Tractable Probabilistic Inference

(Bayesian Networks: Approximate Inference)

Monte Carlo Estimation: The Basic Idea

- What is the probability that 10 dice throws add up exactly to 34?

  Conscientious Way. Calculate this exactly by counting all possible ways of making 34 from 10 dice. This is the exact answer.

  Lazy Way. Simulate throwing the dice (say 500 times), count the number of times the results add up to 34, and divide this by 500. This is an estimate of the exact answer.

- In fact, the estimate from the Lazy Way can get quite close to the correct answer quite quickly.

Monte Carlo versus Exact

The exact probability that 10 dice throws to add up to 34 is $4325310/(6^{10}) \approx 0.072$.

n	Approx P
10	0.000
100	0.090
250	0.048
500	0.056
5000	0.062
10000	0.065

Law of large numbers for a consistent estimator:

$$\lim_{n \to \infty} \mathrm{P}(|\hat{\theta} - \theta| < \epsilon) \to 1$$

Using Monte Carlo for Approximate Computation

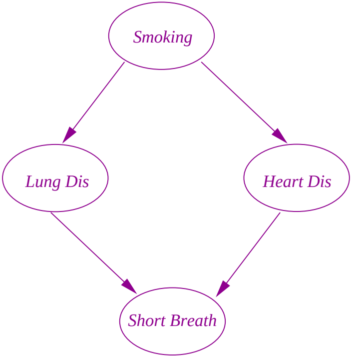
Use Monte Carlo simulation when:

1. We have all the probabilities we need for the computation, but
2. ... actually doing the computation is too hard.

This applies directly to the problem of obtaining probabilities in a Bayesian Network.

A Simple Bayesian Network

P(s)
0.2



S	P(l)
T	0.8
F	0.1

S	P(h)
T	0.6
F	0.1

L	H	P(b)
T	T	0.9
T	F	0.7
F	T	0.8
F	F	0.1

Monte Carlo using Direct Sampling

Start with the ancestral ordering $[S, L, H, B]$

1. Sample from $\mathbf{P}(S) = \langle 0.2, 0.8 \rangle$; suppose this returns *true*.
2. Sample from $\mathbf{P}(L|S = \mathit{true}) = \langle 0.8, 0.2 \rangle$; suppose this returns *true*.
3. Sample from $\mathbf{P}(H|S = \mathit{true}) = \langle 0.6, 0.4 \rangle$; suppose this returns *false*.
4. Sample from $\mathbf{P}(B|L = \mathit{true}, H = \mathit{false}) = \langle 0.7, 0.3 \rangle$; suppose this returns *false*

Sampling thus returns $[\mathit{true}, \mathit{true}, \mathit{false}, \mathit{false}]$. The probability of this event is $0.2 \times 0.8 \times 0.4 \times 0.3 = 0.0192$. That is, on simulation runs, we would expect it to occur approximately 2% of the time (i.e. the Monte Carlo estimate would be ≈ 0.02).

Monte Carlo using Direct Sampling (Again)

- Usually, we are interested in the conditional probability of a variable, given some evidence eg. $\mathrm{P}(H = \mathit{true} | B = \mathit{true})$.
- If the evidence happens to be in the 'top' nodes of the network, we can still use direct sampling procedure.
- Otherwise ...?

Monte Carlo using Rejection Sampling

- Use the direct sampling approach, and just throw away (*reject*) any events that do not agree with the evidence. Then estimate probability from those that are left.
  - For example, to estimate $\mathrm{P}(H = \mathit{true} | B = \mathit{true})$, simulate using direct sampling and reject any events that have $B = \mathit{false}$. Of the remaining, determine the proportion of events for which $H = \mathit{true}$. This is the Monte Carlo estimate of the conditional probability.
- Problem: if the evidence involves many variables, most of the draws will be rejected (very few will agree with the evidence).

Markov Chain Monte Carlo using Gibbs Sampling

1. Generate initial values for all node in the net (evidence nodes values are known)
2. Repeat until we have enough data:
   2.1 For each node X
      2.1.1 Throw away current value for X
      2.1.2 Generate a new value for X according to the probability distribution over X, conditional on current values for other nodes. Actually, sufficient do be conditional on current values of nodes in the Markov Blanket for X. (This is the Gibbs Sampling step)
   2.2 Update tallies for each node

MCMC using Gibbs Sampling

- Do the initial values matter?
  - After some number of iterations, the final probability distribution for each variable Y converges to $\mathbf{P}(Y|\mathbf{e})$ irrespective of the starting point.
  - Best to throw away the first, say 1000, iterations.
- Because each new guess for a node depends only on the current values of of other nodes, we have a *Markov Chain*
- Using the conditional probability for transitions is characteristic of *Gibbs Sampling*
- Because this is computation by simulation, it is a Monte Carlo method.

