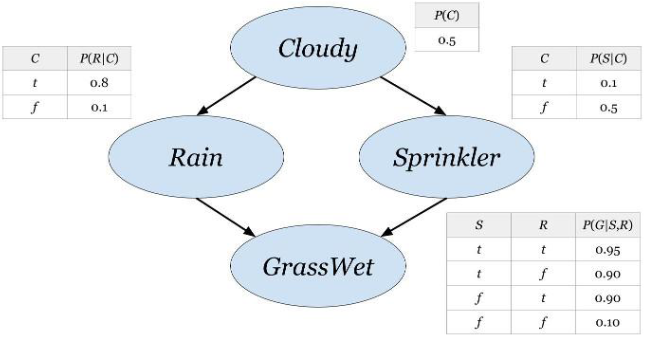
Methods based on randomized sampling are called Monte Carlo algorithms

Next, we’ll consider two classes:

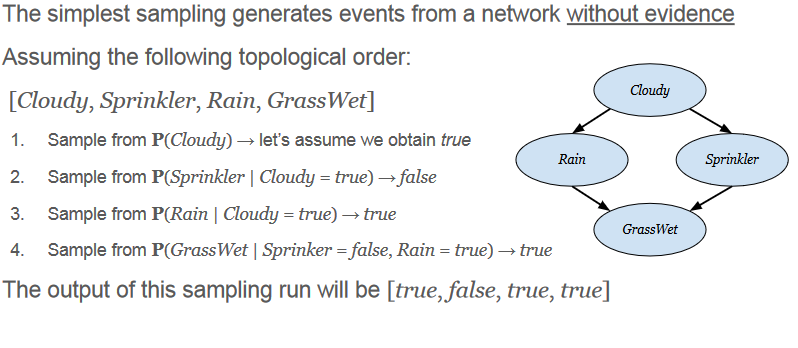
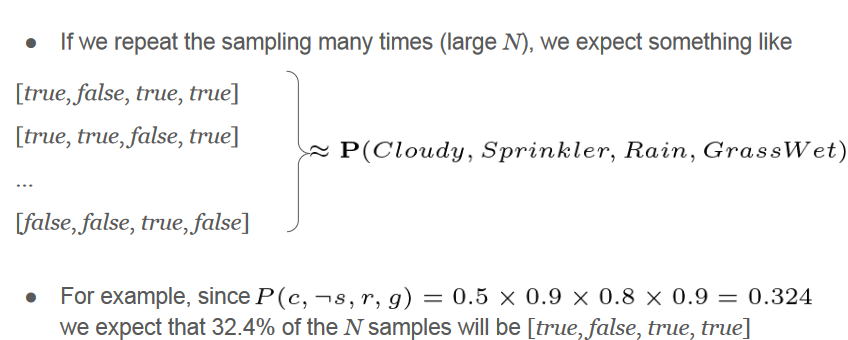
Direct sampling Methods (prior sampling, rejection sampling, likelihood weighting)

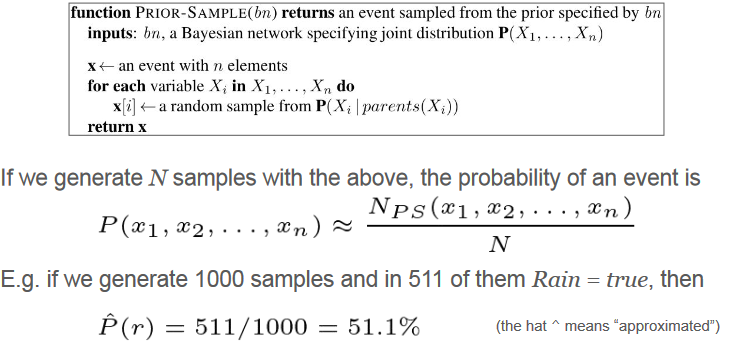
Markov Chain Monte Carlo (MCMC) methods (Gibbs sampling)