Aside: Markov Chains I

1. A Markov Chain is simply a sequence of trials $\langle X_1, X_2, \ldots \rangle$ with the following properties:
   1.1 The outcome of any trial X_i can only be one of the finite set $\{a_1, a_2, \ldots, a_m\}$.
   1.2 The outcome of trial X_i depends at most upon the outcome of trial X_{i-1}.
   1.3 With each outcome pair (a_i, a_j) there is a probability p_{ij} that denotes the probability that the outcome a_j occurs immediately after a_i.
2. The probabilities (a_i, a_j) can be arranged in a matrix called the transition matrix.
3. The transition matrix of a Markov Chain is a stochastic matrix. That is, all its rows are probability vectors (each entry is non-negative and the row sums to 1).
4. All powers of the transition matrix P of a Markov Chain are stochastic matrices.
5. An entry (a_i, a_j) in the n^{th} power of a transition matrix represents the probability that the system changes from a_i to a_j in exactly n steps. In particular, if $\mathbf{p}$ is a probability vector representing the current distribution over the states then $\mathbf{p}P^n$ represents the distribution after exactly n steps.

Aside: Markov Chains II

6. A transition matrix P is regular if all entries of P are non-zero for some power m of P.
7. A transition matrix that has a 1 on the main diagonal is not regular.
8. A transition matrix that contains no 0s or 1s is regular.
9. A regular transition matrix P has a unique fixed point probability vector $\mathbf{t}$ whose entries are all positive. That is:
   - $\mathbf{t} = \mathbf{t}P$;
   - $P, P^2, P^3 \ldots$ approach the matrix T all of whose rows are $\mathbf{t}$
   - For any probability vector $\mathbf{p}$, $\mathbf{p}P, \mathbf{p}P^2, \ldots$ approaches $\mathbf{t}$

Important Results: MCMC with Gibbs Sampling

1. A change of state occurs when a variable Y changes its value. If the corresponding transition probability is $\mathbf{P}(Y|MB(Y))$ then it can be shown that the fixed point probability vector of the stochastic matrix is the posterior probability $\mathbf{P}(Y|\mathbf{e})$.

   $MB(Y)$ is the Markov Blanket of Y

2. It can be shown that

   $\mathrm{P}(Y = y|MB(Y)) =$

   $$\alpha \mathrm{P}(Y = y|Parents(Y)) \times \prod_{Z_j \in Children(Y)} \mathrm{P}(z_j|Parents(Z_j))$$

Example of MCMC with Gibbs Sampling I

- Let the query be $\mathbf{P}(HeartDis|smoking, shortBreath)$? That is, what are the probabilities of a smoker ($Smoking = true$) who is short of breath ($ShortBreath = true$) having/not having heart disease?
  - Evidence variables: $Smoking, ShortBreath$
  - Non-evidence variables: $LungDis, HeartDis$
- Decisions
  - Use pseudorandom number generator for the range $[0, 1]$.
  - Initial values: set all non-evidence variables to *true*. Evidence variables must be set to the values given (ie *true* in this case).
  - We can loop through non-evidence variables in a pre-defined order or randomly. We will do the latter. We will use the pseudorandom number generator for this. If number returned is > 0.5 we will draw a value for *LungDis*, otherwise we will draw a value for *HeartDis*.

Example of MCMC with Gibbs Sampling II

  - Since all variables are boolean, distributions are over $\langle true, false \rangle$ and have the form $\langle p, 1-p \rangle$. When drawing values, if the pseudorandom number returned is $\leq p$ then variable is set to *true* otherwise *false*.
  - When do we tally? Usually after about 1000 iterations. Here we will do it after 2 iterations (for illustrative reasons)

- Let random number generator give: 0.154, 0.383, 0.938, 0.813, 0.273, 0.739, 0.049, 0.233, 0.743, 0.932, 0.478, 0.832,...

Example of MCMC with Gibbs Sampling III

Round 1

- Random number is 0.154. So draw a value for *HeartDis*.
- To draw a value, we must first determine the distribution $\langle p, 1-p\rangle$ of *HeartDis* given its Markov Blanket.
- To find the value of p , we need the probability of $HeartDis = true$ given its Markov Blanket. Recall:

$$\mathrm{P}(Y = y|MB(Y)) =$$
$$\alpha \mathrm{P}(Y = y|Parents(Y)) \times \prod_{Z_j \in Children(Y)} \mathrm{P}(z_j|Parents(Z_j))$$

- So,

$$\mathrm{P}(heartDis|MB(HeartDis)) =$$
$$\alpha \mathrm{P}(heartDis|smoking) \times \mathrm{P}(shortBreath|heartDis, lungDis)$$
$$= \alpha(0.6)(0.9) = 0.54\alpha$$

- Similarly

$$\mathrm{P}(\neg heartDis|MB(HeartDis)) =$$
$$\alpha \mathrm{P}(\neg heartDis|smoking) \times \mathrm{P}(shortBreath|\neg heartDis, lungDis)$$
$$= \alpha(0.4)(0.7) = 0.28\alpha$$

Example of MCMC with Gibbs Sampling IV

- Since $\langle 0.54\alpha, 0.28\alpha \rangle = \langle p, 1 - p \rangle$, we get $p = 0.66$.
- Therefore we will set *HeartDis* = *true* if the next random number is ≤ 0.66. It is 0.383. Therefore *HeartDis* remains *true*.