**Sampling from Bayesian Networks**

****

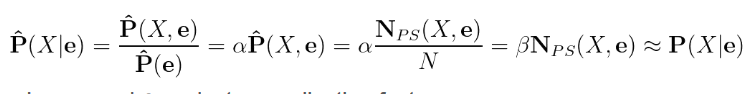
**Prior Sampling**

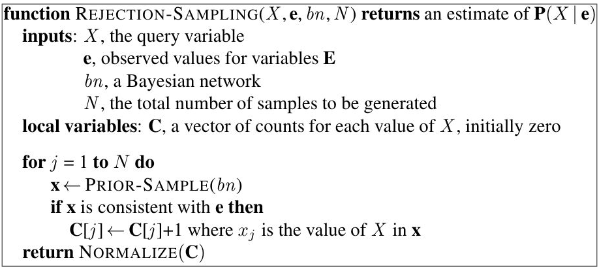


**Rejection Sampling**

Let’s assume we have some evidence e and want to compute P(X | e)  
The idea is to generate all the samples before, but rejecting all those which   
are not consistent with the evidence

  
Where 𝛼 and 𝛽 are just normalisation factors  
E.g. estimate P (Rain | Sprinkler = true) from 100 samples

* 73 have Sprinkler = false, therefore are rejected
* of the 27 with Sprinkler = true, 8 have Rain = true and 19 have Rain = false
* the approximate distribution is



Note the algorithm calls the same Prior-Sample function, but counts only the cases that are consistent with given evidence.

**Likelihood Weighting**

The problem with rejection sampling is that it wastes a lot of samples!

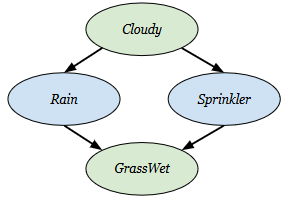
We can improve this by:

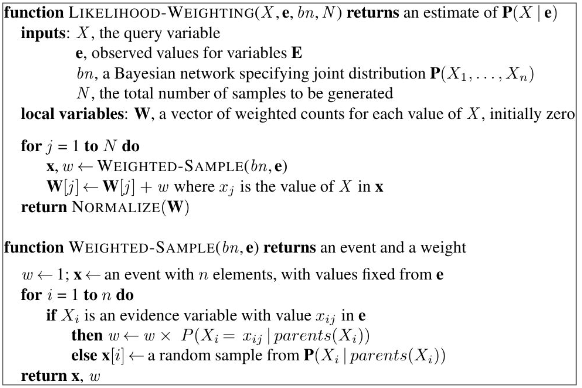
* Generates only sample that are consistent with the evidence e
* Fix the values of the evidence nodes E in the network
* Sample only the non-evidence variables

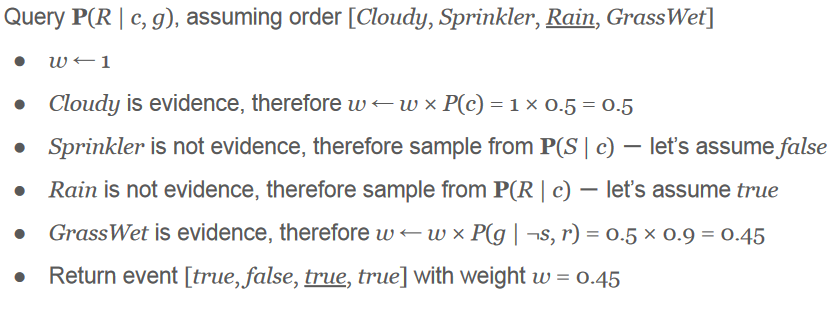
E.g. if the query P(R | c, g), then

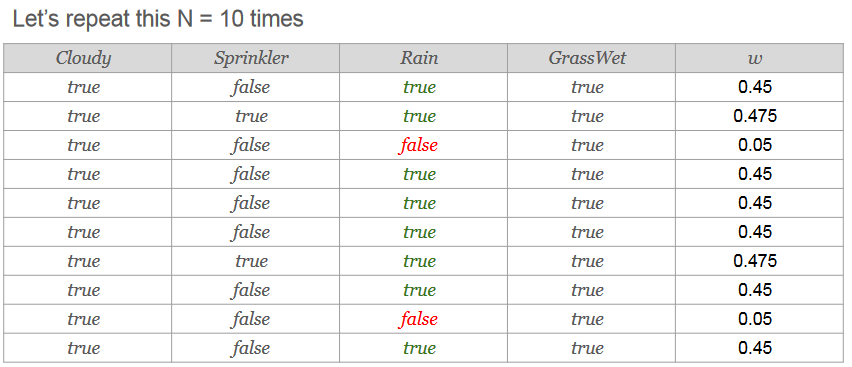
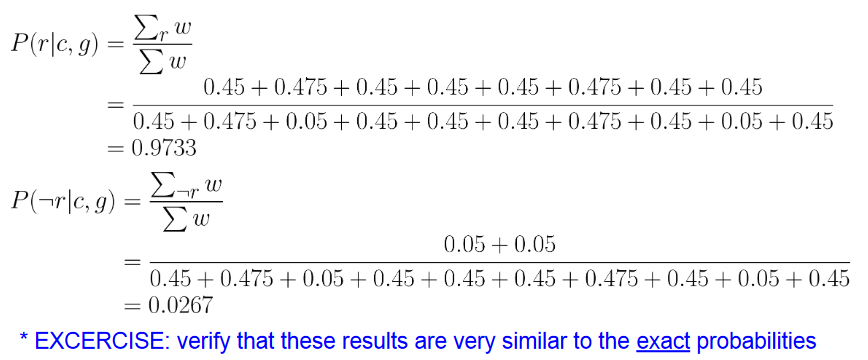
Fix the values Cloudy = true and GrassWet = true