Example of MCMC with Gibbs Sampling V

Round 2

- Random number is 0.938. So draw a value for *LungDis*.
- Proceeding as before:

$$\mathrm{P}(\mathit{lungDis}|\mathit{MB}(\mathit{LungDis})) = \\ \quad \alpha\mathrm{P}(\mathit{lungDis}|\mathit{smoking}) \times \mathrm{P}(\mathit{shortBreath}|\mathit{heartDis}, \mathit{lungDis}) \\ \quad = \alpha(0.8)(0.9) = 0.72\alpha$$

$$\mathrm{P}(\neg\mathit{lungDis}|\mathit{MB}(\mathit{LungDis})) = \\ \quad \alpha\mathrm{P}(\neg\mathit{lungDis}|\mathit{smoking}) \times \mathrm{P}(\mathit{shortBreath}|\neg\mathit{heartDis}, \mathit{lungDis}) \\ \quad = \alpha(0.2)(0.8) = 0.16\alpha$$

- Solving for α as before, $p = 0.82$
- Next random number is 0.813. Therefore, *LungDis* stays at *true* (just)

Example of MCMC with Gibbs Sampling VI

Since 2 iterations have been completed, tallying commences:

heartDis	¬*heartDis*	*lungDis*	¬*lungDis*
1	0	1	0

Example of MCMC with Gibbs Sampling VII

Round 3

- Next random number is 0.273. So, we draw a value for *HeartDis*.
- Since the values of all other variables are the same, the final distribution for *HeartDis* will be the same as after Round 1: $\langle 0.66, 0.34 \rangle$.
- Next random number is 0.739, so *HeartDis* is set to *false*.
- Update tally:

heartDis	$\neg$*heartDis*	*lungDis*	$\neg$*lungDis*
1	1	2	0

Example of MCMC with Gibbs Sampling VIII

Round 4

- Next random number is 0.049. So, we draw a value for *HeartDis*.
- Since the values of all other variables are the same, the final distribution for *HeartDis* will be the same as after Round 1: $\langle 0.66, 0.34 \rangle$.
- Next random number is 0.233, so *HeartDis* is set to *true*.
- Update tally:

heartDis	$\neg$*heartDis*	*lungDis*	$\neg$*lungDis*
2	1	3	0

Example of MCMC with Gibbs Sampling IX

Round 5

- Next random number is 0.743. So, we draw a value for *LungDis*.
- Since the values of all other variables are the same, the final distribution for *LungDis* will be the same as after Round 2: $\langle 0.82, 0.18 \rangle$.
- Next random number is 0.932, so *LungDis* is set to *false*.
- Update tally:

heartDis	$\neg$*heartDis*	*lungDis*	$\neg$*lungDis*
3	1	3	1

Example of MCMC with Gibbs Sampling X

Round 6

- Next random number is 0.478. So, we draw a value for *HeartDis*.
- We have to recompute the distribution for *HeartDis* since the setting for *LungDis* has changed.

  $$\mathrm{P}(\mathit{heartDis}|\mathit{MB}(\mathit{HeartDis})) = \alpha\mathrm{P}(\mathit{heartDis}|\mathit{smoking}) \times \mathrm{P}(\mathit{shortBreath}|\mathit{heartDis}, \neg\mathit{lungDis}) = \alpha(0.6)(0.8) = 0.48\alpha$$

  Similarly, $\mathrm{P}(\neg\mathit{heartDis}|\mathit{MB}(\mathit{HeartDis})) = \alpha(0.4)(0.1) = 0.04\alpha$
- Solving, we get the distribution $\langle 0.92, 0.08\rangle$. Next random number is 0.832, so *HeartDis* is set to *true*.
- Update tally:

heartDis	$\neg$*heartDis*	*lungDis*	$\neg$*lungDis*
4	1	3	2

Example of MCMC with Gibbs Sampling XI

Final Results

- Of course, we have not had nearly enough iterations to obtain an accurate estimate.
- At this stage, the answer to our query

$$\mathbf{P}(HeartDis|smoking, shortBreath)$$

  is

$$\langle 4/5, 1/5 \rangle = \langle 0.8, 0.2 \rangle$$

- Exact answer (using variable elimination) is $\langle 0.7, 0.3 \rangle$