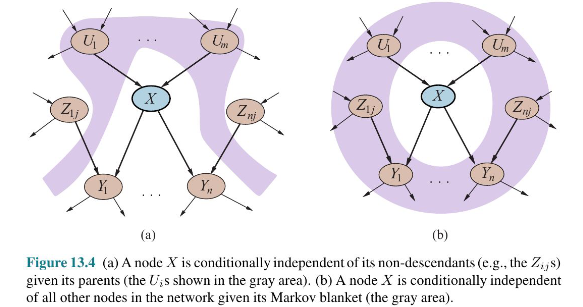
Sample only Rain and Sprinkler





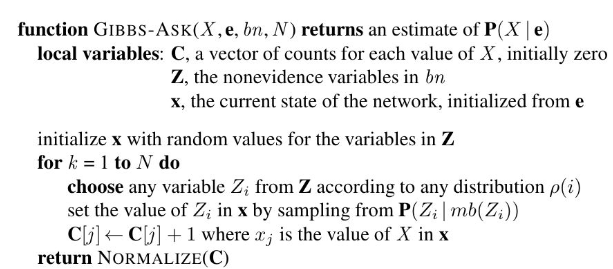
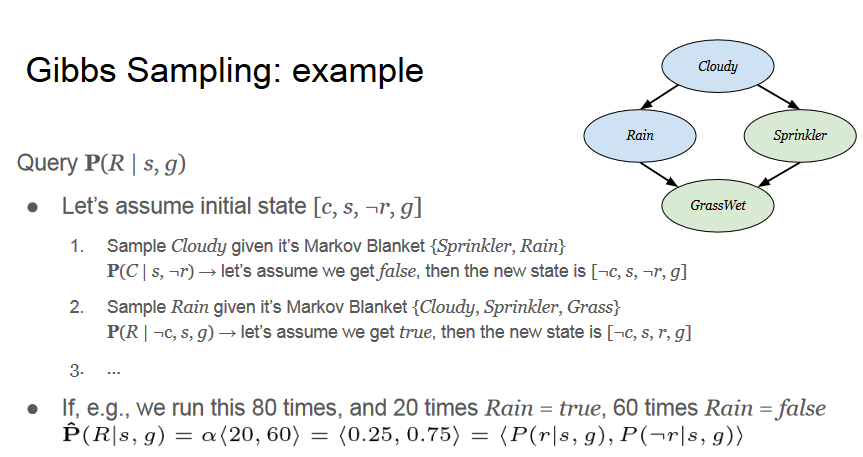


Markov blankets also include co-parents

**Gibbs Sampling**

* Likelihood Weighting’s performance decreases as the number of variable   
  grows because the weights become smaller and smaller.
* Instead of sampling every time from scratch, Markov Chain Monte Carlo   
  methods generate each samples by randomly changing the previous one   
  (that’s why it’s called a “chain”...)
* Gibbs Sampling is a particular MCMC method that works as follows:
  + Starting from arbitrary event (state), and given some evidence e , generate the next state by   
    sampling a value from one of the non-evidence variable Xi given its Markov Blanket
  + Wander around the state space flipping one variable at a time, but keeping the evidence fixed.

Reading

Russell & Norvig “Artificial Intelligence -- A Modern Approach”  
○ Ch 14 - Sec 14.5